ADAPTIVE OPTIMISATION OF ROAD SAFETY STRATEGIC MANAGEMENT

by

Bosong Jiao

A thesis submitted to University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Civil Engineering

School of Engineering

College of Engineering and Physical Sciences

University of Birmingham

September 2024

Abstract

Road crashes cause significant injuries and fatalities, hindering social and economic progress despite the essential role of a well-developed road network. In 2021, road crashes resulted in 1.19 million fatalities globally, underscoring this critical health issue. Various initiatives by the UN, WHO, Global Road Safety Facility (GRSF), and national road entities focus on reducing traffic crashes through strategic road infrastructure safety management. However, current tools and methods used for road safety management suffer from fragmented data analysis, focusing on specific segments instead of the whole network, limiting comprehensive crash insights. There is also poor integration across decision-making levels and reliance on subjective, localised countermeasures rather than adaptive, data-driven investment strategies. This creates a need for a system utilising Data Science and Artificial Intelligence (AI) to enhance road safety management.

To achieve this, the thesis begins with a systematic literature review of cutting-edge technologies and best practices in road safety. It emphasises the role of information technology, particularly big data, in crash data analysis, identifying a gap in the comprehensive understanding of crash databases and the need for advanced techniques from AI and multi-objective optimisation. These insights guide subsequent research aimed at developing an adaptive, data-driven system for strategic management, encompassing both data analysis and resource allocation.

Existing methods for data analysis often suffer from subjectivity and bias, as they rely on simplistic predefined parameters for frequent pattern searches without critical and logical thinking. Furthermore, they face a significant limitation when applied to the infrequent patterns search in road safety analysis, particularly in cases where fatalities are minimal. An analytical framework using association rule mining and K-means clustering analyses the UK STATS 19 database (2018) to identify associations with fatalities and overview crash data characteristics. This mitigates bias by integrating these methods to explore infrequent patterns, offering a comprehensive perspective to inform strategy development.

Additionally, a novel two-stage model for optimising road safety strategies is introduced, integrating multi-objective optimisation and association rule mining. This model enhances decision-making and resource allocation, validated through a case study in Utrecht. Improvements across objectives demonstrate its efficiency, allowing local authorities to tailor investment plans to their road network characteristics.

In conclusion, this thesis presents a novel conceptual framework for road safety strategic management using advanced data information technology. The adaptable analytical tools developed provide valuable insights for decision-makers at various administrative levels, significantly contributing to road safety. This research establishes a foundation for future studies and practical applications aimed at reducing road traffic crashes.

Acknowledgements

Completing this PhD has been a long and challenging journey, and I am profoundly grateful to everyone who has supported and encouraged me throughout this process.

First and foremost, I want to express my deepest gratitude to my parents. Despite not having had the opportunity for higher education themselves, they have devoted everything they could to ensure I could complete my studies. Although they do not speak any foreign languages, they have always encouraged me to learn and explore opportunities abroad. Their understanding and support during the difficult times of the COVID-19 pandemic enabled me to make the tough decision to pursue a PhD. I am eternally thankful for their unwavering belief in me and their generous sponsorship, which has made my academic achievements possible.

To my supervisor, Dr Harry Evdorides and Dr Sakdirat Kaewunruen, I offer my heartfelt thanks. Their invaluable guidance, constructive feedback, and continuous support have been instrumental in the completion of this thesis. Our journey began during my MSc, and I cannot express how fortunate I am to have had the opportunity to conduct my research under their mentorship. they have provided a supportive environment that illuminated the path ahead without exerting pressure, offered structural guidance rather than word-for-word assistance, and maintained a presence when needed without daily supervision. I cherish the way they have nurtured me into an independent researcher, and I am deeply honoured to be considered one of their colleagues.

I am immensely apricated to my partner Dr Kuok Sin Un. His persistent support has been the cornerstone of my journey towards completing my PhD. Throughout every step of this challenging path, his encouragement, understanding, and belief in me have been a constant

source of strength. His patience during long hours of study, his willingness to listen to my ideas and concerns, and his unwavering faith in my abilities have all played an integral role in my success. Today, as I reflect on my accomplishments, I am profoundly grateful for his presence in my life. He is not just my partner, but my rock, my confidant, and my greatest supporter. I am forever indebted to him for his commitment, his encouragement, and his belief in me.

I am filled with immense gratitude for my dance partner, Dr Kirsty Mary Davies, and our coach, Mr David Mallabone. Their support has extended far beyond the dance floor, encompassing every aspect of my life in the UK. They have been a source of inspiration, encouragement, and friendship. Together, we have shared countless hours of dance, facing challenges with determination and grace. Beyond dance, they have been mentors and friends, offering invaluable advice and encouragement every step of the way. Together, Kirsty and David have become my extended family in the UK, offering warmth, encouragement, and a sense of belonging. However, with a heavy heart, I must express my regret for pausing our dance journey due to various reasons, a decision that undoubtedly brought frustration to both Kirsty and David.

Finally, to all my friends and colleagues, your unwavering support has meant the world to me. Thank you all for being a source of motivation and inspiration throughout this journey.

With heartfelt appreciation,

Bosong

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AADT	Average Annual Daily Traffic
AASHTO	American Association of State Highway and Transportation Analyses
AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANWB	Algemene Nederlandse Wielrijders Bond (Dutch) Royal Dutch Tourist Association (English)
ARM	Assoicaition Rule Mining
BAAC	Accident Analysis Bulletin Corporal
BCA	Benefit Cost Analysis
BCR	Benefit Cost Ratio
BRON	Bestand geRegistreerde Ongevallen in Nederland (Dutch) National Road Crash Registration (English)
CASP	Critical Appraisal Skills Program
CEA	Cost Effectiveness Analysis
CER	Cost Effectiveness Ratio
CRF	Crash Reduction Factor
DEA	Data Envelopment Analysis
DMU	Decision-Making Units
DRSA	Dominance-based Rough Set Approach
EB	Empirical Bayes
EPPI	Evidence for Policy and Practice Information and Coordinating
EU	European Union
FHWA	Federal Highway Administration
FP-Growth	Frequent Pattern Growth
GPS	Government Policy Statement
HDM-4	Highway Development and Management
HSIP	Highway Safety Improvement Program
ISMO	Importance Segregated Multi-objective Optimisation
KSI	Killed or Seriously Injured
LHS	Left Hand Side of rules (Antecedent)
LMICs	low- and middle-income countries
MCA	Multi-Criteria Analysis
MILP	Mixed Integer Linear Programming
NB	Negative Binomial
NPOs	Non-Profit Organisations
NPV	Net Present Value
NSGA-II	Non-dominated Sorting Genetic Algorithm II
NZ	New Zealand
ONF	One Network Framework
PCA	Principal Component Analysis
PDO	Property Damage Only

PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
iRAP	International Road Assessment Programme
RF	Random Forest
RHS	Right Hand Side of rules (Consequent)
RPS	Road Protection Scores
SHSP	Strategic Highway Safety Plans
SII	Safety Improvement Indices
SRIP	Safer Road Investment Plan
SRN	Strategic Road Network
SSL	Short Service Life
SVM	Support Vector Machine
SWOV	Stichting Wetenschappelijk Onderzoek Verkeersveiligheid (Dutch) Dutch Institute for Road Safety Research (English)
UK	United Kingdom
UN	United Nations
US	United States

Chapter 1. Introduction

1.1 Background

As transportation networks have expanded globally and vehicle ownership has risen, road traffic collisions have become a significant challenge. While the expansion of road networks promotes economic growth and facilitates daily life, it also leads to negative impacts on individuals, families, communities, and economies (Haghighi et al., 2020). Road traffic collisions remain a major public health concern, causing millions of deaths or life-altering injuries annually, especially among young people aged 5-29, according to the World Health Organisation (2023). These collisions not only result in physical injuries but also impose emotional, psychological, and economic burdens on survivors and their families, disrupt livelihoods, strain healthcare systems, and incur substantial societal costs (Blincoe et al., 2015).

To address this issue, the World Health Organisation et al. (2021) developed the 'Global Plan Decade of Action for Road Safety 2021-2030', aiming to prevent at least 50% of road traffic deaths and injuries by 2030. Despite efforts to improve road safety through engineering, education, and enforcement measures, significant challenges remain. These challenges include inadequate safer road infrastructure, insufficient enforcement capacity, low compliance with traffic laws, and the increasing complexity of modern transportation systems. Consequently, there is an urgent need to adopt an integrated approach to road safety, pursuing long-term and sustainable solutions while strengthening national intersectoral collaboration, including engagement with non-governmental organisations, civil society, academia, businesses, and industry (United Nations, 2020).

The World Health Organisation et al. (2021) highlighted importance of addressing the interconnection of the four pillars of the global plan—safe user, safe vehicle, safe road, and effective post-crash response—with a particular emphasis on safe roads and effective post-crash response, which involve engineering and management. Notably, these organisations highlighted the significance of understanding the road environment and suggested looking for supportive technologies to improve road safety. In alignment with these recommendations, this research originates from a civil engineering perspective, focusing on road infrastructure engineering. The aim is to optimise the road safety strategic management processes and road safety outcomes by exploring the inherent explanations of road traffic crashes on road networks and developing adaptive investment plans.

1.2 Strategic Management

The use of strategy has existed for centuries, originating from military contexts, and only in the last six decades has it gained prominence in the realm of management. Born out of military conflicts, the use of superior strategy enabled one warring party to defeat another. The concept of strategy has since been adapted by modern profit-driven organisations, where it plays a pivotal role in securing competitive advantage (Henry, 2021). Notably, numerous non-profit organisations (NPOs) did not participate in strategic planning procedures due to their lack of competitive orientation. Unlike private entities that rely on fee structures for financial support, public organisations procure funding through budget allocations or taxation, offering either free services or charging minimal fees to partially cover operational expenses. Therefore, NPOs are encouraged to explore competitive advantages beyond monetary gains, including considerations of environmental sustainability, market efficacy, and alternative revenue streams (Katsioloudes & Abouhanian, 2009).

According to Johnson et al. (2008), strategic management comprises three core components:

strategic analysis; strategic choice, and strategy implementation. Katsioloudes and Abouhanian (2009) highlighted strategy must be dynamic. The process of strategic management involves a sequence of activities designed to achieve identified objectives, as illustrated in Figure 1.2.1. These activities include: 1) identifying objectives, representing the desired outcomes by the end of the cycle; 2) conducting strategic data analysis to assess the status, position, or issues; 3) developing strategic choices by formulating potential courses of action; 4) implementing the chosen strategy; and 5) evaluating the overall performance of the strategic management process (Rao, 2009).

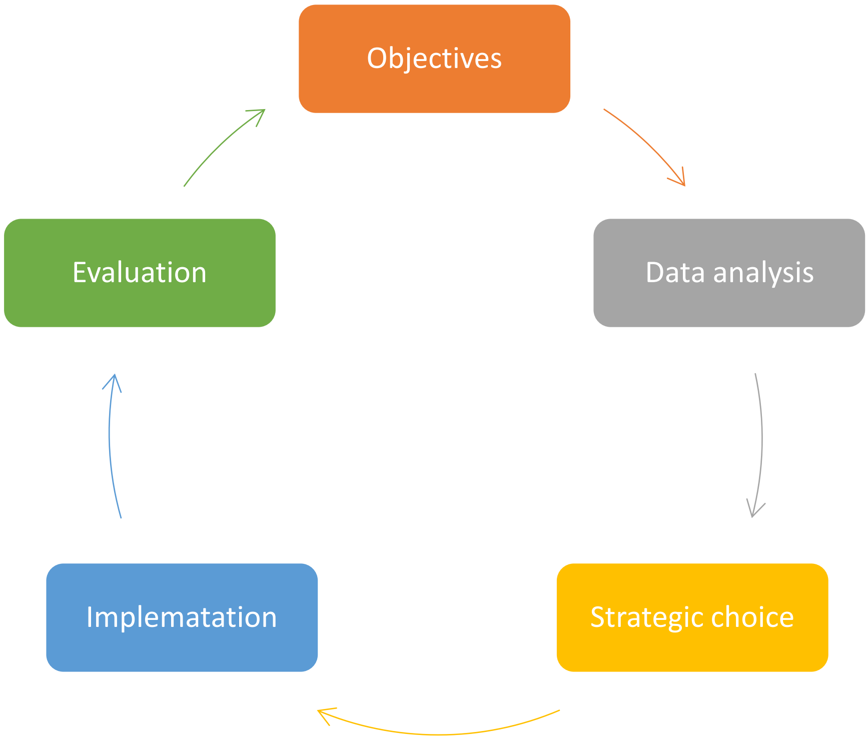

Figure 1.2.1 Process of strategic management (Rao, 2009)

Organisations are commonly structured as hierarchical pyramids with top management, followed by middle management, first-line managers, and non-managerial staff. Typically depicted in Figure 1.2.2, it illustrates the three tiers of organisational strategic planning: corporate, business, and functional levels. This inverted pyramid configuration signifies the varying scope of strategies typically addressed at each level (Katsioloudes & Abouhanian, 2009).

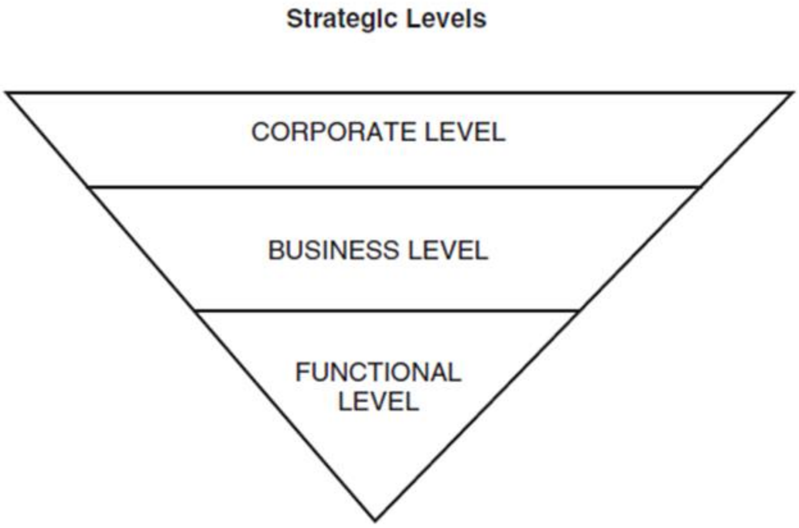

Figure 1.2.2: Pyramid exhibition of strategic levels (Katsioloudes & Abouhanian, 2009)

Thompson et al. (2007) identified two primary reasons why strategy is important in business organisations. Firstly, management is tasked with proactively creating the manner in which the organisation conducts its operations. A well-defined strategy serves as management's directive for conducting business, guiding the organisation towards competitive advantage, satisfying customers, and enhancing financial performance. Secondly, prioritising strategy is more likely to achieve robust financial performance compared to one where strategy is considered secondary to other priorities. Effective formulation and implementation of strategy are shown to significantly enhance revenue growth, earnings, and return on investment.

Correspondingly, strategic management is crucial in road safety management for its ability to establish clear goals, allocate resources efficiently, assess risks, monitor progress, adapt to changing conditions, and promote public awareness. By employing strategic management principles, road safety managers can effectively plan, implement, and evaluate initiatives aimed at reducing road traffic crashes, injuries, and fatalities. A comprehensive approach ensures that efforts are targeted where they can have the greatest impact, ultimately leading to safer road network environments and improved overall road safety outcomes.

1.3 Problem Definition

1.3.1 What are the current problems in road safety strategic management?

Current issues in road safety strategic management stem from high rates of fatalities and injuries, making it a leading cause of death, particularly among children and young adults. The outcome of road traffic collisions exposes the inherent problems within the processes and methods employed for road safety management.

Data fragmentation is a significant issue; existing methods often rely on sectional data analysis, focusing on specific road segments or junctions rather than the entire road network. This fragmented approach hinders the comprehensive understanding of crash databases, which is crucial for developing effective road improvement strategies (Koorey, 2009). Furthermore, there is a lack of strategic integration, evidenced by deficiencies in the uniform application of strategic, tactical, and operational levels of decision-making within road safety management. Additionally, subjectivity and bias in data analysis and resource allocation present challenges, as current methods can be subjective, concentrating on localised countermeasures from expert opinions without adopting an adaptive strategic approach (Varhelyi, 2016).

1.3.2 What are the approaches being used?

In both industry and academia, various methods are employed to execute road safety strategic management. Data mining (Dia et al., 2021; Jain et al., 2016; Kumar & Toshniwal, 2015) and statistical analysis techniques (Karacasu et al., 2014; Ng et al., 2002), including association rule mining and K-means clustering, are used to analyse crash data and identify patterns related to fatalities. Multi-criteria analysis (MCA) (Odoki et al., 2015; J. Yu & Liu, 2012) and optimisation (Byaruhanga & Evdorides, 2022; Saha & Ksaibati, 2016) methods prioritise safety projects by balancing cost and benefit.

Moreover, industrial best practices, such as those advocated by the Federal Highway Administration (FHWA) (Federal Highway Administration, 2010; Gross et al., 2021), recommend using net present value (NPV) and benefit-cost ratio (BCR) techniques to evaluate the efficiency, effectiveness, and feasibility of potential countermeasures. Initiatives utilising risk mapping and iRAP embedded tools (International Road Assessment Programme, 2023; Jane et al., 2011) apply segmentation analysis to identify and rank high-risk road networks, thus providing comprehensive databases for effective safety upgrades. Additionally, the New Zealand Transport Agency (2021a) has introduced three investment criteria to prioritise proposed activities.

These diverse approaches collectively contribute to a comprehensive framework for enhancing road safety and optimising resource allocation. While their advancements are recognised, there remain constraints and areas for potential improvement.

1.3.3 What is needed?

To advance road safety management, it is essential to develop comprehensive methods that integrate diverse datasets, such as crash dataset and road attributes dataset and alternative countermeasures, to provide a holistic view of engineering safety issues. Minimising expert efforts, reducing subjectivity in data analysis with logical data mining, and facilitating objective investment through data-driven solutions in management processes are critical for enhancing strategic road safety. Therefore, the implementation of advanced data mining techniques and multi-objective optimisation becomes necessary. These innovations collectively enable more informed and efficient road safety strategic management, thereby improving overall road safety outcomes.

1.3.4 How will the proposed system address the above?

The proposed system addresses the challenges through an adaptive approach that utilises big data analysis, thereby offering a comprehensive understanding of databases tailored to various legislative objectives, such as fatality analysis, strategic road network analysis, and specific area crash analysis. By employing such methods, the system mitigates bias by uncovering infrequent patterns, resulting in unbiased insights. Furthermore, it enhances strategic resource allocation through a two-stage model that integrates multi-objective optimisation and association rule mining, thus improving decision-making and efficiency. Without subject expert intervention, the system enables local authorities to formulate investment plans that are customised to the specific characteristics of their road networks, ultimately enhancing the overall road safety strategy.

1.4 Research Aim and Objectives

While various frameworks and technologies have been employed for road safety strategic management, there remains a noticeable gap, particularly within the view of engineering, requiring a thorough comprehension of crash databases facilitated by Artificial Intelligence (AI) and Data Science. Additionally, the need for innovative optimisation methodologies to enhance the efficiency of resource allocation based on data-derived insights becomes apparent.

Considering these aspects, this study aims to develop the adaptive solutions to optimise road safety strategic management processes and outcomes.

The objectives of the research may be identified as follows:

- To investigate the role of road infrastructure in reducing road traffic collisions by employing AI and data science techniques and comprehensively interpret the road collision dataset, identifying key factors influencing crash risk and fatalities.
- To apply the proposed data analysis framework to different scenarios evaluating its robustness and applicability in various contexts, encompassing diverse legislation levels, strategic objectives and traffic conditions.
- To optimise resource allocation by leveraging data analytics and optimisation technique providing data-driven solutions.
- To examine the adaptability of models and evaluate their potential impact on road safety strategic management for reducing the incidence and severity of road traffic collisions.
- To advance the field of road safety strategic management by providing robust empirical evidence and developing novel theoretical insights, with a particular emphasis on engineering perspectives.

1.5 Methods and Innovations

The research methodology employed in this study involves two phases of strategic management, of which they are data analysis and resource allocation, aimed at comprehensively investigating the problems of road traffic collisions and optimising safety management processes. Meanwhile, to achieve objectives listed above, a flow map is provided in Figure 1.5.1.

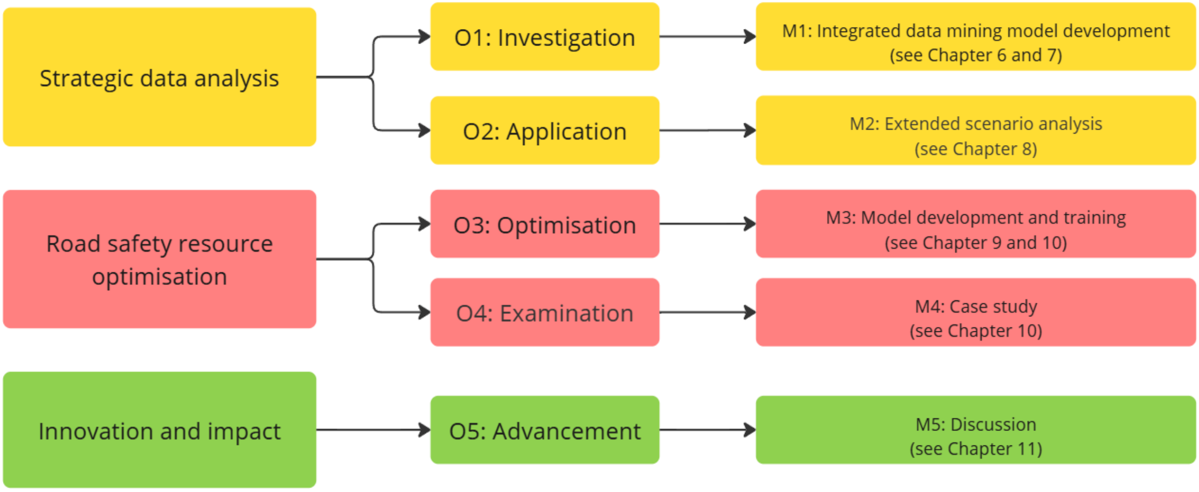

Figure 1.5.1: Flow map of methodology

Initially, the study employs an integrated approach utilising data mining methods, specifically association rule mining and K-means clustering (see Chapter 7), to conducts crash data analysis and reveal patterns and association related to fatalities at crash sites. The UK STATS 19 road traffic collision database (see Chapter 5) serves as the primary data source for prototype development, evaluating the performance of the proposed method in strategic data analysis. The significant innovation is to interpret the results from multiple perspectives and generate evidence-based strategies for reducing road traffic fatalities. This methodology not only contributes to methodological advancements by exploring infrequent patterns in databases rather than relying solely on frequent pattern searches but also validates its adaptability to address diverse objectives at various legislative levels through comprehensive exploration.

Based on the findings of strategic data analysis, a data-driven methodology is proposed, employing multi-objective optimisation techniques. This approach integrates association rule mining with multi-objective optimisation method (see Chapter 10) to allocate funding based on insights from crash data, thereby reducing the reliance on expert intervention in selecting optimal solutions. The implementation of this methodology requires access to various databases, including crash data, road attributes data, and a recommended safer road investment plan (SRIP) (see Chapter 5). To illustrate the effectiveness of the method, a case study of Utrecht is utilised, drawing crash data from the national road crash registration managed by SWOV, and road attributes data and SRIP data obtained from iRAP. The optimisation model demonstrates proficiency in exploring the Pareto front across diverse objectives within budget constraints. Notably, it underscores the practical utility of the model in real-world decision-making contexts and offers empirical support for the efficiency of the non-dominant sorting genetic algorithm in identifying the Pareto front.

1.6 Structure of Thesis

The structure of this thesis encompasses 12 Chapters, each contributing to the comprehensive understanding and exploration of road safety strategic management.

Chapter 1 initiates the introduction by outlining the problem of road traffic collisions and general background of strategic management, while identifying the research objectives, methodologies, innovations, and contributions to existing knowledge.

In Chapter 2, a systematic literature review scrutinises existing methods for road safety strategic management, highlighting insights and identifying gaps in current literature and practices.

Chapter 3 presents the overall methodology highlighting the importance of data and detailing the process of analysing historical collision data and discussing challenges encountered in strategic data analysis. Simultaneously, it explores the data-driven methodology utilising multi-objective optimisation techniques for resource allocation.

Chapter 4 examines the global challenges of road safety and highlights the crucial role of well-designed, well-maintained road infrastructure in minimising accidents. It also reviews strategic initiatives such as the Road Safety Toolkit and Safer Road Safety Investment Plans (SRIP), illustrating ongoing efforts to enhance the effectiveness of road safety management.

In Chapter 5, the datasets considered for prototype development are introduced, providing an overview of various databases, their data collection methods, and applications.

Chapter 6 provides an overview of road collision data analysis, building on previous research related to the application of data mining techniques. It establishes the theoretical framework for the proposed method and assesses various algorithms for association rule mining, aiming to identify the most effective approach for analysing road collision data.

Methodological details are elaborated upon in the Chapter 7, describing the prototype development process, presenting results, and fostering comprehensive discussions on the result interpretation.

Chapter 8 demonstrates the adaptability of integrating association rule mining and K-means clustering for road safety analysis across different governance levels through two scenarios: analysing crash data in London and on the strategic road network (SRN).

Chapter 9 examines the impact of resource allocation on strategic road management, presenting an overview of the practical approaches adopted by various road authorities. It also reviews current optimisation methods used in strategy development and introduces the foundational theory of multi-objective optimisation, suggesting potential research directions for improving resource allocation through advanced optimisation techniques.

Chapter 10 interprets methodologies step by step, aiming to develop a resource optimisation prototype through a case study in Utrecht. An evaluation is conducted to assess the efficiency and feasibility of the proposed methods.

Chapter 11 discusses the implications of the study for road safety strategic management, highlighting how its knowledge discovered align with international evidence on road fatalities. It evaluates the adaptability of the proposed methodology for data analysis and resource allocation, focusing on its application across both stages of strategic management.

Finally, Chapter 12 concludes by summarising key findings and offering recommendations for future research and action within the realm of road safety strategic management.

.

Chapter 2. Systematic Literature Review

2.1 Background

A road network plays an essential role in daily life, and its fundamental objective is to provide safe and accessible services to travellers. However, the number of deaths on roads reached approximately 1.3 million each year, which is unacceptably high. Notably, road traffic fatalities are ranked as the 12th leading cause of death across all age groups (World Health Organisation, 2023). The occurrence of roadway crashes not only poses a tragic threat to the individuals directly involved but also emerges as a significant public health concern, given the ubiquitous use of this essential infrastructure daily. This pervasive concern extends beyond the immediate participants, affecting a broader spectrum of society and emphasising the imperative nature of addressing and mitigating the impacts of road safety on public health.

In the initial stage of society development, road networks were conceptualised and constructed with the aim of establishing a connected societal framework conducive to the commuting and travel needs of individuals. This endeavour reflected a huge demand for the creation of a comfortable and convenient transport system. With the completion of road network construction in developed countries, the focus has shifted from creating new roads to the crucial realm of road management. Roadway crashes are inevitable in the complex transport environment, which result in a devastating toll on human lives, heavy burden on the economy and negative social impact. Consequently, the imperative to prioritise road safety emerges as an elemental consideration during both the initial design and subsequent operational phases of the road network. This underscores the essential need for employing diverse and strategic methods to address and enhance the safety of the overall road network.

In 2020, the United Nations General Assembly on Global Road Safety provided its endorsement, setting a strategic target to achieve a reduction of over 50% in road traffic fatalities and injuries by 2030 (World Health Organisation, 2022). This significant initiative has motivated researchers to study road safety strategic management, which involves the development and implementation of policies, programs, and plans to improve safety on the roads and decrease the occurrence of crashes and injuries. This can be engineering interventions, such as enhancements to road networks and data analysis for the identification and mitigation of high-risk roads. Additionally, measures ranging from educational initiatives to the enforcement of traffic laws are recognised as effective strategies for enhancing overall roadway safety.

Numerous methods and tools are employed to assess road safety. Most of them aim to detect undesirable incidents, predict potential crashes, and identify hazardous locations at specific junctions or sections of road, which provides valuable and useful information for road administration (Elvik, 2010). However, the concept of management functions in terms of strategic, tactical, project and operation level decision making are not fully applied as is the case for general road network management. There is a need therefore to utilise explicitly the concepts of strategic management using appropriate tools and associated data. The practical implication of this shift from road segments to road networks involves a systematic and deliberate effort to integrate strategic considerations into the decision-making processes related to road safety. To learn from existing successful programmes with the view to design enhanced ones, a systematic literature review is carried out. This review seeks to assess and synthesise evidence pertaining to high-level road safety management methods. By doing so, it aims to provide valuable insights and guidance for researchers, practitioners, and policymakers involved in the enhancement of road safety at the strategic level.

2.2 Method

The systematic review is a structured approach to answer a well-defined primary question with the aim of minimising bias and producing more reliable outcomes (Gough et al., 2017). It is a widely used secondary research method conducted by academia on various subjects to identify, scrutinise and synthesise all evidence pertinent to the primary question. The systematic review applies the guidelines outlined by Gough et al. (2017) to validate acquired data and underwent modifications based on the preferred reporting items for systematic reviews and meta-analyses (PRISMA). PRISMA, a reporting guideline crafted for enhanced reporting practices (Page et al., 2021), is employed to refine the review process. This combined approach ensures a comprehensive and standardised methodology in accordance with established guidelines, fostering a more robust and transparent presentation of the research findings.

2.2.1 Definition

A well-defined primary question is essential to systematic review, which defines what subject will be covered and what kinds of studies will be included. PICO process is utilised to formulate primary question as shown in Table 2.2.1. PICO question formulation lies in its role as a structured framework that guides the research process, from question development to evidence synthesis, ultimately contributing to evidence-based practice and informed decision-making (Counsell, 1997).

This review addresses following question:

- What is the evidence supporting the different methods used for decision making in road safety management for strategic purposes?

According to the review questions, the clarification of the research domain and its subsequent targeted exploration require the provision of precise definitions for the terminology employed, as elaborated in Table 2.2.2. This strategic inclusion of definitions aims to ensure conceptual clarity, facilitating a comprehensive and focused examination of the specified subject matter.

Table 2.2.1: Description of PICO

PICO	Description
Problems	Road safety problems include but not restrict different types of crash involving different road users and stakeholders, which have a detrimental impact on public and economic health.
Interventions	Different methods including models and tools etc., are used to improve the safety of road network for decision makers which reduce the number of deaths and mitigate the severity of accident.
Control	Strategic road safety management is a high-level control of road network, particularly focusing on the highway, and develops a strategy to improve the high-level safety not addressing problems of a specific section.
Outcome	All kinds of evidence including academic publications and technical reports gathered to support the road safety strategic management are compiled into a paper to provide a better understanding to whom are interested and develop an innovative method based on the gap from main findings.

Table 2.2.2: Definitions of terminologies

Terminology	Definition
Methods	Any topics related to existing research results or practice are all considered. The methods are not limited to tools, systems and computer software but also include models and AI.
Strategic management	Strategic management is a high-level process involving identifying objectives, analysing data in depth, allocating resources, forming strategies and evaluating implementation (Saloner et al., 2005).

2.2.2 Search strategy

Constructing an effective search strategy is a critical component of a systematic review, ensuring the comprehensive identification of relevant studies. The search strategy involves several key steps:

(a) Selection of Information Sources: Twenty-six bibliographic databases and twenty-four organisational websites (listed in Table 2.2.3) are selected to ensure a wide-ranging search of existing literature on road safety strategic management.

(b) Identification of Search Terms: A comprehensive set of search terms (shown in Table 2.2.4) is identified to broaden the scope of search results. These terms are developed based on the review question and definitions of key terminologies, ensuring an unbiased search for evidence.

(c) Application of Boolean Operators: Boolean operators (AND, OR NOT) are used to refine the search strategy. These operators are fundamental to expanding or narrowing search parameters based on the volume of data needed, thus optimising the search process (Higgins, 2011).

The initial search results are imported into EPPI-Reviewer, where duplicates are automatically removed. The remaining studies are then screened in two stages:

(d) Title and Abstract Screening: Studies are initially screened based on their titles and abstracts to determine their relevance to the review question. Studies that do not meet the inclusion criteria were excluded at this stage.

(e) Full-Text Screening: The full texts of potentially relevant studies are retrieved and assessed against the inclusion criteria. Studies that meet all criteria are included in the final review for quality assessment.

Table 2.2.3: Information sources

Database	**Organisation websites**
ACM Digital Library	Africa Development Bank
Abstracts in New Technology & Engineering	Austroads
ASTM Digital Library	Asian Development Bank
Cambridge Core	Association of National Road Agencies
Compendex (Engineering Village)	Asian Infrastructure Investment Bank
CORE-COnnecting REpositories	Chartered Institution of Highways and Transportation
CIS-Construction Information Service	Development Bank of Japan Inc.
ASCE	Department of Transport
CSA -Cambridge Scientific Abstracts	European Bank for Reconstruction and Development
DOAJ-Directory Open Access Journals	European Union
EBSCO	Federal Highway Administration
Elsevier Science Direct	gTKP-global Transport Knowledge Partnership/Practice
Engineering Handbooks Online	Global Road Safety Facility (GRSF)
Geobase (Engineering Village)	Institution of Civil Engineers, UK
ICE Virtual Library	iRAP-International Road Assessment Programme
IEEE Xplore	IRF-International Road Federation
Inspec (Engineering Village)	REAAA-Road Engineering Association of Asia and Australasia
Knovel	Roads Liaison Group
Oxford University Press Journals	The Organisation for Economic Co-operation and Development

PubMed	Transport Research Laboratory Ltd
Royal Society Journals	Transport Research Board
SAGE Journals Online	United Nations
SciCentral	United Nations Economic Commission for Europe
SpringerLink	World Bank
ProQuest Databases	World Health Organisation
Taylor & Francis online	PIARC-World Road Association
Web of Science	
Wiley online library	

The search strategy is designed to cover various dimensions of road safety and related methodologies. EPPI-Reviewer, a web-based tool suitable for all types of reviews, is employed to facilitate this study. This tool is used for gathering references, removing duplicates, screening studies, coding data, and extracting relevant information based on the inclusion criteria (Thomas et al., 2010).

Table 2.2.4: Search term

Problems	**Interventions**	**Control**	**Outcome**
Road safety	Methods	Strategic management	Research
Accident	Models	High-level management	Projects
Crash	Tools and equipment	Highway/motorway	Manuals
Road deaths	Computer technology	Road network	Reports

2.2.3 Eligibility criteria

A set of inclusion criteria are defined regarding the evidence being sought. All studies will be screened out for relevance based on the criteria given below. These criteria are categorised according to the primary question as defined above.

- Countries: studies from any countries will be included without geographical restriction.
- Road safety: focusing on the safety on highway/motorway.
- Language: all considered studies are likely to be carried out in English.
- Year: studies after 2000 are considered.

Inclusion criteria extend to encompass tools, models, equipment, and methods utilised for the optimisation of both performance and safety within the road network. Additionally, the review considers research endeavours exploring the correlation between crash data and road infrastructure, thereby facilitating informed decision-making by managerial personnel for the enhancement of road safety.

2.2.4 Weight of evidence

A weight of evidence framework is a process of assessing the quality of studies (i.e. evidence) and the extent of their relevance. Even though a study has met all the inclusion criteria, it may not meet the relevance and quality requirements for a review.

In general, the assessment is developed based on three criteria: 1) robustness and quality of execution of the study, 2) appropriateness of the research design, and 3) relevance to answer the question. The framework used for quality assessment of the evidence gathered is given in Table 2.2.5 (Gough, 2007). A study achieving high ratings in at least two criteria was included in the synthesis of the review.

Table 2.2.5: Weight of evidence (WoE)

Robustness and quality of execution	H	The study is transparent and clearly clarifying the purpose. The data analysis and results are accurate and detailed. The overall interpretation is understandable and explicit.
	M	The study is acceptably clarifying the purpose. The data analysis and results are satisfactory. The overall interpretation is not easy to understand.
Appropriateness of research design	H	It covers the whole or partial network of roads or motorways, and it aims at reducing deaths and severity mitigation to improve overall road safety
	M	It covers other types of roads or road sections, and methods are applied to road safety improvement.
Relevance to answer the question	H	It develops a strategic plan or provides a data-based method to support management staff in decision making.
	M	It focuses on the efficiency of tools or methods but does not concern the practice at high level.

2.2.5 Data synthesis

The systematic review conducts data synthesis through a narrative approach, enabling a descriptive analysis of the various methods employed in road safety strategic management. This process involves understanding of existing research and scholarly literature to identify how these methods were practically implemented in strategy development to achieve road safety goals. By systematically collating and analysing relevant evidence, researchers will gain valuable insights into the strengths and limitations of existing methodologies, as well as identifying areas where specialised knowledge is lacking.

This process empowers researchers to pave the way for future research initiatives that can lead to more effective road safety strategies and interventions. The complete search process is illustrated in Figure 2.2.1.

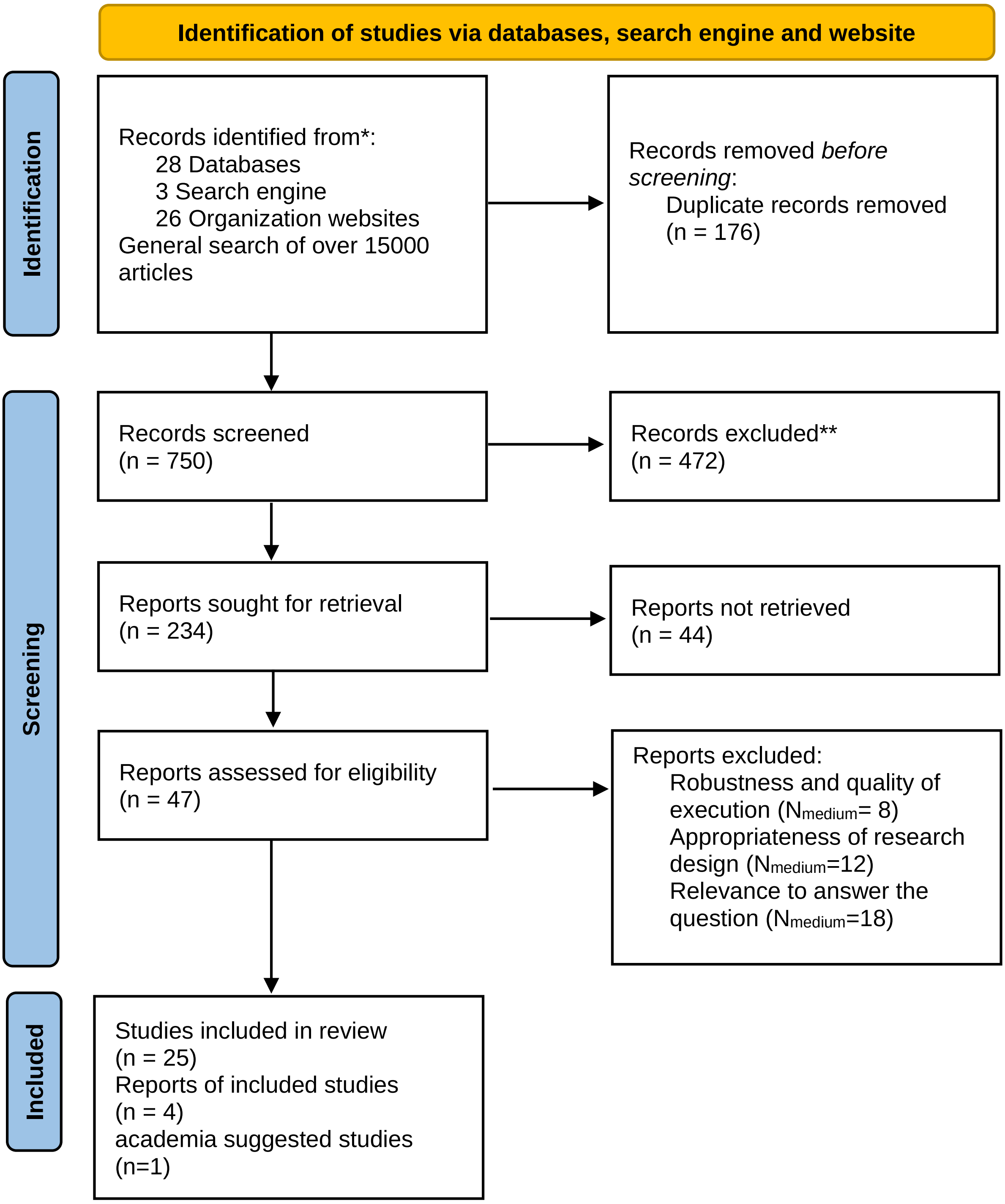

Figure 2.2.1: Searching process

2.3 Research Findings

According to the 30 articles identified shown in Table 2.3.1, there are different methods developed for strategic road safety purposes, with several of them having practical applications. These methods include data mining, statistical analysis, multi-criteria models, data envelopment analysis (DEA), and optimisation techniques. Most of methods were implemented for crash data analysis and resource allocation for strategy development. This section demonstrates how these various methods can be employed to achieve proficient strategic road safety management, ultimately leading to the reduction of road traffic collisions.

Table 2.3.1: Summary of 30 reference identified

Data mining	Jain et al. (2016), Castro & Kim (2016), Maji et al. (2018), Saranyadevi et al. (2019), El Mazouri et al. (2019), Dia et al. (2021), Momeni Kho et al. (2021), El Abdallaoui et al. (2019).
Statistical analysis	Ng et al. (2002), Karacasu et al. (2014), Asare & Mensah (2020).
Multi-criteria analysis	J. Yu & Liu (2012), Odoki et al. (2015).
Data envelopment analysis	Shen et al. (2012), Fancello et al. (2020), Dadashi & Mirbaha (2019).
Optimisation	Chassiakos et al. (2005), Saha & Ksaibati (2016), Byaruhanga & Evdorides (2022).
Others	Y. Han et al. (2011), Hogg & Mathie (2011), Park &

	Young (2012), Coll et al. (2013), Hasan et al. (2021)
Practice	Herbel et al. (2010), Federal Highway Administration (2010), Gross et al. (2021), International Road Assessment Programme (2023), Waka Kotahi NZ Transport Agency (2022), Waka Kotahi NZ Transport Agency (2021b)

2.3.1 Data mining

Data mining is a technique to cope with the occurrence of big data in order to uncover various patterns and discover internal relationships among data items in order to extract knowledge (Azzalini & Scarpa, 2012; Tan et al., 2006). By analysing historical crash data, these methods (summarised in Table 2.3.2) assisted in identifying high-risk areas, contributing factors, and potential countermeasures.

Jain et al. (2016) investigated the data from all the union territories and states of India, analysed 58 attributes of crash to identify the crash-prone reas through K-means clustering. They discovered the crashes contributors in various states of India using decision trees which provided insightful assistance in outlining their potential relationships. In another study, Castro and Kim (2016) employed three data mining classifications – Bayesian network, decision trees and artificial neural networks – to figure out the underlying elements influencing the severity of a vehicle crash and predict crash severity in the UK. The study focused on the relevant nominal attributes sourced from STATS 19 between 2010 and 2012 and took advantage of the Waikato Environment for Knowledge Analysis (WEKA) to apply the data mining process. Maji et al. (2018) adopted a different approach to develop strategies by comparing various regions to benchmark regions. They used hierarchical clustering analysis based on fatal crash percentage of causation factors to group different regions. Three distinct criteria were investigated as follows: road crash causation factors related average fatal crash, average severity index and average severity index growth rate. By comparing the results with predefined threshold values established by experts, safety improvement strategies for regions with poor performance could be derived, drawing insights from regions with outstanding performance.

Furthermore, association rule mining (ARM), a data mining technique to extract knowledge in the form of rules, was applied to Indian crash database (Saranyadevi et al., 2019). To enhance the rule extraction process, they integrated ARM with fuzzy theory and context-free grammar, formulating a recommendation system that aids road administrators in comprehending rules derived from vast amounts of data attributes. In a case study conducted in France, El Mazouri et al. (2019) utilised ARM with multi-criteria aggregation ELECTRE II method to rank different rules from the best to the worst. Crash data collected from the Accident Analysis Bulletin Corporal (BAAC) in 2016 were used in this study. This work examined and addressed national level issues and found a solution to screen out redundant and unnecessary rules.

The ARM with K-modes was developed in the study of Dia et al. (2021) to achieve a strategic level of road safety. They classified crash cases based on recorded variables and extract implicit characteristics' rules for each group so that preventative policies and measures could be developed. Momeni Kho et al. (2021) investigated three years of collision data from eight states in the US (vehicle collisions only) between 2007 and 2009. They created and evaluated the prediction of risk map from three different methods: support vector machine (SVM), artificial neural network (ANN) and random forest (RF). The results showed that RF had a high degree of similarity to the actual risk map. The ARM method was then applied to extract significant and accurate decision rules based on the dataset predicted by the RF model. These studies collectively demonstrate the feasibility of utilising association rule mining by road agencies to define effective road safety strategies and policies.

Table 2.3.2: Summary of data mining

Data analysis			
Input	**Procedure**	**Output**	**Reference**
Crash database	K-means; Decision Tree	Accident-prone areas Dominant factors of accidents	Jain et al. (2016)
	Bayesian Network; Decision Tree; ANN	Prediction of accident severity Influential factors of severity	Castro & Kim (2016)
	Hierarchical Clustering	Identification of benchmark States and union territories	Maji et al. (2018)
	Association Rule with fuzzy theory	Rules related to accident attributes Comparison of attributes	Saranyadevi et al. (2019)
	Association Rule; Multi-criteria aggregation (ELECTRE II)	Rules linking conditions and victims of crash Ranking of rules by importance	El Mazouri et al. (2019)
	K-modes; Association Rule	Clusters of various situations Rules for policy development	Dia et al. (2021)
	ANN, SVM, RF; Association Rule	Prediction of accident severity Extraction of rules by importance	Momeni Kho et al. (2021)
Crash database, crowdsourcing data	PCA, K-means; Association Rule	Rules for strategy development Information for report	El Abdallaoui et al. (2019)

Due to the large amount of data being investigated into different aspects of road safety management, which makes data analysis complicated and nearly impossible. El Abdallaoui et al. (2019) were inspired to propose a system for big data gathering and analysis. The scope of databases extended beyond police reports to encompass information sourced from crowdsourcing data concerning the individuals involved in accidents. To streamline the data, a process of data cleaning and pre-processing was conducted using principal component analysis (PCA). Subsequently, K-means clustering was utilised to classify accidents according to records, while association rule mining was employed to identify the most valuable and common attributes within each cluster. The overall process could provide preventative strategies and statistic and prediction reports for road safety administration.

2.3.2 Statistical methods

Statistical analysis is an alternative approach to interpret and describe data to explore hidden patterns and trends (Ott & Longnecker, 2010; Peck et al., 2020), which are crucial in understanding the relationships between various factors influencing road safety. They aid in identifying risk factors and guiding decision-making processes where three studies are screened out and summarised in Table 2.3.3.

Ng et al. (2002) established an accident risk assessment algorithm to estimate crash events and determined the factors that have the greatest impact on crashes for high-risk zone identification. They used the SPSS (statistical software) to conduct clustering analysis on crash data. In order to explain and estimate crashes occurrence resulting from the probable contributing factors, negative binomial (NB) regression models were developed, and the coefficients of the NB models were obtained using the maximum likelihood estimation method. Based on the observed crash data and contributory factors, the risks of zones could be assessed through empirical Bayes (EB).

There were two other studies focusing on applying regression models to crash severity analysis, which had policy and practical implications for road administrators in the long run. Karacasu et al. (2014) studied injured and non-injured types of crashes and investigated their contributing factors by using logistic regression. Discriminant analysis was developed to classify the crash types and determine the prominent factors. Asare and Mensah (2020) also presented an ordinal logistic regression to study the factors which affect the severity of crashes.

Table 2.3.3: Summary of statistical method

Data analysis			
Input	**Procedure**	**Output**	**Reference**
Crash database	Clustering analysis; NB; regression models; EB	Factors on accidents High-risk zones	Ng et al. (2002)
	Logistic regression; Discriminant analysis	Significance of variables Function for classification of accident type	Karacasu et al. (2014)
Crash database	Ordinal logistic regression	Significant predictors	Asare & Mensah (2020)

2.3.3 Multi-criteria analysis methods

Multi-criteria analysis (MCA) is an approach to evaluate results, assess performance, estimate impacts, and balance trade-offs so that many alternatives may be identified and compared with the criteria of preference (Converse, 2020; Triantaphyllou & Triantaphyllou, 2000). Two studies are identified and summarised in Table 2.3.4.

Table 2.3.4: Summary of multi-criteria analysis methods

Resource allocation			
Input	**Procedure**	**Output**	**Reference**
Indicators	Fuzzy-AHP Model	Determination of weights and ranking score	J. Yu & Liu (2012)
Performance index	HDM-4; Utility-cost ratio; NPV-cost ratio	Ranking of projects	Odoki et al. (2015)

(Yu & Liu, 2012) proposed a multi-criteria model for project prioritisation in highway safety enhancement based on a novel analytical hierarchy process with fuzzy logic, which takes into account a number of indicators, concerning technical, economic and social criteria and weighs them properly for decision making. Fuzzy scaling was developed by employing two fuzzy functions to classify indicators into two types: ‘‘the-lower-the-better’’ and ‘‘the-higher-the-better’’ based on the characteristics of each indicator. Comparison matrix was constructed by using standard deviation to determine the relative importance of pair-wise indicators. Through a non-linear optimisation method to obtain the weights of criteria, the final ranking of each candidate project can be synthesised.

Odoki et al. (2015) also considered the economic, environmental, and social impacts of various investment alternatives. The utility/cost ratio technique was employed in this research in order to propose an MCA methodology to enable road agencies to prioritise and optimise road network investments under different budgets. This method expanded the application of high-level resource allocation and programming to the network level.

2.3.4 Data envelopment analysis (DEA)

DEA is a method for identifying the performance of decision-making units (DMUs). DEA has found application in road safety management to assess the performance of different regions or jurisdictions (summarised in Table 2.3.5). By comparing inputs and outputs, it helps identify best practices and areas requiring improvement (Cooper et al., 2011; Ramanathan, 2003). With the growing number of indicators being created, a composite road safety index is developed for evaluation.

Shen et al. (2012) developed a DEA-based road safety model (DEA-RS) which aimed at minimising output levels given input levels. They defined input as indicators of exposure: the population size, the number of registered vehicles, distance travelled, and output as the number of road fatalities to assess the overall risk situation in terms of road safety. Road safety performance of 27 European Union (EU) countries was ranked by the efficiency scores, and it enabled to identify the benchmark for improving the operations of those underperforming countries.

Table 2.3.5: Summary of data envelopment analysis

Data analysis			
Input	**Procedure**	**Output**	**Reference**
Indicators of exposure	DEA-RS	Risk ranking of countries	Shen et al. (2012)
Average annual daily traffic (AADT), conflict points, social cost	DEA	Score of road sections and urgent improvement need ranking of sections	Fancello et al. (2020)
Resource allocation			
Cost and crash frequency	DEA with Monte Carlo simulation	Ranking of projects	Dadashi & Mirbaha (2019)

Fancello et al. (2020) proposed a DEA-based decision support method to identify which roads most urgently needed safety improvements. The average number of conflict points at junctions and traffic flow were taken into account as inputs in this model, which used the social cost of crashes as its only output indicator. Two different DEA models constant returns to scale based and variable returns to scale based were carried out to validate the rankings of road sections that needed improvements immediately.

Dadashi and Mirbaha (2019) presented a ranking approach that combined DEA with Monte Carlo simulation to consider uncertainty to prioritise road safety improvement projects. Instead of running a DEA model only, a Monte Carlo simulation was used to provide stochastic values as the inputs and outputs and two ranking indicators were established based on the efficiency scores of each DMU.

2.3.5 Optimisation methods

Optimisation methods are used to identify solutions that maximise or minimise a certain objective function, which play a vital role in strategic road safety purposes (Adby, 2013; Ciarlet et al., 1989; McKeown et al., 2022). They assist in determining the optimal distribution of resources to maximise safety outcomes within budgetary constraints. Three studies are outlined in Table 2.3.6.

Regarding resource allocation, Chassiakos et al. (2005) proposed a computerised system to provide road agencies with valuable suggestions based on crashes data characteristics for setting improvement priorities and likely countermeasures. A priority list was developed by index of accident and fatality rates as well as repetitive occurrence of a particular accident type. A weighted-objective linear optimisation was employed to assist decision makers in justifying safety improvement alternatives.

Table 2.3.6: Summary of optimisation methods

Resource allocation			
Input	**Procedure**	**Output**	**Reference**
Performance index	Weighted-objective Linear Optimisation	Maintenance priority list	Chassiakos et al. (2005)
Countermeasures, cost and CRF	Linear Programming	Best combination of safety improvements	Saha & Ksaibati (2016)
Countermeasures, cost and benefits	Mixed Integer Linear Optimisation	Prioritisation of safety improvements	Byaruhanga & Evdorides (2022)

Saha and Ksaibati (2016) developed a traffic safety management system (TSMS) that sought the ideal combination of safety enhancement initiatives that would reduce crashes and benefit society. In order to minimise the overall predicted crash, particularly fatal-and-injured crash, on the segments with high traffic volume, linear optimisation was used to determine the best combination of safety improvement projects based on the cost and crash reduction factor (CRF) of each countermeasure.

Byaruhanga & Evdorides (2022) applied mixed integer linear optimisation to determine the most reasonable expenditure between capital and maintenance works. It enabled to maximise the economic benefits created from the number of all crashes saved including property damage only (PDO) crashes under a constrained budget. They considered the effectiveness of countermeasure alternatives over their life span together with budget constraints to achieve high-level safety plan development.

2.3.6 Other methods

Building on the commonly used approaches discussed earlier, Table 2.3.7 highlights additional methods that contribute to road safety evaluation. These methods incorporate a range of techniques and tools designed to assess road safety performance, analyse risk factors, and optimise resource allocation.

Han et al. (2011) established a high-level index system and developed a road safety assessment method for regional road safety. They employed six representative indicators: fatality per 10 thousand vehicles (U1), traffic fatality (U2), fatality per 100 thousand population (U3), fatality per million passenger transport (U4), accident rate per hundred kilometres (U5), accident rate per million tons freight transport (U6). These indicators were merged into a comprehensive road safety index using a combination of analytic hierarchy process (AHP) and fuzzy logic theory.

Table 2.3.7: Summary of other methods

Data analysis			
Input	**Procedure**	**Output**	**Reference**
Evaluation index	AHP Fuzzy theory	Weight of each evaluation index Road safety index	Y. Han et al. (2011)
Crash database, Vehicle miles Working hours	Safety Risk Model	Standard report of various dimensions Risk analysis	Hogg & Mathie (2011)
Crash database	Overdispersion test; Beta-binomial test	Ranking of each jurisdiction	Park & Young (2012)
Crash database, indicators	Non-linear aggregation method	Composite road safety index Identification and ranking of policing areas	Coll et al. (2013)
Resource allocation			
Raw crash data, project limits, cost and treatment	Phase 1: treatment algorithm Phase 2: sorting, identification and ranking algorithms	Calculation of safety improvement index List of projects recommended to improve	Hasan et al. (2021)

Hogg and Mathie (2011) developed a comprehensive safety risk model to enhance road safety decision-making processes. Their model integrated three distinct types of data analysis: standard and bespoke outputs, bow-tie modelling, and what-if analysis to facilitate segmentation and factor scaling. Standard and bespoke outputs provided a foundation of reliable data tailored to specific needs, while bow-tie modelling visually represents potential accident scenarios and their preventive measures. The what-if analysis further aided in evaluating hypothetical situations, enhancing the robustness of risk calculations. This approach not only enabled precise risk calculations but also offered deep insights into various risk factors, ultimately aiding in informed decision-making for road safety.

A supplementary tool called the beta-binomial test was applied to prioritisation of critical areas. Park and Young (2012) demonstrated the practical application of the beta-binomial test in their study, where they used it to rank various jurisdictions based on their safety performance. Their work illustrated how this tool could be instrumental in developing strategic highway safety plans.

Coll et al. (2013) introduced a novel non-linear aggregation method for evaluating road safety, which culminated in the creation of a composite safety performance index. This index aggregated multiple safety performance indicators into a single measure, allowing for a comprehensive evaluation of road safety across different policing areas. The non-linear nature of this method ensured that the index could capture complex interactions between various factors affecting road safety. By using this composite index, authorities could more effectively identify and rank policing areas.

In terms of resource allocation, Hasan et al. (2021) proposed a computational toolkit including sorting, identification, treatment, ranking algorithms to parse crash database. It enabled to identify crash hotspots, suggest proper remedies, and estimate potential safety benefits. They used safety improvement indices (SII), a benefit-cost ratio, to compare project safety benefits to implementation costs.

2.4 Road Safety Strategic Management in Practice

In addition to academic research findings, there are five works created by road-related organisations for the better understanding of the current situation of road safety network management.

Herbel et al. (2010) developed a technical report to describe the Highway Safety Improvement Program (HSIP) and provided a roadway safety management process and strategic approach to improving highway safety under the support of the Federal Highway Administration (FHWA). It is applicable to the United States of America. The authors introduced benefit-cost analysis for countermeasure prioritisation by converting benefits to a monetary value. They advocated the adoption of the net present value (NPV) technique as well as the benefit-cost ratio (BCR) to assess the efficiency, effectiveness, and feasibility of potential countermeasures. According to the report, project prioritisation can be established by using ranking, incremental benefit-cost analysis and optimisation methods.

Subsequently, in 2021, the FHWA produced a comprehensive guide titled "Selecting Projects and Strategies to Maximise Highway Safety Improvement Program Performance" (Gross et al., 2021) to explain in detail how to conduct BCR method in USA. It represented an in-depth examination of the BCR in relation to fatal and severe-injury accidents exclusively, with the aim of methodically optimising the safety impact of countermeasures in terms of mitigating loss of life and severe injuries. It introduced a countermeasure score to represents how much it costs to generate 1% of those potential benefits. Moreover, the guide clarified the limitation and advantages of diverse approaches—ranging from site-specific and systemic to systematic—in order to assist road safety personnel in selecting the most fitting approach for devising optimal strategies.

Apart from the HSIP, the FHWA established a guidance for Strategic Highway Safety Plans (SHSP) and created an implementation process model (Federal Highway Administration, 2010). It was also a data-driven process, and the integration of all types of data was regarded as the fundamental element serving the identification of emphasis areas, crash types and strategy development. Performance management, strategic goals and objectives and emphasis areas were critical components of the SHSP process. It considered all aspects of road safety, including engineering, management, operation, education, enforcement, and emergency service elements, by coordinating different road-related plans from the corresponding agencies, as illustrated in Figure 2.4.1. It assisted state entities in identifying and prioritising their most urgent road safety needs and developing a programme or strategy that had the best chance of preventing fatalities and serious injuries.

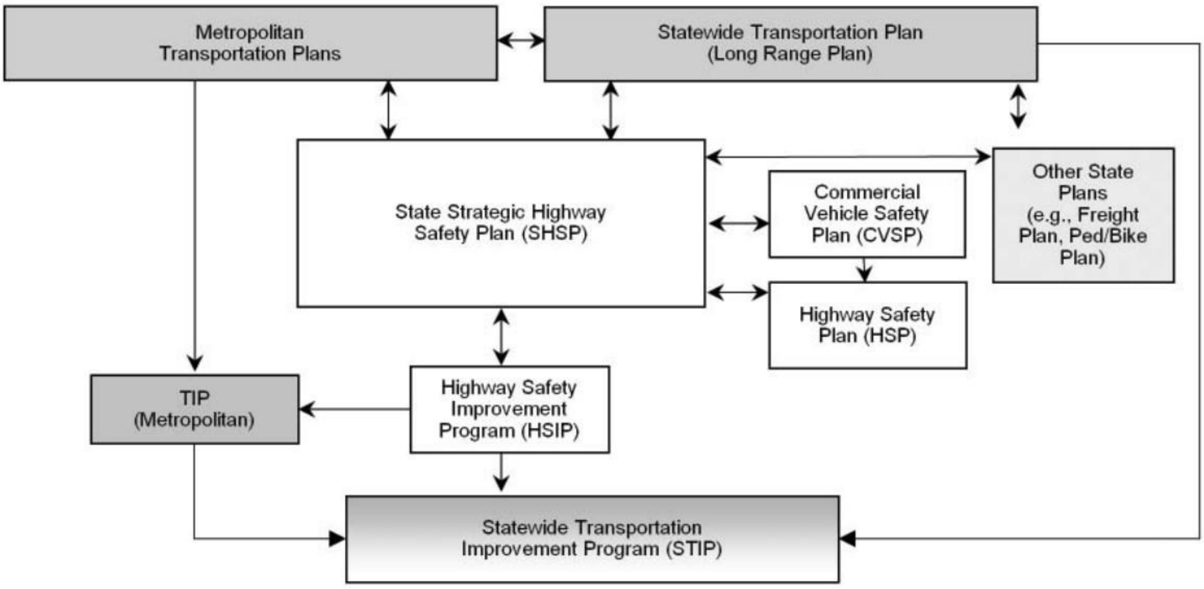

Figure 2.4.1: Coordinated transportation safety planning (Federal Highway Administration, 2010)

International Road Assessment Programme (iRAP) aims at mitigating the risk of road around the world, which is a world standard road safety amassment organisation. The programme introduced two methods, risk maps and star rating maps, which enabled road safety agencies to develop investment strategies based on ranking risk level of road and identify high-potential risk area based on crash related data. Jane et al. (2011) verified the feasibility and suitability of the RAP in Manitoba, Canada. They obtained the international jurisdictional survey regarding the application of RAP and discussed different aspects of RAP implementation in terms of the effectiveness, strengths, weaknesses and so on. Risk mapping methodology is the process of determining the time period of data, road classification, road segmentation and risk level development, and star rating mapping is used to realise overall risk ranking. In addition to Canada, United Kingdom started the programme in 2020 where an enhanced visualisation of risk mapping in London, is shown in Figure 2.4.2.

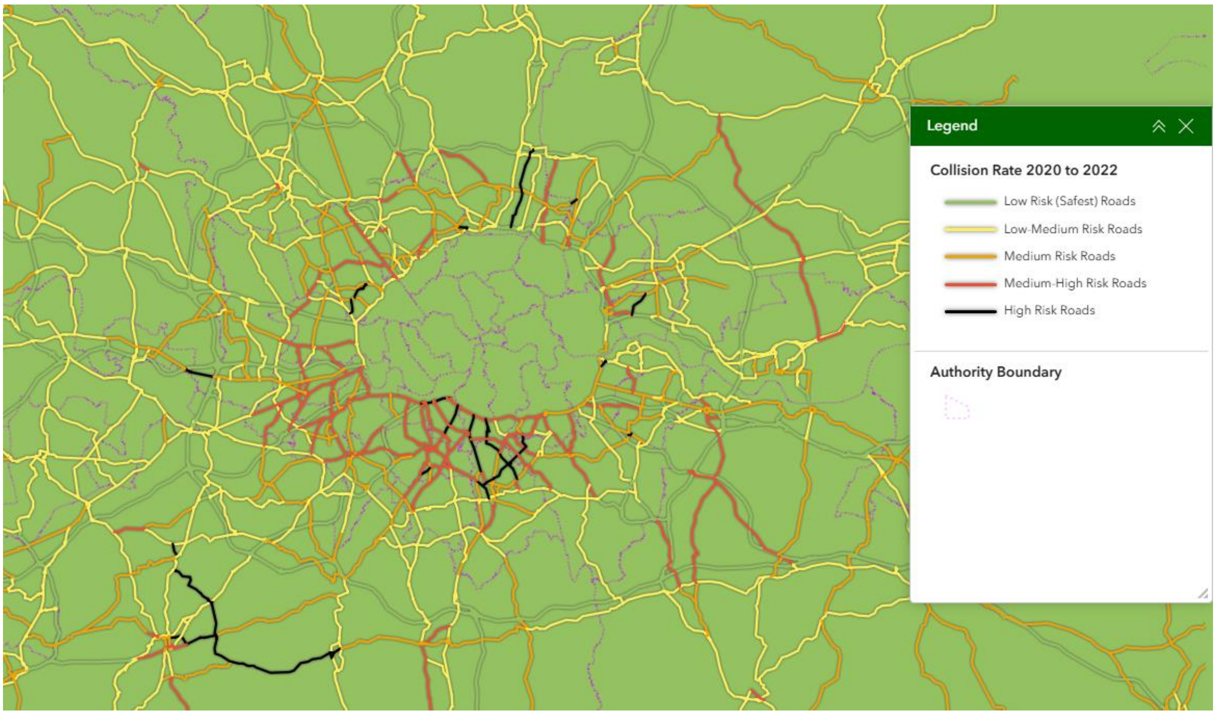

Figure 2.4.2: London Risk mapping visualisation (Road Safety Foundation, 2024)

In addition, iRAP star rating and investment plan was updated by iRAP in 2023 and ViDA web software was developed for safety analyses (International Road Assessment Programme, 2023). A list of safety countermeasures that can improve star ratings and lower the level of risk was called a Safer Roads Investment Plan. It was a crash-based model to appraise safety countermeasure according to benefit cost ratio. More than 90 road treatment options were considered to produce feasible and economically sound investments, which was designed to provide a network-level assessment of risk and cost-effective countermeasures.

The New Zealand Transport Agency introduced a novel One Network Framework, which integrated the movement function, coordinating diverse transportation modes, and the place function, aimed at improving accessibility and travel efficiency in specific geographic areas. This framework served as a comprehensive approach to facilitate strategic and well-informed decision-making (Waka Kotahi NZ Transport Agency, 2022). Furthermore, in the 2021-24 National Land Transport Programme, the agency utilised three investment criteria—government policy statement (GPS) on land transport alignment, scheduling, and efficiency—to prioritise proposed activities. It was an interesting and decisive transport funding allocation framework aligning with strategic objectives. In terms of its alignment with GPS priorities, it assessed how proposed activities correlated with strategic goals and underscores their potential contributions towards achieving these goals. The national land transport programme (Waka Kotahi NZ Transport Agency, 2021b) outlined four strategic priorities—safety, enhanced travel options, improved freight connectivity, and addressing climate change. Additionally, scheduling within this framework denoted the importance and interconnectedness of proposed activities within a program or package. Efficiency indicated anticipated returns on investment, considering comprehensive cost-benefit analyses, with the benefit-cost ratio (BCR) serving as a key metric for evaluation. The prioritisation process involved the application of a matrix to rate the three factors explained above to determine the order of precedence for these activities and all details of rating criteria within three factors were elaborated by Waka Kotahi New Zealand (NZ) Transport Agency.

2.5 Discussion

The main findings of the review describe various methods used for strategy development, categorizable into two groupings. One is data analysis considering different sources of data based on a big data background, and the other is allocation of resources, involving tasks such as project ranking, prioritisation, and investment combinations.

2.5.1 Interpretation and knowledge gaps

Concerning the practical applications of methods in various countries aimed at enhancing strategic road safety, this review identifies a few articles that employed data mining, statistical techniques, and risk modelling to analyse crash-related data. These methods aim to uncover hidden patterns of fatalities, injuries, crash occurrence to develop strategies for improving overall safety of road network.

Data mining has been widely applied to predict the severity of crashes, explore the influential factors of severity, and investigate the contributory factors of crash occurrence. However, the prediction model developed by Castro and Kim (2016) only considered road-related attributes to forecast severity, of which evidence was insufficient for strategic purposes, given the recognised relationship between crash severity, drivers, vehicles, and the environment. Jain et al. (2016) managed to classify different levels of crash-prone states and find their dominant crashes contributors.

Nevertheless, from a strategic perspective, there exists a gap in comprehending the different safety performance of these states under analogous jurisdictional contexts, where Maji et al. (2018) offered an intriguing approach suggesting that an effective safety strategy and policy could be developed by exploring strategies of those benchmarking countries and validating these well-performance strategies in the context of one's own safety situation.

Among the data mining methodologies, association rule mining (ARM) (Agrawal et al., 1993), an unsupervised technique, stands out by creating comprehensive rules based on data characteristics to bolster strategy development. The basic definition and principles of ARM is illustrated in Figure 2.5.1. It looks at data from a grocery store and identifies that if customers buy item A (like ketchup), they are likely to also buy item B (like chips). It used to discover "if-then" rules based on customer behaviour. For instance, in purple rule, if people buy cheese, lettuce, ketchup and beef together, then they are likely to buy bread.

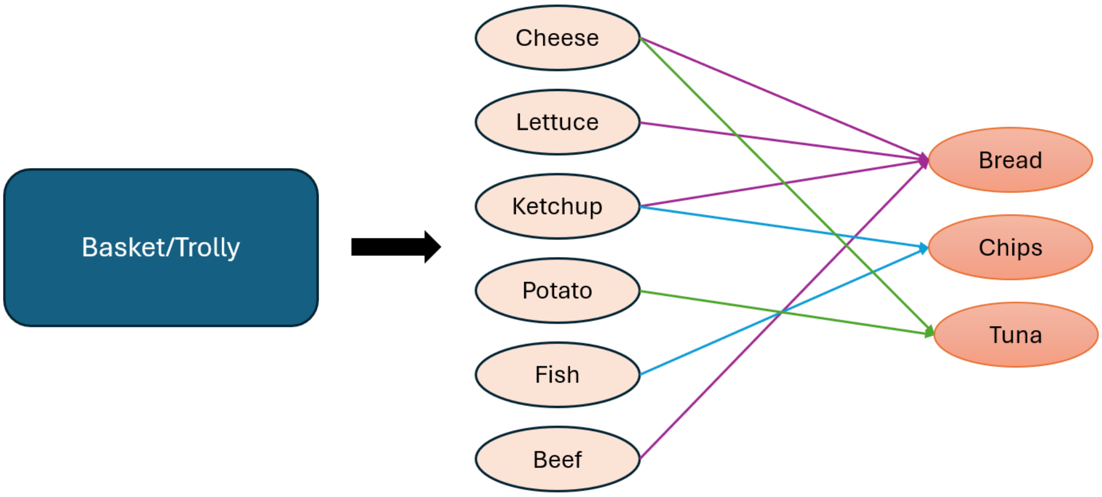

Figure 2.5.1: Illustration of ARM

In road safety, the efficacy of ARM has been validated by Saranyadevi et al. (2019) and El Mazouri et al. (2019), yet a more exhaustive comprehension of the rules, avoiding biases induced by minimal support and confidence thresholds, remains a subject warranting further investigation. It holds significance that infrequent occurrences in crash databases, such as fatal incidents, which are a main concern for road administrators Moreover, obtaining a comprehensive knowledge of a multitude of rules becomes imperative within the road safety context, where integrating a clustering model into the analysis of rules could provide a systematically classified framework, enhancing comprehensive understanding of roadway crashes and enabling a more thorough grasp of their reality.

From a strategic management standpoint, the application of ARM could be extended to extract rules pertinent to fatal crashes, road class and regional areas for diverse strategic purposes. This potential research direction will enable authorities to formulate relevant policies and strategies by thoroughly comprehending the rules related to fatal crashes, road class and regions. This adaptability caters to the varying administrative requisites across different levels.

In terms of resource allocation, various techniques such as multiple criteria analysis (MCA), data envelopment analysis (DEA), and optimisation methods have been employed to prioritise safety enhancement projects. Yu and Liu (2012) and Odoki et al. (2015) utilised MCA, whereas Gross et al. (2021) and Herbel et al. (2010) employed benefit-cost ratio (BCR) to convert unaccountable benefits into monetary value, considering different impacts on projects. It is important to note that while BCR proves inadequate for comparing projects across multiple sites and identifying optimal project combinations, multi-criteria method Odoki et al. (2015) and optimisation approach (Byaruhanga & Evdorides, 2022; Saha & Ksaibati, 2016) were capable of accommodating project combinations. The differences between MCA and optimisation are their focus and their outputs. Multi-Criteria Analysis (MCA) evaluates alternatives by considering multiple criteria, each assigned a specific weight, to rank or identify the most preferred option. In contrast, optimisation focuses on determining the optimal solutions for a defined objective function while adhering to given constraints.

There are only three articles used linear optimisation method to achieve resource allocation. To establish the accuracy and efficacy of the optimisation method, there arises a necessity to formulate alternative optimisation methods that facilitate the pursuit of the most optimal solutions and enable result comparison for multiple objectives.

Concerning the methodologies presently employed within the road safety sector, all of them were considering data analysis on a road-segment level, and resources allocation was being based on benefit and cost analyses of individual countermeasures. Notably, while the US published a strategic highway safety plan (SHSP) with an implementation process model (Federal Highway Administration, 2010), it relied on the safety performance function (AASHTO, 2010) and highway safety improvement program (Herbel et al., 2010) for plan implementations. An intriguing and innovative approach was exemplified by iRAP, which initiated with segmentation analysis and culminates in a comprehensive risk map and a ranking of road networks. It developed a software named ViDA to provide a valuable countermeasures database for professionals engaged in road safety strategic management, aiding them in identifying the most cost-effective road safety upgrades.

While Waka Kotahi NZ Transport Agency (2021) formulated a comprehensive strategic management framework for the transportation system to align with governmental policies, there remains a need to explore methodologies that can assist decision-makers in the automatic development of user-defined investment plans for local authorities within budgetary constraints, utilising diverse datasets such as crash data and road inventory data etc. Data-driven methods for road safety strategy development will provide a set of recommendations for road agencies to reduce human efforts and subjectivity. It is essential to develop a practical procedure or programme for high-level road safety management from data, thus there is a gap of using data mining methods to conduct strategic data analysis to enable a systematic and automated resource allocation approach that caters to the diverse objectives of a strategy.

2.5.2 Limitations of review process

However, there are limitations to every systematic review. Resource constraints may have limited the authors' ability to access a complete set of relevant articles, potentially influencing the overall scope of the synthesis. As mentioned in Figure 2.2.1, in this review there are 44 articles to which the authors have no access due to license constraint. In addition, the subjective nature of the synthesis process is a main concern with regards to reliability and objectivity of the synthesised information as the selective inclusion of studies may introduce inadvertent biases. In an effort to mitigate these limitations, the review diligently employs reasonable inclusion criteria, combines different tools, and follows search strategy to structure a transparent process. In terms of methodology, emerging tools like PRISMA and the critical appraisal skills program (CASP) hold promise for future applications in systematic reviews. These tools, aligning with the principles and theories elaborated by Gough et al. (2017), offer a more detailed and transparent approach in the selection and appraisal processes.

2.6 Conclusion

The objective of this study is to perform a systematic review of the existing methods and proposed potential methods of improving road safety performance within the realm of strategic management and to obtain a wider and more in-depth understanding of their applications. This review set out to address the question:

- What is the evidence of supporting the different methods used for decision making in road safety management for strategic purposes?

In order to answer the question, 30 articles are identified, which demonstrate different methods applied to strategic road safety management, including:

- Data mining
- Statistical analysis
- Multi-criteria analysis
- Data envelope analysis

- Optimisation methods

Additionally, the study identifies practical guidance from road-related entities, which contributes to a better understanding of methodologies for road safety managers. Regarding the countermeasure selection, existing practical methods were restricted to project evaluation based on benefit cost ratio and cost effectiveness analysis. Nonetheless, it is imperative to acknowledge that strategic management involves a higher-level process in the endeavour to mitigate crash occurrences, in contrast to sectional analysis and crash prediction. Consequently, the development of effective strategies needs a stronger foundation of objective data-driven evidence and an automated decision-making tool to construct a safer road network system.

2.7 Summary

The systematic review outlined above serves as a foundational exploration into various methodologies employed in strategic road safety management. It lays the groundwork by identifying key methods such as data mining, statistical analysis, multi-criteria analysis, data envelope analysis, and optimisation methods, all of which contribute to informed decision-making in road safety. The review also underscores the practical guidance offered by road-related entities, providing insights into the application of these methodologies in real-world scenarios. The identified knowledge gaps highlight the need to address subjectivity and bias in data analysis to develop a comprehensive and systematic framework. Additionally, a data-driven approach is required to create an effective and evidence-based investment plan in road safety strategic management. These gaps drive the necessity for this research to contribute to the advancement of road safety strategic management.

Moving forward to the methodology Chapter, this Chapter sets the stage by highlighting the existing methodologies and their applications in road safety management. It encourages a deeper exploration into how these methodologies can be effectively utilised to develop an innovative system. The methodology Chapter will focus on the specific methods selected for detailed investigation to construct an innovative system. Additionally, it underscores the critical importance of objective data-driven evidence and automated decision-making tools in the development of a robust and safer road network system.

In essence, the systematic review acts as a bridge between the introductory discussion of methodologies in road safety management and the detailed exploration that follows in the methodology Chapter.

Chapter 3. Methodology

3.1 Introduction

Building upon the insights discussed in Chapter 2, there exists a persistent need to enhance strategic management practices in road safety by developing a novel approach that minimises reliance on expert intervention, achieves multiple objectives, and offers data-driven solutions for engineering practitioners and governing authorities. Therefore, the goal of this Chapter is to present the methodology adopted in this research, developing an overall optimisation procedure framework. Specifically, the research methodology aligns with two fundamental attributes: Data-Driven Decision-Making and Adaptive Optimisation. These serve as guiding principles in selecting the most appropriate methods for data collection, processing, and analysis.

As this research is primarily quantitative in nature, it systematically applies data science and mathematical modelling techniques to analyse road safety issues objectively. Quantitative analysis provides empirical evidence, reduces subjectivity, and enhances reproducibility, making it particularly well-suited for strategic decision-making in road safety infrastructure management. The integration of data-driven approaches and optimisation techniques ensures that the findings are based on measurable trends rather than expert assumptions, improving the reliability and validity of recommendations.

With the development of data science, data-driven decision-making is introduced to improve the performance based on analysis of data rather than human intervention. Brynjolfsson et al. (2011) conducted research to investigate the impact of data-driven approaches on company performance. Their findings indicate a positive correlation between the adoption of data-driven methods and increased productivity, as well as higher returns on assets, equity, and asset utilisation, along with enhanced market value. Moreover, the study suggests a relationship between the implementation of data-driven approaches and improved company performance, highlighting the diverse advantages of data-driven decision-making.

3.2 Research Design

The research follows a two-stage process, integrating both data mining methods and optimisation method to enhance the reliability and validity of findings. The rationale for this methodology is to integrate the strengths of both approaches to enhance road safety strategic management while minimising the need for human intervention. The key research questions include:

- What are the most suitable and effective data-driven techniques for improving decision making?
- How does the adaptive optimisation process improve the strategic management of road safety in the view of engineer?
- How can the proposed methods be scaled and adapted to different geographic or socio-economic contexts?
- What are the strengths and limitations of the integrated approaches in this context?

For road infrastructure engineers, access to diverse databases is essential for understanding the complexities of the road network, including crash records, road attributes, and inventory details. By utilising the valuable information stored in these databases, data mining methods become powerful tools for gaining a comprehensive understanding of road safety issues. These methods integrate quantitative trends with qualitative insights, facilitating a holistic approach to road safety strategic management.

Moreover, recognising the diverse objectives in selecting countermeasures, an optimisation model is developed to assist decision-makers in identifying the optimal alternatives based on the knowledge extracted from these databases. This approach reduces reliance on expert opinions, enhancing both the effectiveness and efficiency of resource allocation.

The conceptual framework of overall methodology is illustrated in Figure 3.2.1, which outlines the interconnected steps of the procedure. The methodological steps for achieving Strategic Data Analysis which serves as one of the foundational inputs are discussed in Section 3.4, while the steps for Resource Optimisation which draws upon the data analytics are further explored in Section 3.5.

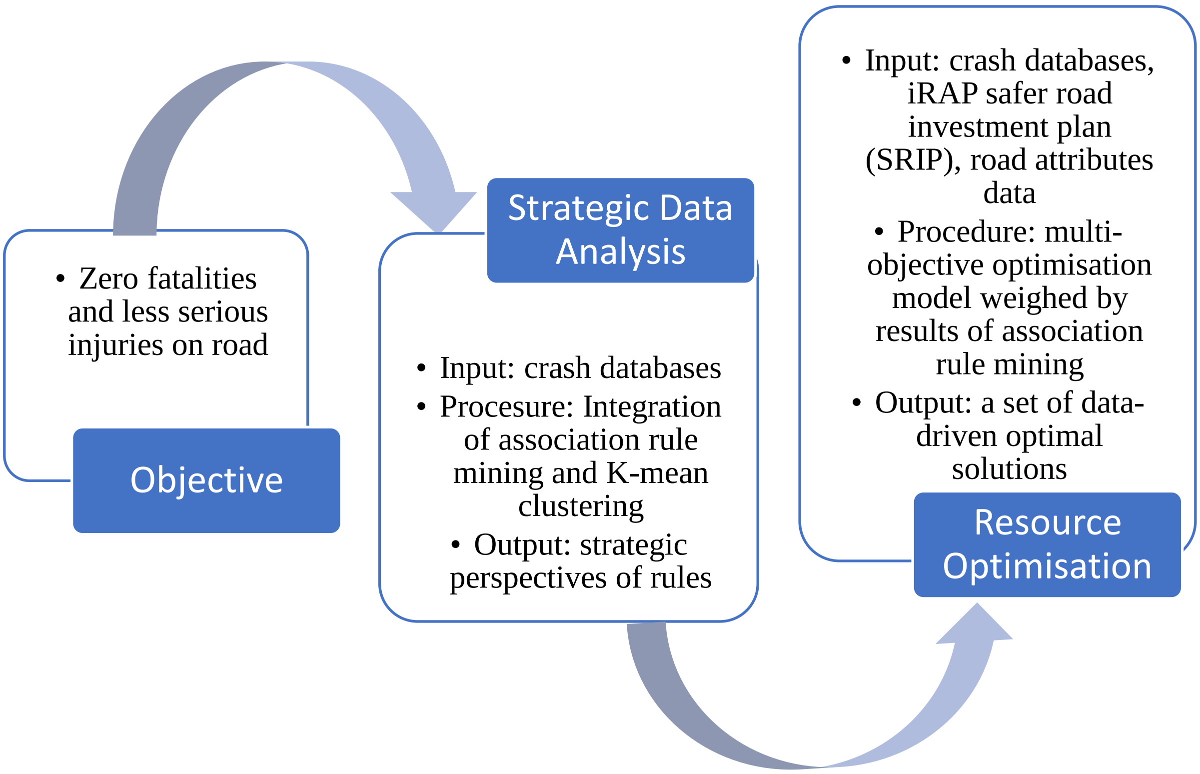

Figure 3.2.1: Conceptual framework of methodology

Overall, this study seeks to demonstrate the positive impact of integrating data mining techniques within road safety strategic management. By navigating through the complexities of data-driven decision-making, a multi-objective optimisation method is proposed to facilitate a more adaptive and responsive methodology for ensuring optimal solutions aimed at improving road network safety.

3.3 The Importance of Data

In contemporary organisational strategies, the role of data has evolved into a critical asset that underpins adaptive optimisation across various sectors. Data-driven decision-making enables organisations to dynamically adjust their strategies in response to changing conditions, customer preferences, and operational challenges (McAfee et al., 2012; Provost & Fawcett, 2013). Similarly, in road safety management, data plays a crucial role in informing engineering decisions and optimising outcomes from safety measures. Data-driven decision-making in road safety management allows engineers and policymakers to formulate their strategies on empirical evidence rather than intuition. By analysing historical collision data, traffic patterns, and environmental factors, decisions can be tailored to address specific safety objectives.

Despite its transformative potential, data presents inherent challenges that can hinder effective adaptive optimisation:

1) Data quality concerns such as inaccuracies, incompleteness, or biases can have significant impacts on both enterprise management and road infrastructure safety. From an enterprise perspective, these issues may result in customer dissatisfaction, increased operational costs, impaired decision-making, and hindered strategic execution (Redman, 1998). Similarly, in the perspective of road infrastructure safety, inaccurate or outdated data pertaining to road conditions, traffic patterns, or maintenance histories can lead to misunderstanding of road safety and inefficient resource allocation.

2) The volume of data generated and collected, often referred to as "big data," poses a challenge in terms of storage, processing, and analysis. The exponential growth in data volume overwhelms traditional data management systems and requires scalable infrastructure and advanced analytical tools to extract meaningful insights efficiently(Manyika et al., 2011).

3) Data comes in diverse formats and structures where countries or authorities have their own registration systems that are not always compatible with each other (European Commission, 2024a). Variety complicates integration efforts and requires sophisticated techniques for data harmonisation and normalisation which is essential for comprehensive analysis and accurate decision-making in adaptive optimisation scenarios.

Therefore, the quality of data is determined the effectiveness of adaptive optimisation. High-quality data ensures that decisions are made by reliable information, minimising the risk of inaccurate conclusions and inefficient strategies. Organisations must invest in data governance practices, including data cleansing, validation processes, to maintain data quality.

The UK and the Netherlands are recognised for their robust road traffic crash data collection systems, which serve as foundational resources for understanding and improving road safety through data-driven approaches. These datasets, namely the STATS 19 crash database in the UK and the BRON database in the Netherlands, are critical in facilitating evidence-based decision-making and adaptive optimisation of strategic management in road safety.

By addressing challenges related to data quality and leveraging datasets from regulated environments like the UK and the Netherlands, it guarantees the exploration of strategic data analysis and resource optimisation.

3.4 Strategic Data Analysis

Central to the methodology employed for data analysis is the utilisation of data mining methods, which serve as invaluable tools for comprehending large and complex datasets. Unlike traditional statistical modelling, data mining not only delves into historical data to discern past trends but also induces logical implications for future occurrences (Azzalini & Scarpa, 2012). Based on the type of knowledge discovery, data mining functions can be divided into unsupervised and supervised algorithms (Chen & Liu, 2004). The former recognises relationships in non-classified data, while the latter requires the data to be pre-classified to explain those relationships.

According to these two main function types, data mining algorithms can be categorised into association, clustering, classification, and regression. Given the inherent challenge posed by categorical data in processing, this study has determined to utilise association and clustering techniques. These techniques are particularly well-suited for dealing with categorical data as they allow for the discovery of patterns and relationships within the dataset without the need for converting categorical variables into numerical ones. By retaining the original form of the data, association and clustering methods ensure that the integrity of categorical information is maintained, enabling more accurate and meaningful analysis. The rationale behind this choice is to preserve the categorical nature of the data while extracting meaningful insights.

To comprehensively analyse annual crash data within a road network, this work integrates association rule mining and K-means clustering techniques to uncover hidden insights. Given the potential drawbacks of simple parameter setting constraints in rule mining, the frequent pattern search method is adapted to identify infrequent patterns, thereby catering to the requirements of infrequency of fatalities and serious injuries in road safety analysis. This approach mitigates bias and subjectivity stemming from expert interference. The STATS 19 dataset, renowned for its reliability and official status in UK road safety, serves as the primary data source, offering a substantial volume of crash-related information. The association rule mining algorithm is then employed to extract rules based on recorded characteristics of crash sites. Subsequently, the results of rule mining are applied to K-means clustering to group rules according to measures of interestingness, facilitating an overview of the dataset. Finally, visualisations are employed to comprehensively understand the revealed knowledge within the database. This integrated methodology, as illustrated in Figure 3.4.1, provides a robust framework for analysing and interpreting road crash data.

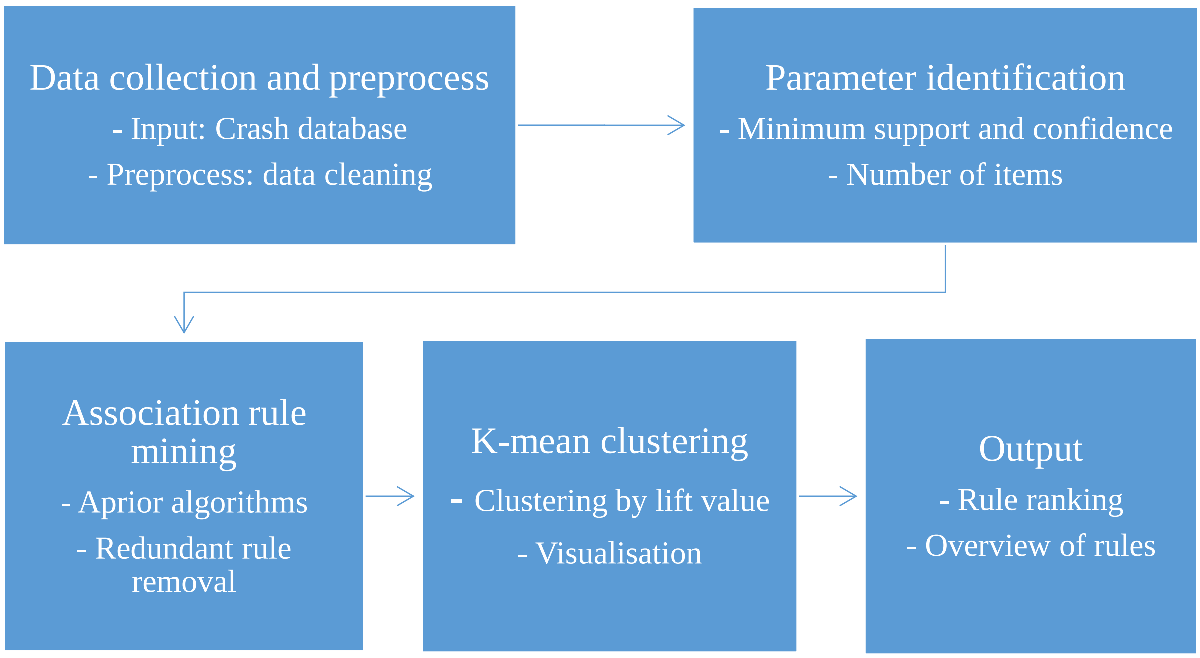

Figure 3.4.1: Strategic data analytics procedure

3.5 Resource Optimisation

The crucial component of strategic management involves developing and implementing an investment plan to achieve the strategic objectives set by stakeholders, rooted in data analysis as demonstrated in previous section. In the realm of road safety, the investment plan entails the identification, selection, combination of countermeasure for road infrastructure improvement, aiming to enhance the overall safety of road network.

To maximise the benefits derived from these investments, optimisation methods play a vital role. These methods, grounded in mathematical optimisation, are utilised to search for the best or optimal solutions within defined constraints. This systematic approach allows for the evaluation of various factors, such as cost-benefits, cost-effectiveness, feasibility, and impact, to identify the most efficient allocation of resources. Consequently, investments are directed towards interventions that yield the greatest improvement in road safety within the constraints of available resources. By leveraging optimisation techniques, stakeholders can make informed decisions about resource allocation, intervention prioritisation, and fund allocation to achieve the most significant impact on road safety.

In the context of decision-making, the consideration of divergent objectives poses a challenge, as single-objective optimisation may overlook critical aspects of different objectives, leading to practical dilemmas. To address this challenge, a data-driven approach is proposed for road safety authorities, utilising multi-objective optimisation techniques. This method aims to maximise the value derived from road safety enhancements while adhering to budget constraints. The proposed approach integrates association rule mining and optimisation methods to allocate funding based on insights extracted from crash data, minimising the need for expert intervention in seeking optimal solutions. The methodology comprises two main steps: firstly, understanding the associations between crash contributors and fatalities through association rule mining; and secondly, developing a data-driven multi-objective optimisation model to assist decision-makers in selecting the most suitable solutions from a set of Pareto optimal options.

This approach necessitates access to various databases, including crash data, road attributes data, and a recommended safer road investment plan (SRIP). To exemplify this methodology, a case study of Utrecht is employed. Crash data from Utrecht is sourced from the national road crash registration (BRON) managed by SWOV, the national scientific institute for road safety research in the Netherlands. Road attributes data and SRIP data are obtained from iRAP. The process of methodology is illustrated in Figure 3.5.1, which includes all data and tools that the author considered for application to improve road safety.

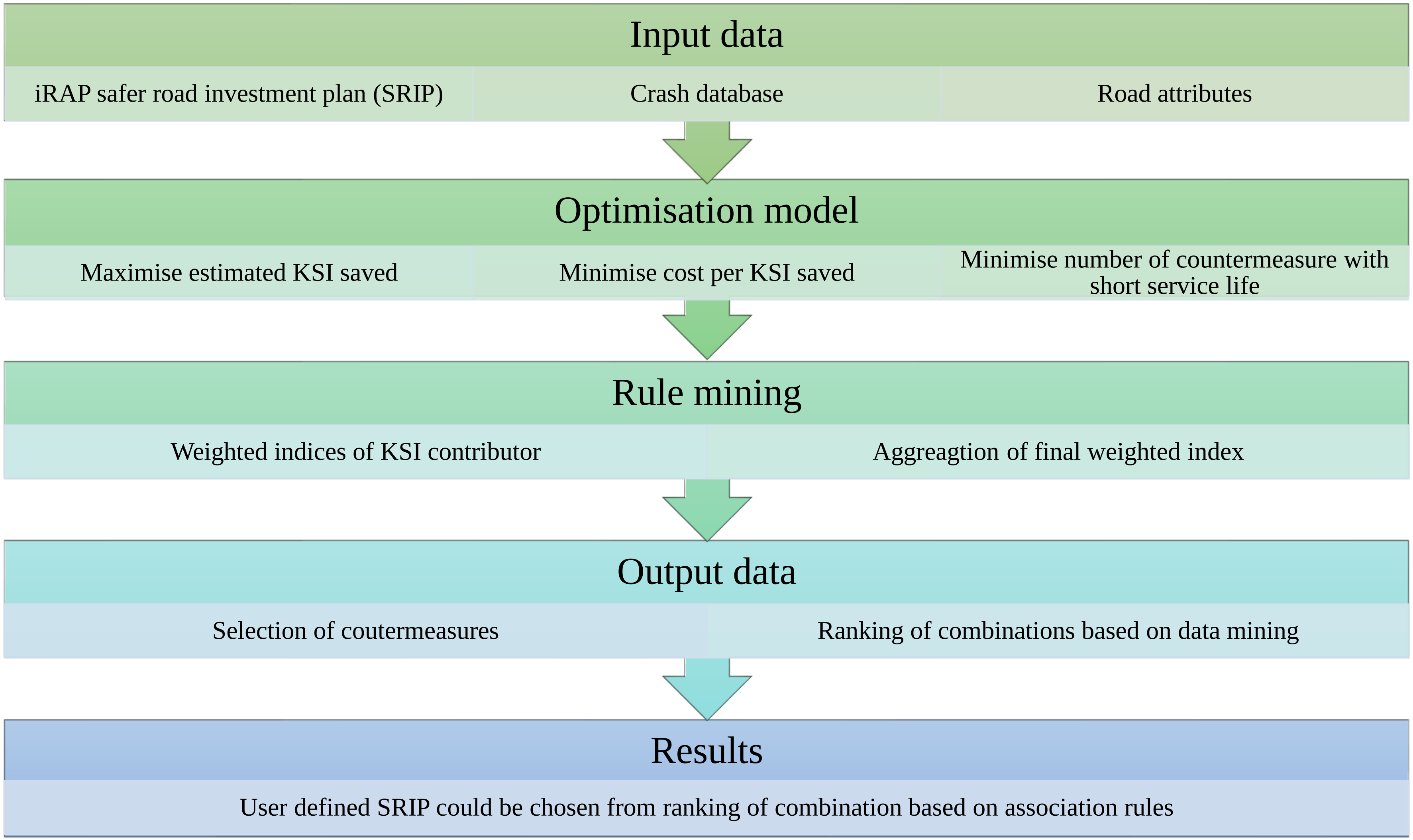

Figure 3.5.1: Multi-objective optimisation procedure

3.6 Strengths and Limitations

This research adopts a data-driven decision-making approach to enhance objectivity and reduce human biases in road safety management. Through adaptive optimisation, strategies can be dynamically adjusted in response to emerging data trends, ensuring a flexible and responsive framework. The integrated-methods approach facilitates a comprehensive evaluation by combining quantitative rigor (optimisation) with qualitative insights (data mining), resulting in a more holistic analysis. Additionally, Furthermore, the integration of multiple data sources enhances the robustness of findings by leveraging diverse datasets. By prioritising fatal severity, the study focuses on the most critical aspects of road safety, enabling more effective intervention prioritisation.

However, this research is subject to certain limitations, primarily related to underlying assumptions. It assumes that the collected data accurately represents real-world road safety conditions, that data integration from various sources is feasible without introducing inconsistencies, and that the chosen analytical methods effectively capture relevant correlations. In practice, data availability and quality vary significantly across different sources and countries, posing a major challenge, particularly in low- and middle-income countries where crash data may be insufficient or unreliable for data-driven decision-making. Additionally, the study primarily focuses on crash report analysis, potentially overlooking other critical dimensions of road safety, such as behavioural and vehicle-related factors. The integration of multiple datasets further introduces complexity and potential compatibility issues, which may impact the effectiveness of the proposed approach.

3.7 Summary

This chapter outlines a data-driven framework for optimising road safety strategic management by leveraging diverse datasets and minimising human bias. The methodology integrates data mining and optimisation techniques to enhance decision-making and achieve multiple safety objectives. Key data sources include crash databases (e.g., UK STATS19, Netherlands BRON), road attributes data, and SRIP data, which provide basic information for safety interventions and resource allocation.

In summary, the purpose of this methodology section is to:

- Explain the rationale behind adopting a purely data-driven decision-making approach, minimising human bias and influence.
- Justify the selected approach for addressing the posed research questions, particularly the use of adaptive data mining and optimisation techniques.
- Outline the procedures used to optimise the process and the results of road safety strategic management.
- Briefly discuss key methodological considerations and challenges encountered during the research process.

Data analysis techniques form a central part of the methodology. Data mining methods, including association rule mining and K-means clustering, are employed to uncover patterns and relationships within the crash data. Association rule mining extracts rules based on crash site characteristics considering infrequent patterns. K-means clustering groups the extracted rules according to measures of interestingness, facilitating a comprehensive understanding of the dataset. Visualisations are used to interpret and present the results of data mining, aiding in the communication of findings.

Resource optimisation is another crucial component. Optimisation methods are essential for developing investment plans that maximise road safety improvements. These methods evaluate factors such as cost-benefit, cost-effectiveness, and public impact. A multi-objective optimisation approach addresses the conflicting objectives of road safety interventions, using insights from data mining to allocate resources effectively within budget constraints. This approach is exemplified using a case study of Utrecht, where crash data from BRON and iRAP are integrated to improve road safety.

The overall process follows a structured approach: data collection, data analysis using mining techniques, clustering of results, and optimisation of resource allocation. Each step is designed to enhance the understanding of road safety issues and inform strategic decisions. This methodology aims to demonstrate the positive impact of integrating data mining and optimisation techniques in road safety strategic management. By employing a data-driven approach, it provides a more adaptive and responsive framework for identifying and implementing effective safety measures, ultimately contributing to the improvement of road network safety.

Having established the innovation of a data-driven framework for enhancing road safety, it is essential to understand the foundational aspects of road safety and the importance of road infrastructure. The next Chapter will provide a comprehensive background on road safety, highlighting the importance of road infrastructure in mitigating risks and reducing the incidence of traffic accidents.

Chapter 4. Road Infrastructure Safety

4.1 Introduction

This chapter builds upon the methodology established in the previous section, presenting a data-driven methodology for optimising road safety strategic management. While the methodology outlined key approaches for leveraging data mining and optimisation techniques, this chapter shifts focus to the role of road infrastructure in enhancing road safety outcomes revealing the importance of this research in contributing to the enhancement of road safety, particularly from an engineering perspective.

This Chapter explores the global road safety crisis, discussing its widespread impact and the essential role that road infrastructure plays in mitigating traffic-related risks. The alarming statistics on road traffic fatalities and the substantial economic impact of traffic crashes worldwide underscore the urgent need for well-designed and properly maintained road infrastructure to prevent road collisions.

Furthermore, the section introduces the concept of "forgiving" and "self-explanatory" roads, which aim to reduce the likelihood of driver errors and enhance overall safety. It also explores the critical role of infrastructure-related countermeasures, such as those outlined in the Road Safety Toolkit and iRAP, to guide the selection of effective safety interventions. The process of developing Safer Road Safety Investment Plans (SRIP) is discussed in detail, showcasing how these plans prioritise safety measures and allocate resources to reduce fatalities and injuries.

4.2 Global Road Safety

Road safety is a critical issue affecting both developed and developing nations globally. According to the World Health Organisation (WHO), there were approximately 1.19 million road traffic fatalities in 2021, equating to a death rate of 15 per 100,000 people. Economically, the impact is significant, with countries losing an estimated 3% of their gross domestic product due to road traffic crashes (World Health Organisation, 2023).

Without new initiatives, forecasting suggests that over the first half of the 21st century, the world could witness more than 50 million fatalities and 500 million serious injuries from road traffic accidents (Bhalla et al., 2008). In comparison, the likelihood of more than 40 million people being killed by mega-wars or a severe influenza epidemic over the same period is estimated at only 1%, and approximately 4 million deaths are expected from volcanic eruptions or tsunamis (Smil, 2011).

Economic growth is a major driver of increased mobility and motorisation. It is anticipated that in the first three decades of the 21st century, global motor vehicle production will exceed the total output of the previous century. The majority of these vehicles are expected to be used in low- and middle-income countries (LMICs) (Bliss & Breen, 2012). However, with the rapid increase in motorisation rates in LMICs comes a significant rise in deaths and disabilities, where 92% of global road traffic fatalities occur in these countries (World Health Organisation, 2023). The difference in road safety performance between high-income and middle-lower-income nations is estimated to widen further. By 2030, it is predicted that about 96% of global road traffic deaths will occur in LMICs, compared to just 4% in high-income countries (HICs).

4.3 Road Infrastructure

Road infrastructure encompasses all physical components that support and facilitate the movement of vehicles, pedestrians, and cyclists across a network of roads. It includes various structures and features crucial for ensuring safe and efficient transportation. The road infrastructure significantly influences the risk of traffic crashes by shaping how road users perceive and interact with the road environment. Roads provide guidance and instructions to users, influencing their behaviour and reducing or increasing crash likelihood based on design and maintenance quality (Ahmed, 2013).

Safer roads are one of the five pillars of the Global plan Decade of Action for Road Safety 2021-2030 (World Health Organisation et al., 2021), of which pillar emphasises the need to raise the inherent safety and protective quality of road networks for the benefit of all road users. This goal involves implementing improved safety-conscious planning, design, construction, operation and maintenance practices to benefit all road users. Research by Pembuain et al. (2019) highlighted that elements within road infrastructure significantly affect traffic crash risks. Factors such as narrow road widths, non-compliance with safety standards, sharp curves, steep gradients, pavement damage, and inadequate lighting contribute to increased crash rates.

Forgiving and self-explanatory road infrastructure have been advocating to enhance road safety recently concerning road design, maintenance and improvement. A forgiving road is specifically referring to a road network context could prevent or mitigate driving errors, allowing drivers to regain control without causing significant harm or damage (Bekiaris et al., 2011). For example, roads with wide shoulders, clear recovery zones, and barriers designed to reduce injury can prevent fatal collisions. Approximately 25–30% of fatal accidents involve collisions with fixed roadside objects, often resulting from lane or road departure due to driving errors. Thus, the primary objective of forgiving road environments is to minimise crashes to instances of minor injuries or damage, aiming to avoid undesirable injuries and fatalities to the greatest extent possible (iRAP, 2022a).

Similarly, a self-explanatory road is designed to guide road users effectively, thereby reducing the likelihood of driver errors and promoting safe driving behaviours (Bekiaris et al., 2011). For example, well-placed signs, consistent lane markings, and appropriate visibility around curves and intersections all contribute to creating a self-explanatory road environment. This conception ensures that the road layout, signage, and overall roadside environment communicate clear and intuitive messages to users, enabling them to navigate safely and predictably. The goal of a self-explanatory road is to align the road environment with user expectations, encouraging safer speeds, manoeuvres, and interactions among road users (iRAP, 2022b). Consistency in road types and design elements, including layout, road furnishings, and signage, plays a crucial role in achieving this objective.

By focusing on these two principles, road infrastructure can enhance safety by upgrading road network and promoting corresponding countermeasures to create a friendly driving environment that aligns with user behaviour. This approach not only reduces the risk of crashes but also encourages safer driving through thoughtful design. Investing in forgiving and self-explanatory road infrastructure enhancement is vital for improving road safety worldwide.

4.4 Infrastructure Related Countermeasures

The Road safety toolkit is originally the result of collaboration between the International Road Assessment Programme (iRAP), the Global Transport Knowledge Partnership (gTKP) and the World Bank’s Global Road Safety Facility (GRSF), which is a platform providing free information on the causes and prevention of road crashes that result in death and injury. In safer road section, it classifies all countermeasure into five sections: intersections, mid-block, roadsides, vulnerable road user, speed management and traffic calming. Each section contains alternatives with detailed information including benefits, weakness, cost level, service life and potential casualty reduction, exemplified in Table 4.4.1 . It is a valuable source for road engineer and authorities to understand the countermeasures (iRAP, 2022c).

Table 4.4.1: Example of countermeasures (iRAP, 2022c)

	Treatment	**Costs**	**Treatment life**	**Potential casualty reduction**
Intersection	Delineation	Low	1 year - 5 years	10-25%
	Grade Separation	High	20 years +	25-40%
	Roundabout	Medium to high	10 - 20 years	60% or more
	Signalise	Medium	10 - 20 years	25-40%
	…			
Midblock	Delineation	Low	1 year - 5 years	10-25%
	Duplication	High	20 years +	25-40%
	Lane Widening	Medium to high	5 - 10 years	25-40%
	Median	Medium to high	10 - 20 years	60% or more
	…			
Roadsides	Paved Shoulder	Medium	5 - 10 years	25-40%
	Barriers	Medium	10 - 20 years	40-60%
	Hazard Removal	Low to medium	5 - 10 years	25-40%
	Side slope Improvements	Medium	20 years +	10-25%
	…			

Vulnerable road user	Bicycle Facilities	Low to medium	10 - 20 years	25-40%
	Motorcycle Lanes	Medium	10 - 20 years	25-40%
	Pedestrian Crossing – Grade Separation	High	20 years +	60% or more
	Pedestrian Crossing – Signalised	Medium	10 - 20 years	25-40%
	…			
Speed Management and Traffic Calming	Speed limits	Medium	5 - 10 years	25-40%
	Traffic calming	Medium	5 - 10 years	25-40%
	Vertical deflection	Medium	5 - 10 years	25-40%
	Speed humps	Medium	5 - 10 years	25-40%
	…			

iRAP methodology was developed based on road safety toolkit to select appropriate countermeasures. A total of 94 countermeasures can be used in the iRAP standard countermeasure pack of the iRAP model. Users can also access an urban-enhanced pack with 23 additional countermeasures specially designed for urban areas (International Road Assessment Programme, 2023).

4.5 Safer road Safety Investment Plans (SRIP)

An investment plan constitutes a structured list of safety measures prioritised to enhance Star Ratings and mitigate infrastructure-related risks efficiently in iRAP methodologies. These plans are formulated through an economic evaluation where the cost of implementing each measure is calculated and the potential reduction in crash is estimated. It involves detailed engineering data, including road attributes, proposals for safety measures, and economic analyses applied to 100-meter segments across the road network. It is essential to emphasise that investment plans are designed to offer a comprehensive solution to risks at the network level and recommend cost-effective safety enhancements (International Road Assessment Programme, 2023).

The process of SRIP development consists of six steps:

(1) Firstly, triggers, the prerequisite conditions to select countermeasures, of which iRAP has defined more than 300 triggers, are tested for further consideration of investment plan. A trigger must be satisfied before that countermeasure can be considered for a 100-metre segment of road, which is a function of star rating, road attribute and traffic flow. As a very simple example, the trigger for the ‘Improve Delineation’ countermeasure is: if delineation is coded as ‘poor’, then the ‘improve delineation’ countermeasure is applied, regardless of what the road’s Star Rating or volume flow.

(2) Secondly, the number of deaths and serious injuries prevented is calculated.

$$KSI_{saved} = KSI_{before} \times \frac{Star\ rating\ score_{before}}{star\ rating\ score_{after}} \tag{4.5-1}$$

where:

KSI_{before} = number of killed and serious injuries that occur on the road before the installation of the countermeasure, where calibration factors are used to ensure that the total estimated number of killed and serious injuries (KSI) on the network is equal to the actual number of KSI on that network.

(3) Thirdly, benefit cost ratio (BCR) is calculated as follows:

$$BCR = \text{present value economic benefit / present value economic cost} \tag{4.5-2}$$

Any countermeasures with a BCR below the predefined threshold will be screened out.

(4) Fourthly, minimum length, minimum spacing and hierarchy rules are considered by analysis team to define their preference for selecting countermeasures. An example of a minimum length rule is that additional lanes (such as overtaking lanes or 2+1 cross section) must be required for a minimum one kilometre (that is a combined ten 100-metre segments of road) within a section (which is defined during the coding stage) of road before they are considered viable.

(5) Fifthly, the impact of all viable countermeasures is determined at the 100-metre level, and the combined deaths and serious injuries after all viable countermeasures have been applied are calculated for each crash type.

(6) Lastly, the total number of deaths and serious injures prevented by the individual countermeasure at that location is adjusted to account for multiple countermeasures affecting the same crash type. A multiple countermeasure correction factor (MCCF) and adjustment is calculated for each 100-metre segment of road and for each crash type. The factor is calculated as follows:

$$MCCF = \frac{Reduction_{Multiplied}}{Reduction_{Additive}} \tag{4.5-3}$$

$$Reduction_{Multiplied} = KSIs_{before} \times [1 - (1 - ER_1)(1 - ER_2) \ldots (1 - ER_n)]$$

(4.5-4)

$$Reduction_{Additive} = \sum_{i=1}^{n} KSIs_{before} \times ER_i$$

(4.5-5)

Where ER_n indicates the expected reduction of KSI with the n^{th} countermeasures implementation. n denotes the number of countermeasures proposed at a given location.

At the completion of the countermeasure selection process, a final economic analysis for all the countermeasures selected across all roads is undertaken, and SRIP is produced.

4.6 Summary

In summary, addressing the global road safety crisis needs well-maintained road infrastructure supported by strategic and effective countermeasures. Initiatives such as the Road Safety Toolkit and iRAP emphasise the importance of a systematic approach to enhancing road safety. Safer Road Safety Investment Plans (SRIP) illustrate the critical role of selecting and implementing appropriate countermeasures to improve road infrastructure and global safety outcomes. The structured SRIP development exemplifies the comprehensive approach for effective road safety management.

By setting the foundation for understanding the relationship between road infrastructure and safety, this chapter paves the way for the next section, which will focus on the datasets considered for model development. Understanding the data behind road safety strategies is essential for informed decision-making and the development of effective, data-driven safety measures. The precision and dependability of any data-driven approaches are dependent on the quality and comprehensiveness of the data employed. A thorough understanding of the data sources, and types used in the analysis is fundamental for developing robust models that can accurately maximise the impact of various safety measures. This understanding is crucial for guiding strategic decision-making in road safety management.

Chapter 5. Data Sets Considered

5.1 Introduction

Data are raw elements that are measured, recorded, and stored through various means, representing observable phenomena. They comprise elements derived from observation, experimentation, computation, and documentation (Kitchin, 2014). Data can be categorised into quantitative and qualitative forms. Quantitative data primarily consists of numerical values (e.g., age, height, and weight), whereas qualitative data encompass non-numeric formats (e.g., text, images, and music). A data-driven approach to making fact-based decisions as opposed to relying on individual opinions and feeling, requires high-quality of data, as poor-quality data can lead to serious consequences in data analysis, resulting in inconsistency and inaccuracy.

When applied to the domain of road infrastructure safety, the strategic management framework can be delineated as follows: 1) identifying objectives aimed at enhancing road safety; 2) identifying road attributes and their multi-attribute impact on road safety; 3) determining road infrastructure improvements and selecting a combination of various countermeasures; 4) implementing the investment plan; 5) evaluating the outcomes of road safety initiatives.

This project focuses specifically on the first three steps to develop data-driven tools for strategy development. The primary objective of this project is identified to decrease the number of Killed or Seriously Injured (KSI) crashes. Concerning the requirement of data for strategic analysis and countermeasures identification, crash databases are utilised to comprehend existing problems and issues within the road network. Road attribute data are employed to interpret the road context and provide precise information about road infrastructure. Additionally, a recommended Safer Road Investment Plan (SRIP), formulated based on a comprehensive analysis of various countermeasures, serves as a crucial guiding direction for investment decisions.

The selection and utilisation of road safety databases play a crucial role in this study, as they provide the foundation for quantitative analysis and optimisation. Various databases, including STATS19 (UK), BRON (Netherlands), and iRAP, have been assessed to determine their applicability and transferability in road safety research. STATS19 and BRON are highly context-specific, designed to reflect the road safety environments, reporting standards, and data collection methodologies of their respective countries. STATS19 relies on police reports and self-reported incidents, while BRON integrates data from police reports, hospital records, and insurance claims, potentially providing a more comprehensive perspective on crash severity. Conversely, iRAP operates on an international scale, offering standardised methodologies such as risk mapping, star ratings, and safer road investment planning (SRIP), making its approaches more transferable across regions.

The transferability of findings derived from these databases depends on several factors:

- Data collection: differences in police reporting standards, crash definitions, and data availability can influence comparability across different regions.

- Geographic and socioeconomic contexts: variations in road design, traffic laws, driver behaviour, and enforcement affect the applicability of findings in different settings.

- Infrastructure effectiveness differences: the effectiveness of countermeasures in one context may not be directly replicable in another due to differences in infrastructure investment, regulatory frameworks, and enforcement practices.

- Data integration and adaptation: while some data elements, such as collision trends and risk factors, may offer transferable insights, applying findings from one country's dataset to another requires careful calibration to account for local conditions.

By acknowledging these factors, this research ensures that the data-driven methodologies proposed can be adapted effectively to various contexts while maintaining robustness and reliability in road safety strategic management. In the subsequent subsections, all considered databases utilised for research purposes are introduced and elaborated upon in details.

5.2 STATS 19 Crash Database

STATS19 is a UK national database containing detailed information about road traffic collisions. It is maintained by the Department for Transport (DfT) and compiled from reports submitted by police officers who attend road traffic incidents. Self-reported crashes are also included, although with limited information. The quality of STATS19 data can vary depending on factors such as the accuracy of reporting by police officers, the completeness of information provided, and any biases in the reporting process. Despite these variations, STATS19 data provides a comprehensive overview of road traffic collisions in the UK, making it a valuable resource for researchers, policymakers, and practitioners in the field of road safety. It offers a wealth of information that can be used to analyse trends and patterns in road traffic collisions, identify high-risk areas, and evaluate the effectiveness of interventions and policies aimed at improving road safety.

The data collected in STATS19 includes various details classified into four types: collision data, vehicle data, casualty data and contributory factors. The collision data, which form the foundation of this research, includes 36 detailed variables. These variables encompass temporal factors (e.g., date, day of the week, and time), spatial factors (e.g., local authority district, first road class, and road type), and contextual factors (e.g., weather conditions, light conditions, and road surface conditions). By focusing on collision data, researchers can identify patterns and correlations between various factors and road traffic collisions, which is crucial for developing targeted interventions.

5.3 BRON Crash Database

The national road crash registration database in the Netherlands, often referred to as BRON (Bestand geRegistreerde Ongevallen in Nederland), serves as a comprehensive source of data related to road traffic collisions across the country. BRON collects information from various sources, including police reports, hospital records, and insurance claims. SWOV (Stichting Wetenschappelijk Onderzoek Verkeersveiligheid), the Dutch Institute for Road Safety Research, manages the crash database with the mission of contributing to safer road traffic through scientific research. This research aims to address the questions that policymakers and other traffic professionals must deal with.

One of the notable features of BRON is its web platform, which allows users to access a wide range of information related to road traffic collisions. This platform supports data exportation, enabling researchers and policymakers to conduct detailed analyses and generate reports tailored to specific needs. The high quality of BRON data, underpinned by rigorous reporting standards and comprehensive data collection protocols, makes it a reliable resource for developing and evaluating road safety initiatives.

5.4 iRAP

The International Road Assessment Programme (iRAP) is a global initiative dedicated to enhancing road safety and mitigating severe injuries worldwide. It offers a range of methodologies, including risk mapping, star rating, and safer road investment planning (SRIP), to adopt a comprehensive and evidence-driven approach toward ensuring safer road infrastructure for all users. In the process of risk mapping, iRAP considers different data sources involving traffic flow, crash density, crash rate and crash cost. Vida, an online platform developed by iRAP, is fundamental for conducting star rating evaluations through on-site surveys and data coding. Vida facilities the calculation, management, analysis, and presentation of star ratings and SRIP whiles providing in-depth insights into road networks, encompassing their attributes and operational characteristics.

A total of 78 attributes, coded and stored in Vida, are collected during road inspections conducted via road survey and road coding procedures. Road surveys require the acquisition of detailed geo-referenced image data of a road network, while road coding involves recording road attribute categories using survey images. As per SRIP requirements, attribute data are extracted based on the locations or road segments of each proposed countermeasure. Additionally, Vida integrates countermeasures databases from the road safety toolkit, facilitating the development of SRIP. Given that SRIP serves as a crucial data source in the proposed methodology, it forms the basis for decision-makers to develop user-defined investment plan. An example of data with performance of each countermeasure is summarised in Table 5.4.1.

Table 5.4.1: Example of safer road investment plan data

Countermeasure	Length / Sites	KSIs saved	PV of safety benefit	Estimated Cost	Net monetary benefit	Cost per KSI save	life
Additional lane (2 + 1 road with barrier)	15.10 km	80	25660031	20655000	5005031	258188	medium
Bicycle Lane (off-road)	3.50 km	3	1122681	392156	730525	130719	medium
Central hatching	5.70 km	0.6	184481	34173	150308	56955	medium
Central median barrier (1+1)	34.80 km	41	13286137	6288200	6997937	153371	medium
Central median barrier (no duplication)	0.70 km	1	365701	95182	270519	95182	medium
Centreline rumble strip / flexi-post	1.80 km	0.3	108423	19512	88911	65040	low
Clear roadside hazards (bike lane)	1.20 km	0.5	160633	216000	-55367	432000	low
Clear roadside hazards - driver side	0.10 km	0.1	34232	20000	14232	200000	low
Delineation and signing (intersection)	8 sites	0.5	149378	88582	60796	177164	low
Duplication with median barrier	1.20 km	26	8323159	6480000	1843159	249231	high
Footpath provision driver side (>3m from road)	25.40 km	28	9060036	2922040	6137996	104359	low
Footpath provision driver side (adjacent to road)	27.90 km	35	11319681	4388280	6931401	125379	low
Footpath provision driver side (informal path >1m)	4.70 km	1	307139	114747	192392	114747	low
Footpath provision passenger side (>3m from road)	26.80 km	29	9267974	3085368	6182606	106392	low
Footpath provision passenger side (adjacent to road)	44.10 km	52	16843531	6939120	9904411	133445	low
Footpath provision passenger side (informal path >1m)	5.80 km	1	406768	141451	265317	141451	low
Improve curve delineation	0.40 km	0.5	148419	7460	140959	14920	low
Improve Delineation	45.70 km	13	4190892	847446	3343446	65188	low
Lane widening (>0.5m)	0.10 km	0.6	186272	152529	33743	254215	low
Lane widening (up to 0.5m)	1.00 km	2	684333	656756	27577	328378	low
Overtaking lane	0.30 km	1	339394	405000	-65606	405000	medium
Parking improvements	1.50 km	0.3	99666	18900	80766	63000	low
Pedestrian fencing	27.20 km	9	2976955	179606	2797349	19956	medium

5.5 Summary

In conclusion, the integration of high-quality data is essential for developing effective strategies to improve road safety. Accurate data is imperative for making evidence-based decisions in road safety management, as poor-quality data can lead to inconsistencies and inaccuracies in analysis. Each database provides unique insights into road traffic collisions, offering valuable information for analysis and model development.

The strategic management framework for road infrastructure safety comprises several critical steps: identifying objectives, strategic data analysis, determining infrastructure improvements, implementing investment plans, and evaluating outcomes. This thesis specifically focuses on data analysis and investment plan development, utilising crash databases to understand current road network issues and employing road attribute data to provide detailed information about road infrastructure.

The STATS19 database, with its extensive collision data, enables the identification of patterns and correlations essential for targeted interventions. As we transition to the next Chapter, "Strategic Data Analysis using Road Crash Data," the theoretical foundations and analytical methodologies used to interpret these datasets will be explored. This analysis forms the basis for developing robust models to understand the impact of various safety measures, thereby guiding strategic decision-making in road safety management. By introducing relevant background information and examining the theory and methods behind strategic crash data analysis, a comprehensive framework is established for understanding and addressing road safety challenges effectively.

Chapter 6. Strategic Data Analysis using Road Crash Data

6.1 Introduction

Road crashes are a significant public health issue estimated to be one of top ten leading killers of death globally. Each year, approximately 1.35 million deaths occur worldwide due to road crashes involving cars, buses, motorcycles, bicycles, trucks, or pedestrians (World Health Organisation, 2018). From an economic perspective, Chen et al. (2019) calculated global macroeconomic cost of road injuries to be 1.8 trillion dollars (in 2010 USD) from 2015 to 2030, equivalent to a yearly tax of 0.12% on global GDP. Crashes are inevitable in complex operation environment involving the interaction of road infrastructure, drivers, vehicles, and surroundings, but road infrastructure is an essential and controllable element to improve road safety. Therefore, a high-level road infrastructure safety management tool is required to provide strategic insights to road authorities and policymakers.

Building upon the understanding of various sources of road crash data discussed in the previous chapter, this chapter explores the application of data mining techniques, specifically integrating Association Rule Mining (ARM) and K-means clustering, to enhance road safety analysis. It begins by introducing the core concepts of ARM and K-means clustering, detailing their individual applications in road safety analysis. While ARM is widely used for discovering patterns in large datasets, it has limitations, such as sensitivity to parameter settings and challenges in detecting rare but critical events, such as fatal crashes. To address these limitations, K-means clustering is proposed as a complementary technique for refining and grouping crash data patterns.

The integration of ARM and K-means clustering allows for a more comprehensive and nuanced analysis of crash data, enabling the identification of rare but significant fatality patterns while improving the overall visualisation and interpretation of high-risk crash scenarios. This combined approach provides deeper insights into the relationships between road infrastructure attributes and crash severity, supporting the development of more effective road safety management strategies.

Followed by theoretical framework, the chapter reviews related works that demonstrate the application of ARM in road safety studies, highlighting the strengths and weaknesses of various methodologies. The chapter concludes by providing a detailed explanation of the basic definitions and algorithms used in ARM, along with a discussion of clustering methods, including K-means clustering, to justify the proposed methodology. These foundational concepts will lay the groundwork for understanding the novel methodology designed in this study.

6.2 ARM and K-mean clustering

In recent years, there has been a growing interest in using data mining techniques to analyse road crash data. There are two types of data mining tasks: prediction and description. Data mining is a process to extracting previously unknown and potentially useful information from data, which can provide valuable knowledge for decision making, prediction, and estimation (Pujari, 2001). Association rule mining (ARM) is one of such techniques developed by Agrawal et al. (1993), to describe patterns and their interactions between products in large-scale transaction data from supermarket point-of-sale (POS) systems. These rules operate based on conditional statements, often expressed through an "if, then" framework, where the antecedent denotes the conditional patterns leading to an outcome, known as the consequent.

The ARM process consists of three main steps: (1) Frequent Itemset Generation, which identifies commonly co-occurring attributes using algorithms such as Apriori; (2) Rule Generation, which extracts association rules based on predefined confidence thresholds; and (3) Rule Evaluation and Pruning, where rules are assessed using support, confidence, and lift metrics to ensure their significance and relevance.

In road crash analysis, ARM has been employed to uncover key contributors to crashes by analysing historical crash databases. The main advantages of ARM include its ability to discover hidden patterns, process large datasets efficiently, and provide interpretable results for decision-making. Despite its strengths, ARM has several limitations. It is computationally intensive, can generate redundant or spurious rules, lacks causal inference, and is sensitive to parameter selection. These drawbacks present challenges when analysing crash data, where rare but critical events such as fatalities must be identified and examined effectively. To enhance effectiveness of ARM in road crash analysis, integrating clustering techniques—specifically K-means clustering—is a justified and strategic choice.

The concept of K-mean clustering was firstly introduced by MacQueen (1967). It described a process to group several dimensions of dataset into a pre-defined number of clusters. Hartigan & Wong (1979) developed a variation of K-means algorithm to search for the local minimum of distance sum of square within each cluster. K-means clustering has emerged as a prevalent methodology employed across various academic disciplines for the purpose of feature learning and cluster analysis. A further enhancement of K-mean clustering was introduced by Hahsler & Karpienko (2017), providing a method for enhancing the visualisation of association rules by grouping rules efficiently. This technique offers an innovative approach to visualising association rules, allowing for improved comprehension and analysis of complex datasets, which enhances the interpretability of association rules, making it a valuable tool for data analysts and researchers in various fields.

The integration of K-means clustering with ARM offers several advantages. K-means is highly efficient for large datasets, easy to implement, and capable of uncovering meaningful patterns that might otherwise be obscured. However, it also has challenges, such as sensitivity to the initial selection of cluster centroids and potential convergence to local minima. Despite these drawbacks, its ability to refine association rules and reveal rare yet critical fatality patterns make it an indispensable tool in road crash analysis. By combining ARM with K-means clustering, road safety analysts can enhance pattern recognition, improve the detection of high-risk factors, and develop more effective crash mitigation strategies. This integrated approach provides a robust framework for data-driven decision-making in road safety management, supporting targeted interventions that ultimately contribute to reducing crash risks and improving roadway safety outcomes.

This study conducts a rigorous data analysis to elucidate the relationships between road infrastructure attributes and fatal crashes. Adopting a strategic approach, it comprehensively examines road network safety from multiple perspectives by integrating Association Rule Mining (ARM) with K-means clustering. Instead of relying solely on simplistic parameter settings like minimum support and confidence levels, which can overlook valuable insights, this methodology strengthens the detection of complex and rare crash patterns.

The primary contribution of this study is the novel combination of ARM and K-means clustering to analyse road crash data. While ARM is traditionally used for rule-based pattern discovery and K-means clustering for segmentation, their integration in road safety analysis offers a multi-dimensional understanding of how infrastructure attributes and crash severity interact. This novel analytical framework enhances the ability to uncover rare but critical fatality patterns, offering a structured method to visualise high-risk crash scenarios. Although ARM has been applied to road safety in previous studies and K-means clustering is a well-established technique, their combined application to systematically analyse road safety data, particularly for identifying rare crash patterns, is relatively uncommon.

The proposed procedure involves several steps as illustrated in Figure 3.4.1. First, the road crash data is pre-processed to remove any irrelevant data and deal with missing data. Next, ARM is used to generate rules based on predefined minimum support and confidence values, describing the association between antecedent and consequent. Then, by applying the K-means clustering to rules, stakeholders will obtain an overall understanding of the rules concerning roadway crashes. Finally, rules are visualised by different clusters with support and lift illustration to inform road infrastructure safety strategic management from three different perspectives: quantitative analysis, qualitative analysis, and clustering analysis.

The subsequent sections provide a comprehensive review of existing literature on road safety studies utilising association rule mining techniques and elaborate the theoretical framework of the two data mining methods following an appraisal of various algorithms to determine the most appropriate one based on the characteristics of the crash database.

6.3 Related works

Association rule mining is one of the prevalent methods used to generate rules and figure out the hidden information of database, which has been applied to road safety analysis. Some studies integrating association rules mining and multi criteria analysis to find out the most interesting and relevant rules based on crash data analysis. Addi et al. (2016) used Apriori algorithm with setting a threshold of support, confidence, and lift to extract more relevant rules. At the same time, they presented ELECTRE TRI (multi criteria analysis method) to eliminate redundant and non-interesting rules, but it was sensitive due to the preference of decision makers. Furthermore, they (Ait-Mlouk et al., 2017) applied Frequent Pattern Growth algorithm to perform association rules mining and ranked the rules by PROMEHEE (multi criteria analysis method) calculating the degree of preference. Another case study was carried out by El Abdallaoui et al. (2019). They employed Apriori algorithm and ELECTRE II to rank all the association rules by outranking of each rule based on three criteria: support, confidence, lift.

Kumar & Toshniwal (2015) combined K-modes clustering used to classify different types of crash and association rules mining used to understand factors in each cluster. Das et al. (2019) investigated the vehicle-pedestrian crashes by Apriori algorithm to discover vehicle-pedestrian crash patterns based on eight years of Louisiana crash data. Xu et al. (2018) studied the contributory factors that have high co-occurrence in serious casualty crashes by using association rules mining. They used relatively higher threshold values for support and confidence to increase the accuracy and strength and the results enables policymakers to develop corresponding policy initiatives and engineering safety improvement countermeasure.

Geurts et al. (2003) conducted a comparative analysis of the rules sets of the high-frequency accident locations and the low-frequency accident locations. They explained the relationship between results from chi-squared test introduced by Silverstein et al. (1998) and lift to identify the dependency between antecedent and consequent based on rules mined by setting minimum support and confidence.

Pande & Abdel-Aty (2009) applied Apriori algorithm to investigate crash data in United States as first attempt based on five variables. They evaluated rules based on lift value as it is more important for determining the strength of an association rule than the other two criteria: support and confidence and they also empathised that the rare events of interest such as fatal crashes were needed to consider. Yu et al. (2019) considered more variables involving environment, management, drivers, and vehicles, which achieved a big size of database analysis using association rule mining. They conducted a validity test of rules, but the result indicated that three criteria: support, confidence, and lift threshold can do an excellent job of filtering out most rules that are not statistically significant.

Montella (2011) explored the crash contributory factors in 15 urban roundabouts in Italy. They summarised contributory factors based on site inspections and knowledges from specialists and applied those factors as input to obtain rules, which was helpful to identify most important co-occurrence of crash contributory factors by ranking lift of rules.

Yang et al. (2019) used Analytic Hierarchy Process (AHP) based association rules mining to improve the operation efficiency and extract essential association rules. By AHP calculation, the input data was cut down from 481 to 200 which was considered as most influential attribute to crash. Three different regions of crash attributes were investigated, and the results provided a better understanding of crash risk in different areas.

Current studies in this field prioritise rules based on predefined criteria, often using minimum support and confidence thresholds, which can introduce bias and subjectivity into outcomes. This issue is particularly critical in datasets involving infrequent events like fatal crashes, where important patterns may not meet minimum support thresholds and could be excluded from analysis (Tan et al., 2006). High thresholds can also obscure subtle relationships, favouring common attributes over potentially insightful but less frequent ones (Freitas, 2002). Tactical management of these thresholds is essential for producing actionable insights, as high thresholds limit rule diversity and practicality, hindering effective interpretation for strategic decision-making.

Despite extensive literature on association rule mining, gaps remain in understanding the tactical setting of minimum support and confidence thresholds, especially in datasets with infrequent events. Existing studies often overlook adaptivity and applicability of data mining in strategic management, particularly in data analysis in line with different objectives. Addressing these gaps can enhance analytical approaches and broaden the comprehension of infrequent patterns in fields such as road crash analysis, extending the utility of association rule mining.

6.4 Theoretical Framework

6.4.1 Basic definitions of association rule mining

Support and confidence are two measures to assess the significance of association rules. Support represents how frequently an itemset appears in dataset (Agrawal et al., 1993). For association rules, support corresponds to the occurrence frequency of both the antecedent and consequent item sets. With regards to road safety, support is defined as the percentage of crash-recorded attributes that appear together with a specific consequent in database denoted as:

$$Supp(X \rightarrow Y) = P(X \cup Y) = \frac{N(X \cup Y)}{N(transactions)} \tag{6.4-1}$$

where X and Y represent distinct item sets in RHS and LHS of a rule respectively that co-occur.

Confidence serves as an additional measure that quantifies the extent to which the right-hand side (RHS) of rules consistently appears in a database that already contain items from the left-hand side (LHS). In the context of road safety, confidence is defined as the percentage of occurrences where crash attributes and a specific consequent are observed together in the dataset, relative to the rules containing the same crash attributes. Mathematically, it can be represented as follows:

$$Conf(X \rightarrow Y) = P(Y|X) = \frac{Supp(X \rightarrow Y)}{Supp(X)} = \frac{P(X \cup Y)}{P(X)}$$

(6.4-2)

Association rule mining aims to illustrate the relationships between antecedents and consequents within frequent patterns. However, two constrained parameters are configured to be as minimal as feasible, with the intention of comprehending the entire databases, rather than solely focusing on frequent patterns. To facilitate the elucidation of associations, the parameter of "lift" (Berry & Linoff, 2004), used to validate rule robustness, is also incorporated to extract interesting and strongly associated rules. Lift serves as a reliable indicator of the rule's performance and interestingness. It quantifies the relationship between the observed support and the support that would be anticipated under the assumption of independence between the left-hand side (LHS) and right-hand side (RHS). This parameter measures the robustness of the association between specific consequents and attributes within the domain of road safety. Its representation is as follows:

$$Lift(X \rightarrow Y) = \frac{P(Y|X)}{P(Y)} = \frac{Supp(X \rightarrow Y)}{Supp(X) * Supp(Y)} = \frac{P(X \cup Y)}{P(X)P(Y)}$$

(6.4-3)

In cases where the lift is greater than 1, it signifies a mutual reliance between two instances of the left-hand side (LHS) and right-hand side (RHS), suggesting the rule is potentially useful and LHS has positive effect on RHS. In simpler terms, the occurrence of LHS substantially increases the likelihood of RHS occurrence. Conversely, when the lift is less than 1, it indicates

LHS and RHS are alternatives to each other, suggesting the rule is not interesting and useful. A lift value of 1 implies that the probability of occurrence of the antecedent and that of the consequent are independent to each other, suggesting that no rule can be drawn involving these two item sets.

6.4.2 Rule mining algorithm selection

There are different algorithms introduced since 1993, which can be applied to discover hidden patterns and knowledge. These mainly includes AIS algorithm, SETM algorithm, Apriori algorithm, Frequent Pattern Growth (FP-Growth) algorithm and other variation of these algorithms. With regards to road safety data analysis, it is essential to select an appropriate algorithm to explore the crash databases so that knowledge could be provided for strategic level management.

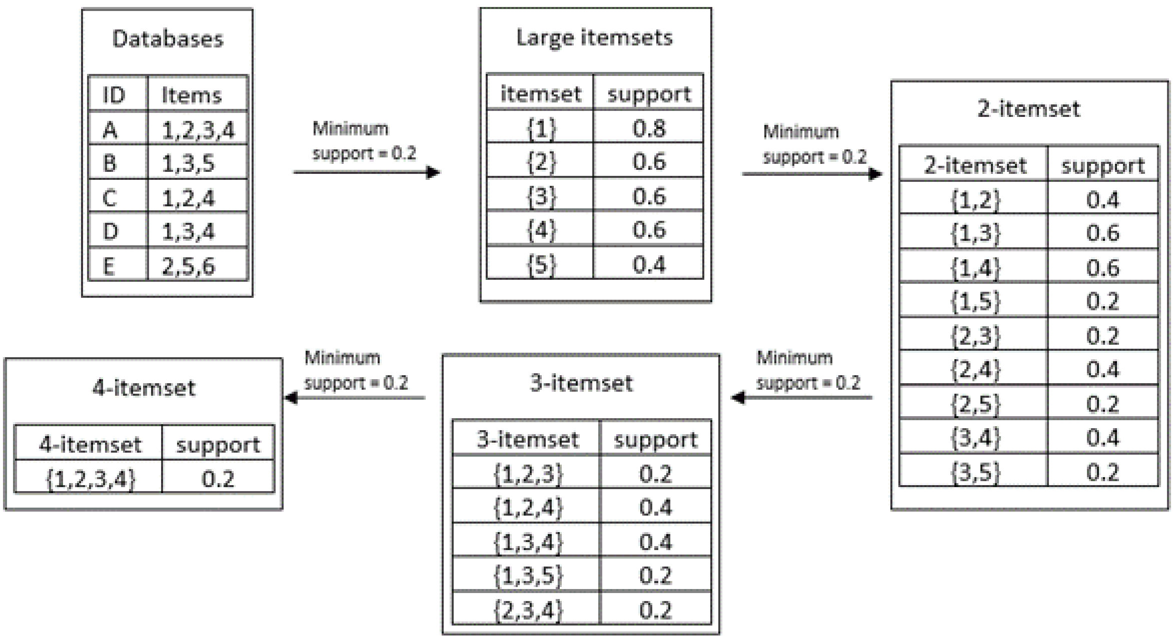

Figure 6.4.1: Process of AIS algorithm

AIS algorithm was first introduced by Agrawal et al. (1993) for mining association rules. It generates rules that satisfy two additional constraints: syntactic constraint and support constraints. Syntactic constraint refers to restricting items involved by predefining a specific item in antecedent or consequent. Support constraints include minimum support and confidence. The process of AIS is constructed by two phases as shown in Figure 6.4.1. All items combination of which support is above the predefined minimum support, called large item sets, are generated by scanning transaction database multiple times. Rules are extracted based on predefined confidence by combining items in large item sets with a specific item in antecedent or consequent, and the iteration of item sets generation terminates when none of itemset is obtained under the same constraints. However, AIS only can extract rules with one item in the consequent and only consider large item sets as interest of research. The iteration of going passed databases reduces the efficiency of rule mining. Consequently, it requires more space with highly computational cost.

A set-oriented mining algorithm, called SETM, based on sort-merge strategy to was introduced by Houtsma and Swami (1995). It is similar to AIS that both need multiple passes over databases. It consists of two sort operations and one merge-scan join as shown Figure 6.4.2. Firstly, transactions are sorted by individual item with transaction identifier and determine whether they are frequent items in database. Then, it expands the item sets based on previous pass so that, ultimately, rules can be generated under constraints of predefined support and confidence. It saves the transaction identifier with item sets, but it has the same problem to AIS that every iteration needs pass database resulting in low efficiency.

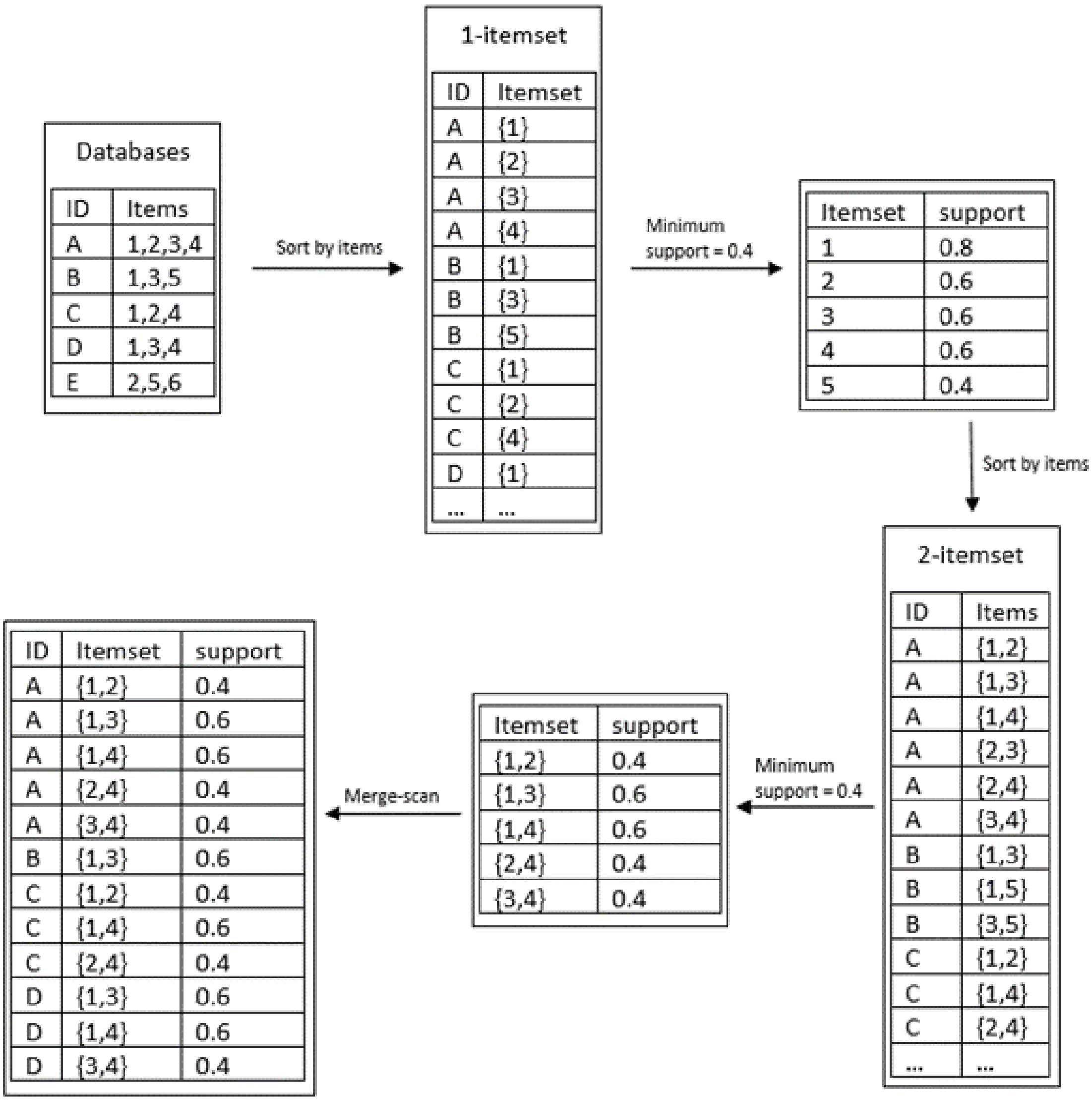

Figure 6.4.2: Process of SETM algorithm

In both the AIS and SETM algorithms, candidate item sets are generated on-the-fly during the pass as data is being read. Specifically, after reading a transaction, it is determined which of the item sets found large in the previous pass are present in the transaction. New candidate item sets are generated by extending these large item sets with other items in the transaction. This results in unnecessarily generating and counting too many candidate item sets that turn out to be small.

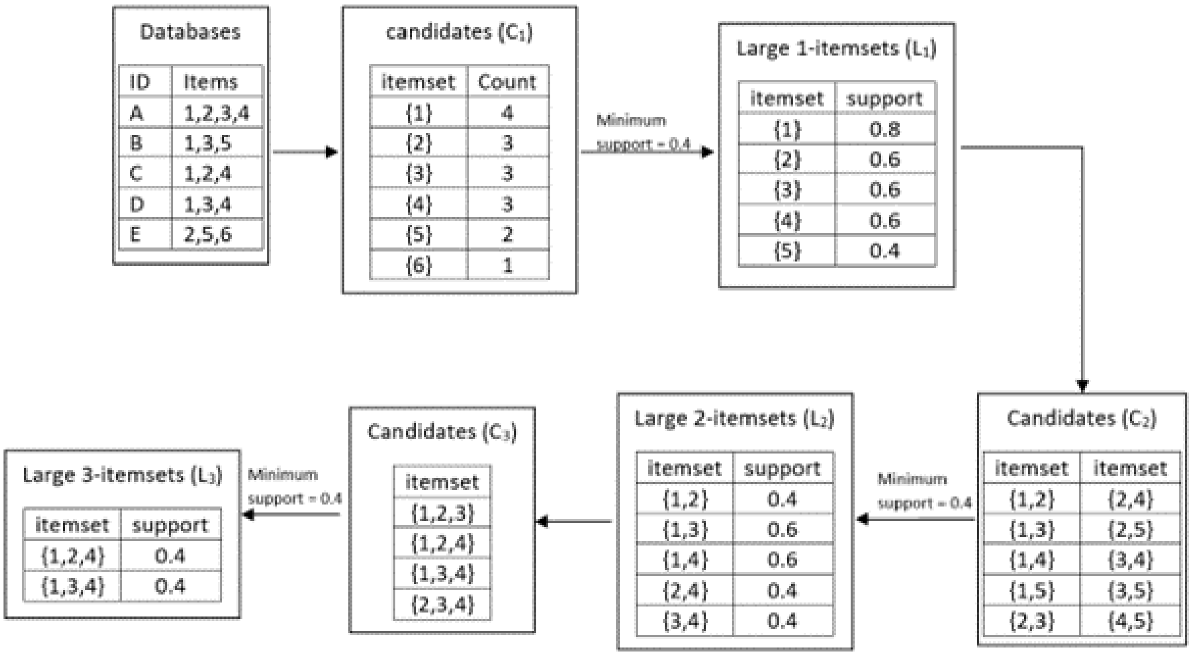

Figure 6.4.3: Process of Apriori algorithm

The Apriori and AprioriTid algorithms (Agrawal et al., 1994) generate the candidate item sets to be counted in a pass by using only the item sets found large in the previous pass without considering the transactions in the database. Figure 6.4.3 illustrates the process of Apriori algorithm, where two phases are built for Apriori. First, the large 1-itemsets (L_1) are determined by counting item occurrences. Then, the candidates of large 2-itemsets (C_2) are found from last item sets (1-itemsets) and the large 2-itemsets (L_2) are obtained by scanning transaction database and calculating support of candidates (C_2). The second phase will repeat for k-item sets rules mining.

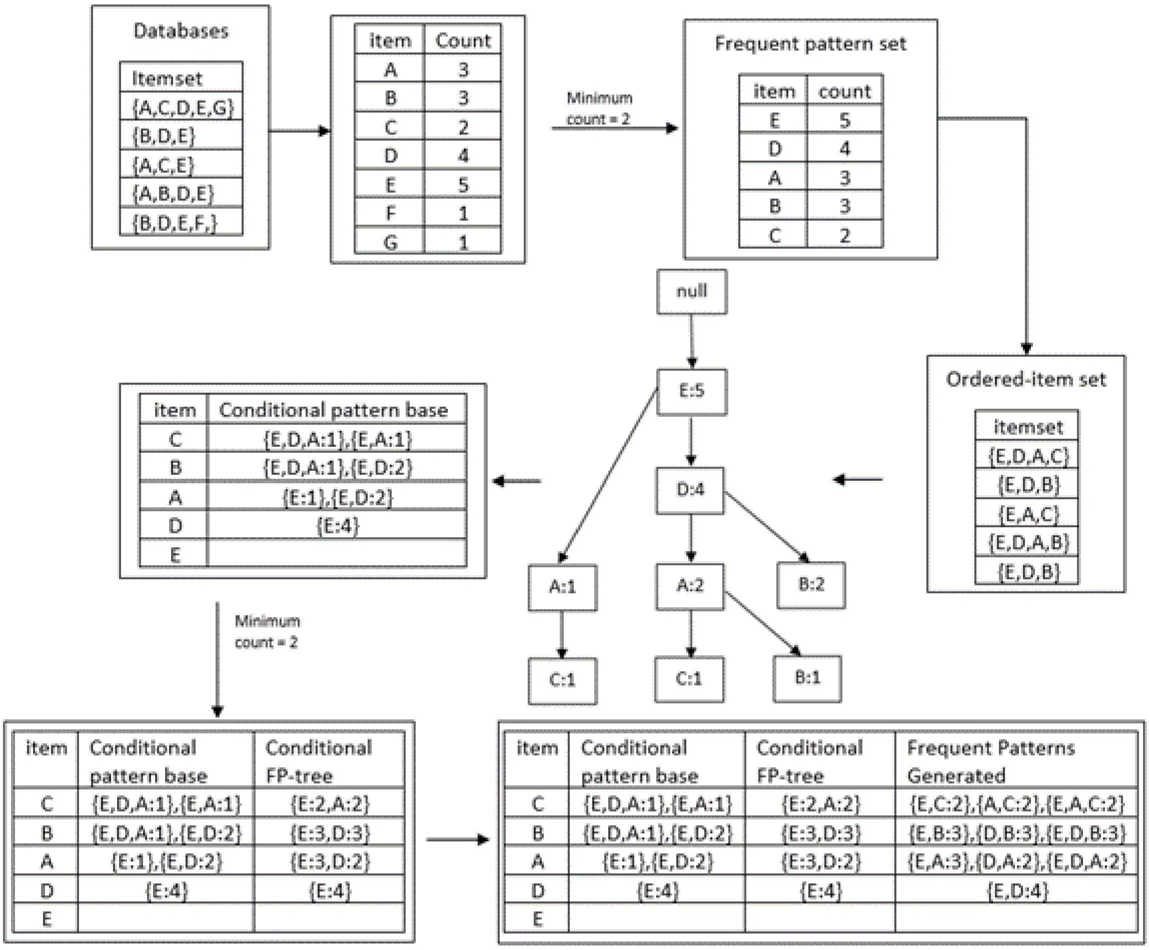

Figure 6.4.4: Process of FP-Growth algorithm

FP-growth is an alternative algorithm utilised in mining a comprehensive collection of frequent patterns through pattern fragment growth. Han et al. (2000) introduced this method as a solution to address the limitations of the Apriori technique, which faces challenges in handling a vast number of candidate-sets and repeatedly scanning the database for pattern matching. The FP-growth algorithm was designed in three distinct stages. Initially, the algorithm scans the database, counting the occurrences of each item and selects the frequent items based on a minimum support threshold. These frequent items are then sorted and used to construct the FP-tree. A node list is generated from the frequent items, and these nodes are connected to corresponding nodes within the tree. Subsequently, the database is re-examined, and if an item in a transaction is deemed frequent, it is incorporated into the tree structure. Finally, starting from the least frequent item, a recursive procedure is employed to discover frequent patterns. The algorithm calculates the support count of each pattern and presents the frequent ones. Figure 6.4.4 illustrates the step-by-step execution of the FP-Growth algorithm.

The four prominent algorithms prove to be valuable tools in the process of rule mining. We have previously outlined the principles and procedures associated with each algorithm, and a concise comparison has been presented in Table 6.4.1.

Table 6.4.1: Four rules mining algorithms comparison

Algorithm	**Advantages**	**Disadvantages**
AIS	Easy to understand and use.	Only one item allowed in the consequent. Multiple scans of database Large size of candidate set.
STEM	Separates generation from counting.	Large execution time. Large size of candidate set.
Apriori	Easy to implement. Less candidate sets.	Many scans of database
FP-Growth	Only twice scanning databases. Faster than others (Győrödi et al., 2004)	More memory consumption. Tree structure creates complexity.

It is imperative to select an appropriate algorithm to optimise the data analysis efficiency, particularly in the context of road safety investigations. The yearly crash data in STAT19 under investigation is medium-sized, consisting of approximately less than 200,000 records. Each record represents a unique traffic crash, capturing various attributes such as date of occurrence, location, weather conditions, vehicle types, and other related information. The traditional mindset of association rules mining aims at extracting frequent patterns, which may overlook the significance of rare attributes in understanding crash patterns comprehensively.

The decision to employ the Apriori algorithm for this crash database analysis is based on several key factors. Firstly, the algorithm excels in handling large datasets, making it suitable for the medium-sized crash database under examination. Furthermore, the Apriori algorithm can handle exceptionally low minimum support and confidence values without introducing excessive complexity, a characteristic that holds significant importance in achieving a comprehensive comprehension of crash patterns. This capability sets Apriori apart from FP-growth, making it particularly suitable for analysing the crash database under investigation. Thus, by identifying and analysing infrequent attributes, the algorithm can unveil previously unnoticed insights, contributing to a more complete analysis.

6.4.3 Clustering methods

Clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics. Clustering is the process of grouping similar objects into different groups, or more precisely, the partitioning of a data set into subsets. There are mainly three types of clustering methods are employed in data analysis: hierarchical, partitional and density-based clustering (Han et al., 2011).

Hierarchical clustering encompasses agglomerative (bottom-up) and divisive (top-down) approaches, constructing a hierarchy of clusters represented as dendrograms. This method iteratively merges or splits clusters based on proximity measures between data points. Hierarchical clustering does not require the prior specification of the number of clusters and offers a visual representation of cluster relationships through dendrograms. It is beneficial for understanding nested structures within data. However, this method can be computationally intensive and sensitive to noise and outliers. Additionally, once clusters are merged or split, these decisions are irreversible (Madhulatha, 2012; Rokach & Maimon, 2005).

Partitional clustering methods, such as K-means, are widely used in data analysis to partition data into k clusters, where each data point is assigned to the cluster with the nearest centroid. K-means iteratively updates cluster centroids to minimise the variance within clusters, making it computationally efficient and straightforward to implement (A. K. Jain et al., 1999). This efficiency makes K-means particularly suitable for handling large datasets, where it can efficiently scale with the size of the data (Xu & Tian, 2015). However, K-means does have limitations. It requires the prior specification of the number of clusters k, which can be challenging without prior knowledge of the data structure.

Density-based clustering identifies clusters as regions of high data point density separated by low-density areas. Unlike partitional methods, it does not require specifying the number of clusters beforehand. It can discover randomly shaped clusters and is robust to noise and outliers. It adapts well to varying cluster densities within a dataset and does not require strict assumptions about cluster shapes. However, Performance can depend heavily on parameter settings, such as the radius and minimum points required to define a cluster. It can also be computationally expensive for large datasets with varying density levels (Ester et al., 1996; Madhulatha, 2012).

K-means clustering is chosen for its efficiency, simplicity, and scalability in handling the large volumes of association rules generated in data mining. Its ability to produce clear and interpretable clusters aligns well with the objective of identifying distinct groups of association rules based on their lift and support metrics. While hierarchical and density-based clustering methods offer distinct advantages, the straightforward implementation and robust performance of K-means clustering make it the preferred choice for data analysis. Its capacity to handle large datasets and derive meaningful clusters from association rule data support the analytical objectives of this research.

Hahsler & Karpienko (2017) developed a method for grouping the rules using K-means clustering to manage the substantial number of rules. They took advantage of Lift as a measure of interest to group antecedent containing substitutes and used K-means clustering to partition the set of antecedents into K groups $S = \{S_1, S_2, S_3, \dots, S_k\}$ while minimising the within-cluster sum of squares. Formally, the objective is to find:

$$\arg min_s \sum_{i=1}^{k} \sum_{m_j \ni S_i} \left\| m_j - \mu_j \right\|^2 \tag{6.4-4}$$

where μ_i is the mean (also called centroid) of vectors in S_i.

6.5 Summary

In this Chapter, the background information of data mining methods specifically association rule mining as well as related work to crash data analysis have been demonstrated. The theoretical foundations of association rule mining and clustering methods within the context of road crash data analysis have been explored. The strategic analysis on road crash data involves leveraging these methodologies to uncover hidden patterns and associations that inform road infrastructure safety management. Specifically, association rule mining using the Apriori algorithm is proposed to reveal intricate relationships between road attributes and fatal crash occurrence. K-means clustering has been selected for its efficiency in organising association rules into distinct clusters based on similarity in lift and support metrics.

The forthcoming Chapter will extend this theoretical foundation into a practical application through the prototype development of a data analysis framework. This prototype will integrate association rule mining results with K-means clustering, aiming to provide stakeholders with valuable insights for enhancing road safety strategies. By building a bridge between theory and implementation, the prototype development phase will demonstrate how these analytical techniques can be applied to real-world crash data to support informed decision-making and strategic management of road infrastructure safety.

Chapter 7. Development of the Prototype Strategic Tool

7.1 Methodology

This research aims to employ association rule mining, a data mining technique, to comprehensively analyse the entire crash database in terms of fatal crash contributors and association analysis. To facilitate a holistic view of rules from three dimensions of strategic perspectives, K-means clustering is utilised for grouping and visualising the rules. The research process involves several essential steps, including identifying minimum support and confidence thresholds, determining the number of items, applying rule mining, applying K-means clustering, and finally, visualising the results as illustrated in Figure 3.4.1. Detailed explanations of each step are elaborated in the following subsections to provide a comprehensive understanding of the overall procedure.

7.1.1 Data preparation

In the context of road safety, the primary concern is fatal crash, which result in substantial economic losses and have a profound impact on public health. To comprehend the factors contributing to fatal crash and their underlying causes, association rule mining combined with K-means clustering is employed. The dataset used in this study is derived from the STATS 19 road accident injury statistics database for the year 2018. STATS20 is an introduction document used for completion of road accident report, which provides a detailed explanation of the information to understand the attributes. The definitions of killed were elaborated in STATS20 corresponding to the severity of crash. A fatal crash is when deaths occur within a period of fewer than 30 days following the accident, which excludes fatalities resulting from natural causes or suicide. Serious incident is defined as any detention in hospital as an in-patient, either immediately or later, and injuries to casualties who die 30 or more days after the accident from injuries sustained in that accident. Slight injury includes but not limited to whiplash or neck pain, shallow cuts/lacerations/abrasions, sprains and strains (not necessarily requiring medical treatment), bruising and slight shock requiring roadside attention.

To identify the infrastructure related engineering problems, some irrelevant variables are eliminated, and collision data was specifically considered. There are in total 36 variables in STATS 19 collision dataset, which furnishes comprehensive details such as accident location, road type, junction details, junction control, lighting conditions, and more. Repetitive and unnecessary information are excluded from consideration, while missing values are deemed to have minimal impact on the overall dataset. Consequently, 20 relevant and meaningful attributes are identified as shown in Table 7.1.1 for utilisation in the data mining process.

Table 7.1.1: Roadway crash attributes data (derived from STATS 20)

Roadway crash attributes	**Detailed information**
Accident severity	1 (Fatal); 2 (Serious); 3 (Slight).
Month	January to December
Day of week	1 (Sunday); 2 (Monday); 3 (Tuesday); 4 (Wednesday); 5 (Thursday); 6 (Friday); 7 (Saturday).
Time	0-24
Local authority ons district	Names and codes for a wide range of geographies (See STATS 20)
Regional area	North East; North West; Yorkshire and the Humber;

	East Midlands; West Midlands; East of England. London; South East; South West.
First road class	1 (Motorway); 2 (A(M)); 3 (A); 4 (B); 5 (C); 6 (Unclassified).
Road type	1 (Roundabout); 2(One way street); 3 (Dual carriageway); 6 (Single carriageway); 7 (Slip road); 12 (One way street/Slip road).
Speed limit	20,30,40,50,60,70 are the only valid speed limits on public highways
Junction detail	0 (Not at junction or within 20 metres); 1 (Roundabout); 2 (Mini roundabout); 3 (T or staggered junction); 5 (Slip road); 6 (Crossroads); 7 (More than 4 arms not roundabout); 8 (Private drive or entrance); 9 (Other Junction).
Junction control	0 (Not at junction or within 20 metres); 1 (Authorised person); 2 (Auto traffic signal); 3 (Stop sign); 4 (Give way or uncontrolled)
Pedestrian crossing human control	0 (None within 50 metres);

	1 (Control by school crossing patrol); 2 (Control by other authorised persons)
Pedestrian crossing physical facilities	0 (None within 50 metres); 1 (Zebra); 4 (Pelican, puffin, toucan, or similar non-junction pedestrian light crossing); 5 (Pedestrian phase at traffic signal junction); 7 (Footbridge or subway); 8 (Central refuge)
Light condition	1 (Daylight); 4 (Darkness - lights lit); 5 (Darkness - lights unlit); 6 (Darkness - no lighting); 7 (Darkness - lighting unknown)
Weather condition	1 (Fine no high winds); 2 (Raining no high winds); 3 (Snowing no high winds); 4 (Fine + high winds); 5 (Raining + high winds); 6 (Snowing + high winds); 7 (Fog or mist); 8 (Other)
Road surface condition	1 (Dry); 2 (Wet or damp); 3 (Snow); 4 (Frost or ice); 5 (Flood over 3cm deep); 6 (Oil or diesel); 7 (Mud)
Special condition at site	0 (None); 1 (Auto traffic signal - out); 2 (Auto signal part defective);

	3 (Road sign or marking defective or obscured); 4 (Roadworks); 5 (Road surface defective); 6 (Oil or diesel); 7 (Mud);
Carriageway hazard	0 (None); 1 (Vehicle load on road); 2 (Other objects on road); 3 (Previous accident); 4 (Dog on road); 5 (Other animals on road); 6 (Pedestrian in carriageway - not injured); 7 (Any animal in carriageway (except ridden horse))
Urban or rural area	1 (Urban); 2 (Rural)l; 3 (Unallocated)
Trunk road flag	1 (Trunk (Roads managed by Highways England)); 2 (Non-trunk)

Additionally, the dataset contains some missing values and meaningless entries, such as "nonspecial condition at site" and "unknown information (self-reported)." As these information cannot provide valuable insights for decision making, these entries are treated as missing values. The proportion of missing values for each variable is calculated and presented in the Table 7.1.2. Each row corresponds to a variable, with the accompanying number indicating the percentage of missing values for that variable. Ten variables (1, 2, 3, 4, 5, 6, 7, 8, 9, and 14) exhibit no missing values, indicating that the dataset is complete and reliable for these fields. Their completeness ensures robustness in any analysis involving these variables. While some variables are entirely observed, others display minimal (less than 5%) to moderate (less than 30%), various levels of missing data, and a few exhibits significant (more than 30%) proportions of missing data.

Table 7.1.2: Percentage of missing values for each variable

No.	Variable	Percent (%)	No.	Variable	Percent (%)
1	accident severity	0	11	junction control	44.7197
2	date	0	12	pedestrian crossing human control	2.587353
3	day of week	0	13	pedestrian crossing physical facilities	2.32397
4	time	0	14	light conditions	0
5	Local authority Ons district	0	15	weather conditions	0.015493
6	regional areas	0	16	road surface conditions	0.1533
7	first road class	0	17	special conditions at site	96.71301
8	road type	0	18	carriageway hazards	97.25119
9	speed limit	0	19	urban or rural area	0.000815
10	junction detail	0.000815	20	trunk road flag	8.685938

Numerous methods exist for handling missing categorical data, including imputation techniques such as Missing Categorical Data Imputation Based on Similarity (MIBOS) (Wu et al., 2012), k-nearest neighbours (KNN) (Batista & Monard, 2002), and various prediction models. Imputation aims to estimate and fill in missing values using identified relationships within the dataset's observed values. The goal is to leverage known patterns to make informed estimates of the missing data. However, when employing association rule mining, which focuses on uncovering associations based on the frequency and conditional probability of different patterns, imputation can fundamentally alter the data, potentially leading to skewed or inaccurate analysis results. Consequently, in the context of association rule mining, the optimal strategy for handling missing data may be to exclude it altogether. This approach preserves the integrity of the analysis by avoiding the introduction of potentially misleading imputed values.

Retaining variables with a high proportion of missing data, such as special conditions at site and carriageway hazards, is particularly crucial. Despite their missingness, these variables can provide invaluable insights into infrequent patterns that are essential for reflecting the true state of road infrastructure conditions. By considering these variables, even with their substantial missing data, the analysis remains comprehensive and representative of all potential scenarios, ensuring that critical yet uncommon patterns are not overlooked. The analysis does not ignore the missing data problem; instead, it acknowledges and systematically validates the findings through robustness check elaborated in Section 7.4.2.

7.1.2 Measurement of interest identification.

Confidence is a crucial metric for assessing the strength of an association rule (Agrawal et al., 1993). However, it is essential to recognise that confidence operates within the context of support. When mining association rules involving rare consequents, the confidence value may be significantly low due to frequent patterns in the database. Determining an appropriate confidence threshold can be challenging. To extract interesting rules, it’s common to set the minimum confidence value equal to the minimum support value. This approach aims to capture a maximal number of potential rules.

To gain a comprehensive understanding of the attributes contributing to fatal crashes, frequency of fatalities is calculated. Based on this frequency, minimum support and confidence thresholds are established, corresponding to 10% of the total killed crash occurrences. This approach ensures that a significant number of rules, even those with low support and confidence, are considered in the analysis. By doing so, the focus extends beyond individual attributes to encompass broader patterns.

7.1.3 Item number identification

The Apriori algorithm was selected for rule mining from the STATS 19 database. However, this database contains numerous attributes, potentially resulting in extended programming execution time when experimenting with various rule lengths. Additionally, the number of rules is influenced by the items within each rule. To capture all interesting rules, a correlation analysis was conducted to establish the maximum number of items within the rules. The objective of this analysis is to avoid bias introduced by varying item counts in the rules and to obtain the majority of potentially interesting rules aligned with the essential mining criteria as mentioned in section 7.1.2.

7.1.4 Rule mining and redundant rules removal

In the process of analysing the data, Rstudio is utilised, specifically the 'arules' package based on Apriori algorithms, to extract rules while adhering to predefined constraints. However, due to the aim of comprehensively understanding the entire database rather than focusing exclusively on frequent patterns, the resulting rules often become excessively numerous, thereby hindering efficient mining and comprehension by end users. To address this issue, Bayardo et al. (2000) proposed an algorithm integrated to eliminate redundant rules that can be simplified to yield an equally or more predictive rule.

In frequent pattern mining, identifying a suitable pair of constraints, minimum support and confidence, is vital important to filter out extraneous and less predictive rules. However, there is a dilemma that discovering association rules with a high support tends to overlook rare item rules, whereas setting a low support gives rise to an excessive abundance of rules, leading to redundancy and consequently reducing the effectiveness of the exploration. Usually, a redundant subgroup collection is several tens to hundreds of times greater in size than a subgroup collection (Li et al., 2014). Removing redundant rules simplifies the analysis, making it easier for analysts to understand and interpret the remaining rules. This clarity aids decision-makers in identifying actionable insights and making informed decisions, which generally enhances the clarity, efficiency, and quality of data analysis, aiding in effective decision-making

Bayardo et al. (2000) introduced a conception of minimum improvement which signifies the minimum difference between its confidence and the confidence of any of its simplifications. If the improvement of a rule is positive, removing any subset from antecedent decreases the confidence by at least its improvement. Conversely, a rule with negative improvement is undesirable because the rule can be simplified to yield a proper sub-rule which is more predictive. Briefly, a parent rule is defined as redundancy, if a more general rule with the same or a higher confidence exists. It can be explained as follow: a rule X→Y is redundant, if

$$\exists X' \in X, Conf(X' \rightarrow Y) \geq Conf(X \rightarrow Y) \tag{7.1-1}$$

7.1.5 K-mean clustering and visualisation

To avoid the usage of a few rules or ranking of rules from some biased weight assignments to represent the underlying knowledge within the database, a balloon plot is utilised for the purpose of illustrating the organised rules. In the visual representation, antecedent groups are depicted as columns, while consequents are displayed as rows, as described by Hahsler &

Karpienko (2017) in their study. The colouration of each balloon in the plot corresponds to the collective interest measure within the group, whereas the size of the balloon indicates the combined support. Additionally, the grouped matrix visualisation offers an interactive option for zooming in on specific groups and examining rules in a finely detailed manner. Groups are labelled by the most interesting item in the group. Hahsler M et al. (2023) advocated using highest ratio of support in the group to support in all rules as a metric to measure the relative significance of a particular item within a specific group compared to its prevalence in the entire dataset. This indicates that the item is more prevalent and significant within the specific group compared to its overall occurrence in the entire dataset suggesting that the item is remarkable characteristic of the that group. The ratio is denoted as follows:

$$Ratio = \frac{Support\ in\ the\ group}{Support\ in\ all\ rules} \tag{7.1-2}$$

7.2 Prototype Application

To comprehensively understand the attributes contributing to fatal crashes, the frequency of such crashes in the database is calculated to establish minimum support thresholds corresponding to 10% of its frequency. For instance, in 2018, the STATS 19 dataset recorded a total of 122,635 incidents, including 1,671 fatal cases. Based on the rule support definition, the highest support values for rules related to fatal outcomes are 0.0136. Consequently, the minimum support value is set at 0.001. In this study, specifying confidence is less critical because the primary parameter for assessing interestingness is the lift value. Confidence is the conditional probability of the support value of a rule. Therefore, the minimum confidence is set to match the minimum support value, which is very low, indicating patterns with relatively weak associations and frequencies in the crash database. This approach ensures that a significant number of rules with low support and confidence levels regarding fatal crashes are considered in the analysis, rather than focusing exclusively on attributes that may not capture broader patterns.

Concerning the quantity of roadway crash attributes within rules, an investigation was conducted on rules with varying attribute number, ranging from 1 to 10 shown in Figure 7.2.1. illustrates how the increasing number of attributes affects the total count of rules. Notably, flatten at the point of seven items, suggesting that further increments in attribute count do not yield a significant impact on the total number of rules. Thus, it is determined that mining rules with seven attributes is s suitable choice for obtain potential rules without excessive efforts in both scenarios.

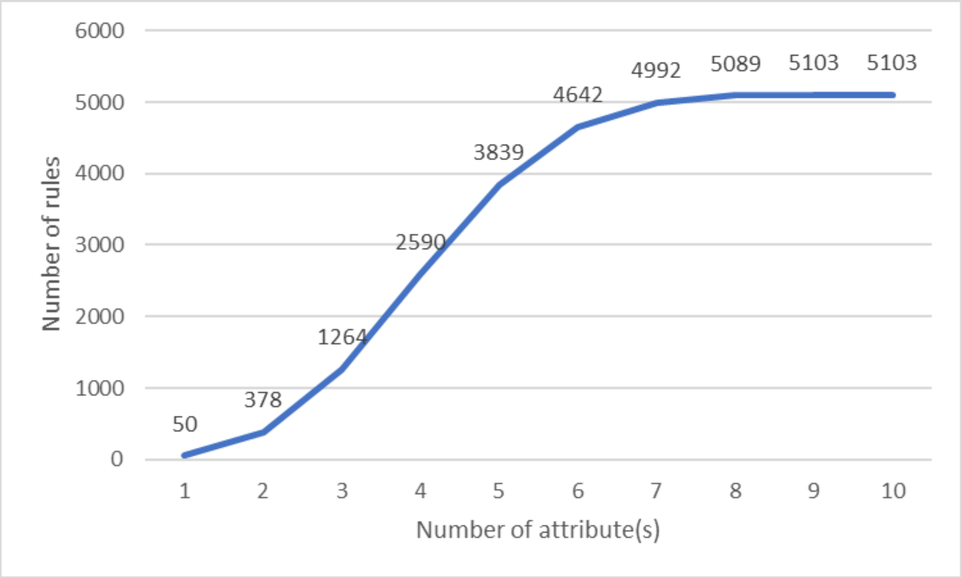

Figure 7.2.1: Correlation between number of attributes and rules

7.3 Result

In compliance with the mining constraints, a comprehensive set of rules amounting to 4992 rules are produced from the STATS 19 dataset, corresponding to fatal cases. To identify meaningful rules, two crucial steps: sorting based on lift and eliminating redundancy are employed. Consequently, a subset of 643 rules as displayed in Appendix A is extracted pertaining to fatal crashes, which are subsequently utilised in K-means clustering and visualisation analyses.

The outcomes of rule mining using K-means clustering are visually presented in Figure 7.3.1. To elucidate concealed insights regarding contributing attributes, ten groups are established. The size of each circle in the visual representation signifies the combined support of all rules within a specific group. A larger circle indicates that all rules within that group have relatively high support, suggesting those attributes presented in these rules have high frequency of occurrence to fatal crash. Furthermore, the colour of each circle signifies the common lift value of all rules within that particular group. A darker red shade indicates that the collective lift of all rules within the group is relatively higher, suggesting a highly likelihood that these attributes, when occurring together, contribute significantly to fatal crashes.

Figure 7.3.1 summaries all significantly affected roadway crash attributes to fatal crash, where left-hand side (LHS) indicates the road attributes and right-hand side (RHS) refers to the fatal results. These includes Light condition=6 (Darkness: no street lighting); Road surface condition=1 (Dry); Speed limit=60; Junction detail=0 (Not at or within 20 metres of junction); Regional areas=East Midlands; Urban or rural area=2 (Rural); Trunk road flag=1 (Trunk roads managed by Highways England); Day of week=7 (Saturday); Speed limit=70; Speed limit=50; Regional areas=Scotland; Light condition=4 (Darkness: streetlights present and lit); Regional areas=South West; Regional areas=South East; Month=August; Month=July.

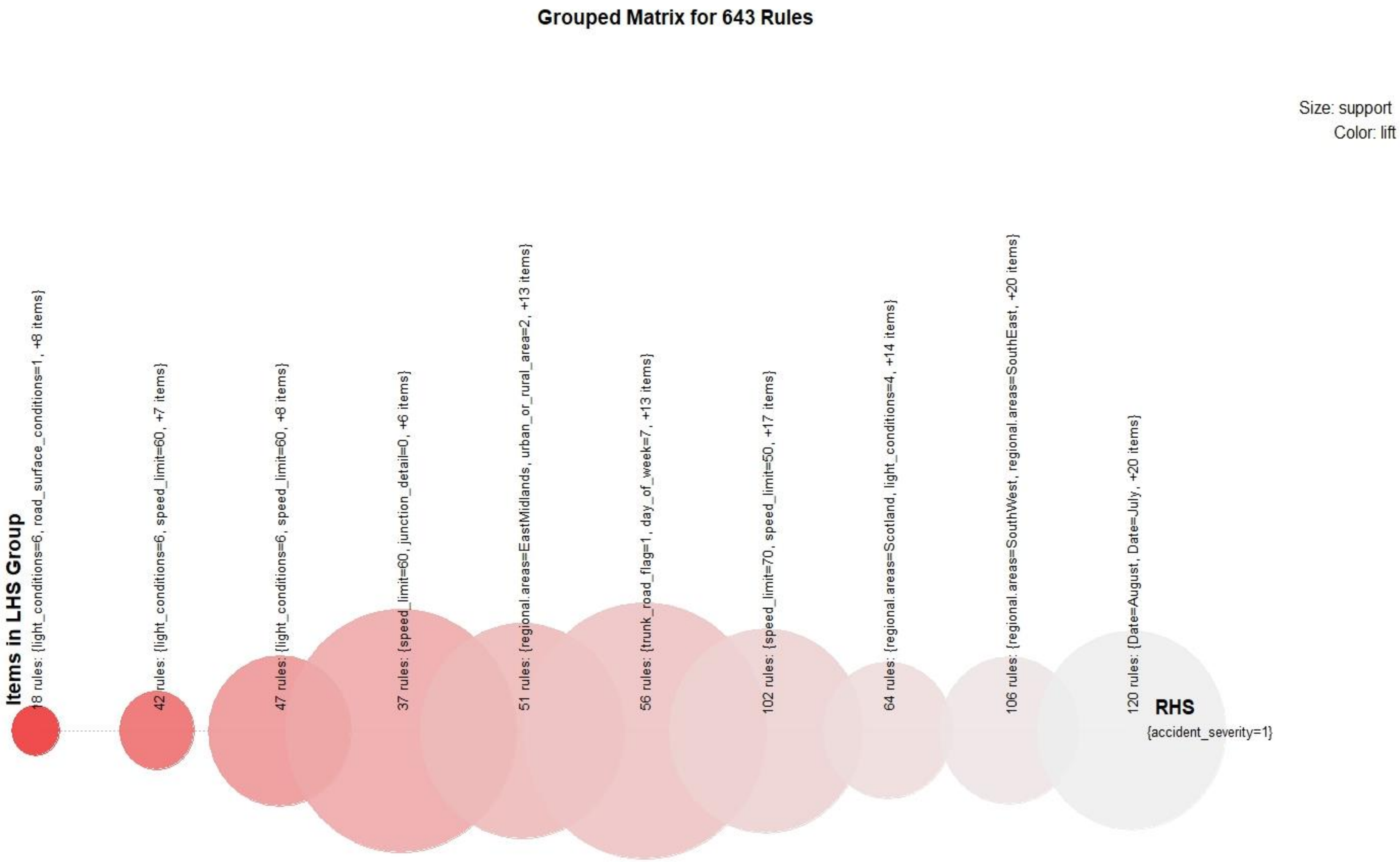

Figure 7.3.1: Fatal crash rules visualisation

In order to extract noteworthy rules, ranking the rules based on different criteria such as support and lift, can help prioritise rules and identify the most relevant and influential ones. These two measures, introduced in Section 6.4.1, represent the frequency with which the rule appears in the dataset and the extent to which its occurrences are statistically associated. Table 7.3.1 and Table 7.3.2 and are shown as below to demonstrate the top ten rules ranking by support and lift respectively.

According to the findings presented in Table 7.3.1, fatal road crashes predominantly occur on roads without human control within 50 metres and on roads without pedestrian crossing facilities within the same distance. Additionally, the data reveals that adverse weather conditions are not a common factor in these incidents, as fatal crashes typically occur under fine weather conditions without high winds. It is evident that single carriageways are the road type where fatal crashes occur most frequently.

Furthermore, Table 7.3.2 reveals a significant correlation between specific road characteristics and fatalities. Notably, parent rules D, H, and J, which will be discussed in the subsequent section 7.4, identify scenarios with a higher likelihood of fatal incidents.

Rule D suggests that fatalities are more likely to occur on roads that are not at or within 20 metres of a junction, are in darkness without street lighting, have dry road surfaces, and are in rural areas. Rule H emphasises that fatalities are more associated on A roads under similar lighting conditions and in fine weather without high winds. Additionally, it has been observed that fatalities more likely occur on A roads, without human control at pedestrian crossings, occur, in darkness without street lighting, under fine weather without high winds, and are situated in rural areas

Table 7.3.1: Ranking of rules by support

No.	Rules	Support	Lift
1	pedestrian crossing human control=0 (No human control within 50 metres)	0.013512	1.031237
2	pedestrian crossing physical facilities=0 (No physical facilities within 50 metres)	0.011652	1.105374
3	Pedestrian crossing human control=0 (No human control within 50 metres), pedestrian crossing physical facilities=0 (No physical facilities within 50 metres)	0.011636	1.109275
4	weather conditions=1 (Fine no high winds)	0.011171	1.013339
5	pedestrian crossing human control=0 (No human control within 50 metres), weather conditions=1 (Fine no high winds)	0.011098	1.039322
6	road type=6 (Single carriageway)	0.010217	1.041155
7	road type=6 (Single carriageway), pedestrian crossing human control=0 (No human control within 50 metres)	0.010136	1.06375
8	Pedestrian crossing physical facilities=0 (No physical facilities within 50 metres), weather conditions=1 (Fine no high winds)	0.009565	1.114218
9	Pedestrian crossing human control=0 (No human control within 50 metres), pedestrian crossing physical facilities=0 (No physical facilities within 50 metres), weather conditions=1 (Fine no high winds)	0.009557	1.11845

10	road type=6 (Single carriageway), pedestrian crossing physical facilities=0 (No physical facilities within 50 metres)	0.008888	1.141241

Table 7.3.2: Ranking of rules by confidence (lift)

No.	Rules	Support	Lift
A	first road class=3 (A road), pedestrian crossing human control=0 (No human control within 50 metres), light conditions=6 (Darkness no street lighting), weather conditions=1 (Fine no high winds), urban or rural area=2 (Rural area).	0.001085	5.823923
B	first road class=3 (A road), light conditions=6 (Darkness no street lighting), weather conditions=1 (Fine no high winds), urban or rural area=2 (Rural area).	0.001085	5.810056
C	junction detail=0 (Not at junction or within 20 metres), pedestrian crossing human control=0 (No human control within 50 metres), light conditions=6 (Darkness no street lighting), road surface conditions=1 (Dry), urban or rural area=2 (Rural area).	0.001239	5.608501
D	junction detail=0 (Not at junction or within 20 metres), light conditions=6 (Darkness no street lighting), road surface conditions=1 (Dry), urban or rural area=2 (Rural area).	0.001239	5.597244
E	first road class=3 (A road), Pedestrian crossing human control=0 (No human control within 50 metres), pedestrian crossing physical facilities=0	0.001093	5.594019

	(No physical facilities within 50 metres) light conditions=6 (Darkness no street lighting), weather conditions=1 (Fine no high winds).		
F	first road class=3 (A road), pedestrian crossing physical facilities=0 (No physical facilities within 50 metres), light conditions=6 (Darkness no street lighting), weather conditions=1 (Fine no high winds).	0.001093	5.581319
G	first road class=3 (A road), Pedestrian crossing human control=0 (No human control within 50 metres), light conditions=6 (Darkness no street lighting), weather conditions=1 (Fine no high winds).	0.001109	5.579131
H	first road class=3 (A road), light conditions=6 (Darkness no street lighting), weather conditions=1 (Fine no high winds).	0.001109	5.545036
I	first road class=3 (A road), junction detail=0 (Not at junction or within 20 metres), Pedestrian crossing human control=0 (No human control within 50 metres), light conditions=6 (Darkness no street lighting), urban or rural area=2 (Rural area).	0.00115	5.527786
J	first road class=3 (A road), junction detail=0 (Not at junction or within 20 metres), light conditions=6 (Darkness no street lighting), urban or rural area=2 (Rural area).	0.00115	5.513061

7.4 Discussion

STATS20 is an introduction document used for completion of road accident report, which provides a detailed explanation of the information to understand the result. The definitions of killed were elaborated in STATS20 corresponding to the severity of crash. A fatal crash is when deaths occur within a period of fewer than 30 days following the accident, which excludes fatalities resulting from natural causes or suicide.

7.4.1 Strategic prospectives

This part will analyse the results of fatal crashes as example from three different perspectives, elucidating the accuracy and impact of tools are elaborated through a comparison of reports sourced from the Department of Transport.

(a) Quantitative analysis

Quantitative analysis in Association Rule Mining is essential for ensuring that the discovered rules are meaningful, reliable, and applicable to the dataset. It helps in making informed decisions and extracting valuable insights from large datasets. Table 7.3.1 and Table 7.3.2 are displayed the top 10 rules in terms of support and lift, which is explained for quantitative analysis.

Rule 1 has a support of 1.35%, indicating that 1.35% of all crashes involve “no human control within 20 metres” crash on-site attributes. Although this support rate is relatively low, when compared with the occurrence of fatal crashes at 1.36% of the entire crash dataset, the rule suggests that more than 99% of fatal crashes observable with this specific attribute, which could still be significant in a large dataset when analysing the fatal crashes. Moreover, rule 4 is evidenced that over 80% of fatal crashes occur under fine without high winds weather condition. According to rule 6, rule 6 has a support of 1%. Correlating with fatal crash, nearly 74% of fatal crashes occur on “singe carriageway”. All these knowledges are based on the calculation from the Equation (6.4-1) reflecting the facts of the data.

In this scenario, rules are acquired with regard to a specific fatal consequent, thus from the view of mathematical definition, confidence and lift are the same measurement but they will be elaborated in separate ways in the context of road safety.

Rule A has the highest confidence and lift within the results derived from Association Rule Mining, registering at 7.9% and 5.82, respectively. The value of 7.9% indicates that when individuals drive vehicles, on “A roads”, in “rural areas”, particularly “in the evening without street lighting”, “in a good weather with no high wind environment” setting, and there is “no pedestrian crossing with human control within 50 metres”, there's a 7.9% chance they will involve in a fatal crash. This suggests a strong association. In the sense of lift, rule A has a lift of 5.82, which is quite high. This signifies a strong association between attributes mentioned above, the LHS of rules, and the consequent of fatal crash. The lift of 5.82 implies that this association is 5.82 times more likely to occur than if the two sides were independent of each other. This finding highlights a robust and statistically significant association between the specified driving conditions and the occurrence of fatal crashes.

Notably, as shown in Figure 7.4.1 Rule H is the parent of rule A, B, E, F and G. Similarly, Rule D serves as the parent rule for Rules C, while Rule J serves as the parent rule for Rule I. What makes this pattern particularly interesting is the observed characteristics in the child rules. Specifically, there is a minor decrease in support, accompanied by a slight increase in both confidence and lift, which suggests that certain conditions may amplify each other's impact to fatal crash.

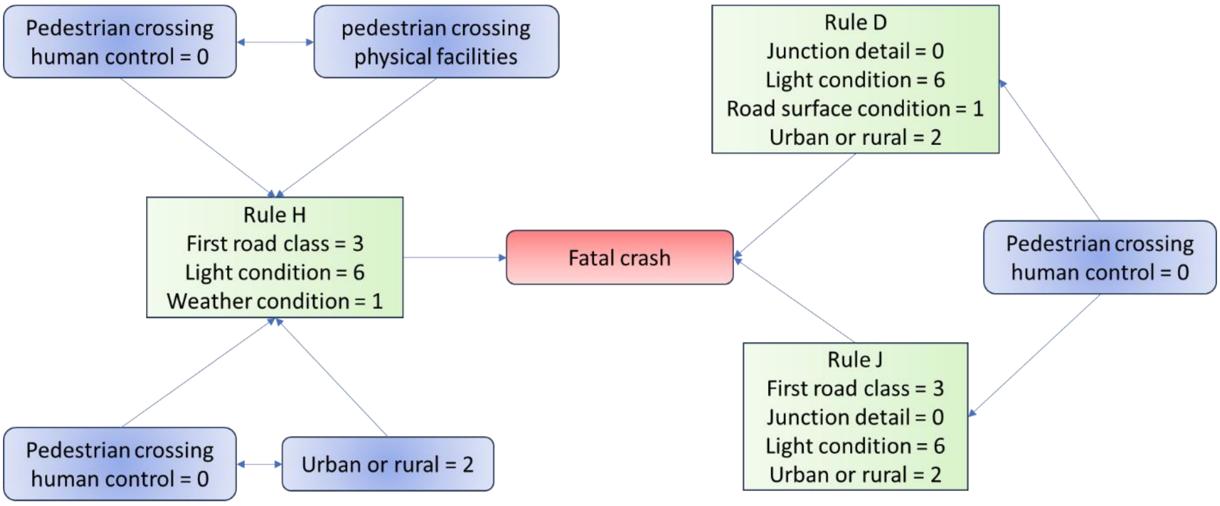

Figure 7.4.1: Visualisation of top 10 fatal rule ranking by lift

(b) Qualitative analysis

Specifically, within the road setting context, the repeatedly presence of “pedestrian crossing human control=0”, “pedestrian crossing physical facilities=0”, are obviously discovered across multiple rules. These factors not only show a relatively high frequency but also demonstrate a robust association with fatal crashes. This underscores the prominent significance of absence of crossing facilities and crossing control as contributory factors to fatal crashes as indicated by the high lift. Conversely, considering the inverse way, no crossing human control and no crossing facilities are frequent fatal roadway crash attributes, which points out the interesting fact that roads were not built for crossing. The absence of pedestrian crossings calls for a deeper investigation on ideal design and strategic placement of crossing facilities. This enhancement should be formulated by an understanding of pedestrian behaviour, aiming to encourage safe road crossing practices. In addition, “road type=6” (single carriageway) is another repeated road feature in fatal incident. To improve the road safety, an in-depth investigation into the attributes of A road and single carriageway, is necessary. This analysis outlines a prospective strategy for enhancing road safety, offering insights to guide precisely targeted interventions with the objective of reducing fatal crashes.

Notably, the presence of "light conditions=6" (Darkness - no lighting) and "weather conditions=1" (Fine no high winds) in several rules suggests a noteworthy impact on fatal crash occurrences, which implies that fatal crashes are more likely under these specific driving environment setting, Darkness - no lighting and fine no high winds. This aligns with common safety knowledge that reduced visibility during darkness can contribute to serious incidents.

In summary, all these rules exhibit similar conditions, indicating consistency in the attributes of fatal crashes and contributors to fatal crashes. The qualitative analysis enables to understand rules in two reversed mindsets based on different meaning of support and confidence/lift. It describes varied attributes of fatal crashes, emphasising the frequent occurrence of diverse characteristics. Simultaneously, it reveals consistent patterns in the rules highlighting the relationship between antecedent of numerous factors such as road class, lighting, and weather conditions, urban or rural areas, pedestrian crossings, junction details and consequent of fatal crashes. These findings can guide further investigation and intervention strategies to enhance road safety.

(c) Overview of visualisation

Hundreds of rules were grouped by K-mean clustering to enhance comprehension of the collective characteristics of the outcomes. The resulting visualisations, depicted in Figure 7.3.1 illustrates distinct metrics for highlighting the most noteworthy rules. As explained in section 7.1, two measurements are used for identifying interesting rules. Concurrently, essential factors for each group are labelled and summarised as outlined below in Table 7.4.1.

Table 7.4.1: Interesting attributes in groups

Group	Attributes	
1	Light condition=6 (Darkness: no street lighting)	Road surface condition=1 (Dry)
2	light conditions=6	Speed limit=60
3	light conditions=6	Speed limit=60
4	Speed limit=60	Junction detail=0(Not at or within 20 meters of junction)
5	Regional areas=East Midlands	Urban or rural area=2 (Rural)
6	Trunk road flag=1 (Trunk roads managed by Highways England)	Day of week=7 (Saturday)
7	Speed limit=70	Speed limit=50
8	Regional areas=Scotland	Light condition=4 (Darkness: streetlights present and lit)
9	Regional areas=South West	Regional areas=South East
10	Month=August	Month=July

Ten clusters were classified by lift, where the attribute labels within each group exerted the most pronounced influence on fatal outcomes. This determination was made by evaluating the relative implication of specific attributes within their respective groups, compared to their occurrence within the entire dataset to understand and extract rules with these outstanding attributes in each group as they could represent a cluster with regards to a specific lift value. All rules considered in this study have a lift value over 1, which consequently provides valuable insights into the underlying patterns of association between roadway crash attributes and fatal outcomes within each cluster.

7.4.2 Robustness check

To verify robustness of the results, 2018 annual report (Department of Transport, 2019) was used for validation. This report conducted statistical analysis utilising crash data from the year 2018. While the report primarily undertook casualty-based analysis, our study focused on mining crash-based rules. In addition, the report presented crash-based statistics, which were adapted for the purpose of conducting a robustness check. However, it is important to note that basic statistics alone provide limited insights, revealing only numerical values such as the count (support of item) without researching for a deeper understanding of the underlying data.

According to the Table 7.4.2, the data reveal an obvious contrast in fatality rates between road classes. Motorways, characterised by higher speeds, exhibit a lower fatality rate (6%) compared to A roads (56%). Notably, the study highlights the importance of considering crash-based rules to acknowledge that A road is not only a frequent pattern to roadway fatalities and has a strong association with fatal crash. In addition, the table underscores a significant rural-urban different impact on fatality rates. Rural roads show evidence of a higher fatality rate (60%) than urban Roads (40%), suggesting potential differences in infrastructure, traffic patterns, or emergency response capabilities between these settings, which aligning with the knowledge from rules.

Furthermore, the statistics across varying speed limits unveils that 30-mph speed limit stands out with a substantial fatality rate of 36%, closely followed by 60 mph at 34%. This challenges the common assumption that higher speed limits directly correspond to higher fatality rates. A critical factor in understanding this phenomenon is the distinction between speed limits and actual driving speeds. Speed limits are regulatory measures set by authorities to manage traffic flow and enhance safety, but they do not necessarily reflect the real-world speeds at which vehicles travel. In urban areas with a 30-mph speed limit, factors such as pedestrian density, frequent intersections, and stop-and-go traffic contribute to a complex driving environment with increased risk of collisions. These conditions can result in a high frequency of accidents, which, despite the lower speeds, can still lead to severe injuries or fatalities.

Table 7.4.2: Reported accidents and rates by severity, road class, and speed limit

Roads	Fatal (count)	Rate (%)
Motorways	92	6
A road	939	56
B road	205	12
Other roads	435	26
Total	1,671	
(Motorway excluded)		
All rural roads	952	60
All urban roads	626	40
Total	1578	
Speed limit		
20 mph	49	3
30 mph	567	36
40 mph	168	11
50 mph	134	8
60 mph	542	34
70 mph	119	8
Total	1579	

As per the data presented in Table 7.4.3, the Southeast exhibits the highest number of fatalities (237). Additionally, fatal incidents are reported in the Southwest, Northwest, and East Midlands regions, indicating a certain degree of conformity with the overview of patterns outlined in grouped fatal rules This observation lends credibility to the consistency and reliability of the established rules.

Table 7.4.3: Reported accidents by region, severity

Region authority	Fatal
Northeast	51
Northwest	184
Yorkshire and The Humber	165
East Midlands	179
West Midlands	167
East of England	163
Southeast	237
Southwest	172

The report demonstrates a noteworthy alignment with the rules derived from the crash database, providing empirical evidence that association rule mining serves as an advanced tool for uncovering hidden information and enhancing insights within the context of road safety. Tailoring interventions based on road class characteristics and critically evaluating the effectiveness of speed limits, especially in rural areas, can contribute to more targeted and impactful safety measures.

The overview provides a clear ranking of how different attributes associated with the consequent and how robust correlation they have. This study provides an integrated methodology to investigate the specific crash factors impact. Meanwhile, the report (Department of Transport, 2019) emphasises that there is no single underlying factor that drives roadway crash. Notably, the available police-reported road casualty data offers only a constrained view. To address this limitation, this methodological approach could be employed involving the integration of diverse data sources related to crashes. This integrated approach aims to provide a strategic perspective on roadway safety to enhance our understanding of road safety dynamics and inform evidence-based policy decisions.

7.5 Summary

This Chapter presented a methodology for identifying and analysing fatal crash contributors using association rule mining and K-means clustering. By examining crash database, critical patterns and associations are extracted that provide a comprehensive understanding of the factors contributing to fatal crashes. The visualisations and rules derived from strategic data analysis highlight specific road attributes, environmental conditions, and temporal factors that significantly influence crash outcomes. This data-driven approach provides valuable insights for tailored investment plan to enhance road safety.

The upcoming Chapter will explore the adaptability and applicability of the proposed methodology in alignment with diverse objectives set forth by various legislative levels. This section will emphasise how this methodology can be applied using same database to achieve a comprehensive understanding of road infrastructure relating to crash catering to a range of legislative goals. By exploring the intersection of methodology and legislative objectives, the Chapter aims to demonstrate the versatility of the approach in addressing different policy aims, thereby enhancing its overall utility and relevance in the field of road safety analysis.

Chapter 8. Exploration of the Adaptive Data Analysis

8.1 Introduction

Integration of association rule mining and K-means clustering offers adaptability and applicability across various levels of governing bodies, catering to diverse strategic objectives. At the local authority level, decision-makers may seek to understand the influence of road infrastructure on crash occurrences, enabling the formulation of tailored road safety enhancement strategies within specific geographic areas. Similarly, at the national road division level, decision-makers focus on the broader network, emphasising considerations such as road class to crashes. This section provides a broad overview of how this methodological framework can be applied for crash data analysis to address distinct strategic inquiries across different administrative levels.

To illustrate the adaptability of the proposed methodologies, the Chapter will present two scenarios with distinct objectives. The first scenario aims to explore the internal relationship between crash occurrences in London and various road infrastructure attributes. The second scenario seeks to understand the association between crash site attributes and road classification. Through these examples, the Chapter will demonstrate how the methodology can be tailored to meet the specific needs of different levels of governance, thereby highlighting its versatility and potential for enhancing road safety analysis.

8.2 London scenario

The choice of London for this extended analysis is primarily driven by its unique combination of high traffic density, diverse road infrastructure, and complex accident patterns. As the capital and a major transportation hub, London provides a rich context for studying the relationship between road infrastructure and crash occurrences. Its varied road types and urban density offer valuable insights into how infrastructure influences accident rates, which can be generalised to other areas with similar characteristics.

However, when considering other areas in the UK, results may vary significantly due to differences in road characteristics, traffic flow, population density, and even weather conditions. For example, rural areas or smaller cities may have less traffic congestion but higher speed limits, which could influence accident severity differently than in urban settings like London. These differences would require the development of adaptive methodologies to address regional differences in influencing factors, thereby highlighting the need for such an extension to facilitate comprehensive exploration and analysis.

The analysis utilises the same year of 2018 data from the STATS 19 database, which recorded 25,662 crash occurrences in London. To conduct the association rule mining, a minimum support and confidence of 0.02 were established, ensuring that 90% of the crash data in London is considered for extracting significant rules. The maximum number of rules was set to 7, determined through the methodology detailed in Chapter 7.

This study employs a combined quantitative and qualitative approach to understand the rules and uncover hidden insights essential for strategic management. However, this section will not provide exhaustive details; instead, it will highlight a few key rules to illustrate the significance of the associations discovered. The results are visualised in Figure 8.2.1, and representative rules from each group are summarised in Table 8.2.1, providing crucial information.

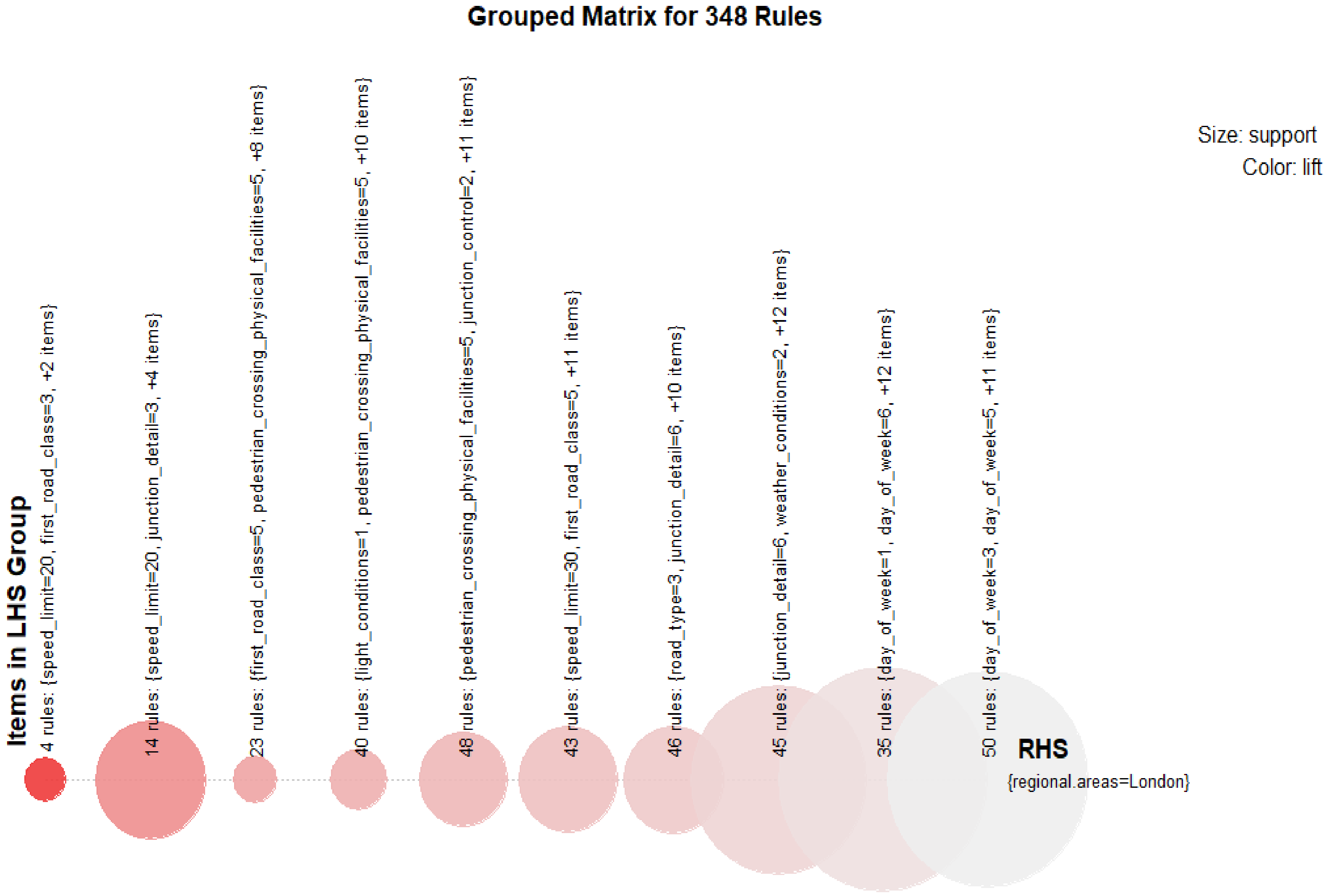

Figure 8.2.1: Overview of rules to crash occurrence in Londo

Table 8.2.1: Representative rules within each cluster

Group	Representative rules
1	First road class=3 (A road), speed limit=20, urban or rural area=1 (Urban), trunk road flag=2 (non-trunk).
2	Speed limit=20, junction detail=3 (T or staggered junction), urban or rural area=1 (Urban), trunk road flag=2 (non-trunk).
3	Accident severity=3 (Slightly injured), first road class=5 (C road), urban or rural area=1 (Urban), trunk road flag=2 (non-trunk). First road class=3 (A road), pedestrian crossing physical facilities=5 (Pedestrian phase at traffic signal junction), road surface conditions=1 (Dry), urban or rural area=1 (Urban), trunk road flag=2 (non-trunk)
4	Pedestrian crossing physical facilities=5 (Pedestrian phase at traffic signal junction), light conditions=1 (Daylight), urban or rural area=1 (Urban), trunk road flag=2 (non-trunk).
5	Pedestrian crossing physical facilities=5 (Pedestrian phase at traffic signal junction), trunk road flag=2 (non-trunk). Accident severity=3 (Slightly injured), junction control=2 (Auto traffic signal), road surface conditions=1 (Dry), urban or rural area=1 (Urban), trunk road flag=2 (non-trunk)
6	First road class=3 (A road), speed limit=30, light conditions=4 (Darkness - lights lit), trunk road flag=2 (non-trunk). First road class=5(C road), trunk road flag=2 (non-trunk).
7	Junction detail=6 (Crossroads), road surface conditions=1 (Dry), urban or rural area=1 (Urban), trunk road flag=2 (non-trunk). Accident severity=3 (Slightly injured), road type=3 (Dual carriageway), weather conditions=1 (Fine), road surface conditions=1 (Dry), trunk road flag=2 (non-trunk).
8	Junction detail=6 (Crossroads), road surface conditions=1 (Dry), urban or rural area=1 (Urban). Weather conditions=2 (Raining no high winds), urban or rural area=1 (Urban), trunk road flag=2 (non-trunk).

9	Day of week=1 (Sunday), urban or rural area=1 (Urban). Day of week=6 (Friday), road surface conditions=1 (Dry), urban or rural area=1 (Urban).
10	Accident severity=3 (Slightly injured), day of week=3 (Tuesday), road surface conditions=1 (Dry), trunk road flag=2 (non-trunk). Day of week=5 (Thursday), road surface conditions=1 (Dry), trunk road flag=2 (non-trunk).

Group 1, identified as the cluster with the strongest association to crash occurrences in London, highlights it significance. Specifically, it reveals that non-trunk roads categorised as A road, featuring a speed limit of 20 in urban areas of London, exhibit an enhanced likelihood of causing crashes. Moreover, the case of road users on non-trunk roads with a 20-mph speed limit, particularly at T or staggered junctions in urban areas, significantly increases the probability of incidents in London.

According to the result, several strategic recommendations and policy initiatives can be formulated to improve road safety in London. Enhanced traffic calming measures, such as speed bumps, should be implemented on non-trunk A roads with a 20-mph speed limit. Additionally, redesigning T and staggered junctions to enhance visibility and reduce conflict points is crucial. There should be increased enforcement of speed limits through the deployment of speed cameras and regular police patrols, accompanied by public awareness campaigns to educate drivers about the dangers of speeding and the importance of adhering to the 20-mph speed limit. Furthermore, infrastructure for pedestrians and cyclists should be improved by creating dedicated bike lanes, enhancing crosswalks with better lighting and signage, and installing pedestrian islands.

8.3 Strategic road network scenario

In the analysis, crash data from the same year 2018 is utilised, encompassing a total of 4566 occurrences on the strategic road network (SRN). This network comprised 1865 miles of motorway and 2571 miles of trunk A-roads. The strategic road network plays a vital role in facilitating both national and local economic growth by fostering connectivity between communities and facilitating essential activities (Nick, 2015). To ensure the inclusion of a significant portion of crash data concerning to incidents on SRN, three parameters are determined, the minimum support and confidence established at 0.004 and the maximum number of rules restricted to 8.

The result of the analysis shown in Figure 8.3.1 offers clear insights can inform strategic management initiatives aimed at enhancing road safety. Notably, the result highlights specific regional areas, such as the Northwest, East of England, and Yorkshire and the Humber, which exhibit high frequencies or associations with crash occurrences. Furthermore, the analysis reveals a noteworthy pattern, indicating that crashes predominantly occur on Thursdays, Fridays, and Saturdays.

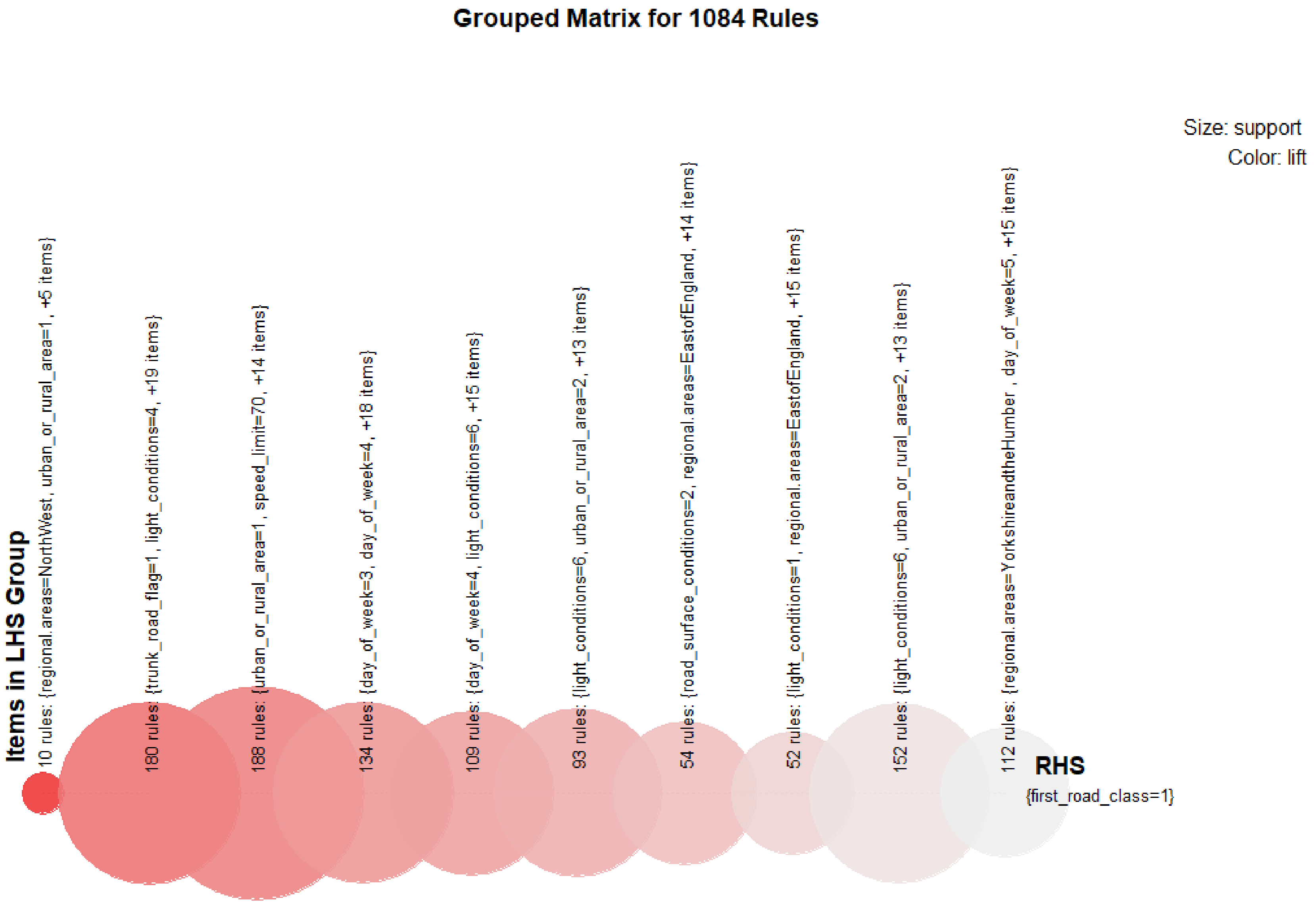

Figure 8.3.1: Overview of rules to crash occurrence on SRN

In response to these findings, road administrators and decision makers can formulate strategy and policy to improve road safety. Enhancing road safety in the identified high-risk regions can be achieved by conducting detailed investigations into the specific factors contributing to crashes in these areas. Furthermore, increasing enforcement and supervision on the strategic road network during Thursdays, Fridays, and Saturdays is essential. Deploying additional traffic police, utilising automated enforcement technologies, and conducting public awareness campaigns focused on high-risk behaviours during these days can help reduce the incidence of crashes. By prioritising resources and interventions based on the spatial and temporal patterns identified in the analysis, authorities can develop a more focused and effective road safety strategy. This approach not only addresses the specific areas and times of heightened risk but also ensures that resources are used efficiently to maximise their impact on reducing crash occurrences and enhancing road safety for all users.

8.4 Summary

Chapter 8 justifies the use of the proposed methodology, integration of association rule mining and K-means clustering, to enhance road safety analysis across various governance levels. Two scenarios illustrate this adaptability by using the same year of data in STATS 19.

In the first scenario, it reveals a higher likelihood of crashes on non-trunk A roads with a 20-mph speed limit, especially at T or staggered junctions. Recommendations include implementing traffic calming measures, redesigning junctions, increasing speed limit enforcement, and improving pedestrian and cyclist infrastructure.

The second scenario focuses on the strategic road network (SRN), identifying high-risk regions (Northwest, East of England, Yorkshire and the Humber) and a pattern of crashes occurring predominantly on Thursdays, Fridays, and Saturdays. Strategies include investigating crash causes in high-risk areas, increasing enforcement on high-risk days, and prioritising resources based on identified patterns.

This Chapter demonstrates the feasibility of proposed methodology for different levels of administrations, enabling decision-makers to identify the current problems on road networks. The next Chapter will introduce the concept of strategic resource optimisation in road safety management. It will explore optimisation techniques by reviewing current research findings in road safety optimisation and providing a comprehensive overview of various techniques, widely acknowledged for their efficacy in resource allocation. Additionally, the Chapter will highlight the importance of data-driven optimisation in road safety management and propose an innovative methodology for improving safety outcomes.

Chapter 9. Road Safety Resource Optimisation

9.1 Background

After identifying problems within the road network—where incidents lead to not only substantial human suffering but also considerable economic burdens, such as medical costs, lost productivity, and infrastructure damage—policymakers and road authorities bear the primary responsibility for allocating resources to enhance overall road safety. The insights gained from data analysis (see Chapter 7) provide an evidence-based foundation for these investment plans. At the core of an effective strategy is the development of a data-driven investment plan aimed at achieving optimal safety outcomes for society.

Certain studies have demonstrated the ability of enhancing road safety through the implementation of structured safety management processes. For instance, the Highway Safety Improvement Program (HSIP) outlined a three-phase approach to allocate resources effectively for road safety, which involves identifying high-risk areas, proposing appropriate countermeasures, and allocating funds within budgetary constraints (Herbel et al., 2010). This methodical approach ensures a systematic assessment of safety priorities and efficient resource allocation.

Similarly, international road assessment programme developed a suite of tools based on iRAP methodology to support road infrastructure safety management. It includes four protocols: risk mapping, star rating, safer road investment plans and performance tracking (International Road Assessment Programme, 2023), which aligns with the three-step structure promoted by HSIP.

In 2022, Waka Kotahi NZ Transport Agency (2022) developed a definition of 'one network framework' (ONF) to put people, place and movement at the heart of planning and investment, which allows land use and transport planning to be integrated. The ONF aligns with strategic transport planning at all levels, where road to zero strategy is one of the objectives to achieve zero death on road. Meanwhile, they adopted an investment method that utilised three investment criteria—government policy statement (GPS) on land transport alignment, scheduling, and efficiency—to prioritise proposed activities (Waka Kotahi NZ Transport Agency, 2021a). It is an interesting and decisive transport funding allocation framework aligning with strategic objectives.

In recent years, scholars and practitioners have investigated diverse methodologies and frameworks aimed at optimising the allocation of resources for road safety. These approaches involve traditional cost-benefit analyses to prioritise countermeasures, as well as more advanced optimisation techniques for selecting combinations of interventions. However, the search for an optimal investment plan in road safety strategic management is hindered by a variety of practical constraints which are often conflict, leading to trade-offs between benefits, efficiency, and costs. Additionally, the unmeasurable impacts on road network operations exacerbate these challenges (Odoki et al., 2015; Augeri et al., 2021; Byaruhanga & Evdorides, 2022). These conflicting objectives and unaccountable effects emphasise the complexities and gaps in current road safety strategic management practices. Decision-making subjectivity and expert preferences often overlook those pertinent road network issues, highlighting the need for advanced and comprehensive optimisation methodologies that accommodate diverse objectives with less human intervention.

This study seeks to overcome these limitations by introducing a novel two-stage optimisation model that integrates multi-objective optimisation with association rule mining. By systematically analysing crash databases, this model shifts decision-making from expert-driven heuristics to a data-driven, computationally optimised framework, reducing human bias while enhancing strategic planning efficiency. The approach also studies unmeasurable impacts of construction by considering the lifecycle of countermeasure, making optimisation more objective and quantifiable. Moreover, using the non-dominated sorting genetic algorithm, the model optimally balances conflicting road safety investment objectives, lifting the constraints of conventional trade-offs.

The effectiveness of this methodology is demonstrated through a case study from Utrecht, developed by iRAP, where significant improvements across multiple road safety objectives are achieved. While this approach marks a substantial advancement in strategic road safety management, further refinements are required to fully capture the interactive effects of multiple road attributes in risk assessment. By addressing long-standing limitations in road safety optimisation, this study provides a scalable, data-driven framework that enables road authorities to design targeted, evidence-based investment plans tailored to their specific network characteristics.

This study exclusively examines investment planning for road safety, focusing on the third step of the Highway Safety Improvement Program (HSIP). The identification of high-risk areas and the proposal of countermeasures will not be addressed within the scope of this investigation. The subsequent sections provide a comprehensive review of existing literature on optimisation techniques applied to road safety investment plan and elaborate the theoretical framework of the multi-objective optimisation following an elaboration of Non-dominated Sorting Genetic Algorithm II (NSGA II), which is applied to search the optimal solutions in this research.

9.2 Related Works

The optimisation of road safety resource allocation is a critical endeavour in road infrastructure engineering, where the aim is to minimise crash or severity for improving overall safety outcomes within constrained budgets. Over the years, researchers have proposed various methodologies and frameworks to address this complex challenge.

The mixed integer linear programming (MILP) model was developed by Banihashemi & Dimaiuta (2005) to maximise the benefits obtained by predefined alternatives for different segments of the highway which was achieved by minimising the expected number of crashes. Their subsequent work (Banihashemi, 2007) extended this model to incorporate travel time considerations, seeking to minimise both crash and delay costs within budget constraints.

Chowdhury et al. (2000) introduced a multi-objective approach that emphasises keeps the objectives in their respective units and provides a set of solutions rather than a single optimal. The core to their methodology is the introduction of disutility values, which estimated the relative severity of different crash levels. By identifying causal factors on road segments and suggesting appropriate countermeasures, the study aims to optimise resource allocation while minimising expected loss of disutility at all road types.

Lambert et al. (2003) proposed a multi-objective framework tailored for guardrail resource allocation, which facilitated interpretation of variety of benefits and costs in their own units. Their approach allows for subjective evaluations of different objectives, accommodating the varied needs and preferences of stakeholders.

Importance segregated multi-objective optimisation (ISMO) model was proposed by Yu et al. (2012). They separated objectives into two levels: decision and supporting level, so that fewer optimal solutions would be obtained from decision level and the process avoided overrating less important objectives.

Mishra et al. (2015) presented optimal multiple resource allocation strategies to improve urban intersections under policy and budget constraints. focused on optimal resource allocation strategies for urban intersections, with a particular emphasis on equity in safety benefits and allocation within counties. Their approach considers policy and budget constraints to maximise safety benefits while ensuring fair distribution within communities.

Augeri et al. (2021) proposed an interactive multi objectives optimisation method using dominance-based rough set approach (DRSA) to provide an explicit structure for road safety decision makers. They applied DRSA to formulate decision rules after acquirement of pareto solutions, iteratively refining constraints until an optimal solution is reached.

While the literature review provides a brief overview of various methodologies and frameworks for optimising road safety resource allocation, a notable gap lies in the limited discussion on the integration of emerging technologies and data-driven approaches. Artificial intelligence, machine learning, and big data analytical methods offer promising opportunities for enhancing road safety initiatives. There is therefore a need for further exploration of holistic and data-driven optimisation procedure that consider a broader socio-economic impacts of road safety interventions.

9.3 Theoretical Framework

9.3.1 Multi-objective optimisation techniques

According to Chang (2015), the primary objective of multi-objective optimisation is to capture the preferences of the designer concerning the prioritisation importance of objectives. Hence, multi-objective optimisation techniques are classified based on how these preferences are articulated: those with a priori articulation and those with a posteriori articulation.

Methods involving a priori articulation of preferences require specifying preferences first in terms of goals or the relative significance of different objectives. Many of these approaches incorporate parameters such as coefficients, exponents, or constraint limits, which can be adjusted to reflect the preferences or dynamically modified to discover multiple solutions. Among these methods, the weighted-sum method is the simplest and most widely used, where multiple objectives are combined into a single scalar function for optimisation.

Expressing preferences for objective functions can sometimes be challenging, yet choosing from a set of Pareto-optimal solutions remains feasible. In contrast, methods with a posteriori articulation of preferences, also known as generate-first-choose-later approaches, defer preference articulation until after the generation of potential solutions. Non-dominated sorting genetic algorithms was proposed by Srinivas & Deb (1994) and revised to NSGA II (Deb et al., 2002) is one of this kind of method.

This thesis aims to minimise human influence on decision-making processes by employing the NSGA II method, a posteriori articulation of preferences, to derive Pareto-optimal solutions. These solutions will be ranked based on data analysis results to provide an unbiased ranking list. To facilitate comprehension of NSGA II and for the development of the model, it is essential to establish fundamental definitions.

9.3.2 Basic definitions of Pareto-based approaches

Obviously, the multi-objective optimisation implies the exploration of optimal solutions, constituting a process essentially based on trade-off analysis. It is impossible to obtain a solution that is the best for all considered objectives. In comprehending the mechanics of multi-objective optimisation in facilitating trade-off analysis, it becomes imperative to understand several foundational definitions. As interpreted by Chang (2015), these definitions serve as crucial conceptual foundations within multi-objective optimisation frameworks.

Definition 1: Nondominated and Dominated. A Pareto optimal point has no other point that improves at least one objective without detriment to another; that is, it is non-dominated, vice versa.

Definition 2: Pareto front. A set of solutions that are non-dominated to each other but are superior to the rest of solutions in the search space.

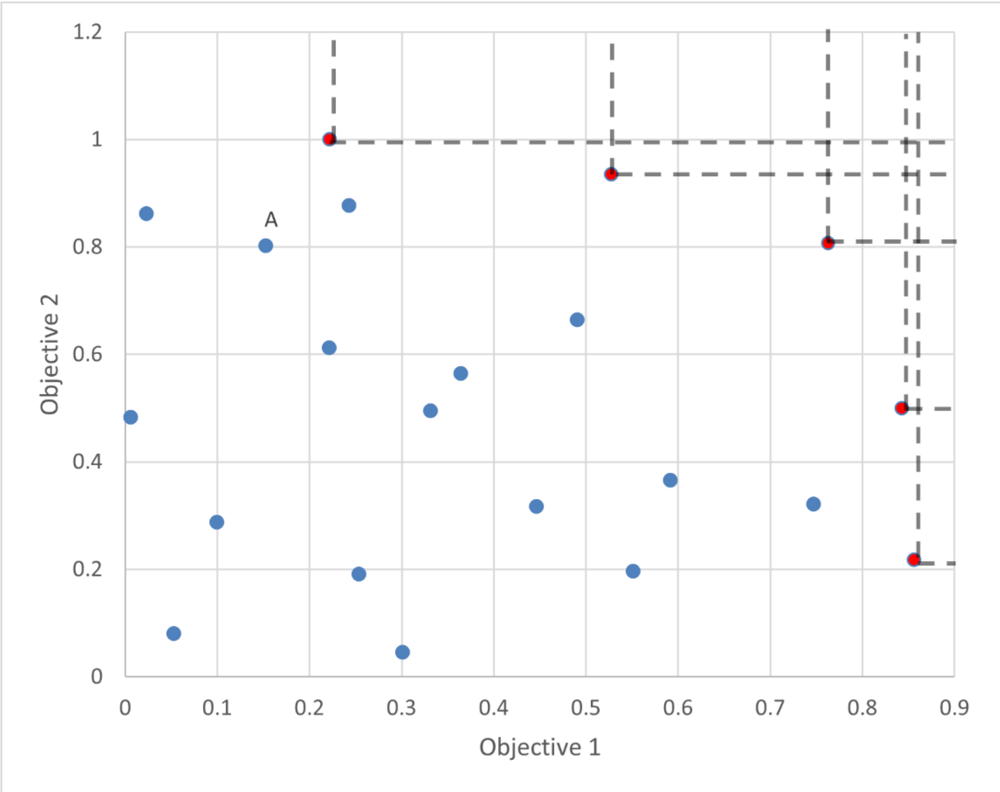

Figure 9.3.1: Random data set for definition illustration

By elucidating these fundamental concepts, researchers gain a general understanding of the operational mechanisms governing trade-off analysis within the context of multi-objective optimisation. If a random data set is considered with two objectives aiming at maximisation a plotted may be created as shown in Figure 9.3.1.

As depicted in the above figure, the red dots are the set of non-dominated points called pareto front, each representing a unique trade-off between competing objectives. The Pareto front is deemed Pareto optimal (combination of countermeasures), as they offer the best achievable compromises across the objectives under consideration. Conversely, points located below or within the Pareto front as shown as blue dots are considered dominated, as they can be surpassed by other solutions in at least one objective without any corresponding improvement in other objectives.

9.4 Summary

Chapter 9 focuses on the critical issue of reducing road traffic accidents and their impacts through effective road safety resource optimisation. It emphasises the global significance of road accidents in terms of human lives lost and economic costs, underscoring the need for efficient resource allocation strategies.

The background section surveys strategic frameworks and methodologies used worldwide to enhance road safety, while related works review various optimisation models aimed at maximising safety benefits within specific constraints. The theoretical framework focuses on multi-objective optimisation techniques and some basic definitions are introduced to elucidate trade-off analysis in multi-objective optimisation.

This Chapter lays the groundwork for Chapter 10, which will develop and validate a comprehensive methodology for a multi-objective resource optimisation prototype. Building upon the theoretical foundations established in this Chapter, the subsequent Chapter aims to demonstrate the practical application and effectiveness of the proposed model in automating and enhancing decision-making processes for allocating road safety resources

Chapter 10. Prototype for Resource Optimisation

10.1 Methodology

In this section, a multi-objective optimisation integrated with association rule mining is introduced for road safety strategic resource allocation considering estimated net present value (NPV) of benefit, cost per KSI saved and service life of countermeasure.

Since the objectives and the constraints involved are often heterogeneous and conflicting, a multi-objective optimisation methodology is proposed. Decision makers are trying to minimise the overall cost per KSI saved and minimise the construction impact to road network to maximise the overall benefits for road network and reduce KSI crash. This method is a data driven procedure composed of two methods to avoid subjectivity and bias in selection of countermeasures. The Pareto optimal set generates a sample of solutions (e.g. combination of countermeasures) where the non-dominated soring genetic algorithm (NSGA II) was employed. The results of association rule mining are applied to weigh the optimal solutions through aggregation method, where a rule-based model expressed in terms of easily understandable “if…, then…” decision rules is used. Figure 10.1.1 illustrates how these two different methods are integrated to facilitate decision maker developing investment strategy with minimum subjective interference.

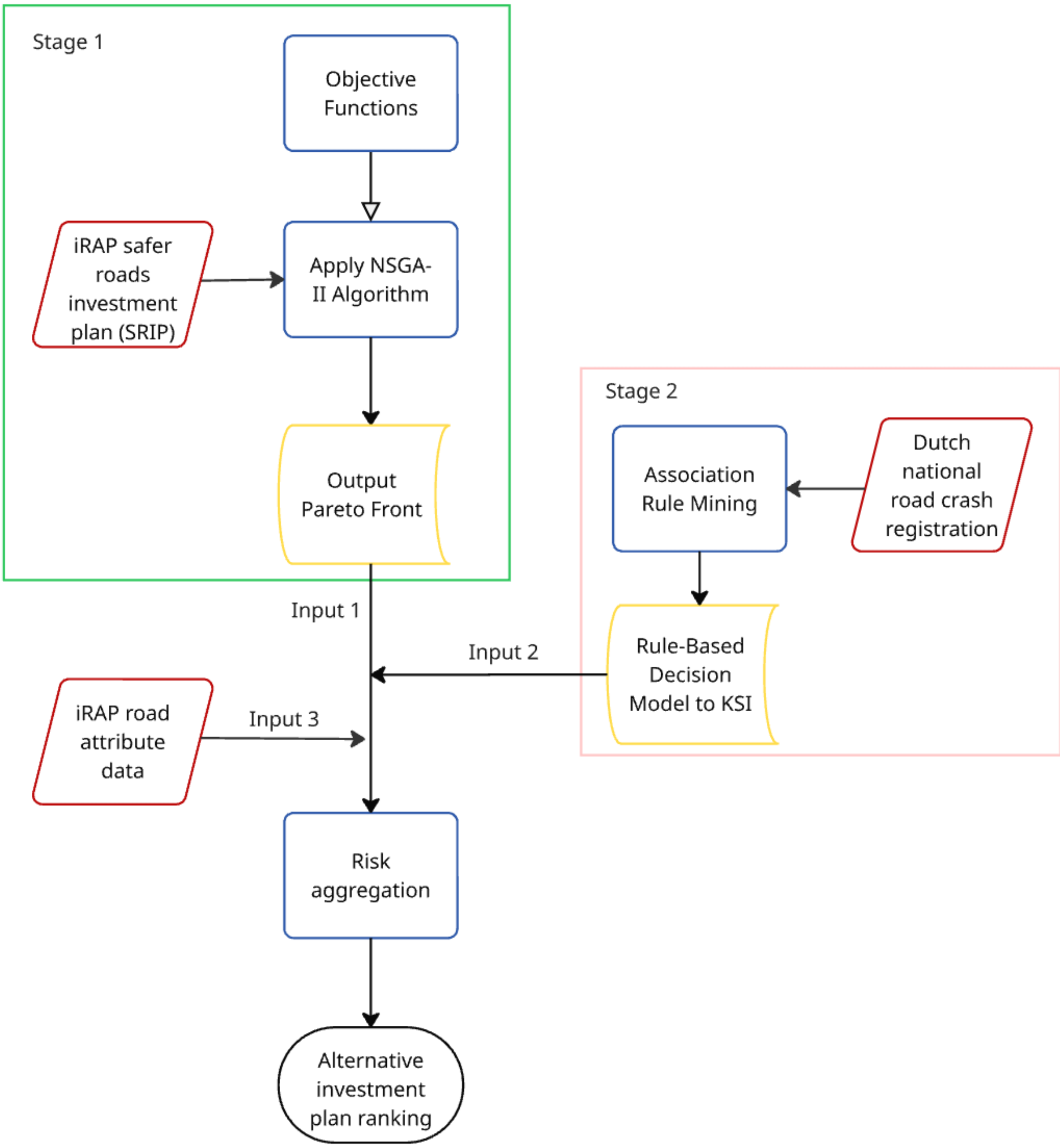

Figure 10.1.1. Flowchart of integrated optimisation method

10.1.1 Stage 1: multi-objective optimisation

The first stage involves generating a Pareto optimal set of solutions using the Non-Dominated Sorting Genetic Algorithm II (NSGA II) (Deb et al., 2002), implemented in R Studio. NSGA II is chosen for its ability to provide a better spread of solutions and convergence compared to other multi-objective evolutionary algorithms.

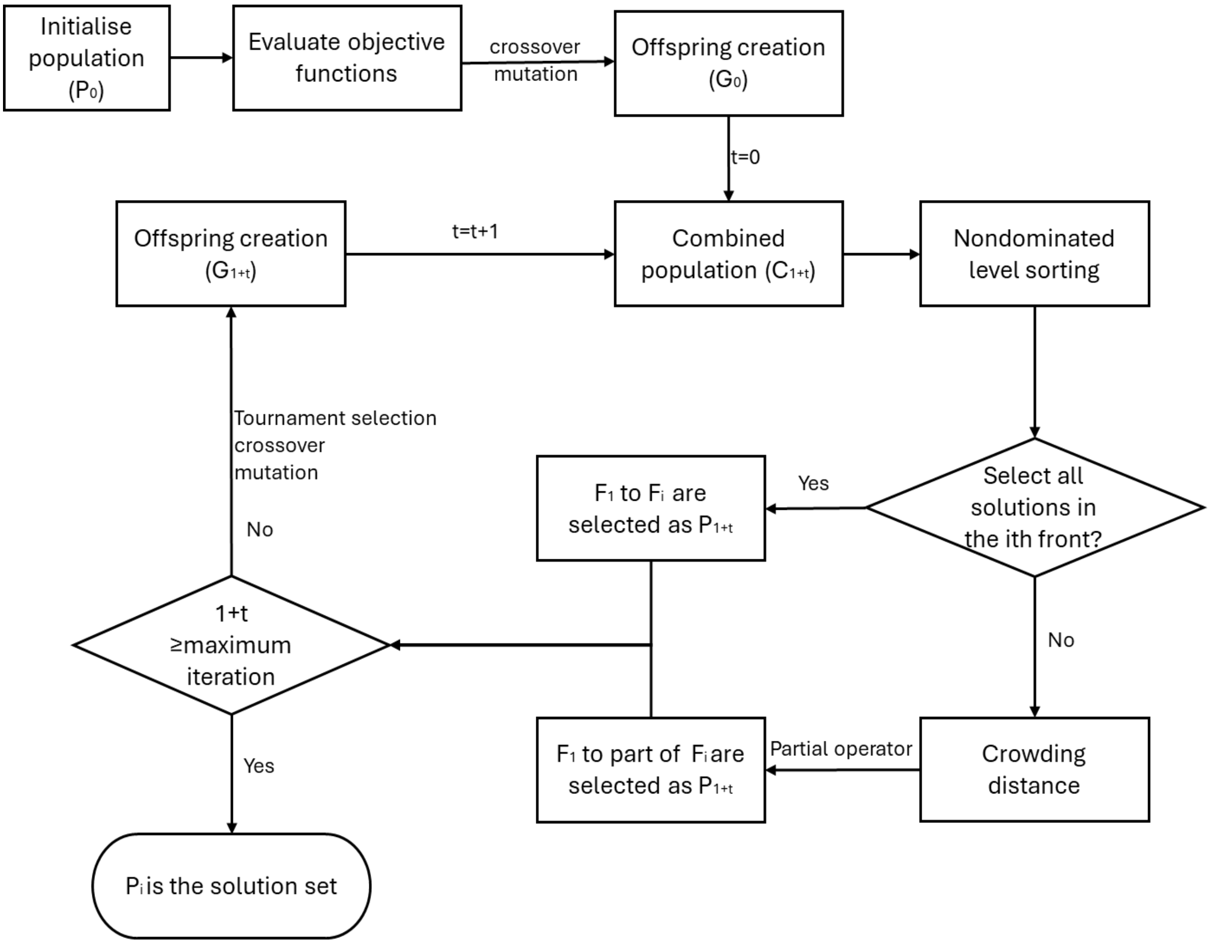

Figure 10.1.2: Flow chart of NSGA II

It consists of two main steps: nondominated levels sorting and diversity preservation to search optimal solutions. The overall procedure of NSGA II is illustrated in Figure 10.1.2 and demonstrated as followed. First, a random parent population (P_0) of size N is initialised evaluating objective functions to rank their front tier level. Offspring (G_0) of size N is created by elitism election recombination, and mutation operators. The loop will begin once the first initial generation ends. The combined population of C_1 is composed of P_0 and G_0 with size of 2N. Nondominated levels sorting is used to rank the front tier level (F_i) and search best solutions in the combined population of C_1. Next, in order to identify the new population (P_2) with size of N, which is composed of F_1 to F_i, crowding distance is calculated for each solution in ith front tier level to guarantee a good spread of solutions, where a partial operator was identified for binary tournament selection. Between two solutions, solution with a lower rank of front tier level is preferred; solution in a less crowded region if preferred if they are ranked as the same front tier level. The new population (P_1) of size N is now used for selection, crossover, and mutation to create an offspring (G_1) to make up the new combined population of C_2. The process will iterate until the maximum generation is achieved (Deb et al., 2002).

10.1.2 Stage 2: association rule mining

Once solutions are searched out, it is necessary to identify a most suitable solution based on different weight criteria, where some studies (Augeri et al., 2021; Yu & Liu, 2012) tried to employ multi criteria and interactive method to select the best solution based on experts' preference and predefined criteria. Existing methods are subjective resulting in bias and neglect of the importance of historical data. Thus, association rule mining is introduced to provide data-driven evidence in the context of road engineering.

Association rule mining is a widely used data mining method to extract hidden information for road safety, which reflects the overall safety performance and potential problems of road network in the form of "if …then…". There are some studies focusing on employing association rule mining to carry out crash data analysis, where the rules were extracted by simple setting measure of interestingness such as minimum support and confidence. There is a crucial parameter named lift indicating the robustness of association between antecedent (e.g. road type) and consequent (e.g. fatalities) in a rule (Berry & Linoff, 2004). It quantifies the relationship between the observed support and the support that would be anticipated under the assumption of independence between the left-hand side (LHS) and right-hand side (RHS). The value of lift implies how strong association of a rule, where a value of 1 indicates the items in rules are independent, a value of over 1 indicates items are more likely to occur than if items were independent to each other. Therefore, road attributes data at all sites and road sections for proposed road improvements are required to carry out lift weight aggregation.

For holistic road safety improvement, prioritising the reduction of KSI incidents is a fundamental goal in road safety. EU Road Safety Policy Framework 2021-2030 aims at halving the number of fatalities and serious injuries on European roads by 2030, as a milestone on the way to 'Vision Zero' (European Commission, 2019), supporting the importance of understanding of KSI incidents to improve overall public safety and well-being. Therefore, this study employs lift values as indicators to gauge the strength of associations contributing to KSI crashes. A higher lift value in a rule signifies a greater likelihood of resulting in fatal or serious injury outcomes. Given that the performance of investment plan is based on overall countermeasure effectiveness, resulting in various characteristics within the same road attribute category, the traits with higher lift values are used for lift aggregation. Ultimately, the lift aggregation countermeasure is calculated by summing the lifts of all associated road attributes (see eq. (10.1-1)). Similarly, the lift aggregation of combinations is the sum of lift of all selected countermeasures, which stands for the risk level of roads where these selected countermeasures are required to enhance road safety.

$$Lift\ aggregation\ of\ combination = \sum_{i=1}^{i} k_i \left(\sum_{n=1}^{n} l_{n,i} \right) \tag{10.1-1}$$

Where n indicates the nth road attributes associated to fatal crash, i denotes the i^{th} countermeasure proposed for road investment plan, $l_{n,i}$ is the lift value of the rule specified the strongness of association between nth attribute and KSI outcome for i^{th} countermeasure option, and k_i is a binary parameter denoting that k_i=1, if the i^{th} countermeasure is selected in the combinations, otherwise, k_i= 0.

10.1.3 Multi objectives identification

The optimisation model is structured to address multiple objectives, and the objectives are considered from various perspectives. Benefit cost analysis (BCA) and cost effectiveness analysis (CEA) stand as two fundamental approaches for evaluating the efficacy of countermeasures. Within the iRAP methodology, the benefit cost ratio (BCR) serves as a key metric for assessing countermeasures in economic terms. This involves computing the present value of safety benefits, considering the prevention of fatalities and serious injuries alongside the economic valuation of human life and severe injury. The costs associated with countermeasures are derived from construction expenses and data on service life. While BCR effectively quantifies the monetary implications of each countermeasure, decision-makers, operating under budgetary constraints, often aim to maximise the net present value of benefits, where the first objective is determined.

Behnood & Pino (2020) proposed an optimisation framework grounded in cost effectiveness analysis. Their procedure seeks to minimise the costs of countermeasures while simultaneously maximising the reduction in the number of crashes. A critical aspect of this approach lies in understanding the resource allocation required for each additional unit of prevented fatalities and serious injuries. Accordingly, the second objective is identified as the minimisation of the cost effectiveness ratio (CER), which quantifies the resources expended per additional unit of prevented fatalities and serious injuries.

From the economical perspective, this model has chosen the service life as the third objective. Service life is a sensible measure in the way that it indicates the duration of the countermeasure after the implementation. Thus, the longer the service life, the more economical benefits (larger net present values) the countermeasures would yield for the same cost of implementation (Yu et al., 2012). In addition, according to the study of Elvik (2014) and Byaruhanga & Evdorides (2022), the service life could affect the selection of countermeasures. In instances where the budget is sufficiently robust, prioritising countermeasures characterised by long service life becomes instrumental in optimising economic advantages and ensuring enduring safety benefits. This strategic emphasis on countermeasures with continued effectiveness contributes to a reduction in treatment recurrence at specific locations, thereby minimising disruptions to the operational efficiency of the road network.

Therefore, three objectives optimisation model is formulated focusing on: 1) maximise the net present value of benefits (see Eq. (10.1-2)); and 2) minimise the cost per KSI saved; (see Eq. (10.1-3)); and 3) minimise the total count of countermeasures characterised by a short service life (SSL) (see Eq. (10.1-4)).

$$MAX\ \ f_1 = \sum_{i=1}^{n} (B_i k) \tag{10.1-2}$$

$$MAX\ \ f_2 = \sum_{i=1}^{n} (E_i k) \tag{10.1-3}$$

$$MIN\ \ f_3 = \sum_{i=1}^{n} k_{i,\ \ SL=short} \tag{10.1-4}$$

In these equations, n represents the total number of recommended improvement activities outlined in the iRAP safer road investment plan. The parameter i signifies the i^{th}

countermeasure proposed for road enhancement, where B_i denotes the NPV of benefits by implementing the i^{th} countermeasure, and E_i indicates the cost per KSI saved after implementing the i^{th}countermeasure. SL is an abbreviation for service life, representing the duration for which the countermeasure remains effective. The variable $k_{i,\ SL=short}$ signifies the decision to implement the i^{th}countermeasure with SSL (where k =1 if it is selected and it has short service lift, otherwise k =0), and the main constraint of the model is the total budget (B) given by:

$$\sum_{i=1}^{n} C_i k \leq B \tag{10.1-5}$$

Where, C_i denotes the cost of i^{th}countermeasure, k is a binary parameter, where k=1 indicates the i^{th}countermeasure is selected, otherwise k=0.

10.1.4 Data collection

This research utilises three distinct databases, with a focus on Utrecht-related data for prototype development. The first database is the national road crash registration (BRON), housed within SWOV, the Dutch national scientific institute for road safety research. The remaining two databases consist of the safer roads investment plan (SRIP) and road attribute data, both stored within the Vida web software integrated into the iRAP methodology.

The BRON database is a collection of all road traffic crashes in the Netherland. It offers a great variety of characteristics and attendant circumstances and the results, which are used extensively for research work and for guidance in the improvement of road safety in relation to roads, road users, vehicles and traffic movement. Crash data records the onsite situation, where in this study, information about year, accident severity, mode of transport, inside/outside built-up area, light condition, road condition road surface, road type, situation of road are applied to association rule mining to understand the relationship between context of road infrastructure and KSI crashes.

SRIP are the results of iRAP embedded in the Vida application, a web online tool, which has been elaborated in the methodology of iRAP (International Road Assessment Programme, 2023). There are plenty of road related data required to develop SRIP in ViDA, which also includes a calibration of fatalities estimation.

Road attribute data, totalling 78 attributes, are coded and stored in ViDA. According to the SRIP methodology, these attribute data are coded based on the locations or road segments associated with each proposed countermeasure. Through a comparative analysis between road crash data in the Netherlands and the road attribute data embedded in ViDA, four specific representative attributes have been identified. These attributes are consistently recorded in both databases and expected to have significant implications for road infrastructure safety. A summary of road attribute data pertaining to countermeasures is presented in Table 10.1.1.

Table 10.1.1: Example of road attribute data related to countermeasures

Countermeasure	Length / Sites	Area	Street lighting	Speed limit	Junction detail
Additional lane (2 + 1 road with barrier)	15.10 km	rural	partly	80	not
Bicycle Lane (off-road)	3.50 km	both	present	50/80	not
Central hatching	5.70 km	urban	present	50/60/80	not
Central median barrier (1+1)	34.80 km	rural	partly	60/70/80	not
Central median barrier (no duplication)	0.70 km	rural	present	70/80	not
Centreline rumble strip / flexi-post	1.80 km	rural	present	50/60/70/80	not
Clear roadside hazards (bike lane)	1.20 km	both	present	60/80	not
Clear roadside hazards - driver side	0.10 km	urban	present	50	not
Delineation and signing (intersection)	8 sites	both	present	50/60/80	at 3/4 legs
Duplication with median barrier	1.20 km	rural	present	80	not
Footpath provision driver side (>3m from road)	25.40 km	rural	partly	50/60/80	not
Footpath provision driver side (informal path >1m)	4.70 km	rural	partly	50/60	not
Footpath provision passenger side (>3m from road)	26.80 km	rural	partly	50/60/70/80	not
Footpath provision passenger side (informal path >1m)	5.80 km	rural	partly	50/60/70	not
Improve curve delineation	0.40 km	rural	present	80	not
Improve Delineation	45.70 km	both	partly	50/60/80/100	not

10.2 Results

10.2.1 Result of optimisation model

R studio, an integrated development environment for R which is a programming language for statistical computing and graphics, is used for searching front pareto solutions. Three objectives were identified in section 3: 1) maximise the NPV of benefits, 2) minimise the cost per KSI saved, 3) minimise the number of countermeasures with SSL selected. Initial population was identified as 50, and the probability of crossover and mutation was set to 0.8 and 0.1 by default. Maximum iteration was determined to 1000 times.

Budget constraint used in this study were identifies from the user-defined investment plan determined by the Royal Dutch Tourist Association (ANWB). They ultimately chose 30 countermeasures out of 47 countermeasures recommended by iRAP (presented in Appendix B) to improve the provincial road in Utrecht, which estimated saving 777 fatalities and serious injuries, 250 million euro, with benefit cost ratio of 2.2 by costing 113 million euro. Therefore, the budget constraint was set to 113 million euro to validate the advancements of the methodology proposed in this study.

After eliminating duplicates and solutions exceeding the budget constraint, 45 optimal solutions of combination emerged where alternatives are depicted in detail in Appendix C. These solutions are assessed based on their objective values, as detailed in the Table 10.2.1, and the findings were visually presented in the accompanying Figure 10.2.1.

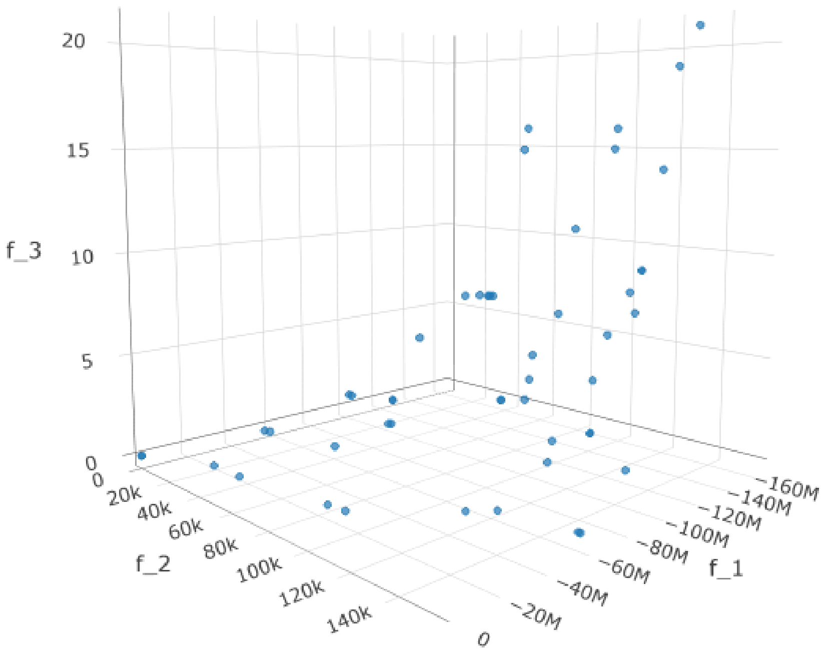

Figure 10.2.1: Visualisation of Pareto front

The net present value of benefits, as determined by ANWB, amounts to 135.7 million euros. Among the 45 solutions, 8 of them (designated as solutions 1, 12, 20, 23, 24, 25, 28, and 43) outperform ANWB's investment decision, specifically in maximising the net present value of benefits. It costs 147,281 euros to save each KSI occupant in their investment plan, with only two solutions (designated as solutions 6 and 8) performing less favourably compared to ANWB's decision, while the majority of solutions prove to be more cost-effective. According to investment strategy determined by ANWB, which involves the selection of 19 countermeasures with short service life, it is noted that over half of the improvements necessitate reconstruction or rebuilding within a period of one to five years. This would result in periodic efforts and potentially disrupt the standard functioning of the road network significantly. Among the 45 solutions evaluated, only two investment combinations opt for at least 19 countermeasures with short service lives, while the remaining alternatives exhibit superior performance with respect to the third specified objective.

Table 10.2.1: Objective value of optimal solutions

Solutions	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Obj. 1: NPV of benefits	163745935	62355466	106094212	61624941	45647564	62128373	57182693	62187724
Obj. 2: cost per KSI saved	129629	69897	125084	69145	58020	153058	125139	153890
Obj. 3: Number of SSL Countermeasures	21	2	2	2	1	0	0	0
Solutions	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Obj. 1: NPV of benefits	102935572	58817531	88288245	151958418	88295800	37919102	95286182	49288692
Obj. 2: cost per KSI saved	100203	50570	82046	123103	120504	25673	89228	118680
Obj. 3: Number of SSL Countermeasures	4	3	8	14	1	1	8	0
Solutions	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Obj. 1: NPV of benefits	93173174	89325513	21084157	141657702	108866915	101958742	145938452	139524884
Obj. 2: Cost per KSI saved	84663	121505	92654	108621	96799	98967	118656	120238
Obj. 3: Number of SSL Countermeasures	8	2	0	15	5	3	9	7
Solutions	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**
Obj. 1: NPV of benefits	142835876	131532448	129387034	155634348	106692602	101194832	94390118	20143355
Obj. 2: cost per KSI saved	108972	98601	115371	126531	95753	143154	87614	84887
Obj. 3: Number of SSL Countermeasures	16	11	6	19	16	1	8	0
Solutions	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
Obj. 1: NPV of benefits	14086220	94960805	95066993	105514428	13875943	95120643	65815468	38649627
Obj. 2: cost per KSI saved	43457	93587	93634	95033	27584	87935	66995	28086
Obj. 3: Number of SSL Countermeasures	0	3	3	15	0	8	3	1
Solutions	**41**	**42**	**43**	**44**	**45**			
Obj. 1: NPV of benefits	117578318	74322784	139007051	118748037	58905038			
Obj. 2: cost per KSI saved	101829	72487	118310	116720	51984			
Obj. 3: Number of SSL Countermeasures	7	6	8	4	3			

10.2.2 Result of integrated method

Once the optimal solutions are searched, all alternative combinations are employed to weight assignment process. As mentioned in methodology, the lift value is used as a weight metric in the form of rule to indicate the association between road attributes to KSI result, where association rule mining is applied to explore their association. The overall result of rule mining of Utrecht crash database is shown in Appendix D, where the rules concerning the four road attributes data are extracted and demonstrated in Table 10.2.2.

Table 10.2.2: Decision rules associated to KSI crash

LHS	RHS	Support	Lift
Inside outside built-up areas=Urban area	KSI	0.005359057	3.513389
Inside outside built-up areas=Rural area	KSI	0.005716327	2.839102
Speed limit road =60 km/h	KSI	0.011432655	1.873808
Situation of road =Bend	KSI	0.016434441	1.395647
Situation of road =Intersection - 3 legs	KSI	0.019649875	1.314533
Road surface=Concrete	KSI	0.00785995	1.238695
Road surface=Porous asphalt	KSI	0.034297964	1.230071
Speed limit of road =100 km/h	KSI	0.016791711	1.228641
Speed limit of road =120 km/h	KSI	0.010360843	1.195872
Situation of road =Intersection - 4 legs	KSI	0.03965702	1.160673
Street lighting=Not burning	KSI	0.123258307	1.108781

According to the attributes derived from ViDA database and the result of crash databases rule mining analysis, four distinct risk attributes are used for lift aggregation method involving the road area, speed limit, situation of road, and street lightning. To highlight the urgency and significance of countermeasures recommended by ViDA across diverse locations and sections, the level of risk associated with roads will be signified by greater lifts of attributes among varying road segments. For instance, in the scenario where the construction of refuge islands is recommended for both three-leg and four-leg intersections, the lift value observed for the three-leg intersection will serve as a measure for urgency due to its comparatively higher lift. The top ten solutions ranked by lift aggregation with their objective value are demonstrated in Table 10.2.3.

Table 10.2.3: Ranking of investment plan

Investment alternatives	Obj. 1 (Maximise)	Obj. 2 (Minimise)	Obj. 3 (Minimise)	Weight	Ranking
1	163745935	129628.5	21	163.85	1
29	155634348	126531.3	19	145.06	2
26	142835876	108971.9	16	134.27	3
13	151958418	123103	14	132.94	4
30	106692602	95752.54	16	128.32	5
21	141657702	108620.5	15	119.24	6
37	105514428	95032.68	15	113.29	7
25	139524884	120238	7	96.42	8
24	145938452	118656	9	93.05	9
16	95286182	89227.85	8	89.47	10
Utrecht	135739970	135739970	19	-	-

10.3 Discussion

It is impossible to obtain a solution that is the best for all considered objectives, whereas it is imperative to discover the pareto front where a set of optimal solutions could be identified for decision makers. This approach enables an evidence-based assessment of prioritisation concerning differential risk levels identified across distinct road infrastructure configurations, thereby informing the allocation of resources and interventions in road safety management. Contrasted with the investment plan outlined in the Utrecht report (details shown in Appendix E), which involved the selection of 33 countermeasures within a budgetary allocation of

114,378,840 euros, their strategy selected for 19 countermeasures characterised by a limited operational lifespan of fewer than 5 years. This selection was estimated to generate a net present value of benefits amounting to 135,739,970 euros, with an associated cost of 147,282 euros per KSI saved. However, our model diverged in its selection.

The top 10 alternatives are ranked by lift aggregation. In other word, these alternatives concern road locations and sections with high risky road attributes that are very likely to cause KSI crash. Each objective will be discussed separately to validate the advancement of our methodology. With regards to the objective 1, aimed at maximising the NPV of benefits, it is noteworthy that alternative 1 shows a visible enhancement of about 20% when compared with the NPV estimation described in the Utrecht report. This improvement implies the advanced ability of our proposed methodology in searching the best solution for maximising monetary gains.

As for objective 2, which strives to minimise the cost per KSI saved, alternative 16 emerges as a standout performer, with an average cost of 89,227.85 euros per KSI saved. This represents a significant improvement from the Utrecht plan, where the cost per KSI saved is approximately 147,281.5 euros, thereby signifying a substantial 39% reduction in anticipated expenditures. This marked reduction underscores a valuable advancement by applying the proposed method in cost-effectiveness measures.

Regarding the objective 3, trying to minimise the number of countermeasures with a short lifespan of less than 5 years selected, the Utrecht report identified a total of 19 such countermeasures. In contrast, Alternative 25 selected only seven short-lived road improvement interventions. This plan is anticipated to result in reduced frequency of construction activities and decreased disruption to the operation of the road network. This discrepancy also highlights the capacity to identify optimal solutions in quantitative terms. In addition, the proposed methodology facilitates the proficiency in simultaneously optimising various objectives across different metrics as it is a critical consideration given that real-world objectives typically extend beyond financial concerns to encompass diverse units of measurement.

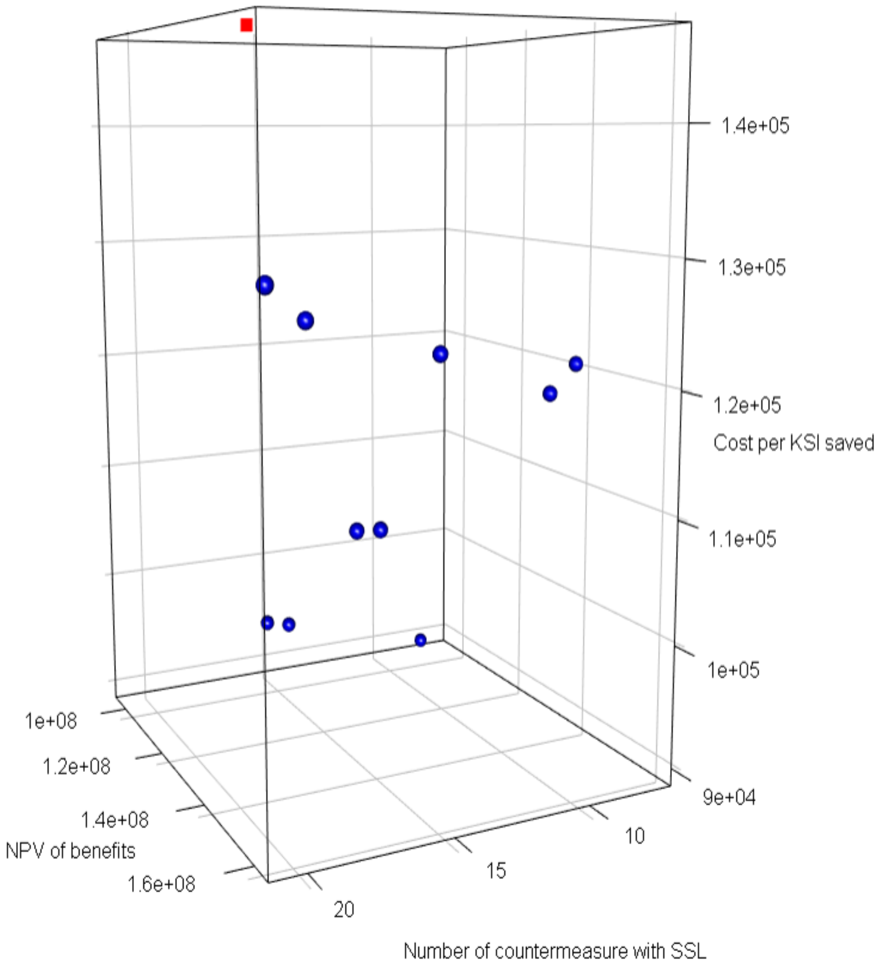

Figure 10.3.1: 3D visualisation of top 10 investment plan

Despite the presence of trade-offs among the three objectives, the model identifies top 10 optimal countermeasure combinations based on lift aggregation. The 3D visualisation depicts the 10 alternatives as blue circles, comparing with the investment plan of Utrecht represented by red squares as show in Figure 10.3.1. Intriguingly, all solutions outperform in the second objective, minimising the cost per KSI saved, which suggests that these investment plans are particularly effective in reducing costs associated with preventing KSIs. Half of these solutions show superior performance across all three objectives compared to the investment plan formulated by ANWB. This indicates that these solutions not only excel in cost-effectiveness but also in achieving other objectives related to safety and efficiency

Due to their exceptional performance regarding the second objective, a 2D visualisation is presented for a clearer understanding of the performance and compromises of all solutions considering the first and third objectives as illustrated in Figure 10.3.2.

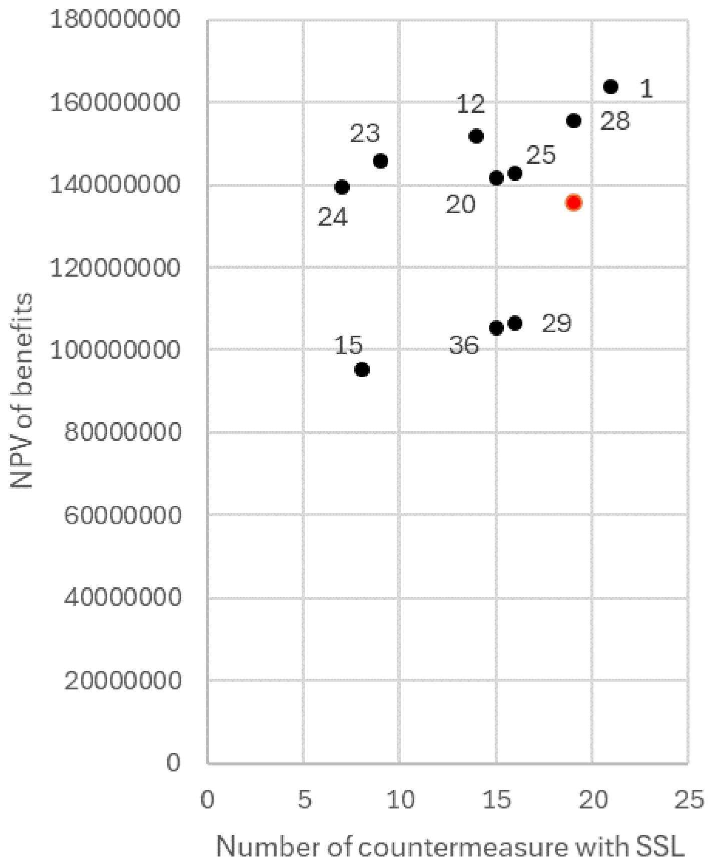

Figure 10.3.2: Two objectives analysis

The analysis of the figure interprets the comparative performance of various solutions in terms of optimising the objective. Notably, alternatives 15, 29, and 36 exhibit a strategic inclination towards reducing the number of Short Service Lift (SSL) countermeasures, with sacrifice of the

Net Present Value (NPV) of benefits. Conversely, combinations 1 and 28 prioritise the maximisation of NPV of benefits, even employing an equal or greater number of SSL countermeasures. Remarkably, the methodology employed in this study has found five solutions (12, 20, 23, 24, and 25) which simultaneously optimise three objectives to varying extents. This optimisation method is contributory in realising substantial value within resource constraints, featuring the multiple considerations and diverse measurement metrics involved. This highlights the value of the proposed methodology in achieving a balance between competing priorities and underscores its potential to outperform conventional planning strategies in optimising outcomes across multiple dimensions

10.4 Summary

This Chapter introduces an innovative methodology for optimising resource allocation in road safety, specifically focusing on minimising the cost per KSI (killed or seriously injured) saved, reducing construction impacts on the road network, and maximising overall benefits. The methodology combines multi-objective optimisation with association rule mining, ensuring a data-driven approach to strategic decision-making.

The methodology unfolds in two main stages. The first stage involves generating a Pareto optimal set of solutions using the Non-Dominated Sorting Genetic Algorithm II (NSGA II), selected for its superior performance in spreading and converging solutions. The second stage employs association rule mining to weigh the optimal solutions. Lift values are used as a metric to determine the strength of associations between road attributes and KSI outcomes, thus providing an objective basis for selecting countermeasures.

This Chapter also emphasises the value of the proposed methodology in providing an evidence-based assessment for road safety interventions. The approach not only achieves a balance between competing priorities but also demonstrates potential superiority over conventional planning strategies by optimising outcomes across multiple dimensions. This Chapter sets the stage for a detailed discussion in the subsequent Chapter.

In the subsequent Chapter, methodology and implementation of strategic data analysis and resource optimisation techniques, building upon the foundation laid by the systematic literature review will be discussed. This next phase of the thesis aims to translate theoretical insights into practical applications, supported by sophisticated data-driven approaches to enhance road safety outcomes. By leveraging advanced optimisation algorithms and association rule mining, how these techniques can be effectively utilised to allocate resources strategically will be explored. This transition from theoretical evaluation to practical application underscores the core objective of this research: to provide adaptive tools for strategies that road authorities and policymakers can implement to achieve tangible improvements in road safety management.

Additionally, the next Chapter will explore the implications of these findings and propose avenues for future research. Special attention will be given to addressing identified limitations, exploring new data sources, and refining optimisation models. This forward-looking perspective ensures the continuous enhancement and relevance of strategic management practices in achieving optimal road safety outcomes.

Chapter 11. Discussion

11.1 Introduction

In this section, a critical analysis and appraisal of the findings presented in the preceding sections, focusing on the innovation, implications, and advantages of technology integration for optimising road infrastructure safety strategic management. The research question and summary of the key findings are recalled before exploring their significance in the context of road safety.

This study aims to optimise road safety strategic management processes and outcome through assessing the impact and feasibility of integrating data-related technologies. Utilising integrated data mining techniques, the research investigates the frequency and correlations between fatalities and various roadway crash characteristics, such as road class, type, and junction details. This integrated approach to data mining enables a comprehensive analysis of crash databases, addressing the challenge of identifying rare patterns. Additionally, leveraging strategic data analysis, the study proposes a data-driven multi-objective optimisation method. This method accommodates diverse unit metrics within objectives and generates optimal combinations of countermeasures based on the risk, aggregated weight of roads and intersections.

In the following discussion, the innovation of the proposed methodology will be investigated, highlighting how the integration of data-related technologies has enhanced the efficiency and effectiveness of road safety strategic management processes. Furthermore, the significant and impact of the research for road safety practice will be discussed. By elucidating the potential benefits and advantages associated with technology integration, valuable insights will be provided for road authorities, policymakers, and researchers. It is essential to acknowledge the challenges and limitations of this research. By doing so, areas for improvement and suggest avenues for future research will be identified in the field of road safety strategic management.

The primary goal of this discussion is to critically evaluate the implications of data-related technologies for road safety practice. By examining the feasibility and effectiveness of technology integration in road safety initiatives, the discussion will convince the usage of evidence-based approaches which will contribute to the continual progress within the field.

11.2 Innovation

This research covers two distinct phases of strategic management aimed at addressing the challenges posed by road crash databases and optimising resource allocation. Adhering to the principles of strategic management, the research presents two interconnected models designed to achieve strategic objectives through comprehensive data analysis (see Chapter 6 and 7) and resource optimisation (see Chapter 9 and 10). Positioned within the framework of strategic management principles, this study introduces an innovative two-stage methodology that ultimately realises optimisation of outcomes.

In terms of strategic data analysis, the integration of data mining methods, association rule mining and k-mean clustering, provides robust evidence of descriptive analysis of databases. While previous studies (Addi et al., 2016; Ait-Mlouk et al., 2017; Geurts et al., 2003; Xu et al., 2018) have utilised association rule mining for crash database analysis, their findings have often lacked specificity and clear objectives. For instance, (Lambert et al., 2003) attempted to refine the analysis by specifying crash severity as the consequent variable, aiming to elucidate the correlation between onsite crash attributes and severity levels.

However, simple setting high parameter thresholds employed in prior studies have failed to provide a comprehensive understanding of crash databases, particularly in capturing infrequent factors associated with fatalities. In contrast, the integrated methodology employed in this study demonstrates a novel approach. Beginning with the logical identification of minimum parameters and extending to clustering and visualisation techniques, it effectively mitigates bias and subjectivity in rule extraction. This advancement represents a significant improvement in the application of association rule mining, or frequent pattern search methods, as it allows for the exploration of infrequent patterns through the integration of k-means clustering. Furthermore, this study offers a multi-dimensional perspective to strategy development, which aids decision-makers in gaining a comprehensive understanding of databases. By considering frequency, association, and an overview of rules, this approach enhances the interpretability and utility of the analytical insights derived from crash databases.

With regards to resource optimisation, a data-driven weighted approach is applied to multi-objective optimisation. The research attempts to employ nondominated sorting genetic algorithms (NSGA II) to conduct optimal solutions search and utilise the lift value from association rule mining as weighted indicator to demonstrate the risk level of road and the urgency of road improvements. In recent years, there has been a prevalence of single objective optimisation methods in the road context, each focusing on different aspects. These methods commonly revolve around optimising specific performance indices (Chowdhury et al., 2000), minimising the number of crashes and time delays (Banihashemi, 2007), or considering the service life of countermeasures (Byaruhanga & Evdorides, 2022). In addition, some multi-objective optimisation models have been developed, all rely heavily on expert input, often involving subjective evaluations of stakeholder needs and preferences (Lambert et al., 2003), two-level optimisation processes by classifying decisions and supporting objectives (J. Yu et al., 2012), or the formulation of decision rules by experts (Augeri et al., 2021).

In contrast, the proposed two-stage model not only considers diverse objectives, even when they differ in units, but also provides data-driven evidence of road context characteristics with minimal expert efforts to allocate resource. Illustrated by a prototype case study of the Utrecht safer road investment plan, the model achieves significant improvements individually in optimising objectives, with the net present value (NPV) of benefits improving by approximately 20%, and the cost per killed or seriously injured (KSI) saved reduced by approximately 39%. The optimisation model effectively explores the Pareto front across various objectives within the constrained budget. Specifically, it underscores the practical applicability of the model in real-world decision-making scenarios and provides empirical evidence of the effectiveness of the non-dominant sorting genetic algorithm in identifying the Pareto front. Furthermore, this study highlights the potential of data-driven preferences to guide resource allocation and offers a framework for integrating data-driven insights with less expert efforts to enhance decision-making in complex road management scenarios.

11.3 Impact

The primary aim of this study is to propose adaptive strategies for optimising road safety management, with a focus on adaptation as the core element driving its effectiveness. The research outlines the strategic management process, emphasising the importance of strategic data analysis and resource allocation. This section seeks to extend the international insights gained from the strategic fatality analysis discussed in earlier chapters, aiming to demonstrate the adaptability of these models for different levels of authorities, each with unique objectives.

11.3.1 Knowledge discovered about fatal crashes

The findings from the strategic data analysis on road fatalities reveal key patterns. Fatal road crashes predominantly occur on roads where human control is absent within 50 meters, and on roads lacking pedestrian crossing facilities. Additionally, adverse weather conditions are rarely a factor in these incidents, as most fatal crashes occur under good weather conditions without strong winds. Single carriageways are identified as the most common road type where fatal crashes occur.

Certain road infrastructure attributes are individually significant in understanding the contributors to fatalities on the road network, as supported by international evidence. Regarding light conditions, darkness is another influential factor in fatal crashes, particularly on roads without street lighting. The Royal Society for the Prevention of Accidents (2018) published a report examining the evidence on the road safety benefits of street lighting. Research in New Zealand (Jackett & Frith, 2013), which compared the quality of road lighting with accident reduction, found that improved lighting does enhance safety. Before-and-after studies have indicated reductions in collisions of around 30% or more where lighting has been improved. Additionally, the International Road Assessment Programme (2013b) established risk factors related to specific road attributes, highlighting that installing streetlights could improve overall road safety and mitigate crash severity.

In terms of speed limit, 70 mph (approximately 112.7 km/h) and 60 mph (approximately 96.6 km/h) emerge as significant patterns in crash fatalities. This factor is recognised as one of the key elements contributing to fatal crashes (European Commission, 2024b). Evidence from Queensland Government (2023) as shown in Figure 11.3.1 demonstrates indicates that higher speeds lead to reduced reaction times in unexpected situations, as the distance travelled before a driver can respond increases with speed, given the same reaction time. Additionally, increased speed extends the braking distance, which is proportional to the square of the speed.

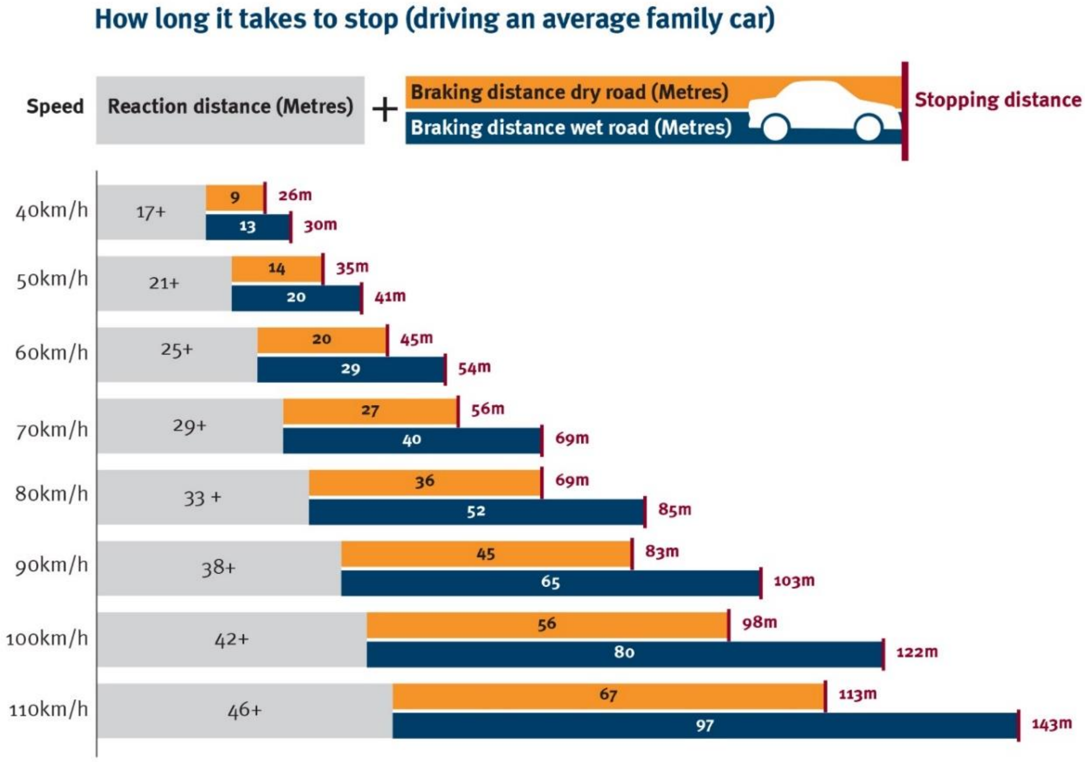

Figure 11.3.1: Reaction distance and stopping distance for different driving speeds (Queensland Government, 2023)

Rural areas also emerge as a significant road attribute associated with crash fatalities. The data analysis is crash-based rather than casualty-based. The percentage of fatalities by crash area and by country for car occupants, as reported by the European Commission (2024a) shown in Table 11.3.1, underscores the significance of this international road safety knowledge. Statistical analysis indicates that the majority of road crash fatalities occur in rural areas, regardless of whether the occupants are in cars, lorries, or heavy goods vehicles. According to the International Transport Forum (2023) rural roads are the deadliest in almost all countries, as depicted in Figure 11.3.2. In 17 countries, more than half of road deaths occurred on rural roads. In Finland, Ireland, and New Zealand, two-thirds of road deaths occurred on this type of road. Only in Korea, the Netherlands, Japan, and Portugal are urban roads deadlier than other road types.

Table 11.3.1: Percentage of fatalities by crash area and by country, occupants of cars (European Commission, 2024a)

Country	Year	Urban roads	Rural roads	Motorways	Total
Belgium	2022	23%	56%	21%	215
Bulgaria	2022	19%	74%	7%	323
Czechia	2022	19%	75%	6%	278
Denmark	2022	10%	81%	9%	68
Germany	2022	14%	71%	15%	1,192
Estonia	2022	17%	83%	0%	24
Ireland	2019	20%	77%	4%	81
Greece	2021	35%	58%	7%	226
Spain	2022	10%	68%	23%	681
France	2022	21%	69%	11%	1,565
Croatia	2022	48%	39%	14%	143
Italy	2022	26%	61%	13%	1,375
Cyprus	2022	36%	55%	9%	11
Latvia	2020	23%	77%	0%	64
Lithuania	2022	29%	70%	2%	59
Luxembourg	2022	26%	61%	13%	23
Hungary	2022	15%	77%	8%	273
Malta	2021	-	-	-	3
Netherlands	2022	21%	53%	25%	225
Austria	2022	16%	73%	11%	180
Poland	2022	23%	74%	3%	913
Portugal	2022	43%	39%	19%	211
Romania	2022	44%	53%	3%	698
Slovenia	2022	14%	38%	48%	21
Slovakia	2022	19%	75%	6%	132
Finland	2021	11%	86%	3%	127
Sweden	2020	21%	77%	3%	102
EU	2022	23%	66%	11%	9,213
Iceland	2022	-	-	-	3
Norway	2022	14%	86%	0%	58
Switzerland	2022	23%	63%	14%	87

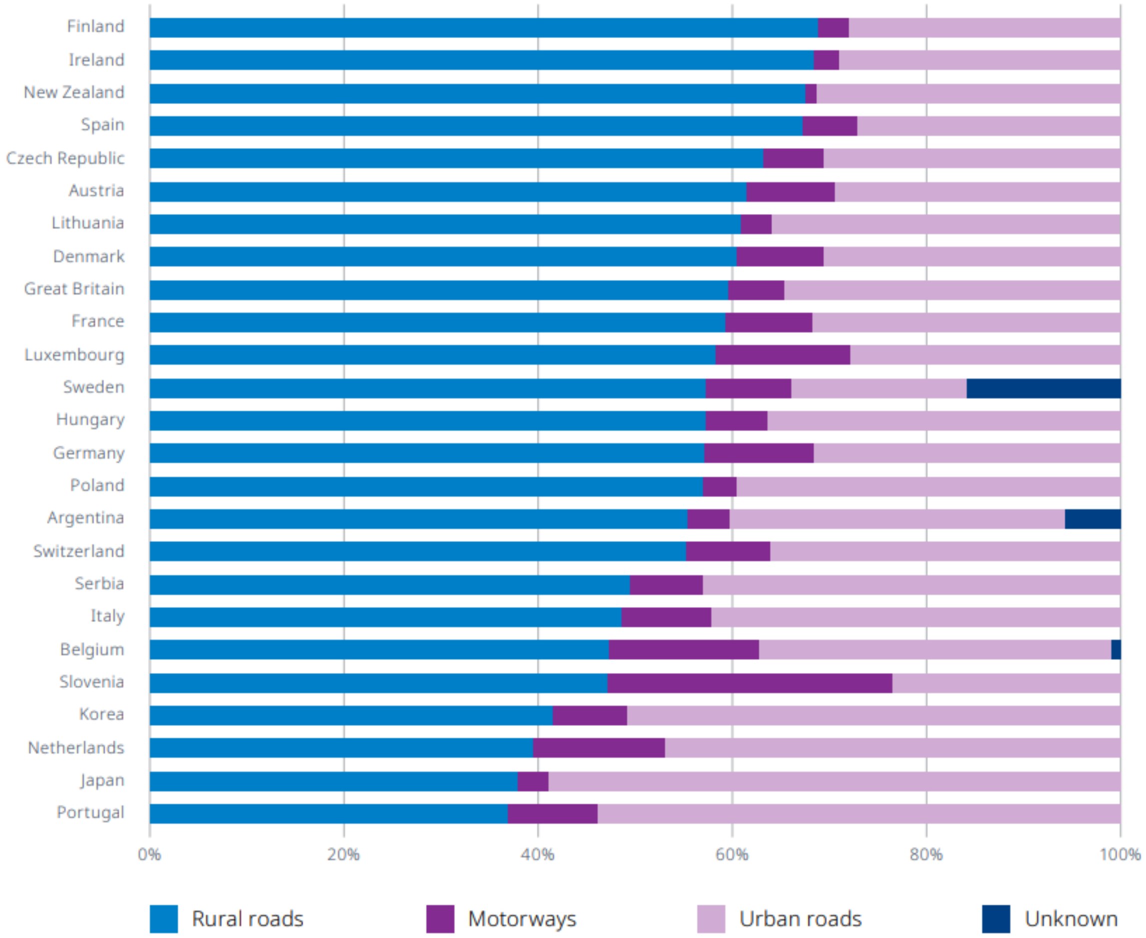

Figure 11.3.2: Road death by road area (International Transport Forum, 2023)

The location of fatal crashes shows that deaths rarely occur at junctions, indicating that on-carriageway locations are the primary sites of crash fatalities. This finding is supported by data from the European Commission (2024a) as shown in Table 11.3.2.

Table 11.3.2: Percentage of fatalities by junction and by country (European Commission 2024a)

Country	Year	Junction	Not at junction	Total
Belgium	2022	18%	82%	539
Bulgaria	2022	18%	82%	531
Czechia	2022	18%	82%	527
Denmark	2022	24%	76%	154
Germany	2022	20%	80%	2,788
Estonia	2022	13%	88%	48
Ireland	2019	15%	85%	140
Greece	2021	9%	92%	624
Spain	2022	21%	79%	1,746
France	2022	19%	81%	3,260
Croatia	2022	14%	86%	275
Italy	2022	18%	82%	3,159
Cyprus	2022	46%	54%	37
Latvia	2020	29%	71%	139
Lithuania	2022	10%	90%	120
Luxembourg	2022	3%	97%	36
Hungary	2022	18%	82%	537
Malta	2021	-	-	8
Netherlands	2022	39%	62%	644
Austria	2022	27%	73%	370
Poland	2022	14%	86%	1,896
Portugal	2022	12%	88%	611
Romania	2022	12%	88%	1,633
Slovenia	2022	9%	91%	85
Slovakia	2022	7%	93%	266
Finland	2021	20%	80%	225
Sweden	2020	23%	77%	195
EU	2022	18%	82%	20,593
Iceland	2022	-	-	9
Norway	2022	11%	89%	116
Switzerland	2022	11%	89%	241

Single carriageways are also highlighted as a road attribute with a high frequency of fatalities. The International Road Assessment Programme (iRAP) provides an overview of different median types to assess road infrastructure risks. The separation of opposing traffic flows reduces the likelihood of severe crashes. Physical barriers can prevent errant vehicles from crossing the median, and physical medians reduce the potential for head-on impacts by making it less likely that vehicles will reach opposing traffic before recovering. The risk factors associated with single carriageways are presented in the Table 11.3.3.

Table 11.3.3: Risk factors by road attribute category, road user type and crash type(International Road Assessment Programme, 2013a)

Median Type	**Vehicle occupant and motorcyclist loss of control (head-on)**	**Vehicle occupant and motorcyclist overtaking (head-on)**	**Pedestrian crossing road**	**Vehicle occupant and motorcyclist property access point crash**
Continuous central turning lane	77	25	3.0	1.0
Centreline rumble strip (or flexipost)	90	0	2.7	1.0
Central hatching (>1m)	83	82.5	2.4	1.0
Centre line	100	100	3.0	1.0
Wide centre line (0.3m to 1.0m)	95	100	2.7	1.0

In addition to the individual risks associated with specific road infrastructure attributes, the findings also indicate that certain combinations of attributes increase the likelihood of fatalities. Fatalities are more likely to occur on roads that are not within 20 meters of a junction, are in darkness without street lighting, have dry road surfaces, and are in rural areas. A roads, in particular, are associated with higher fatality rates under these conditions. Fatalities are also more likely to occur on A roads that lack pedestrian crossings, are in darkness without street lighting, are under fine weather conditions without high winds, and are situated in rural areas.

While it is challenging to find global efforts that support the significance of these rules, the proposed methodology—an integration of association rule mining and K-means clustering—could validate the knowledge gained from conditional probability analyses.

11.3.2 User defined investment plan

The optimisation model was developed based on the outcomes of a standard investment plan derived from the iRAP methodology, which encompasses all feasible safety measures along a given route or network, optimised to minimise fatalities and serious injuries (International Road Assessment Programme, 2023). Customisation of investment plans is feasible to meet specific client requirements or contextual considerations in road assessment, where International Road Assessment Programme (2023) offers various methods for customisation known as user-defined investment plans, involving the identification of budget thresholds, economic returns, preferences for mass-action programs, maintenance-only approaches, and consideration of construction lead times. The proposed optimisation method empowers local authorities to develop user defined investment plans tailored to the specific characteristics of their local road network environment, thereby addressing urgent issues and minimising subjective influence on preferences. This approach provides decision-makers with a range of solutions, affording flexibility in determining the final investment plan. While not advocating for complete elimination of expert control, this model empowers experts with greater insights from natural data pertaining to diverse road contexts, thereby reducing the pressure associated with searching for optimal solutions.

Considering the adaptable nature of the model, some potential ways will be discussed to enhance its ability in accommodating diverse requirements. When identifying the objectives of optimisation, the methodology is capable to address multiple objectives, even when they are measured in different unit metrics. This ensures that various aspects of road safety and infrastructure improvement can be simultaneously considered, ranging from minimising fatalities and serious injuries to optimising traffic flow and enhancing environmental sustainability. Furthermore, the flexibility of the model allows adjustments in defining the consequent of rules to meet various needs, as exemplified in the exploration of strategic data analysis. This adaptability enables decision-makers to tailor the optimisation process to suit specific contexts and priorities. For instance, the rules can be extracted to figure out what type of the roadway attributes affect the occurrence of crash in a specific area, where aggregation of weight could be used to prioritise certain safety measures or select optimal combination of countermeasures based on the road infrastructure characteristics of the road network.

By embracing such flexibility and adaptability, the model enables stakeholders to devise comprehensive and specific strategies for road safety and infrastructure development. This approach not only enhances the effectiveness of investment plans but also fosters greater awareness to evolving needs and objectives in the road transport sector. Ultimately, it enables decision-makers to make informed and strategic choices that maximise the benefits of limited resources while effectively addressing the complex challenges of modern road management.

11.4 Summary

This Chapter presents a critical analysis and appraisal of the findings related to the integration of data-related technologies in optimising road infrastructure safety strategic management. The research emphasises the innovation, implications, and advantages of this integration, aiming to enhance the efficiency and effectiveness of road safety management processes. The study employs integrated data mining techniques to analyse crash databases, focusing on identifying rare patterns and providing a three dimensions framework for rule understanding. In addition, the research utilises data analysis result to propose a multi-objective optimisation method for resource allocation.

The innovative two-stage methodology integrates data mining methods such as association rule mining and k-means clustering, offering a comprehensive approach to analysing fatal collision. The methodology addresses the challenge of infrequent pattern identification and minimises bias and subjectivity in rule extraction. Furthermore, the study demonstrates the practical applicability of the data-driven optimisation model through a case study, showcasing significant improvements in resource allocation and decision-making.

The strategic data analysis and resource optimisation approaches are shown to be adaptable and applicable to various levels of governing bodies, each with distinct strategic objectives. The findings of strategic fatal crash analysis present some valuable insight according to international knowledge. Additionally, the Chapter discusses the potential of the user-defined investment plan model, highlighting its flexibility and adaptability to address specific local road network characteristics and needs.

The next Chapter will synthesise the findings from three key stages of the process of research: the systematic literature review, the strategic data analysis, and the resource optimisation processes. This synthesis will demonstrate how the integration of data-related technologies enhances road safety strategic management, offering evidence-based approaches for continual progress in the field.

Chapter 12. Conclusions

12.1 Overview

This thesis focused on road safety strategic management and existing critical challenges, emphasising the imperative of data use. By conducting a thorough systematic literature review, the study evaluates existing academic findings and practical implementations aimed at enhancing road safety performance within the strategic management domain, thereby fostering a comprehensive understanding of their application. The review advocates for further exploration of research concerning data analysis and resource allocation for road safety enhancement, charting a clear trajectory for future research.

Following the process of strategic management, the research begins with data analysis, aiming to extract key patterns and correlations between road traffic fatalities and road infrastructure. This analysis underscores the significance of understanding rules from crash database and offers invaluable insights for road authorities and policymakers. Subsequently, resource allocation emerges as a critical step, introducing cutting-edge optimisation techniques tailored to diverse objectives. The emphasis on rule mining applications for solution searching highlights the importance of hidden knowledge involvement in the process, facilitating a less interfering approach. Additionally, the thesis also explores and validates the adaptability of the proposed methodology within the framework of pyramid strategic levels, in relation to road networks. This approach seeks to achieve diverse objectives across different administrative levels.

This Chapter concludes the essence of the research, across distinct sections focused on the systematic literature review, strategic data analysis, and strategic resource optimisation, emphasising its contributions and paving the way for further scholarly exploration in the field of road safety strategic management.

12.2 Systematic Literature Review

As part of the study's objectives, a systematic review was conducted to assess existing methods and potential approaches for enhancing strategic road safety management. Moreover, insights gathered from road-related entities provided practical guidance for road safety managers, expanding comprehension of methodological approaches. The main conclusions are:

1) Diverse methodologies are applied in strategic road safety management, including data mining, statistical analysis, multi-criteria analysis, data envelope analysis, and optimisation methods.

2) Practicing methods primarily focuses on project evaluation via benefit-cost ratio and cost-effectiveness analysis for countermeasure selection.

3) Strategic management needs a higher-level and data integrated approach in mitigating crash occurrences, distinct from sectional analysis and crash prediction.

4) The development of effective strategies requires a robust foundation of objective data-driven evidence.

5) The implementation of automated decision-making tools cultivates a safer road network system.

12.3 Strategic Data Analysis

With regards to data analysis, this study has provided valuable insights into road strategic management through the utilisation of data mining techniques, specifically integrating association rule mining and K-means clustering. The research highlights that while fatalities resulting from crashes are infrequent, they hold significance in uncovering hidden knowledge due to the profoundly negative impacts and their strong association with multiple contributing factors. The main conclusions and implications of the study are:

1) Fatal road crashes predominantly occur on roads without human control.

2) Fatalities are more likely to occur when road users are traveling on A roads, without human-controlled pedestrian crossings, in conditions of darkness without street lighting, under good weather conditions without high winds, within rural areas.

3) All significantly roadway crash attributes leading to fatal crashes are summarised as followed: Darkness: no street lighting, Road surface dry condition, Speed limit of 60 mph, Not at or within 20 meters of junction, East Midlands, Rural area, Trunk roads managed by Highways England, Saturday, Speed limit of 70 and 50 mph, Scotland, Darkness: streetlights present and lit, South Wes, South East, August and July.

4) The findings of the data mining process were consistent with the statistical analyses conducted by the Department of Transport.

5) The suggested methodology, from identification of minimum parameters to clustering, and visualisation, has effectively mitigated bias and subjectivity in the extraction of rules.

6) The study offers a three-dimensional perspective to strategy development in terms of quality, quantity and significance of knowledge extracted from the data.

7) The potential application of the methodology involves tailored strategies catering to the

distinct needs and objectives of different administrative levels.

12.4 Resource Optimisation

To optimise resource allocation, this study highlights the novelty and importance of its methodology in addressing the complexities of multi-objective optimisation for road safety strategy development. The main conclusions are:

1) The proposed methodology presents a data-driven approach to develop the investment plan.

2) This equips road agencies and stakeholders with a valuable tool for more efficient resource allocation and road safety improvement, requiring less human intervention.

3) The optimisation model demonstrates a visible maximum enhancement of about 20% compared with the NPV estimation of Utrecht investment plan.

4) It also achieves a maximum 39% reduction with regards to cost per KSI saved

5) The integrated data-driven optimisation method ranks top 10 solutions of investment plan outperforming in minimising the cost per KSI saved suggesting its effectiveness.

6) Half of top 10 solutions show superior performance across all three objectives compared to the investment plan formulated by Utrecht.

7) It provides optimal alternatives for decision makers to choose the most suitable one for implementation.

8) The study emphasises the practical implications of the methodology by providing a systematic framework for diverse objectives and decision rule mining.

9) By overcoming challenges associated with simultaneous optimisation of three objectives, this approach offers a novel solution that promises to improve decision-making processes and optimise resource allocation in road infrastructure management.

12.5 Knowledge Discovered

This thesis highlights the critical road infrastructure attributes contributing to fatal road crashes and emphasises the importance of adaptive strategies in road safety management. Some of the attributes are supported by international statistical analysis and established work in different organisations and countries. The main knowledge discovered is summarised below:

1) Fatal road crashes primarily occur on single carriageways.

2) Fatal crashes occur frequently on road in darkness without street lighting.

3) High speed limit road, specifically at 60 mph and 70 mph, holds high frequency of fatal crashes.

4) The absence of pedestrian crossing facilities within 50 meters is one of frequent road attributes of fatalities.

5) Junctions are fewer common sites for fatal crashes, with most fatalities occurring on open roads carriageway.

6) Fatalities are more likely to occur on roads that are not within 20 meters of a junction, are in darkness without street lighting, have dry road surfaces, and are in rural areas.

7) Fatalities are also more likely to occur on A roads that lack pedestrian crossings, are in darkness without street lighting, are under fine weather conditions without high winds,

and are situated in rural areas.

12.6 Challenges, Limitations, and Future Improvements

While this study has made noteworthy contributions to advancing road safety management and the research community, it is crucial to acknowledge the challenges and limitations encountered summarised in Table 12.6.1. One significant challenge lies in the scope of the study, which primarily focused on crash attributes related to fatalities, thereby neglecting other critical dimensions of crash occurrence, such as the involvement of casualty and vehicles. However, combining and analysing diverse datasets related to road infrastructure, traffic patterns, historical crash data, and other relevant factors pose challenges due to differences in data formats. A more comprehensive study is warranted to demonstrate how this methodology can be applied in practice to address all dimensions of crash occurrence and inform strategy development.

Furthermore, the methodology employed in this study relies heavily on the availability of high-quality crash databases, which presents a significant challenge, particularly in many developing countries where such data may be limited or of poor quality. Poor data quality or insufficient data coverage can significantly compromise the accuracy and reliability of the analysis, thereby reducing the effectiveness of the methodology. This constraint underscores the need for robust data collection and interpretation processes to ensure the reliability and accuracy of the analysis.

Table 12.6.1: Summary of challenges and future improvements

	Challenges/limitations	**Future improvements**
Research design	Research scope only focuses on severity	Combining and analysing diverse datasets related to road infrastructure.
	General evaluation of	Segmentation approach for

	countermeasure performance	detailed evaluation.
	Overestimation of risk and urgency of countermeasure	Considering multi-factors association for risk assessment
Data	Collection and quality	Develop strategies to address data quality issues, particularly in regions with limited or poor-quality crash databases.
	Integration analysis	Refining analytical approaches based on state-of-the-art technologies

Moreover, the study assesses the collective effectiveness of individual countermeasures without considering the varied characteristics of road sections and junctions within the same category of road attributes. To improve the precision and detail of investment plans, there is a need to divide the road network into homogeneous segments and junctions, each exhibiting distinct patterns of countermeasure performance. This segmentation approach would enable a clearer understanding of the varying safety needs across different sections of the road network, facilitating the development of tailored strategies to address specific challenges. Additionally, such an approach would enhance the accuracy of investment prioritisation by aligning countermeasure interventions with the unique characteristics and safety requirements of each road segment or junction.

The methodology uses lift values from association rule mining to weight risk levels, assuming road attributes independently indicate countermeasure urgency. However, this may overlook interactions among attributes, leading to overestimations of risk levels for certain road sections or junctions, and exaggerating the urgency of addressing them.

12.7 Summary

The conclusion of this thesis underscores the significant efforts made in advancing road safety strategic management through academic research. Beginning with a comprehensive systematic literature review, the study identified various methodologies and highlighted the importance of data-driven approaches in mitigating road traffic fatalities. The subsequent focus on strategic data analysis revealed valuable insights into crash occurrences and emphasised the need for integrated data mining techniques to uncover hidden patterns. Moreover, the innovative approach was proposed to resource optimisation offered a promising framework for multi-objective decision-making.

Despite the notable contributions, the conclusion acknowledges several limitations, including the need for more detailed analyses beyond fatalities and the challenges posed by data quality in certain contexts. Additionally, while the methodology presents advancements, further refinement is necessary to address interactive effects among multiple road attributes and to enhance the accuracy of risk assessment.

Overall, the thesis advocates for continued research efforts to refine methodologies, improve data collection processes, and develop more sophisticated models to fully realise the potential of data-driven approaches in enhancing road safety outcomes. By addressing these challenges, researchers can contribute to the promotion of positive road networks and the well-being of road users in safer environments.

Bibliography

AASHTO. (2010). Highway safety manual. In *National Research Council (US). Transportation Research Board. Task Force on Development of the Highway Safety Manual and Transportation Officials. Joint Task Force on the Highway Safety Manual* (Vol. 1). AASHTO.

Adby, P. (2013). *Introduction to optimization methods*. Springer Science & Business Media.

Addi, A.-M., Tarik, A., & Fatima, G. (2016). An approach based on association rules mining to improve road safety in Morocco. *2016 International Conference on Information Technology for Organizations Development (IT4OD)*, 1–6. https://doi.org/10.1109/IT4OD.2016.7479311

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 207–216.

Agrawal, R., Srikant, R., & others. (1994). Fast algorithms for mining association rules. *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, *1215*, 487–499.

Ahmed, I. (2013). Road infrastructure and road safety. *Transport and Communications Bulletin for Asia and the Pacific*, *83*, 19–25.

Ait-Mlouk, A., Agouti, T., & Gharnati, F. (2017). Mining and prioritization of association rules for big data: multi-criteria decision analysis approach. *Journal of Big Data*, *4*(1), 42. https://doi.org/10.1186/s40537-017-0105-4

Asare, I. O., & Mensah, A. C. (2020). Crash severity modelling using ordinal logistic regression approach. *International Journal of Injury Control and Safety Promotion*, *27*(4), 412–419.

Augeri, M. G., Greco, S., Distefano, N., & Leonardi, S. (2021). Road safety resource allocation using interactive multiobjective optimization. *European Transport*, *84*.

Azzalini, A., & Scarpa, B. (2012). *Data analysis and data mining: An introduction*. OUP USA.

Banihashemi, M. (2007). Optimization of Highway Safety and Operation by Using Crash Prediction Models with Accident Modification Factors. *Transportation Research Record*, *2019*(1), 108–118. https://doi.org/10.3141/2019-14

Banihashemi, M., & Dimaiuta, M. (2005). Maximizing safety improvement benefits in crash prediction models with accident modification factors. *Transportation Research Record*, *1908*(1), 9–18.

Batista, G. E., & Monard, M. C. (2002). A study of K-nearest neighbour as an imputation method. *His*, *87*(251–260), 48.

Bayardo, R. J., Agrawal, R., & Gunopulos, D. (2000). Constraint-Based Rule Mining in Large, Dense Databases. *Data Mining and Knowledge Discovery*, *4*(2), 217–240. https://doi.org/10.1023/A:1009895914772

Behnood, H. R., & Pino, S. (2020). Road safety resource allocation using the cost-effectiveness approach; Case study in Iran. *Proceedings of the Institution of Civil Engineers-Transport*, *176*(4), 213–223.

Bekiaris, E., Wiethoff, M., & Gaitanidou, E. (2011). *Infrastructure and safety in a collaborative world: road traffic safety*. Springer Science & Business Media.

Berry, M. J. A., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management* (Second edition). John Wiley & Sons.

Bhalla, K., Shahraz, S., Naghavi, M., & Murray, C. (2008). Estimating the potential impact of safety policies on road traffic death rates in developing countries. *Ninth World Conference on Injury Prevention and Safety Promotion, Merida, Mexico*.

Blincoe, L. M., Ted R Zaloshnja, Eduard Lawrence, & Bruce A. (2015). *The Economic and Societal Impact of Motor Vehicle Crashes*.

Bliss, T., & Breen, J. (2012). Meeting the management challenges of the Decade of Action for Road Safety. *IATSS Research*, *35*(2), 48–55. https://doi.org/https://doi.org/10.1016/j.iatssr.2011.12.001

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? *Available at SSRN 1819486*.

Byaruhanga, C. B., & Evdorides, H. (2022). A budget optimisation model for road safety infrastructure countermeasures. *Cogent Engineering*, *9*(1), 2129363. https://doi.org/10.1080/23311916.2022.2129363

Castro, Y., & Kim, Y. J. (2016). Data mining on road safety: factor assessment on vehicle accidents using classification models. *International Journal of Crashworthiness*, *21*(2), 104–111.

Chang, K.-H. (2015). Chapter 19 - Multiobjective Optimization and Advanced Topics. In K.-H. Chang (Ed.), *e-Design* (pp. 1105–1173). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-382038-9.00019-3

Chassiakos, A. P., Panagolia, C., & Theodorakopoulos, D. D. (2005). Development of decision-support system for managing highway safety. *Journal of Transportation Engineering*, *131*(5), 364–373.

Chen, S. Y., & Liu, X. (2004). The contribution of data mining to information science. *Journal of Information Science*, *30*(6), 550–558.

Chowdhury, M. A., Garber, N. J., & Li, D. (2000). Multiobjective methodology for highway safety resource allocation. *Journal of Infrastructure Systems*, *6*(4), 138–144.

Ciarlet, P. G., Miara, B., & Thomas, J.-M. (1989). *Introduction to numerical linear algebra and optimisation*. Cambridge university press.

Coll, B., Moutari, S., & Marshall, A. H. (2013). Hotspots identification and ranking for road safety improvement: An alternative approach. *Accident Analysis & Prevention*, *59*, 604–617.

Converse, S. J. (2020). Introduction to multi-criteria decision analysis. *Structured Decision Making: Case Studies in Natural Resource Management*, 51–61.

Cooper, W. W., Seiford, L. M., & Zhu, J. (2011). *Handbook on data envelopment analysis*.

Counsell, C. (1997). Formulating Questions and Locating Primary Studies for Inclusion in Systematic Reviews. *Annals of Internal Medicine*, *127*(5), 380–387. https://doi.org/10.7326/0003-4819-127-5-199709010-00008

Dadashi, A., & Mirbaha, B. (2019). Prioritizing highway safety improvement projects: A Monte-Carlo based Data Envelopment Analysis approach. *Accident Analysis & Prevention*, *123*, 387–395.

Das, S., Dutta, A., Avelar, R., Dixon, K., Sun, X., & Jalayer, M. (2019). Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures. *International Journal of Urban Sciences*, *23*(1), 30–48. https://doi.org/10.1080/12265934.2018.1431146

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197.

Department of Transport. (2019). *Reported road casualties in Great Britain: 2018 annual report*.

Dia, Y., Seck, T. C., Diallo, S. M., Sall, O., Bousso, M., & Sarr Ngor Edouard and Mboup, L. M. (2021). Senegalese Road Crash Data Analysis Based On Clustering And Association Rule Mining. *2021 4th International Conference on Data Storage and Data Engineering*, 70–76.

El Abdallaoui, H. E. A., El Fazziki, A., Ennaji, F. Z., & Sadgal, M. (2019). A system for collecting and analyzing road accidents Big Data. *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 663–671.

El Mazouri, F. Z., Abounaima, M. C., & Zenkouar, K. (2019). Data mining combined to the multicriteria decision analysis for the improvement of road safety: case of France. *Journal of Big Data*, *6*(1), 1–30.

Elvik, R. (2010). *Assessment and applicability of road safety management evaluation tools: Current practice and state-of-the-art in Europe*. Institute of Transport Economics.

Elvik, R. (2014). Problems in determining the optimal use of road safety measures. *Research in Transportation Economics*, *47*, 27–36.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd*, *96*(34), 226–231.

European Commission. (2019). *EU Road Safety Policy Framework 2021-2030 - Next steps towards 'Vision Zero'*.

European Commission. (2024a). *Annual statistical report on road safety in the EU, 2024*.

European Commission. (2024b). *Road safety thematic report – Main factors causing fatal crashes*.

Fancello, G., Carta, M., & Serra, P. (2020). Data Envelopment Analysis for the assessment of road safety in urban road networks: A comparative study using CCR and BCC models. *Case Studies on Transport Policy*, *8*(3), 736–744.

Federal Highway Administration. (2010). *The Essential Eight Fundamental Elements and Effective Steps for SHSP Implementation*.

Freitas, A. A. (2002). *Data mining and knowledge discovery with evolutionary algorithms*. Springer Science & Business Media.

Geurts, K., Wets, G., Brijs, T., & Vanhoof, K. (2003). Profiling of high-frequency accident locations by use of association rules. *Transportation Research Record*, *1840*(1), 123–130.

Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22*(2), 213–228.

Gough, D., Oliver, S., & Thomas, J. (2017). An introduction to systematic reviews. London: Sage. In *London: Sage*.

Gross, F., Harmon, T., Cynecki, M. J., Dittberner, R., Chestnutt, C., & others. (2021). *Selecting Projects and Strategies to Maximize Highway Safety Improvement Program Performance*.

Győrödi, C., Győrödi, R., & Holban, S. (2004). A comparative study of association rules mining algorithms. *SACI 2004, 1st Romanian-Hungarian Joint Symposium on Applied Computational Intelligence*, 213–222.

Haghighi, M. R. R., Sayari, M., Ghahramani, S., & Lankarani, K. B. (2020). Social, economic, and legislative factors and global road traffic fatalities. *BMC Public Health*, *20*, 1–12.

Hahsler, M., & Karpienko, R. (2017). Visualizing association rules in hierarchical groups. *Journal of Business Economics*, *87*(3), 317–335. https://doi.org/10.1007/s11573-016-0822-8

Hahsler M, Tyler G, & Chelluboina S. (2023). *Visualizing Association Rules and Frequent Itemsets*.

Han, J., Kamber, M., & Pei, J. (2011). *Cluster Analysis: Basic Concepts and Methods*. United States: Elsevier Science & Technology.

Han, J., Pei, J., & Yin, Y. (2000). Mining Frequent Patterns without Candidate Generation. *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, 1–12. https://doi.org/10.1145/342009.335372

Han, Y., Guan, H. Z., Zhao, K., & Wu, J. H. (2011). Synthetic Evaluation Method of Regional Road Safety Level. In *ICCTP 2011: Towards Sustainable Transportation Systems* (pp. 1904–1913).

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108. http://www.jstor.org/stable/2346830

Hasan, M., Han, Z., & Machemehl, R. (2021). Development of Crash Database Parsing Tools to Support the Highway Safety Improvement Program in Texas. *International Conference on Transportation and Development 2021*, 297–308.

Henry, A. (2021). *Understanding strategic management*. Oxford University Press.

Herbel, S., Laing, L., & McGovern, C. (2010). *Highway Safety Improvement Program (HSIP) Manual*.

Hogg, V., & Mathie, E. (2011). Development of a safety risk model for the highways agency road network. *6th IET International Conference on System Safety 2011*.

Houtsma, M., & Swami, A. (1995). Set-oriented mining for association rules in relational databases. *Proceedings of the Eleventh International Conference on Data Engineering*, 25–33. https://doi.org/10.1109/ICDE.1995.380413

International Road Assessment Programme. (2013a). *iRAP Road Attribute Risk Factors Median Type*.

International Road Assessment Programme. (2013b). *iRAP Road Attribute Risk Factors Street Lighting*.

International Road Assessment Programme. (2023). *iRAP Star Rating and Investment Plan Manual*.

International Transport Forum. (2023). *Road Safety Annual Report 2023*.

iRAP. (2022a). *Forgiving Roads Concept*. Https://Toolkit.Irap.Org/Safer-Road-Treatments/Forgiving-Roads-Concept/.

iRAP. (2022b). *Self-explaining Road (SER) Concept*. Https://Toolkit.Irap.Org/Safer-Road-Treatments/Self-Explaining-Road-Ser-Concept/.

iRAP. (2022c). *The Road Safety Toolkits*. Https://Toolkit.Irap.Org/.

Jackett, M., & Frith, W. (2013). Quantifying the impact of road lighting on road safety—A New Zealand Study. *IATSS Research*, *36*(2), 139–145.

Jain, A., Ahuja, G., Mehrotra, D., & others. (2016). Data mining approach to analyse the road accidents in India. *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, 175–179.

Jane, M., Henry, H., Craig, M., Thomas, B., Maryam, M., Keenan, P., Montufar, J., & Leanna, B. (2011). Risk mapping and a physical-attribute-based star rating system for road safety network screening purposes on Canadian prairie region highways. *2011 Conference and Exhibition of the Transportation Association of Canada - Transportation Successes: Let's Build on Them*.

Johnson, G., Scholes, K., & Whittington, R. (2008). *Exploring corporate strategy: text & cases*. Pearson Education.

Karacasu, M., Ergül, B., & Altin Yavuz, A. (2014). Estimating the causes of traffic accidents using logistic regression and discriminant analysis. *International Control and Safety Promotion*, *21*(4), 305–313.

Katsioloudes, M., & Abouhanian, A. K. (2009). *Strategic management*. Routledge.

Kitchin, R. (2014). *The Data Revolution : Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE Publications, Limited. http://ebookcentral.proquest.com/lib/bham/detail.action?docID=1712661

Koorey, G. (2009). Road Data Aggregation and Sectioning Considerations for Crash Analysis. *Transportation Research Record*, *2103*(1), 61–68. https://doi.org/10.3141/2103-08

Kumar, S., & Toshniwal, D. (2015). Analysing road accident data using association rule mining. *2015 International Conference on Computing, Communication and Security (ICCCS)*, 1–6. https://doi.org/10.1109/CCCS.2015.7374211

Lambert, J. H., Baker, J. A., & Peterson, K. D. (2003). Decision aid for allocation of transportation funds to guardrails. *Accident Analysis & Prevention*, *35*(1), 47–57. https://doi.org/https://doi.org/10.1016/S0001-4575(01)00085-9

Li, J., Liu, J., Toivonen, H., Satou, K., Sun, Y., & Sun, B. (2014). Discovering statistically non-redundant subgroups. *Knowledge-Based Systems*, *67*, 315–327. https://doi.org/https://doi.org/10.1016/j.knosys.2014.04.030

MacQueen, J. (1967). Classification and analysis of multivariate observations. *5th Berkeley Symp. Math. Statist. Probability*, 281–297.

Madhulatha, T. S. (2012). An Overview on Clustering Methods. *CoRR*, *abs/1205.1117*. http://arxiv.org/abs/1205.1117

Maji, A., Velaga, N. R., & Urie, Y. (2018). Hierarchical clustering analysis framework of mutually exclusive crash causation parameters for regional road safety strategies. *International Journal of Injury Control and Safety Promotion*, *25*(3), 257–271.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10), 60–68.

McKeown, J., Meegan, D., & Sprevak, D. (2022). *An introduction to unconstrained optimisation*. Routledge.

Mishra, S., Golias, M. M., Sharma, S., & Boyles, S. D. (2015). Optimal funding allocation strategies for safety improvements on urban intersections. *Transportation Research Part A: Policy and Practice*, *75*, 113–133. https://doi.org/https://doi.org/10.1016/j.tra.2015.03.001

Momeni Kho, S., Pahlavani, P., & Bigdeli, B. (2021). Classification and association rule mining of road collisions for analyzing the fatal severity, a case study. *Journal of Transport & Health*, *23*, 101278. https://doi.org/https://doi.org/10.1016/j.jth.2021.101278

Montella, A. (2011). Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. *Accident Analysis & Prevention*, *43*(4), 1451–1463. https://doi.org/https://doi.org/10.1016/j.aap.2011.02.023

Ng, K. S., Hung, W. T., & Wong, W. G. (2002). An algorithm for assessing the risk of traffic accident. *Journal of Safety Research*, *33*(3), 387–410.

Nick, P. (2015). *Strategic Road Network Statistics*.

Odoki, J. B., Di Graziano, A., & Akena, R. (2015). A multi-criteria methodology for optimising road investments. *Proceedings of the Institution of Civil Engineers-Transport*, *168*(1), 34–47.

Ott, R. L., & Longnecker, M. (2010). *An introduction to statistical methods and data analysis*. Cengage Learning Inc.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 89. https://doi.org/10.1186/s13643-021-01626-4

Pande, A., & Abdel-Aty, M. (2009). Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. *Safety Science*, *47*(1), 145–154. https://doi.org/https://doi.org/10.1016/j.ssci.2007.12.001

Park, P. Y., & Young, J. (2012). Investigation of a supplementary tool to assist in the prioritization of emphasis areas in North American strategic highway safety plans. *Accident Analysis & Prevention*, *45*, 392–405.

Peck, R., Short, T., & Olsen, C. (2020). *Introduction to statistics and data analysis*. Cengage Learning.

Pembuain, A., Priyanto, S., & Suparma, L. (2019). The Effect of Road Infrastructure on Traffic Accidents. *Proceedings of the 11th Asia Pacific Transportation and the Environment Conference (APTE 2018)*, 176–182. https://doi.org/10.2991/apte-18.2019.27

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51–59.

Pujari, A. K. (2001). *Data mining techniques*. Universities press.

Queensland Government. (2023). *Stopping distances: speed and braking*. Https://Www.Qld.Gov.Au/Transport/Safety/Road-Safety/Driving-Safely/Stopping-Distances.

Ramanathan, R. (2003). *An introduction to data envelopment analysis: a tool for performance measurement*. Sage.

Rao, P. S. (2009). *Strategic Management*. Global Media. http://ebookcentral.proquest.com/lib/bham/detail.action?docID=3011189

Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM*, *41*(2), 79–82.

Road Safety Foundation. (2024). *Road safety foundation crash risk mapping results 2024*. Road Safety Foundation.

Rokach, L., & Maimon, O. (2005). Clustering Methods. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 321–352). Springer US. https://doi.org/10.1007/0-387-25465-X_15

Saha, P., & Ksaibati, K. (2016). An optimization model for improving highway safety. *Journal of Traffic and Transportation Engineering*, *3*(6), 549–558.

Saloner, G., Shepard, A., & Podolny, J. (2005). *Strategic management*. John Wiley & Sons.

Saranyadevi, S., Murugeswari, R., & Bathrinath, S. (2019). Fuzzy logic based recommendation system for road accident data analysis. *2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, 1–6.

Shen, Y., Hermans, E., Brijs, T., Wets, G., & Vanhoof, K. (2012). Road safety risk evaluation and target setting using data envelopment analysis and its extensions. *Accident Analysis & Prevention*, *48*, 430–441. https://doi.org/https://doi.org/10.1016/j.aap.2012.02.020

Silverstein, C., Brin, S., & Motwani, R. (1998). Beyond Market Baskets: Generalizing Association Rules to Dependence Rules. *Data Mining and Knowledge Discovery*, *2*(1), 39–68. https://doi.org/10.1023/A:1009713703947

Smil, V. (2011). Global catastrophes and trends: the next fifty years. *Canadian Studies in Population*, *38*(1–2), 179–188.

Srinivas, N., & Deb, K. (1994). Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, *2*(3), 221–248.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Data mining introduction. *People's Posts and Telecommunications Publishing House, Beijing*.

The Royal Society for the Prevention of Accidents. (2018). *Road Safety information: Street Lighting and Road Safety*.

Thomas, J., Brunton, J., & Graziosi, S. (2010). EPPI-Reviewer 4.0: software for research synthesis. *EPPI-Centre Software. London: Social Science Research Unit, Institute of Education*.

Thompson, A., Strickland, A. J., & Gamble, J. (2007). *Crafting and Executing Strategy: The Quest for Competitive Advantage: Concepts and Cases*. McGraw-Hill Companies,Incorporated. https://books.google.co.uk/books?id=6vNGPgAACAAJ

Triantaphyllou, E., & Triantaphyllou, E. (2000). Introduction to multi-criteria decision making. *Multi-Criteria Decision Making Methods: A Comparative Study*, 1–4.

United Nations. (2020). *UN General Assembly resolution 74/299 "Improving global road safety"*.

Varhelyi, A. (2016). Road safety management–the need for a systematic approach. *The Open Transportation Journal*, *10*(1), 137–155.

Waka Kotahi NZ Transport Agency. (2021a). *2021-24 NLTP Investment Prioritisation Method*. Waka Kotahi NZ Transport Agency. https://www.nzta.govt.nz/planning-and-investment/planning-and-investment-knowledge-base/202124-nltp/2021-24-nltp-investment-prioritisation-method/

Waka Kotahi NZ Transport Agency. (2021b). *National Land Transport Programme*.

Waka Kotahi NZ Transport Agency. (2022). *One Network Framework (ONF) Classification Guidance*.

World Health Organisation. (2023). *Global status report on road safety 2023*.

World Health Organization. (2022). *High-level Meeting of the UN General Assembly on Global Road Safety*.

World Health Organization, United Nations Regional Commissions, & Road Safety Collaboration. (2021). *Global plan Decade of Action for Road Safety 2021-2030*.

Wu, S., Feng, X., Han, Y., & Wang, Q. (2012). Missing categorical data imputation approach based on similarity. *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2827–2832. https://doi.org/10.1109/ICSMC.2012.6378177

Xu, C., Bao, J., Wang, C., & Liu, P. (2018). Association rule analysis of factors contributing to extraordinarily severe traffic crashes in China. *Journal of Safety Research*, *67*, 65–75. https://doi.org/https://doi.org/10.1016/j.jsr.2018.09.013

Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, *2*, 165–193.

Yang, Y., Yuan, Z., Sun, D., & Wen, X. (2019). Analysis of the factors influencing highway crash risk in different regional types based on improved Apriori algorithm. *Advances in Transportation Studies*, *49*, 165–178. https://doi.org/10.4399/978882552809113

Yu, J., & Liu, Y. (2012). Prioritizing highway safety improvement projects: A multi-criteria model and case study with SafetyAnalyst. *Safety Science*, *50*(4), 1085–1092.

Yu, J., Liu, Y., & Mao, J. (2012). An Importance Segregation Based Multi-Objective Optimization Model for Traffic Safety Improvement Countermeasure Selection. *International Journal of Transportation Science and Technology*, *1*(2), 205–218.

Yu, S., Jia, Y., & Sun, D. (2019). Identifying Factors that Influence the Patterns of Road Crashes Using Association Rules: A case Study from Wisconsin, United States. *Sustainability*, *11*(7). https://doi.org/10.3390/su11071925

Appendices

Appendix A: Result of fatal rules in 2018

Group	No.		support	lift
1	2619	first road class=3 pedestrian crossing human control=0 light conditions=6 weather conditions=1 urban or rural area=2	0.001085	5.823923
1	1286	first road class=3 light conditions=6 weather conditions=1 urban or rural area=2	0.001085	5.810056
1	2611	junction detail=0 pedestrian crossing human control=0 light conditions=6 road surface conditions=1 urban or rural area=2	0.001239	5.608501
1	1280	junction detail=0 light conditions=6 road surface conditions=1 urban or rural area=2	0.001239	5.597244
1	2645	first road class=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 light conditions=6 weather conditions=1	0.001093	5.594019
1	1316	first road class=3 pedestrian crossing physical facilities=0 light conditions=6 weather conditions=1	0.001093	5.581319
1	1318	first road class=3 pedestrian crossing human control=0 light conditions=6 weather conditions=1	0.001109	5.579131
1	402	first road class=3 light conditions=6 weather conditions=1	0.001109	5.545036
1	2604	first road class=3 junction detail=0 pedestrian crossing human control=0 light conditions=6 urban or rural area=2	0.00115	5.527786
1	1278	first road class=3 junction detail=0 light conditions=6 urban or rural area=2	0.00115	5.513061
1	1303	first road class=3 junction detail=0 pedestrian crossing human control=0 light conditions=6	0.001191	5.392535
1	394	first road class=3 junction detail=0 light conditions=6	0.001191	5.36822
1	1310	junction detail=0 pedestrian crossing human control=0 light conditions=6 road surface conditions=1	0.001297	5.36261
1	396	junction detail=0 light conditions=6 road surface conditions=1	0.001297	5.342967
1	1295	pedestrian crossing human control=0 light conditions=6 road surface conditions=1 urban or rural area=2	0.00146	5.1376
1	389	light conditions=6 road surface conditions=1 urban or rural area=2	0.00146	5.123574

1	4652	first road class=3 road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.001403	5.043193
1	3875	first road class=3 road type=6 speed limit=60 junction detail=0 road surface conditions=1 urban or rural area=2	0.001403	5.035146
2	3999	first road class=3 road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0 road surface conditions=1	0.001411	4.921125
2	2801	first road class=3 road type=6 speed limit=60 junction detail=0 road surface conditions=1	0.001411	4.913507
2	1287	first road class=3 pedestrian crossing human control=0 light conditions=6 urban or rural area=2	0.001345	4.898617
2	387	first road class=3 light conditions=6 urban or rural area=2	0.001345	4.884784
2	1330	pedestrian crossing human control=0 light conditions=6 weather conditions=1 road surface conditions=1	0.001435	4.789274
2	411	pedestrian crossing human control=0 light conditions=6 road surface conditions=1	0.001517	4.781287
2	410	light conditions=6 weather conditions=1 road surface conditions=1	0.001435	4.759275
2	2615	junction detail=0 pedestrian crossing human control=0 light conditions=6 weather conditions=1 urban or rural area=2	0.001541	4.756771
2	58	light conditions=6 road surface conditions=1	0.001517	4.751331
2	1282	junction detail=0 light conditions=6 weather conditions=1 urban or rural area=2	0.001541	4.74538
2	4655	first road class=3 road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.001582	4.736426
2	3877	first road class=3 road type=6 speed limit=60 junction detail=0 weather conditions=1 urban or rural area=2	0.001582	4.728561
2	3886	first road class=3 speed limit=60 junction detail=0 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.001451	4.716048
2	2718	first road class=3 speed limit=60 junction detail=0 road surface conditions=1 urban or rural area=2	0.001451	4.707551
2	403	first road class=3 pedestrian crossing human control=0 light conditions=6	0.001386	4.706274
2	1314	junction detail=0 pedestrian crossing human control=0 light conditions=6 weather conditions=1	0.001647	4.683987
2	56	first road class=3 light conditions=6	0.001386	4.671034
2	398	junction detail=0 light conditions=6 weather conditions=1	0.001647	4.667764

2	4003	first road class=3 road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0 weather conditions=1	0.001598	4.644648
2	2803	first road class=3 road type=6 speed limit=60 junction detail=0 weather conditions=1	0.001598	4.637162
2	2601	road type=6 speed limit=60 pedestrian crossing human control=0 light conditions=6 weather conditions=1	0.001003	4.538458
2	2812	first road class=3 speed limit=60 junction detail=0 pedestrian crossing human control=0 road surface conditions=1	0.00146	4.528384
2	1273	road type=6 speed limit=60 light conditions=6 weather conditions=1	0.001003	4.524808
2	1519	first road class=3 speed limit=60 junction detail=0 road surface conditions=1	0.00146	4.519038
2	3879	first road class=3 road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0 urban or rural area=2	0.001957	4.465934
2	2716	first road class=3 road type=6 speed limit=60 junction detail=0 urban or rural area=2	0.001957	4.453513
2	3841	road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0 light conditions=6 urban or rural area=2	0.001101	4.444897
2	2591	road type=6 speed limit=60 junction detail=0 light conditions=6 urban or rural area=2	0.001101	4.430982
2	3890	first road class=3 speed limit=60 junction detail=0 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.001631	4.427764
2	2598	road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0 light conditions=6	0.001117	4.421484
2	2720	first road class=3 speed limit=60 junction detail=0 weather conditions=1 urban or rural area=2	0.001631	4.418434
2	2805	first road class=3 road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0	0.00199	4.412815
2	1269	road type=6 speed limit=60 junction detail=0 light conditions=6	0.001117	4.407916
2	1517	first road class=3 road type=6 speed limit=60 junction detail=0	0.00199	4.400886
2	1277	speed limit=60 pedestrian crossing human control=0 light conditions=6 weather conditions=1	0.001019	4.387266
2	384	speed limit=60 light conditions=6 weather conditions=1	0.001019	4.372628
2	1300	pedestrian crossing human control=0 light conditions=6 weather conditions=1 urban or rural area=2	0.001802	4.344825
2	391	light conditions=6 weather conditions=1 urban or rural area=2	0.001802	4.334375

2	2816	first road class=3 speed limit=60 junction detail=0 pedestrian crossing human control=0 weather conditions=1	0.001655	4.297147
2	1521	first road class=3 speed limit=60 junction detail=0 weather conditions=1	0.001655	4.287254
2	1271	speed limit=60 junction detail=0 pedestrian crossing human control=0 light conditions=6	0.001133	4.280838
2	381	speed limit=60 junction detail=0 light conditions=6	0.001133	4.266514
3	1284	junction detail=0 pedestrian crossing human control=0 light conditions=6 urban or rural area=2	0.002022	4.210217
3	386	junction detail=0 light conditions=6 urban or rural area=2	0.002022	4.2005
3	2722	first road class=3 speed limit=60 junction detail=0 pedestrian crossing human control=0 urban or rural area=2	0.002014	4.17874
3	1481	first road class=3 speed limit=60 junction detail=0 urban or rural area=2	0.002014	4.166255
3	416	pedestrian crossing human control=0 light conditions=6 weather conditions=1	0.001908	4.163225
3	412	pedestrian crossing physical facilities=0 light conditions=6 weather conditions=1	0.001884	4.149078
3	60	light conditions=6 weather conditions=1	0.001908	4.142138
3	400	junction detail=0 pedestrian crossing human control=0 light conditions=6	0.002145	4.126923
3	55	junction detail=0 light conditions=6	0.002145	4.111101
3	3934	first road class=3 road type=6 speed limit=60 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.001892	4.085058
3	1523	first road class=3 speed limit=60 junction detail=0 pedestrian crossing human control=0	0.002055	4.081732
3	2745	first road class=3 road type=6 speed limit=60 road surface conditions=1 urban or rural area=2	0.001892	4.078209
3	609	first road class=3 speed limit=60 junction detail=0	0.002055	4.069159
3	2861	first road class=3 road type=6 speed limit=60 pedestrian crossing human control=0 road surface conditions=1	0.001908	4.003101
3	1546	first road class=3 road type=6 speed limit=60 road surface conditions=1	0.001908	3.99658
3	2595	road type=6 speed limit=60 pedestrian crossing human control=0 light conditions=6 urban or rural area=2	0.001272	3.96429
3	1266	road type=6 speed limit=60 light conditions=6 urban or rural area=2	0.001272	3.953339
3	1274	road type=6 speed limit=60 pedestrian crossing human control=0 light conditions=6	0.001288	3.942757
3	382	road type=6 speed limit=60 light conditions=6	0.001288	3.932061

3	3900	road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.0023	3.867694
3	2724	road type=6 speed limit=60 junction detail=0 road surface conditions=1 urban or rural area=2	0.0023	3.862641
3	3939	first road class=3 road type=6 speed limit=60 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.002112	3.859504
3	2747	first road class=3 road type=6 speed limit=60 weather conditions=1 urban or rural area=2	0.002112	3.852464
3	2826	road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0 road surface conditions=1	0.00234	3.828937
3	1525	road type=6 speed limit=60 junction detail=0 road surface conditions=1	0.00234	3.824071
3	4183	first road class=3 road type=6 junction detail=0 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.002006	3.820949
3	393	pedestrian crossing human control=0 light conditions=6 urban or rural area=2	0.002332	3.8066
3	3154	first road class=3 road type=6 junction detail=0 road surface conditions=1 urban or rural area=2	0.002006	3.805646
3	54	light conditions=6 urban or rural area=2	0.002332	3.79696
3	385	speed limit=60 pedestrian crossing human control=0 light conditions=6	0.001305	3.795226
3	2866	first road class=3 road type=6 speed limit=60 pedestrian crossing human control=0 weather conditions=1	0.002136	3.79255
3	1548	first road class=3 road type=6 speed limit=60 weather conditions=1	0.002136	3.78583
3	53	speed limit=60 light conditions=6	0.001305	3.782999
3	2737	speed limit=60 junction detail=0 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.002365	3.733232
3	1484	speed limit=60 junction detail=0 road surface conditions=1 urban or rural area=2	0.002365	3.728
3	4188	first road class=3 road type=6 junction detail=0 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.002267	3.6961
3	3156	first road class=3 road type=6 junction detail=0 weather conditions=1 urban or rural area=2	0.002275	3.696004
3	1538	speed limit=60 junction detail=0 pedestrian crossing human control=0 road surface conditions=1	0.002406	3.65341
3	612	speed limit=60 junction detail=0 road surface conditions=1	0.002406	3.647869

3	2749	first road class=3 road type=6 speed limit=60 pedestrian crossing human control=0 urban or rural area=2	0.002528	3.647741
3	1489	first road class=3 road type=6 speed limit=60 urban or rural area=2	0.002528	3.638989
3	62	pedestrian crossing human control=0 light conditions=6	0.002454	3.630312
3	3158	first road class=3 road type=6 junction detail=0 pedestrian crossing human control=0 urban or rural area=2	0.002854	3.619865
3	2758	first road class=3 speed limit=60 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.002022	3.619162
3	1491	first road class=3 speed limit=60 road surface conditions=1 urban or rural area=2	0.002022	3.613414
3	2004	first road class=3 road type=6 junction detail=0 urban or rural area=2	0.002862	3.612897
3	2	light conditions=6	0.002454	3.60955
4	1550	first road class=3 road type=6 speed limit=60 pedestrian crossing human control=0	0.002569	3.601481
4	617	first road class=3 road type=6 speed limit=60	0.002569	3.593085
4	3905	road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.00265	3.514856
4	2726	road type=6 speed limit=60 junction detail=0 weather conditions=1 urban or rural area=2	0.00265	3.509167
4	2831	road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0 weather conditions=1	0.002699	3.487245
4	1559	first road class=3 speed limit=60 pedestrian crossing human control=0 road surface conditions=1	0.002047	3.485513
4	1527	road type=6 speed limit=60 junction detail=0 weather conditions=1	0.002699	3.481747
4	619	first road class=3 speed limit=60 road surface conditions=1	0.002047	3.478274
4	3167	first road class=3 junction detail=0 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.002732	3.453535
4	2006	first road class=3 junction detail=0 road surface conditions=1 urban or rural area=2	0.00274	3.451239
4	2763	first road class=3 speed limit=60 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.002259	3.4497
4	1493	first road class=3 speed limit=60 weather conditions=1 urban or rural area=2	0.002259	3.443273

4	2742	speed limit=60 junction detail=0 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.002715	3.395239
4	1486	speed limit=60 junction detail=0 weather conditions=1 urban or rural area=2	0.002715	3.389118
4	1543	speed limit=60 junction detail=0 pedestrian crossing human control=0 weather conditions=1	0.002781	3.351105
4	614	speed limit=60 junction detail=0 weather conditions=1	0.002781	3.344835
4	1564	first road class=3 speed limit=60 pedestrian crossing human control=0 weather conditions=1	0.0023	3.342922
4	621	first road class=3 speed limit=60 weather conditions=1	0.0023	3.33538
4	2008	first road class=3 junction detail=0 weather conditions=1 urban or rural area=2	0.003058	3.323429
4	2728	road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0 urban or rural area=2	0.00327	3.320111
4	1482	road type=6 speed limit=60 junction detail=0 urban or rural area=2	0.00327	3.31301
4	1529	road type=6 speed limit=60 junction detail=0 pedestrian crossing human control=0	0.003335	3.299251
4	610	road type=6 speed limit=60 junction detail=0	0.003335	3.292375
4	1495	first road class=3 speed limit=60 pedestrian crossing human control=0 urban or rural area=2	0.002683	3.25234
4	601	first road class=3 speed limit=60 urban or rural area=2	0.002683	3.244473
4	2773	road type=6 speed limit=60 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.003017	3.213154
4	1488	speed limit=60 junction detail=0 pedestrian crossing human control=0 urban or rural area=2	0.003351	3.212286
4	1497	road type=6 speed limit=60 road surface conditions=1 urban or rural area=2	0.003017	3.208598
4	600	speed limit=60 junction detail=0 urban or rural area=2	0.003351	3.205118
4	1574	road type=6 speed limit=60 pedestrian crossing human control=0 road surface conditions=1	0.00309	3.186126
4	2010	first road class=3 junction detail=0 pedestrian crossing human control=0 urban or rural area=2	0.003759	3.18547
4	970	first road class=3 junction detail=0 urban or rural area=2	0.003775	3.182808
4	625	road type=6 speed limit=60 road surface conditions=1	0.00309	3.181025
4	616	speed limit=60 junction detail=0 pedestrian crossing human control=0	0.003433	3.172204
4	173	speed limit=60 junction detail=0	0.003433	3.165055
4	623	first road class=3 speed limit=60 pedestrian crossing human control=0	0.00274	3.164669

4	174	first road class=3 speed limit=60	0.00274	3.156163
5	1602	day of week=7 junction detail=0 pedestrian crossing human control=0 urban or rural area=2	0.001076	3.032083
5	645	day of week=7 junction detail=0 urban or rural area=2	0.001076	3.019796
5	1510	speed limit=60 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.003164	2.993943
5	604	speed limit=60 road surface conditions=1 urban or rural area=2	0.003164	2.989542
5	2778	road type=6 speed limit=60 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.003433	2.974609
5	1499	road type=6 speed limit=60 weather conditions=1 urban or rural area=2	0.003433	2.969749
5	1579	road type=6 speed limit=60 pedestrian crossing human control=0 weather conditions=1	0.003514	2.952046
5	3220	first road class=3 road type=6 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.002781	2.947362
5	627	road type=6 speed limit=60 weather conditions=1	0.003514	2.94682
5	2036	first road class=3 road type=6 road surface conditions=1 urban or rural area=2	0.002781	2.933199
5	638	speed limit=60 pedestrian crossing human control=0 road surface conditions=1	0.003245	2.931777
5	547	day of week=1 junction detail=0 urban or rural area=2	0.00106	2.931098
5	177	speed limit=60 road surface conditions=1	0.003245	2.925903
5	4262	first road class=3 road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.002968	2.852234
5	3221	first road class=3 road type=6 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.002968	2.847067
5	3225	first road class=3 road type=6 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.00309	2.834204
5	2038	first road class=3 road type=6 weather conditions=1 urban or rural area=2	0.003099	2.829286
5	602	road type=6 speed limit=60 urban or rural area=2	0.004118	2.796291
5	1515	speed limit=60 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.003596	2.793946
5	606	speed limit=60 weather conditions=1 urban or rural area=2	0.003596	2.789131

5	3223	first road class=3 road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 urban or rural area=2	0.00362	2.786731
5	2037	first road class=3 road type=6 pedestrian crossing physical facilities=0 urban or rural area=2	0.00362	2.780785
5	2040	first road class=3 road type=6 pedestrian crossing human control=0 urban or rural area=2	0.003775	2.77884
5	629	road type=6 speed limit=60 pedestrian crossing human control=0	0.004208	2.776955
5	175	road type=6 speed limit=60	0.004216	2.776636
5	979	first road class=3 road type=6 urban or rural area=2	0.003784	2.771018
5	548	day of week=1 road type=6 urban or rural area=2	0.001117	2.757667
5	643	speed limit=60 pedestrian crossing human control=0 weather conditions=1	0.003702	2.75092
5	639	speed limit=60 pedestrian crossing physical facilities=0 weather conditions=1	0.003653	2.746308
5	179	speed limit=60 weather conditions=1	0.003702	2.745028
5	3182	road type=6 junction detail=0 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.003319	2.744882
5	1604	day of week=7 road type=6 pedestrian crossing human control=0 urban or rural area=2	0.00115	2.73902
5	2012	road type=6 junction detail=0 road surface conditions=1 urban or rural area=2	0.003319	2.734078
5	646	day of week=7 road type=6 urban or rural area=2	0.00115	2.733954
5	3046	first road class=3 road type=6 junction detail=0 pedestrian crossing human control=0 road surface conditions=2	0.001019	2.713331
5	1399	Regional areas=East Midlands pedestrian crossing human control=0 pedestrian crossing physical facilities=0 urban or rural area=2	0.001076	2.71055
5	468	Regional areas=East Midlands pedestrian crossing physical facilities=0 urban or rural area=2	0.001076	2.704496
5	1859	first road class=3 road type=6 junction detail=0 road surface conditions=2	0.001027	2.685007
5	1426	day of week=1 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.001027	2.677233
5	469	Regional areas=East Midlands pedestrian crossing human control=0 urban or rural area=2	0.001117	2.671216
5	1843	road type=6 junction detail=0 road surface conditions=2 urban or rural area=2	0.001451	2.667644
5	549	day of week=1 road surface conditions=1 urban or rural area=2	0.001027	2.665657

5	82	Regional areas=East Midlands urban or rural area=2	0.001117	2.66555
5	1642	day of week=2 junction detail=0 pedestrian crossing human control=0 urban or rural area=2	0.001027	2.664121
5	3187	road type=6 junction detail=0 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.003865	2.65874
5	2014	road type=6 junction detail=0 weather conditions=1 urban or rural area=2	0.003873	2.6536
5	676	day of week=2 junction detail=0 urban or rural area=2	0.001027	2.651897
5	608	speed limit=60 pedestrian crossing human control=0 urban or rural area=2	0.004289	2.626964
5	1380	first road class=3 pedestrian crossing human control=0 urban or rural area=2 trunk road flag=1	0.001101	2.626637
5	172	speed limit=60 urban or rural area=2	0.004297	2.626596
5	450	first road class=3 urban or rural area=2 trunk road flag=1	0.001101	2.62316
6	551	day of week=1 weather conditions=1 urban or rural area=2	0.001158	2.598207
6	2016	road type=6 junction detail=0 pedestrian crossing human control=0 urban or rural area=2	0.004827	2.596175
6	971	road type=6 junction detail=0 urban or rural area=2	0.004844	2.593169
6	181	speed limit=60 pedestrian crossing human control=0	0.004411	2.586586
6	25	speed limit=60	0.00442	2.585136
6	3426	first road class=3 road type=6 junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.003196	2.572791
6	2203	first road class=3 road type=6 junction detail=0 pedestrian crossing physical facilities=0	0.003196	2.563164
6	1394	first road class=3 pedestrian crossing human control=0 weather conditions=1 trunk road flag=1	0.001003	2.556498
6	460	first road class=3 weather conditions=1 trunk road flag=1	0.001003	2.550719
6	3239	first road class=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=1 urban or rural area=2	0.003629	2.548867
6	2046	first road class=3 pedestrian crossing physical facilities=0 road surface conditions=1 urban or rural area=2	0.003629	2.5443
6	2025	junction detail=0 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.004444	2.527498
6	973	junction detail=0 road surface conditions=1 urban or rural area=2	0.004452	2.522571
6	2049	first road class=3 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.003792	2.521534

6	981	first road class=3 road surface conditions=1 urban or rural area=2	0.0038	2.514878
6	3244	first road class=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.004069	2.494836
6	2050	first road class=3 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.004069	2.490764
6	2923	day of week=7 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.001093	2.484033
6	1605	day of week=7 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.001093	2.482152
6	1607	day of week=7 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.001125	2.470206
6	470	Regional areas=East Midlands junction detail=0 pedestrian crossing human control=0	0.001044	2.464955
6	648	day of week=7 weather conditions=1 urban or rural area=2	0.001125	2.464196
6	83	Regional areas=East Midlands junction detail=0	0.001044	2.456575
6	1728	first road class=3 road type=3 weather conditions=1 urban or rural area=2	0.001011	2.456244
6	2054	first road class=3 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.004216	2.450131
6	983	first road class=3 weather conditions=1 urban or rural area=2	0.004232	2.448541
6	2030	junction detail=0 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.005056	2.443843
6	975	junction detail=0 weather conditions=1 urban or rural area=2	0.005072	2.441498
6	461	first road class=3 pedestrian crossing human control=0 trunk road flag=1	0.001199	2.439701
6	77	first road class=3 trunk road flag=1	0.001199	2.434745
6	871	junction detail=0 road surface conditions=2 urban or rural area=2	0.001843	2.433062
6	2206	first road class=3 road type=6 junction detail=0 pedestrian crossing human control=0	0.003514	2.411648
6	977	junction detail=0 pedestrian crossing human control=0 urban or rural area=2	0.006385	2.408808
6	293	junction detail=0 urban or rural area=2	0.006409	2.406738
6	153	day of week=1 urban or rural area=2	0.001378	2.404603
6	2052	first road class=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 urban or rural area=2	0.00486	2.380827

6	982	first road class=3 pedestrian crossing physical facilities=0 urban or rural area=2	0.00486	2.376558
6	3047	first road class=3 junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=2	0.001191	2.371093
6	1860	first road class=3 junction detail=0 pedestrian crossing physical facilities=0 road surface conditions=2	0.001191	2.363248
6	1059	first road class=3 road type=6 junction detail=0	0.003531	2.360216
6	985	first road class=3 pedestrian crossing human control=0 urban or rural area=2	0.005047	2.349427
6	1861	first road class=3 junction detail=0 pedestrian crossing human control=0 road surface conditions=2	0.001337	2.346654
6	294	first road class=3 urban or rural area=2	0.005064	2.344047
6	649	day of week=7 pedestrian crossing human control=0 urban or rural area=2	0.00137	2.342685
6	647	day of week=7 pedestrian crossing physical facilities=0 urban or rural area=2	0.001321	2.335798
6	182	day of week=7 urban or rural area=2	0.00137	2.3347
6	3447	first road class=3 junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.003539	2.332065
6	3442	first road class=3 junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=1	0.003131	2.330811
6	1611	day of week=7 road type=6 junction detail=0 pedestrian crossing human control=0	0.001305	2.324313
6	2216	first road class=3 junction detail=0 pedestrian crossing physical facilities=0 weather conditions=1	0.003539	2.322542
6	2212	first road class=3 junction detail=0 pedestrian crossing physical facilities=0 road surface conditions=1	0.003131	2.319492
6	879	first road class=3 junction detail=0 road surface conditions=2	0.001345	2.310069
6	1608	day of week=7 road type=6 junction detail=0 pedestrian crossing physical facilities=0	0.001166	2.303006
6	2218	first road class=3 junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.004395	2.300646
6	650	day of week=7 road type=6 junction detail=0	0.001313	2.293443
6	1062	first road class=3 junction detail=0 pedestrian crossing physical facilities=0	0.004395	2.290124
7	101	Regional areas=West Midlands junction detail=0	0.001036	2.220237

7	3286	road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=1 urban or rural area=2	0.004452	2.209353
7	2081	road type=6 pedestrian crossing physical facilities=0 road surface conditions=1 urban or rural area=2	0.00446	2.208043
7	2220	first road class=3 junction detail=0 pedestrian crossing human control=0 weather conditions=1	0.003922	2.201889
7	2084	road type=6 pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.004607	2.199642
7	994	road type=6 road surface conditions=1 urban or rural area=2	0.004615	2.193876
7	1389	junction detail=0 pedestrian crossing human control=0 road surface conditions=1 trunk road flag=1	0.001076	2.192239
7	455	junction detail=0 road surface conditions=1 trunk road flag=1	0.001076	2.187289
7	2693	day of week=1 road type=6 junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001019	2.182153
7	1065	first road class=3 junction detail=0 pedestrian crossing human control=0	0.004868	2.181317
7	1346	road type=3 speed limit=70 pedestrian crossing human control=0 urban or rural area=2	0.001451	2.177605
7	1430	day of week=1 road type=6 junction detail=0 pedestrian crossing physical facilities=0	0.001019	2.175426
7	425	road type=3 speed limit=70 urban or rural area=2	0.001451	2.173619
7	458	junction detail=0 pedestrian crossing human control=0 trunk road flag=1	0.001492	2.163403
7	76	junction detail=0 trunk road flag=1	0.001492	2.158882
7	1063	first road class=3 junction detail=0 weather conditions=1	0.003947	2.1567
7	3291	road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.005113	2.150364
7	2085	road type=6 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.005121	2.148873
7	1431	day of week=1 road type=6 junction detail=0 pedestrian crossing human control=0	0.001085	2.141956
7	2089	road type=6 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.00526	2.133436
7	996	road type=6 weather conditions=1 urban or rural area=2	0.005276	2.131023
7	655	day of week=7 junction detail=0 pedestrian crossing human control=0	0.001582	2.124395

7	2694	day of week=1 junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.001044	2.124365
7	314	first road class=3 junction detail=0	0.004893	2.120287
7	553	day of week=1 road type=6 junction detail=0	0.001093	2.118088
7	1432	day of week=1 junction detail=0 pedestrian crossing physical facilities=0 weather conditions=1	0.001044	2.11623
7	1618	day of week=7 first road class=3 road type=6 pedestrian crossing human control=0	0.001003	2.104194
7	379	speed limit=50 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001093	2.100445
7	652	day of week=7 junction detail=0 pedestrian crossing physical facilities=0	0.001411	2.099289
7	51	speed limit=50 pedestrian crossing physical facilities=0	0.001093	2.097757
7	1434	day of week=1 junction detail=0 pedestrian crossing human control=0 weather conditions=1	0.001125	2.096428
7	454	pedestrian crossing human control=0 urban or rural area=2 trunk road flag=1	0.001655	2.080755
7	995	road type=6 pedestrian crossing physical facilities=0 urban or rural area=2	0.006181	2.0769
7	1349	road type=3 speed limit=70 junction detail=0 pedestrian crossing human control=0	0.001345	2.076012
7	75	urban or rural area=2 trunk road flag=1	0.001655	2.075827
7	183	day of week=7 junction detail=0	0.00159	2.075274
7	430	road type=3 speed limit=70 pedestrian crossing human control=0	0.001598	2.074185
7	426	road type=3 speed limit=70 junction detail=0	0.001345	2.071396
7	4106	road type=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1 road surface conditions=1 urban or rural area=2	0.001256	2.071118
7	2943	day of week=2 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.001068	2.071115
7	555	day of week=1 junction detail=0 weather conditions=1	0.001133	2.070056
7	2973	road type=3 pedestrian crossing physical facilities=0 weather conditions=1 road surface conditions=1 urban or rural area=2	0.001256	2.06998
7	296	road type=6 urban or rural area=2	0.006393	2.069634

7	67	road type=3 speed limit=70	0.001598	2.068816
7	1643	day of week=2 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.001068	2.064887
7	438	speed limit=70 junction detail=0 pedestrian crossing human control=0	0.001394	2.06342
7	69	speed limit=70 junction detail=0	0.001394	2.058681
7	1433	day of week=1 junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001288	2.052327
7	2974	road type=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=1 urban or rural area=2	0.001288	2.05015
7	1730	road type=3 pedestrian crossing physical facilities=0 road surface conditions=1 urban or rural area=2	0.001288	2.048701
7	1731	road type=3 weather conditions=1 road surface conditions=1 urban or rural area=2	0.001288	2.047254
7	1645	day of week=2 pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.001101	2.045772
7	554	day of week=1 junction detail=0 pedestrian crossing physical facilities=0	0.001288	2.044367
7	1644	day of week=2 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 urban or rural area=2	0.001264	2.038251
7	656	day of week=7 first road class=3 road type=6	0.001011	2.036791
7	678	day of week=2 weather conditions=1 urban or rural area=2	0.001101	2.035266
7	435	speed limit=70 pedestrian crossing human control=0 urban or rural area=2	0.001525	2.034084
7	677	day of week=2 pedestrian crossing physical facilities=0 urban or rural area=2	0.001264	2.032787
7	68	speed limit=70 urban or rural area=2	0.001525	2.029872
7	52	speed limit=50 pedestrian crossing human control=0	0.001125	2.02638
7	781	road type=3 road surface conditions=1 urban or rural area=2	0.001321	2.026109
7	2976	road type=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.001427	2.025435
7	1733	road type=3 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.001427	2.023839
7	556	day of week=1 junction detail=0 pedestrian crossing human control=0	0.001378	2.02101
7	1	speed limit=50	0.001133	2.018847

7	679	day of week=2 pedestrian crossing human control=0 urban or rural area=2	0.001297	2.010517
7	3305	pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1 road surface conditions=1 urban or rural area=2	0.005626	2.009334
7	2101	pedestrian crossing physical facilities=0 weather conditions=1 road surface conditions=1 urban or rural area=2	0.005635	2.008261
7	2103	pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=1 urban or rural area=2	0.005879	2.006991
7	1004	pedestrian crossing physical facilities=0 road surface conditions=1 urban or rural area=2	0.005887	2.005591
7	192	day of week=2 urban or rural area=2	0.001297	1.999493
7	783	road type=3 weather conditions=1 urban or rural area=2	0.00146	1.995268
7	1007	pedestrian crossing human control=0 road surface conditions=1 urban or rural area=2	0.006075	1.989437
7	154	day of week=1 junction detail=0	0.001386	1.986044
7	298	road surface conditions=1 urban or rural area=2	0.006091	1.985458
7	419	Regional areas=Scotland pedestrian crossing human control=0 weather conditions=1	0.001052	1.982688
7	1734	road type=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 urban or rural area=2	0.001778	1.978857
7	782	road type=3 pedestrian crossing physical facilities=0 urban or rural area=2	0.001778	1.97739
7	3049	road type=6 junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=2	0.001786	1.974503
7	2108	pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.006695	1.970802
7	1008	pedestrian crossing physical facilities=0 weather conditions=1 urban or rural area=2	0.006703	1.969338
7	1862	road type=6 junction detail=0 pedestrian crossing physical facilities=0 road surface conditions=2	0.001786	1.967011
7	239	road type=3 urban or rural area=2	0.001827	1.962679
7	64	Regional areas=Scotland weather conditions=1	0.001052	1.960516
7	1012	pedestrian crossing human control=0 weather conditions=1 urban or rural area=2	0.00689	1.949351

7	3053	first road class=3 road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=2	0.001231	1.947955
7	300	weather conditions=1 urban or rural area=2	0.006915	1.947274
7	1871	first road class=3 road type=6 pedestrian crossing physical facilities=0 road surface conditions=2	0.001231	1.942833
7	1619	day of week=7 first road class=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001052	1.936456
7	73	speed limit=70 pedestrian crossing human control=0	0.001672	1.927609
7	883	junction detail=0 weather conditions=1 road surface conditions=2	0.001191	1.926113
7	657	day of week=7 first road class=3 pedestrian crossing physical facilities=0	0.001052	1.925821
7	81	pedestrian crossing human control=0 trunk road flag=1	0.001802	1.923305
7	4	speed limit=70	0.001672	1.922436
7	880	road type=6 junction detail=0 road surface conditions=2	0.001949	1.917806
7	5	trunk road flag=1	0.001802	1.917167
7	299	pedestrian crossing physical facilities=0 urban or rural area=2	0.008154	1.913096
7	94	Regional areas=Southwest urban or rural area=2	0.001027	1.904276
7	302	pedestrian crossing human control=0 urban or rural area=2	0.008399	1.899962
7	37	urban or rural area=2	0.00844	1.899161
7	1868	junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=2	0.00234	1.894494
7	882	junction detail=0 pedestrian crossing physical facilities=0 road surface conditions=2	0.00234	1.887533
8	885	junction detail=0 pedestrian crossing human control=0 road surface conditions=2	0.002536	1.862756
8	270	junction detail=0 road surface conditions=2	0.002577	1.85932
8	1456	first road class=4 junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001044	1.850659
8	1620	day of week=7 first road class=3 pedestrian crossing human control=0 weather conditions=1	0.001068	1.844965
8	579	first road class=4 junction detail=0 pedestrian crossing physical facilities=0	0.001044	1.844119
8	1650	day of week=2 first road class=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001093	1.830997

8	558	day of week=1 first road class=3 pedestrian crossing human control=0	0.001215	1.82648
8	684	day of week=2 first road class=3 pedestrian crossing physical facilities=0	0.001093	1.822176
8	418	Regional areas=Scotland pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001068	1.8133
8	63	Regional areas=Scotland pedestrian crossing physical facilities=0	0.001068	1.807164
8	557	day of week=1 first road class=3 weather conditions=1	0.001003	1.799999
8	1455	first road class=4 road type=6 junction detail=0 pedestrian crossing human control=0	0.001068	1.796024
8	578	first road class=4 road type=6 junction detail=0	0.001076	1.783086
8	659	day of week=7 first road class=3 pedestrian crossing human control=0	0.001329	1.782271
8	580	first road class=4 junction detail=0 pedestrian crossing human control=0	0.001133	1.78032
8	2998	junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 light conditions=4 weather conditions=1	0.001011	1.774646
8	1787	junction detail=0 pedestrian crossing physical facilities=0 light conditions=4 weather conditions=1	0.001011	1.769126
8	658	day of week=7 first road class=3 weather conditions=1	0.001068	1.768601
8	155	day of week=1 first road class=3	0.001231	1.766324
8	1791	junction detail=0 pedestrian crossing human control=0 light conditions=4 weather conditions=1	0.001199	1.765978
8	2986	first road class=3 road type=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.001403	1.761528
8	164	first road class=4 junction detail=0	0.001142	1.75815
8	835	junction detail=0 light conditions=4 weather conditions=1	0.001223	1.755746
8	1749	first road class=3 road type=3 pedestrian crossing physical facilities=0 weather conditions=1	0.001403	1.754672
8	1646	day of week=2 road type=6 junction detail=0 pedestrian crossing human control=0	0.001027	1.746725
8	2363	first road class=3 road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.004517	1.735674
8	2984	first road class=3 road type=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=1	0.001239	1.728165
8	1126	first road class=3 road type=6 pedestrian crossing physical facilities=0	0.004517	1.728149

8	65	Regional areas=Scotland pedestrian crossing human control=0	0.001215	1.726146
8	1746	first road class=3 road type=3 pedestrian crossing physical facilities=0 road surface conditions=1	0.001239	1.721233
8	1648	day of week=2 junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001239	1.718316
8	1873	first road class=3 road type=6 pedestrian crossing human control=0 road surface conditions=2	0.001435	1.716045
8	1786	road type=6 junction detail=0 pedestrian crossing human control=0 light conditions=4	0.001109	1.712312
8	681	day of week=2 junction detail=0 pedestrian crossing physical facilities=0	0.001239	1.710674
8	680	day of week=2 road type=6 junction detail=0	0.001027	1.709588
8	3	Regional areas=Scotland	0.001215	1.707281
8	833	road type=6 junction detail=0 light conditions=4	0.001133	1.706749
8	184	day of week=7 first road class=3	0.001337	1.699519
8	1789	junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 light conditions=4	0.001305	1.687131
8	886	first road class=3 road type=6 road surface conditions=2	0.001451	1.68626
8	683	day of week=2 junction detail=0 pedestrian crossing human control=0	0.001362	1.683076
8	834	junction detail=0 pedestrian crossing physical facilities=0 light conditions=4	0.001305	1.680611
8	789	road type=3 junction detail=0 pedestrian crossing human control=0	0.001981	1.679741
8	837	junction detail=0 pedestrian crossing human control=0 light conditions=4	0.001549	1.673765
8	787	road type=3 junction detail=0 pedestrian crossing physical facilities=0	0.001802	1.672603
8	240	road type=3 junction detail=0	0.001998	1.668578
8	476	Regional areas=East Midlands pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001288	1.667
8	86	Regional areas=East Midlands pedestrian crossing physical facilities=0	0.001288	1.662459
8	1750	first road class=3 road type=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001647	1.660114
8	792	first road class=3 road type=3 pedestrian crossing physical facilities=0	0.001647	1.65308
8	193	day of week=2 junction detail=0	0.00137	1.648997
8	2274	road type=6 junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.006108	1.647512

8	259	junction detail=0 light conditions=4	0.001574	1.647204
8	1083	road type=6 junction detail=0 pedestrian crossing physical facilities=0	0.006108	1.640628
8	1875	first road class=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=2	0.001631	1.627096
8	887	first road class=3 pedestrian crossing physical facilities=0 road surface conditions=2	0.001631	1.621345
8	2935	day of week=7 road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.001329	1.606797
8	1439	day of week=1 road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001427	1.605812
8	1098	junction detail=0 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.008089	1.605218
8	1625	day of week=7 road type=6 pedestrian crossing physical facilities=0 weather conditions=1	0.001329	1.601419
8	560	day of week=1 road type=6 pedestrian crossing physical facilities=0	0.001427	1.599612
8	320	junction detail=0 pedestrian crossing physical facilities=0	0.008089	1.5981
8	1086	road type=6 junction detail=0 pedestrian crossing human control=0	0.006605	1.591637
8	317	road type=6 junction detail=0	0.006646	1.569607
9	2930	day of week=7 road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=1	0.001166	1.562423
9	2384	first road class=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.005202	1.557546
9	1621	day of week=7 road type=6 pedestrian crossing physical facilities=0 road surface conditions=1	0.001166	1.55663
9	323	junction detail=0 pedestrian crossing human control=0	0.008774	1.556477
9	1139	first road class=3 pedestrian crossing physical facilities=0 weather conditions=1	0.005202	1.55048
9	1627	day of week=7 road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001598	1.545057
9	1651	day of week=2 first road class=3 pedestrian crossing human control=0 weather conditions=1	0.00106	1.542309
9	661	day of week=7 road type=6 pedestrian crossing physical facilities=0	0.001598	1.539764
9	1141	first road class=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.006328	1.537922

9	89	Regional areas=East Midlands pedestrian crossing human control=0	0.00146	1.532708
9	1629	day of week=7 road type=6 pedestrian crossing human control=0 weather conditions=1	0.001533	1.531678
9	330	first road class=3 pedestrian crossing physical facilities=0	0.006328	1.530359
9	1441	day of week=1 road type=6 pedestrian crossing human control=0 weather conditions=1	0.001313	1.527974
9	6	Regional areas=East Midlands	0.00146	1.527007
9	39	junction detail=0	0.008831	1.526261
9	1751	first road class=3 road type=3 pedestrian crossing human control=0 weather conditions=1	0.001704	1.520776
9	563	day of week=1 road type=6 pedestrian crossing human control=0	0.001615	1.515409
9	561	day of week=1 road type=6 weather conditions=1	0.001329	1.509857
9	686	day of week=2 first road class=3 pedestrian crossing human control=0	0.001231	1.509592
9	1449	day of week=1 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.00146	1.504104
9	662	day of week=7 road type=6 weather conditions=1	0.001541	1.503767
9	1624	day of week=7 road type=6 pedestrian crossing human control=0 road surface conditions=1	0.001362	1.502533
9	569	day of week=1 pedestrian crossing physical facilities=0 weather conditions=1	0.00146	1.497588
9	1404	Regional areas=Southwest pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=1	0.001019	1.496781
9	1129	first road class=3 road type=6 pedestrian crossing human control=0	0.005292	1.495032
9	485	Regional areas=Southwest pedestrian crossing physical facilities=0 road surface conditions=1	0.001019	1.492156
9	685	day of week=2 first road class=3 weather conditions=1	0.001068	1.491717
9	156	day of week=1 road type=6	0.001631	1.491519
9	889	first road class=3 pedestrian crossing human control=0 road surface conditions=2	0.00199	1.487927
9	664	day of week=7 road type=6 pedestrian crossing human control=0	0.001867	1.484136
9	793	first road class=3 road type=3 weather conditions=1	0.001712	1.483343
9	571	day of week=1 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001794	1.479641

9	1760	road type=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.001916	1.47888
9	804	road type=3 pedestrian crossing physical facilities=0 weather conditions=1	0.001916	1.474707
9	159	day of week=1 pedestrian crossing physical facilities=0	0.001794	1.473026
9	760	Regional areas=Southeast first road class=3 weather conditions=1	0.001019	1.466865
9	795	first road class=3 road type=3 pedestrian crossing human control=0	0.00203	1.466508
9	805	road type=3 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.002381	1.464694
9	660	day of week=7 road type=6 road surface conditions=1	0.001362	1.463247
9	245	road type=3 pedestrian crossing physical facilities=0	0.002381	1.460004
9	185	day of week=7 road type=6	0.001884	1.456205
9	271	first road class=3 road surface conditions=2	0.002014	1.454495
9	1403	Regional areas=Southwest Road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001068	1.447473
9	327	first road class=3 road type=6	0.005333	1.445785
9	194	day of week=2 first road class=3	0.001239	1.445363
9	483	Regional areas=Southwest Road type=6 pedestrian crossing physical facilities=0	0.001068	1.443127
9	1638	day of week=7 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.001566	1.439906
9	573	day of week=1 pedestrian crossing human control=0 weather conditions=1	0.001704	1.438618
9	670	day of week=7 pedestrian crossing physical facilities=0 weather conditions=1	0.001566	1.434482
9	791	first road class=3 road type=3 road surface conditions=1	0.0015	1.428368
9	241	first road class=3 road type=3	0.002039	1.425052
9	1406	Regional areas=Southwest pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.001085	1.414008
9	488	Regional areas=Southwest pedestrian crossing physical facilities=0 weather conditions=1	0.001085	1.409719
9	160	day of week=1 weather conditions=1	0.001721	1.408782

9	231	Regional areas=Southeast first road class=3	0.001174	1.406092
9	1408	Regional areas=West Midlands pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.001011	1.404163
9	672	day of week=7 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001933	1.401569
9	162	day of week=1 pedestrian crossing human control=0	0.002071	1.400639
9	496	Regional areas=West Midlands pedestrian crossing physical facilities=0 weather conditions=1	0.001011	1.40049
9	1405	Regional areas=Southwest pedestrian crossing human control=0 weather conditions=1 road surface conditions=1	0.001036	1.398223
9	188	day of week=7 pedestrian crossing physical facilities=0	0.001933	1.395609
9	490	Regional areas=Southwest pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001329	1.394242
9	486	Regional areas=Southwest weather conditions=1 road surface conditions=1	0.001036	1.392166
9	97	Regional areas=Southwest pedestrian crossing physical facilities=0	0.001329	1.390353
9	487	Regional areas=Southwest pedestrian crossing human control=0 road surface conditions=1	0.001076	1.38472
9	96	Regional areas=Southwest Road surface conditions=1	0.001076	1.378808
9	501	Date=December Road type=6 pedestrian crossing human control=0	0.001068	1.375213
9	674	day of week=7 pedestrian crossing human control=0 weather conditions=1	0.001851	1.369017
9	3056	road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1 road surface conditions=2	0.001133	1.365627
9	1885	road type=6 pedestrian crossing physical facilities=0 weather conditions=1 road surface conditions=2	0.001133	1.361617
9	23	day of week=1	0.002087	1.361342
9	1143	first road class=3 pedestrian crossing human control=0 weather conditions=1	0.006254	1.358717
9	95	Regional areas=Southwest Road type=6	0.001117	1.350498
9	1415	Date=October pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.001044	1.350093
9	248	road type=3 pedestrian crossing human control=0	0.00283	1.343873
9	333	first road class=3 pedestrian crossing human control=0	0.007592	1.34363

9	519	Date=October pedestrian crossing physical facilities=0 weather conditions=1	0.001044	1.343335
9	108	Date=December Road type=6	0.001076	1.340089
9	1892	pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1 road surface conditions=2	0.001468	1.339237
9	191	day of week=7 pedestrian crossing human control=0	0.0023	1.338163
9	905	pedestrian crossing physical facilities=0 weather conditions=1 road surface conditions=2	0.001468	1.334906
9	189	day of week=7 weather conditions=1	0.001859	1.334793
9	1460	first road class=4 road type=6 pedestrian crossing physical facilities=0 weather conditions=1	0.001166	1.327951
9	1411	speed limit=40 pedestrian crossing human control=0 weather conditions=1 trunk road flag=2	0.001068	1.327183
9	899	road type=6 weather conditions=1 road surface conditions=2	0.001337	1.325842
9	520	Date=October pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001231	1.325113
9	105	Regional areas=West Midlands weather conditions=1	0.001199	1.321779
9	1457	first road class=4 road type=6 pedestrian crossing physical facilities=0 road surface conditions=1	0.001011	1.320235
9	125	Date=October pedestrian crossing physical facilities=0	0.001231	1.318021
9	32	road type=3	0.002846	1.315317
9	109	Date=December pedestrian crossing physical facilities=0	0.001109	1.31399
9	331	first road class=3 weather conditions=1	0.006303	1.313726
9	491	Regional areas=Southwest pedestrian crossing human control=0 weather conditions=1	0.001142	1.310038
9	98	Regional areas=Southwest weather conditions=1	0.001142	1.30488
9	505	speed limit=40 weather conditions=1 trunk road flag=2	0.001068	1.302196
9	497	Regional areas=West Midlands pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001125	1.299108
9	26	day of week=7	0.002316	1.297001
9	104	Regional areas=West Midlands pedestrian crossing physical facilities=0	0.001125	1.295287
9	102	Regional areas=West Midlands Road type=6	0.001027	1.292045
9	909	pedestrian crossing human control=0 weather conditions=1 road surface conditions=2	0.001721	1.29152

9	839	first road class=3 light conditions=4 road surface conditions=1	0.001003	1.291415
9	898	road type=6 pedestrian crossing physical facilities=0 road surface conditions=2	0.00234	1.289281
9	8	Regional areas=Southwest	0.001403	1.289125
9	40	first road class=3	0.007657	1.288125
9	506	speed limit=40 pedestrian crossing human control=0 weather conditions=1	0.001174	1.286763
9	277	weather conditions=1 road surface conditions=2	0.001745	1.286387
10	1414	Date=October pedestrian crossing human control=0 weather conditions=1 road surface conditions=1	0.001117	1.275299
10	1410	speed limit=40 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 trunk road flag=2	0.001011	1.274385
10	582	first road class=4 road type=6 pedestrian crossing physical facilities=0	0.001378	1.272619
10	503	speed limit=40 pedestrian crossing physical facilities=0 trunk road flag=2	0.001011	1.270826
10	507	speed limit=40 pedestrian crossing human control=0 trunk road flag=2	0.001239	1.265492
10	112	speed limit=40 weather conditions=1	0.001174	1.265045
10	521	Date=October pedestrian crossing human control=0 weather conditions=1	0.001215	1.255181
10	907	pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=2	0.00309	1.251401
10	276	pedestrian crossing physical facilities=0 road surface conditions=2	0.003099	1.250035
10	113	speed limit=40 trunk road flag=2	0.001248	1.244453
10	110	Date=December pedestrian crossing human control=0	0.001329	1.24429
10	518	Date=October pedestrian crossing human control=0 road surface conditions=1	0.00115	1.241663
10	9	Regional areas=West Midlands	0.001362	1.237121
10	517	Date=October weather conditions=1 road surface conditions=1	0.001117	1.232768
10	114	speed limit=40 pedestrian crossing human control=0	0.00137	1.230863
10	128	Date=October pedestrian crossing human control=0	0.001427	1.228317
10	2704	first road class=4 road type=6 pedestrian crossing human control=0 weather conditions=1 road surface conditions=1	0.001109	1.227834
10	590	first road class=4 pedestrian crossing physical facilities=0 weather conditions=1	0.001215	1.226325

10	111	speed limit=40 pedestrian crossing physical facilities=0	0.001117	1.225257
10	480	Date=August pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001027	1.222685
10	1464	first road class=4 road type=6 pedestrian crossing human control=0 weather conditions=1	0.001313	1.22241
10	586	first road class=4 pedestrian crossing physical facilities=0 road surface conditions=1	0.00106	1.220041
10	91	Date=August pedestrian crossing physical facilities=0	0.001027	1.219299
10	10	Date=December	0.001354	1.214875
10	126	Date=October weather conditions=1	0.001215	1.212725
10	12	speed limit=40	0.001378	1.212527
10	901	road type=6 pedestrian crossing human control=0 road surface conditions=2	0.002675	1.211718
10	1458	first road class=4 road type=6 weather conditions=1 road surface conditions=1	0.001117	1.211672
10	1459	first road class=4 road type=6 pedestrian crossing human control=0 road surface conditions=1	0.00115	1.211286
10	274	road type=6 road surface conditions=2	0.002724	1.207385
10	583	first road class=4 road type=6 weather conditions=1	0.001321	1.206047
10	2946	day of week=2 road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.001125	1.204835
10	1657	day of week=2 road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001329	1.19986
10	1655	day of week=2 road type=6 pedestrian crossing physical facilities=0 weather conditions=1	0.001125	1.198136
10	124	Date=October Road surface conditions=1	0.00115	1.197271
10	581	first road class=4 road type=6 road surface conditions=1	0.001158	1.193746
10	689	day of week=2 road type=6 pedestrian crossing physical facilities=0	0.001329	1.193515
10	279	pedestrian crossing human control=0 road surface conditions=2	0.00362	1.186558
10	167	first road class=4 pedestrian crossing physical facilities=0	0.001443	1.179093
10	15	Date=October	0.001427	1.178607
10	704	day of week=2 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001778	1.175883
10	35	road surface conditions=2	0.003686	1.1757

10	198	day of week=2 pedestrian crossing physical facilities=0	0.001778	1.169864
10	585	first road class=4 road type=6 pedestrian crossing human control=0	0.001549	1.167166
10	1418	Date=July pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.001036	1.163905
10	481	Date=August pedestrian crossing human control=0 weather conditions=1	0.001044	1.161036
10	534	Regional areas=Northwest Road type=6 pedestrian crossing human control=0	0.001068	1.159584
10	165	first road class=4 road type=6	0.001574	1.159393
10	526	Date=July pedestrian crossing physical facilities=0 weather conditions=1	0.001036	1.158696
10	122	Regional areas=Yorkshire and the Humber pedestrian crossing human control=0	0.001345	1.157021
10	3026	road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 light conditions=4 weather conditions=1	0.001125	1.15325
10	2573	road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.007331	1.153155
10	14	Regional areas=Yorkshire and the Humber	0.001345	1.150426
10	139	Regional areas=Northwest Road type=6	0.001076	1.149852
10	1239	road type=6 pedestrian crossing physical facilities=0 weather conditions=1	0.007339	1.149456
10	1241	road type=6 pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.008872	1.144339
10	360	road type=6 pedestrian crossing physical facilities=0	0.008888	1.141241
10	479	Date=August pedestrian crossing human control=0 road surface conditions=1	0.001003	1.136901
10	4114	road type=6 pedestrian crossing human control=0 light conditions=4 weather conditions=1 road surface conditions=1 trunk road flag=2	0.001093	1.13442
10	1831	road type=6 pedestrian crossing human control=0 light conditions=4 weather conditions=1	0.001509	1.132659
10	92	Date=August weather conditions=1	0.001044	1.130031
10	509	Date=September pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001011	1.12782
10	144	Regional areas=Northwest pedestrian crossing human control=0	0.001403	1.127768
10	1416	Date=July pedestrian crossing human control=0 pedestrian crossing physical facilities=0 road surface conditions=1	0.001003	1.127388

10	527	Date=July pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.001076	1.12724
10	1840	pedestrian crossing human control=0 pedestrian crossing physical facilities=0 light conditions=4 weather conditions=1	0.001549	1.12353
10	116	Date=September pedestrian crossing physical facilities=0	0.001011	1.12295
10	523	Date=July pedestrian crossing physical facilities=0 road surface conditions=1	0.001003	1.122202
10	130	Date=July pedestrian crossing physical facilities=0	0.001076	1.121758
10	858	road type=6 light conditions=4 weather conditions=1	0.001541	1.121049
10	594	first road class=4 pedestrian crossing human control=0 weather conditions=1	0.001386	1.119757
10	865	pedestrian crossing physical facilities=0 light conditions=4 weather conditions=1	0.001549	1.119471
10	1262	pedestrian crossing human control=0 pedestrian crossing physical facilities=0 weather conditions=1	0.009557	1.11845
10	3024	road type=6 pedestrian crossing human control=0 light conditions=4 road surface conditions=1 trunk road flag=2	0.001133	1.117576
10	20	Regional areas=Northwest	0.001411	1.11706
10	373	pedestrian crossing physical facilities=0 weather conditions=1	0.009565	1.114218
10	93	Date=August pedestrian crossing human control=0	0.001158	1.113636
10	860	road type=6 pedestrian crossing human control=0 light conditions=4	0.00203	1.112514
10	375	pedestrian crossing human control=0 pedestrian crossing physical facilities=0	0.011636	1.109275
10	47	pedestrian crossing physical facilities=0	0.011652	1.105374
10	90	Date=August Road surface conditions=1	0.001003	1.105301
10	589	first road class=4 pedestrian crossing human control=0 road surface conditions=1	0.001215	1.104448
10	1659	day of week=2 road type=6 pedestrian crossing human control=0 weather conditions=1	0.001272	1.10117
10	168	first road class=4 weather conditions=1	0.001394	1.09931
10	869	pedestrian crossing human control=0 light conditions=4 weather conditions=1	0.002104	1.096454
10	692	day of week=2 road type=6 pedestrian crossing human control=0	0.0015	1.096087
10	539	Regional areas=East of England Road type=6 pedestrian crossing human control=0	0.001003	1.094181
10	263	road type=6 light conditions=4	0.002063	1.091833

10	145	Regional areas=East of England Road type=6	0.001011	1.090911
10	166	first road class=4 road surface conditions=1	0.001223	1.082451
10	706	day of week=2 pedestrian crossing human control=0 weather conditions=1	0.001721	1.08002
10	266	light conditions=4 weather conditions=1	0.002145	1.07734
10	690	day of week=2 road type=6 weather conditions=1	0.00128	1.077149
10	7	Date=August	0.001158	1.076703
10	1243	road type=6 pedestrian crossing human control=0 weather conditions=1	0.008391	1.074385
10	170	first road class=4 pedestrian crossing human control=0	0.001647	1.073795
10	201	day of week=2 pedestrian crossing human control=0	0.00203	1.071987
10	195	day of week=2 road type=6	0.001509	1.068481
10	363	road type=6 pedestrian crossing human control=0	0.010136	1.06375
10	864	pedestrian crossing human control=0 light conditions=4 road surface conditions=1	0.001672	1.060178
10	268	pedestrian crossing human control=0 light conditions=4	0.002789	1.05954
10	24	first road class=4	0.001672	1.058761
10	508	Date=September pedestrian crossing human control=0 road surface conditions=1	0.001011	1.057815
10	531	Date=May pedestrian crossing human control=0 weather conditions=1	0.001142	1.056952
10	361	road type=6 weather conditions=1	0.008448	1.053633
10	138	Date=November pedestrian crossing human control=0	0.001248	1.051376
10	199	day of week=2 weather conditions=1	0.001729	1.050059
10	44	road type=6	0.010217	1.041155
10	136	Date=May pedestrian crossing human control=0	0.001231	1.040947
10	135	Date=May weather conditions=1	0.001158	1.040892
10	377	pedestrian crossing human control=0 weather conditions=1	0.011098	1.039322
10	27	day of week=2	0.002047	1.038384
10	117	Date=September pedestrian crossing human control=0	0.001166	1.038266

10	34	light conditions=4	0.002838	1.032077
10	50	pedestrian crossing human control=0	0.013512	1.031237
10	115	Date=September Road surface conditions=1	0.001011	1.02355
10	21	Regional areas=East of England	0.001329	1.017055
10	18	Date=May	0.001248	1.016172
10	48	weather conditions=1	0.011171	1.013339
10	19	Date=November	0.001248	1.005435

Appendix B: **SRIP from iRAP**

No.	Countermeasure	Length / Sites	FSIs saved	PV of safety benefit	Estimated Cost	Cost per FSI saved
1	Additional lane (2 + 1 road with barrier)	15.10 km	80	25660031	20655000	259001
2	Bicycle Lane (off-road)	3.50 km	3	1122681	392156	112392
3	Central hatching	5.70 km	0.6	184481	34173	59602
4	Central median barrier (1+1)	34.80 km	41	13286137	6288200	152286
5	Central median barrier (no duplication)	0.70 km	1	365701	95182	83746
6	Centreline rumble strip / flexi-post	1.80 km	0.3	108423	19512	57905
7	Clear roadside hazards (bike lane)	1.20 km	0.5	160633	216000	432664
8	Clear roadside hazards - driver side	0.10 km	0.1	34232	20000	187986
9	Delineation and signing (intersection)	8 sites	0.5	149378	88582	190806
10	Duplication with median barrier	1.20 km	26	8323159	6480000	250507
11	Footpath provision driver side (>3m from road)	25.40 km	28	9060036	2922040	103774
12	Footpath provision driver side (adjacent to road)	27.90 km	35	11319681	4388280	124736
13	Footpath provision driver side (informal path >1m)	4.70 km	1	307139	114747	114747
14	Footpath provision passenger side (>3m from road)	26.80 km	29	9267974	3085368	107116
15	Footpath provision passenger side (adjacent to road)	44.10 km	52	16843531	6939120	132557
16	Footpath provision passenger side (informal path >1m)	5.80 km	1	406768	141451	111890
17	Improve curve delineation	0.40 km	0.5	148419	7460	16172
18	Improve Delineation	45.70 km	13	4190892	847446	65064
19	Lane widening (>0.5m)	0.10 km	0.6	186272	152529	263475
20	Lane widening (up to 0.5m)	1.00 km	2	684333	656756	308795
21	Overtaking lane	0.30 km	1	339394	405000	383958

(Continue)

22	Parking improvements	1.50 km	0.3	99666	18900	61016
23	Pedestrian fencing	27.20 km	9	2976955	179606	19413
24	Protected turn lane (unsignalised 3 leg)	54 sites	74	23918473	7210571	96999
25	Protected turn lane (unsignalised 4 leg)	3 sites	14	4375323	535399	39373
26	Protected turn provision at existing signalised site (4-leg)	1 site	0.6	204497	237955	374405
27	Refuge Island	14 sites	14	4606949	416422	29084
28	Road surface rehabilitation	0.80 km	0.4	115173	70435	196775
29	Roadside barriers - driver side	206.20 km	199	64041774	27898500	140169
30	Roadside barriers - passenger side	159.20 km	111	35791424	21558000	193804
31	School zone warning - signs and markings	0.10 km	0	2072	1865	289641
32	Shoulder rumble strips	199.30 km	84	27156276	2382592	28230
33	Shoulder sealing driver side (>1m)	103.50 km	40	12712184	9156140	231753
34	Shoulder sealing driver side (<1m)	1.50 km	0.2	68926	66640	311091
35	Shoulder sealing passenger side (>1m)	71.50 km	29	9428574	6334720	216179
36	Shoulder sealing passenger side (<1m)	6.60 km	0.9	292349	294490	324117
37	Side road signalised pedestrian crossing	1 site	0.1	28480	45000	508406
38	Sight distance (obstruction removal)	1.40 km	2	678298	35280	16736
39	Signalised crossing	1 site	10	3233649	45000	4478
40	Street lighting (intersection)	14 sites	21	6599575	504000	24572
41	Street lighting (mid-block)	10.30 km	4	1352702	1483200	352802
42	Street lighting (ped crossing)	2 sites	0.1	26839	36000	431589
43	Traffic calming	2.90 km	1	370148	87581	76132
44	Unsignalised crossing	5 sites	4	1230100	217465	56883
45	Upgrade pedestrian facility quality	43 sites	3	826756	767405	298663

(Continue)

46	Wide centreline	12.50 km	0.6	181898	75710	133924
			940	3.02E+08	1.34E+08	142130

Appendix C: **Optimisation result of countermeasure combinations**

Alternative	ith countermeasure																																												
combinations	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
1	0	✓	✓	✓	✓	0	0	✓	✓	✓	✓	✓	0	✓	✓	✓	0	✓	✓	✓	0	✓	✓	✓	✓	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	0	0	✓	✓	✓	0	0	✓	✓	✓	✓
2	0	✓	✓	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	✓	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
3	0	✓	✓	✓	✓	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
4	0	0	✓	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	✓	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
5	0	✓	✓	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
6	✓	✓	✓	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	0	0	0	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	✓
7	0	✓	✓	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	0	0	0	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
8	✓	✓	✓	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	0	0	0	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
9	0	✓	✓	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	✓	✓	0	0	✓	0	✓	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
10	0	0	✓	0	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	✓	0	0	✓	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
11	0	✓	✓	0	✓	0	0	0	0	0	✓	✓	0	✓	✓	✓	0	✓	0	0	0	0	✓	✓	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
12	0	✓	✓	✓	✓	0	0	✓	✓	✓	✓	✓	0	✓	✓	✓	0	✓	✓	✓	0	✓	✓	✓	0	0	✓	✓	✓	✓	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
13	0	0	✓	0	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
14	0	0	✓	0	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
15	0	✓	✓	✓	✓	0	0	0	0	0	✓	✓	0	✓	✓	✓	0	✓	0	0	0	0	✓	✓	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
16	0	0	✓	0	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	0	0	0	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
17	0	✓	✓	✓	✓	0	0	0	0	0	0	✓	0	✓	✓	✓	0	✓	0	0	0	0	✓	✓	0	0	✓	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
18	0	✓	✓	0	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	✓	✓	✓	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
19	0	✓	✓	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	✓	0	0	0	0	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
20	0	✓	✓	✓	✓	0	0	0	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0	✓	✓	✓	✓	0	✓	✓	✓	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
21	0	✓	✓	0	✓	0	0	0	0	0	✓	✓	0	0	0	0	0	0	0	0	0	0	✓	✓	0	0	✓	0	✓	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	✓
22	0	0	✓	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	✓	0	0	✓	0	✓	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0

Alternative	ith countermeasure																																												
combinations	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
23	0	✓	✓	✓	✓	0	✓	0	0	0	✓	✓	0	✓	✓	✓	0	0	0	0	0	0	✓	✓	✓	0	0	0	✓	✓	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
24	0	✓	✓	✓	✓	0	0	0	0	0	0	✓	0	✓	✓	✓	0	✓	0	0	0	0	✓	✓	0	0	0	0	✓	✓	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
25	0	✓	✓	✓	✓	0	0	0	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0	✓	✓	✓	✓	0	✓	✓	✓	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	✓	✓	✓
26	0	0	✓	0	✓	0	0	0	0	0	✓	✓	0	✓	✓	✓	0	✓	0	0	0	0	✓	✓	✓	0	✓	0	✓	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
27	0	✓	✓	0	✓	0	0	0	0	0	✓	✓	0	✓	0	0	0	0	0	0	0	0	✓	✓	0	0	✓	✓	✓	✓	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
28	0	✓	✓	✓	✓	0	0	0	0	0	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	0	✓	✓	✓	✓	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	0	0	✓	✓	✓	0	0	0	✓	✓	0
29	0	✓	✓	✓	✓	0	0	0	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0	✓	✓	✓	✓	0	✓	✓	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	✓	✓	✓
30	✓	✓	✓	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
31	0	0	✓	✓	✓	0	0	0	0	0	✓	✓	0	✓	✓	✓	0	✓	0	0	0	0	✓	✓	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
32	0	0	✓	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
33	0	✓	✓	0	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	✓	0	0	0	0	0	0	0	0	0	✓	✓	✓	0	0	0	0	✓	✓
34	0	0	✓	0	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	✓	0	0	✓	0	✓	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
35	0	0	✓	0	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	✓	0	0	✓	0	✓	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	✓
36	0	✓	✓	✓	✓	0	0	0	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0	✓	✓	✓	✓	0	✓	✓	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
37	0	✓	✓	0	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
38	0	✓	✓	✓	✓	0	0	0	0	0	✓	✓	0	✓	✓	✓	0	✓	0	0	0	0	✓	✓	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
39	0	0	✓	✓	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	✓	0	0	✓	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
40	0	✓	✓	0	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
41	0	✓	✓	✓	✓	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	0	0	✓	✓	✓	0	✓	0	✓	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	✓	✓	✓
42	0	0	✓	0	✓	0	0	0	0	0	0	0	0	✓	✓	✓	0	✓	0	0	0	0	✓	✓	0	0	0	0	0	0	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
43	0	✓	✓	✓	✓	0	✓	0	0	0	✓	0	0	✓	✓	✓	0	0	0	0	0	0	✓	✓	✓	0	0	0	✓	✓	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
44	0	0	✓	0	✓	0	0	0	0	0	0	0	0	0	✓	0	0	0	0	0	0	0	✓	✓	✓	0	0	0	✓	✓	0	✓	0	0	0	0	0	✓	✓	✓	0	0	0	0	0	0
45	0	✓	✓	0	✓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	✓	✓	0	0	✓	0	0	0	0	✓	0	0	0	0	0	0	✓	✓	0	0	0	0	0	0

Appendix D: **Rule mining result of Utrecht case study**

Rule number	LHS	RHS	support	lift
1	Inside outside built-up areas=Urban area	Accident severity=KSI	0.005359	3.513389
2	Inside outside built-up areas=Rural area	Accident severity=KSI	0.005716	2.839102
4	Speed limit road=60 km/h	Accident severity=KSI	0.011433	1.873808
10	Situation road=Bend	Accident severity=KSI	0.016434	1.395647
13	Situation road=Intersection - 3 arms	Accident severity=KSI	0.01965	1.314533
5	Road surface=Concrete	Accident severity=KSI	0.00786	1.238695
16	Road surface=Porous asphalt	Accident severity=KSI	0.034298	1.230071
11	Speed limit road=100 km/h	Accident severity=KSI	0.016792	1.228641
8	Speed limit road=120 km/h	Accident severity=KSI	0.010361	1.195872
17	Situation road=Intersection - 4 arms	Accident severity=KSI	0.039657	1.160673
20	Location road section or crossroads=Intersection	Accident severity=KSI	0.06538	1.130363
25	Street lighting=Not burning	Accident severity=KSI	0.123258	1.108781
23	Road surface=Asphalt (other)	Accident severity=KSI	0.109325	1.090583
21	Speed limit road=50 km/h	Accident severity=KSI	0.082172	1.087005
15	Speed limit road=80 km/h	Accident severity=KSI	0.020364	1.076684
19	Street lighting=Burning	Accident severity=KSI	0.038585	1.03335
7	Light condition=Twilight	Accident severity=KSI	0.00786	1.014364
18	Light condition=Darkness	Accident severity=KSI	0.038942	1.000416

Appendix E: **Utrecht user-defined safer road investment plan**

No.	Countermeasure	Length / Number	FSIs saved	Estimate cost	PV of safety benefits	NPV of safety benefits	BCR	Service life
1	Additional lane (2 + 1 road with barrier)	15.10 km	80	20655000	25660031	5005031	1.2	medium
3	Central hatching	5.70 km	0.6	34173	184481	150308	5.4	medium
4	Central median barrier (1+1)	34.80 km	41	6288200	13286137	6997937	2.1	medium
5	Central median barrier (no duplication)	0.70 km	1	95182	365701	270519	3.8	medium
6	Centreline rumble strip / flexi-post	1.80 km	0.3	19512	108423	88911	5.6	low
8	Clear roadside hazards - driver side	0.10 km	0.1	20000	34232	14232	1.7	low
9	Delineation and signing (intersection)	8 sites	0.5	88582	149378	60796	1.7	low
10	Duplication with median barrier	1.20 km	26	6480000	8323159	1843159	1.3	high
17	Improve curve delineation	0.40 km	0.5	7460	148419	140959	19.9	low
18	Improve Delineation	45.70 km	13	847446	4190892	3343446	5	low
19	Lane widening (>0.5m)	0.10 km	0.6	152529	186272	33743	1.2	low
20	Lane widening (up to 0.5m)	1.00 km	2	656756	684333	27577	1	low
21	Overtaking lane	0.30 km	1	405000	339394	-65606	0.8	medium
22	Parking improvements	1.50 km	0.3	18900	99666	80766	5.3	low
24	Protected turn lane (unsignalised 3 leg)	54	74	7210571	23918473	16707902	3.3	low
25	Protected turn lane (unsignalised 4 leg)	3 sites	14	535399	4375323	3839924	8.2	low
26	Protected turn provision at existing signalised site (4-leg)	1 site	0.6	237955	204497	-33458	0.9	low
27	Refuge Island	14 sites	14	416422	4606949	4190527	11	low
28	Road surface rehabilitation	0.80 km	0.4	70435	115173	44738	1.6	medium

(Continue)

29	Roadside barriers - driver side	206.20 km	199	27898500	64041774	36143274	2.3	medium
30	Roadside barriers - passenger side	159.20 km	111	21558000	35791424	14233424	1.7	medium
32	Shoulder rumble strips	199.30 km	84	2382592	27156276	24773684	11.4	low
33	Shoulder sealing driver side (>1m)	103.50 km	40	9156140	12712184	3556044	1.4	low
34	Shoulder sealing driver side (<1m)	1.50 km	0.2	66640	68926	2286	1	low
35	Shoulder sealing passenger side (>1m)	71.50 km	29	6334720	9428574	3093854	1.5	low
36	Shoulder sealing passenger side (<1m)	6.60 km	0.9	294490	292349	-2141	1	low
38	Sight distance (obstruction removal)	1.40 km	2	35280	678298	643018	19	medium
39	Signalised crossing	1 sites	10	45000	3233649	3188649	71.9	medium
40	Street lighting (intersection)	14 sites	21	504000	6599575	6095575	13	medium
41	Street lighting (mid-block)	10.30 km	4	1483200	1352702	-130498	0.9	medium
43	Traffic calming	2.90 km	1	87581	370148	282567	4.2	low
44	Unsignalised crossing	5 sites	4	217465	1230100	1012635	5.7	low
46	Wide centreline	12.50 km	0.6	75710	181898	106188	2.4	medium
			776.6	114378840	250118810	135739970	6.62	

Appendix F**: Publications**

Taylor & Francis
Taylor & Francis Group

International Journal of Injury Control and Safety Promotion

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/nics20

Methods of strategic road safety management: a systematic review

Bosong Jiao & Harry Evdorides

To cite this article: Bosong Jiao & Harry Evdorides (01 Apr 2024): Methods of strategic road safety management: a systematic review, International Journal of Injury Control and Safety Promotion, DOI: 10.1080/17457300.2024.2335503

To link to this article: https://doi.org/10.1080/17457300.2024.2335503

INTERNATIONAL JOURNAL OF INJURY CONTROL AND SAFETY PROMOTION
https://doi.org/10.1080/17457300.2024.2335503

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

Methods of strategic road safety management: a systematic review

Bosong Jiao and Harry Evdorides

Department of Civil Engineering, University of Birmingham, Birmingham, United Kingdom

ABSTRACT

A well-developed road network plays a crucial role in fostering social and economic progress within a region. However, road crashes resulting in massive injuries and deaths profoundly affect socioeconomic development. There is a need therefore to identify working approaches used in road safety strategic management which provide evidence and a foundation to achieve safer road transport. This may be achieved through a systematic literature review considering both state-of-the-art technologies and best practice. Such a review is presented in this paper. The review involved searching twenty-six bibliographic databases and twenty-four websites of road-related organizations. Following the EPPI-Reviewer methodology, the researchers identified 30 studies that demonstrated various methods employed in the strategy development process. The review highlighted the prevalence of information technology in crash data analysis, particularly concerning big data applications. Moreover, existing resource allocation methods primarily focus on local countermeasures prioritization and ranking based on benefit cost analysis. However, the review identified a gap in comprehensive crash database understanding, and only a few single-objective optimization methods have been developed for strategy development, while there is a need for data mining methods and multi-objective optimisation methods supported by expert knowledge.

ARTICLE HISTORY

Received 12 September 2023
Revised 27 February 2024
Accepted 23 March 2024

KEYWORDS

Road safety; strategic management; crash data analysis; resource allocation

1. Background

A road network plays an essential role in daily life, and its fundamental objective is to provide safe and accessible services to travellers. However, the number of deaths on roads reached approximately 1.3 million each year, which is unacceptably high. Notably, vehicle crashes have become a leading killer of children and traffic injuries have been suggested as the 8th leading cause of people death (World Health Organization, 2018). The occurrence of roadway crashes not only poses a tragic threat to the individuals directly involved but also emerges as a significant public health concern, given the ubiquitous use of this essential infrastructure daily. This pervasive concern extends beyond the immediate participants, affecting a broader spectrum of society and emphasizing the imperative nature of addressing and mitigating the impacts of road safety on public health.

In the initial stage of society development, road networks were conceptualized and constructed with the aim of establishing a connected societal framework conducive to the commuting and travel needs of individuals. This endeavour reflected a huge demand for the creation of a comfortable and convenient transport system. With the completion of road network construction in developed countries, the focus has shifted from creating new roads to the crucial realm of road management. Roadway crashes are inevitable in the intricate transport environment, which result in a devastating toll on human lives, heavy burden on the economy and negative social impact. Consequently, the imperative to prioritize road safety emerges as an elemental consideration during both the initial design and subsequent operational phases of the road network. This underscores the essential need for employing diverse and strategic methods to address and enhance the safety of the overall road network.

In 2020, the United Nations General Assembly on Global Road Safety provided its endorsement, setting a strategic target to achieve a reduction of over 50% in road traffic fatalities and injuries by 2030 (World Health Organization, 2022). This significant initiative has motivated researchers to study road safety strategic management. It involves the development and implementation of policies, programs, and plans to improve safety on the roads and decrease the occurrence of crashes and injuries. This can be engineering interventions, such as enhancements to road networks and data analysis for the identification and mitigation of high-risk roads. Additionally, measures ranging from educational initiatives to the enforcement of traffic laws are recognized as effective strategies for enhancing overall roadway safety.

Numerous methods and tools are employed to assess road safety. The majority of them aim to detect undesirable incidents, predict potential crashes, and identify hazardous locations at specific junctions or sections of road, which provides valuable and useful information for road administration (Elvik, 2010). However, the concept of management functions in terms of strategic, tactical, project and operation level decision making are not fully applied as is the case for general road network management. There is a need therefore to utilise explicitly the concepts of strategic

CONTACT Bosong Jiao bxj862@student.bham.ac.uk Department of Civil Engineering, University of Birmingham, Birmingham, B15 2TT, United Kingdom

management using appropriate tools and associated data. The practical implication of this shift from road segments to road networks involves a systematic and deliberate effort to integrate strategic considerations into the decision-making processes related to road safety. To learn from existing successful programmes with the view to design enhanced ones, a systematic literature review was carried out. This review sought to assess and synthesize evidence pertaining to high-level road safety management methods. By doing so, it aimed to provide valuable insights and guidance for researchers, practitioners, and policymakers involved in the enhancement of road safety at the strategic level.

2. Method

The systematic review is a structured approach to answer a well-defined primary question with the aim of minimizing bias and producing more reliable outcomes (Gough et al., 2017). It is a widely used secondary research method conducted by academia on various subjects to identify, scrutinize and synthesize all evidence pertinent to the primary question. The systematic review applied the guidelines outlined by Gough et al. (2017) to validate acquired data and underwent modifications based on the preferred reporting items for systematic reviews and meta-analyses (PRISMA). PRISMA, a reporting guideline crafted for enhanced reporting practices (Page et al., 2021), was employed to refine the review process. This combined approach ensures a comprehensive and standardized methodology in accordance with established guidelines, fostering a more robust and transparent presentation of the research findings.

2.1. Definition

A well-defined primary question is essential to systematic review, which defines what subject will be covered and what kinds of studies will be included. PICO process was utilised to formulate primary question as shown in Table 1. PICO question formulation lies in its role as a structured framework that guides the research process, from question development to evidence synthesis, ultimately contributing to evidence-based practice and informed decision-making (Counsell, 1997).

This review addressed following question:

- What is the evidence supporting the different methods used for decision making in road safety management for strategic purposes?

According to the review question, the clarification of the research domain and its subsequent targeted exploration require the provision of precise definitions for the terminology employed, as elaborated in Table 2. This strategic inclusion of definitions aims to ensure conceptual clarity, facilitating a comprehensive and focused examination of the specified subject matter.

Table 1. Description of PICO.

PICO	Description
Problems	Road safety problems includes but not restrict different types of crash involving different road users and stakeholders, which have a detrimental impact on public and economic health.
Interventions	Different methods including models and tools etc., are used to improve the safety of road network for decision makers which reduce the number of deaths and mitigate the severity of accident.
Control	Strategic road safety management is a high-level control of road network, particularly focusing on the highway, and develops a strategy to improve the high-level safety not addressing problems of a specific section.
Outcome	All kinds of evidence including academic publications and technical reports gathered to support the road safety strategic management are compiled into a paper to provide a better understanding to whom are interested and develop an innovative method based on the gap from main findings.

2.2. Eligibility criteria

A set of inclusion criteria are defined with regard to the evidence being sought. All studies will be screened out for relevance based on the criteria given below. These criteria were categorised according to the primary question as defined above.

- Countries: studies from any countries will be included without geographical restriction.
- Road safety: focusing on the safety on highway/motorway.
- Language: all considered studies are likely to be carried out in English.
- Year: Studies after 2000 are considered.

Inclusion criteria extend to encompass tools, models, equipment, and methods utilized for the optimization of both performance and safety within the road network. Additionally, the review considers research endeavours exploring the correlation between crash data and road infrastructure, thereby facilitating informed decision-making by managerial personnel for the enhancement of road safety.

2.3. Search strategy

Constructing an effective search strategy is a critical component of a systematic review, ensuring the comprehensive identification of relevant studies. The process of search involved the selection of information sources, identification of search term, and application of Boolean operators.

Table 2. Definitions of terminologies.

Terminology	Definition
Methods	Any topics related to existing research results or practice are all considered. The methods are not limited to tools, systems and computer software but also include models and Artificial Intelligence (AI) technology.
Strategic management	Strategic management is a high-level process involving identifying objectives, analysing data in depth, allocating resources, forming strategies and evaluating implementation (Saloner et al., 2005).

The review question and definitions of terminologies led to an unbiased search of evidence. Twenty-six bibliographic databases and twenty-four organisation websites (shown in Table 3) were selected to identify studies that had been carried out on road safety strategic management. A set of search terms were identified in Table 4, to broaden the scope of search results, thereby facilitating the identification of pertinent studies and mitigating potential biases. In addition

Table 3. Information sources.

Database	Search engine	Organisation websites
ACM Digital Library	Google.com	Africa Development Bank
ANTE	Goolescholar.com	Austroads
ASTM Digital Library	Findit.bham.ac.uk	Asian Development Bank
Cambridge Core		ASANRA-Association of National Road Agencies
Compendex (Engineering Village)		Asian Infrastructure Investment Bank
CORE-Connecting Repositories		Chartered Institution of Highways and Transportation
CIS-Construction Information Service		Development Bank of Japan Inc.
Civil Engineering Database (ASCE)		Department of Transport
CSA -Cambridge Scientific Abstracts		European Bank for Reconstruction and Development
DOAJ-Directory Open Access Journals		European Union
EBSCO		Federal Highway Administration (USA)
Elsevier Science Direct		Gtkp-global Transport Knowledge Partnership/ Practice
Engineering Handbooks Online		Global Road Safety Facility (GRSF)
Geobase (Engineering Village)		Institution of Civil Engineers, UK
ICE Virtual Library		Irap-International Road Assessment Programme
IEEE Xplore		IRF-International Road Federation
Inspec (Engineering Village)		REAAA-Road Engineering Association of Asia and Australasia
Knovel		Roads Liaison Group
Oxford University Press Journals		The Organisation for Economic Co-operation and Development
PubMed		Transport Research Laboratory Ltd
Royal Society Journals		Transport Research Board
SAGE Journals Online		United Nations
SciCentral		United Nations Economic Commission for Europe
SpringerLink		World Bank
ProQuest Databases		World Health Organisation
Taylor & Francis online		PIARC-World Road Association
Web of Science		
Wiley online library		

Table 4. Search term.

Problems	Interventions	Control	Outcome
Road safety	Methods	Strategic management	Research
Accident	Models	High-level management	Projects
Crash	Tools and equipment	Highway/motorway	Manuals
Road deaths	Computer technology	Road network	Reports

Boolean operators are fundamental to search strategy, allowing for the broadening or narrowing of search parameters based on the quantity of data needed (Higgins, 2011). EPPI-Reviewer, a web tool suitable for all kinds of reviews, facilitated this study and was used for gathering references, deleting duplications, screening, coding, and data extraction (Thomas et al., 2010) based on inclusion criteria.

2.4. Weight of evidence

A weight of evidence framework is a process of assessing the quality of studies (i.e. evidence) and the extent of their relevance. Even though a study has met all the inclusion criteria, it may not meet the relevance and quality requirements for a review. In general, the assessment is developed based on three criteria: 1) robustness and quality of execution of the study, 2) appropriateness of the research design, and 3) relevance to answer the question. The framework used for quality assessment of the evidence gathered is given in Table 5 (Gough, 2007). A study achieving high ratings in at least two criteria was included in the synthesis of the review.

2.5. Data synthesis

The systematic review conducted data synthesis through a narrative approach, enabling a descriptive analysis of the various methods employed in road safety strategic management. This process involved understanding of existing research and scholarly literature to identify how these methods were practically implemented in strategy development to achieve road safety goals. By systematically collating and analysing relevant evidence, researchers will gain valuable insights into the strengths and limitations of existing methodologies, as well as identifying areas where specialized knowledge is lacking. This process empowers researchers to pave the way for future research initiatives that can lead to more effective road safety strategies and

Table 5. Weight of evidence (WoE).

Robustness and quality of execution	High	The study is transparently and clearly clarifying the purpose. The data analysis and results are accurate and detailed. The overall interpretation is understandable and explicit.
	Medium	The study is acceptably clarifying the purpose. The data analysis and results are satisfactory. The overall interpretation is not easy to understand.
Appropriateness of research design	High	It covers the whole or partial network of roads or motorways, and it aims at reducing deaths and severity mitigation to improve overall road safety
	Medium	It covers other types of roads or road sections, and methods are applied to road safety improvement.
Relevance to answer the question	High	It develops a strategic plan or provides a data-based method to support management staff in decision making.
	Medium	It focuses on the efficiency of tools or methods but does not concern the practice at high level.

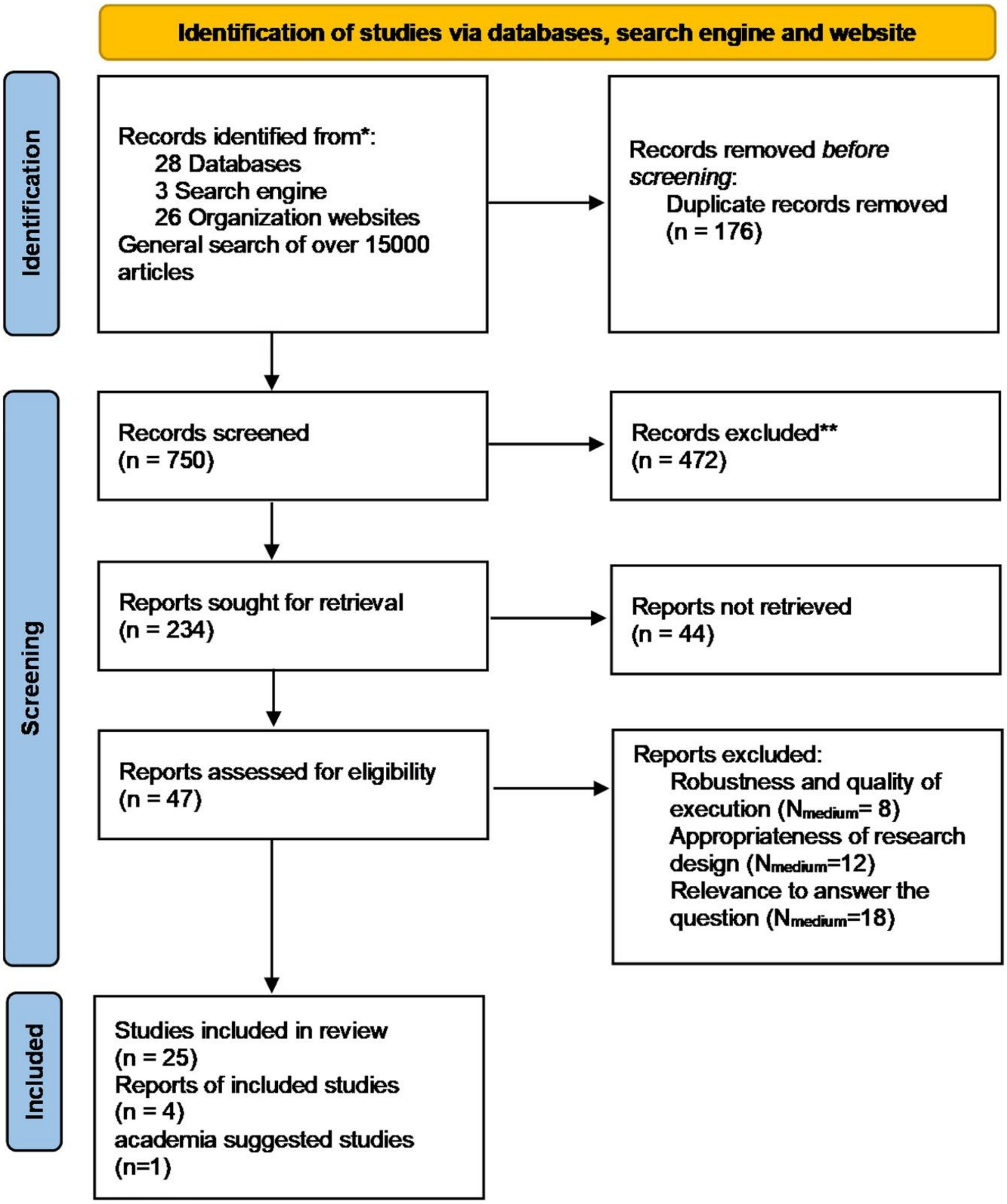

Figure 1. Search process.

interventions. The complete search process is illustrated in Figure 1.

3. Research findings

According to the 30 articles identified, there are different methods developed for strategic road safety purposes, with several of them having practical applications. These methods include data mining, statistical analysis, multi-criteria models, data envelopment analysis (DEA), and optimization techniques. Most of methods were implemented for crash data analysis and resource allocation for strategy development. This section demonstrates how these various methods can be employed to achieve proficient strategic road safety management, ultimately leading to the reduction of road network accidents.

3.1. Data mining

Data mining is a technique to cope with the occurrence of big data in order to uncover various patterns and discover internal relationships among data items in order to extract knowledge. By analysing historical crash data, these methods (summarised in Table 6) assist in identifying high-risk areas, contributing factors, and potential countermeasures.

Jain et al. (2016) investigated the data from all the union territories and states of India, analysed 58 attributes of crash to identify the crash-prone areas through K-means clustering. They discovered the crashes contributors in various states of India using decision trees which provided insightful assistance in outlining their potential relationships. In another study, Castro and Kim (2016) employed three data mining classifications—Bayesian network, decision trees and artificial neural networks—to figure out the underlying elements influencing the severity of a vehicle crash and predict crash severity in the UK. The study focused on the relevant nominal attributes sourced from STATS 19 between 2010 and 2012 and took advantage of the Waikato Environment for Knowledge Analysis (WEKA) to apply the data mining process. Maji et al. (2018) adopted a different approach to develop strategies by comparing various regions to benchmark regions. They used hierarchical clustering analysis based on fatal crash percentage of causation factors to group different regions. Three distinct criteria were investigated as follows: road crash causation factors related average fatal crash, average severity index and average severity index growth rate. By comparing the results with predefined threshold values established by experts, safety improvement strategies for regions with poor performance could be derived, drawing insights from regions with outstanding performance.

Furthermore, association rule mining (ARM), a data mining technique to extract knowledge in the form of rules, was applied to Indian crash database (Saranyadevi et al., 2019). To enhance the rule extraction process, they integrated ARM with fuzzy theory and context-free grammar, formulating a recommendation system that aids road administrators in comprehending rules derived from vast amounts of data attributes. In a case study conducted in France, El Mazouri et al. (2019) utilized ARM with multi-criteria aggregation ELECTRE II method to rank different rules from the best to the worst. Crash data collected from the Accident Analysis Bulletin Corporal (BAAC) in 2016 were used in this study. This work examined and addressed national level issues and found a solution to screen out redundant and unnecessary rules. The ARM with K-modes was developed in the study of Dia et al. (2021) to achieve a strategic level of road safety. They classified crash cases based on recorded variables and extract implicit characteristics' rules for each group so that preventative policies and measures could be developed. Kho et al. (2021) investigated three years of collision data from eight states in the US (vehicle collisions only) between 2007 and 2009. They created and evaluated the prediction of risk map from three different methods: support vector machine (SVM), artificial neural network (ANN) and random forest (RF). The results showed that RF had a high degree of similarity to the actual risk map. The ARM method was then applied to extract significant and accurate decision rules based on the dataset predicted by the RF model. These studies collectively demonstrate the feasibility of utilizing association rule mining by road agencies to define effective road safety strategies and policies.

Due to the large amount of data being investigated into different aspects of road safety management, which makes data analysis complicated and nearly impossible. El Abdallaoui et al. (2019) were inspired to propose a system for big data gathering and analysis. The scope of databases extended beyond police reports to encompass information sourced from crowdsourcing data concerning the individuals involved in accidents. To streamline the data, a process of data cleaning and pre-processing was conducted using principal component analysis (PCA). Subsequently, K-means clustering was utilized to classify accidents according to records, while association rule mining was employed to identify the most valuable and common attributes within each cluster. The overall process could provide preventative strategies and statistic and prediction reports for road safety administration.

3.2. Statistical methods

Statistical analysis is an alternative approach to interpret and describe data to explore hidden patterns and trends, which are crucial in understanding the relationships between various factors influencing road safety. They aid in identifying risk factors and guiding decision-making processes where three studies were screening out and summarised in Table 7. Ng et al. (2002) established an accident risk assessment algorithm to estimate crash events and determined the factors that have the greatest impact on crashes for high-risk zone identification. They used the SPSS (statistical software) to conduct clustering analysis on crash data. In order to explain and estimate crashes occurrence resulting from the probable contributing factors, negative binomial (NB) regression models were developed, and the coefficients of the NB models were obtained using the maximum likelihood estimation method. Based on the observed crash data and contributory casual factors, the risks of zones could be assessed through empirical Bayes (EB).

There were two other studies focusing on applying regression models to crash severity analysis, which had policy and practical implications for road administrators in the long run. Karacasu et al. (2014) studied injured and non-injured types of crashes and investigated their contributing factors by using logistic regression. Discriminant analysis was developed to classify the crash types and determine the prominent factors. Asare and Mensah (2020) also presented an ordinal logistic regression to study the factors which affect the severity of crashes.

3.3. Multi-criteria analysis methods

Multi-criteria analysis (MCA) is an approach to evaluate results, assess performance, estimate impacts, and balance

Table 6. Summary of data mining.

Data analysis			
Input	Procedure	Output	Reference
Crash database	K-means; Decision Tree	Accident-prone areas Dominant factors of accidents	Jain et al. (2016)
	Bayesian Network; Decision Tree; ANN	Prediction of accident severity Influential factors of severity	Castro and Kim (2016)
	Hierarchical Clustering	Identification of benchmark States and union territories	Maji et al. (2018)
	Association Rule with fuzzy theory	Rules related to accident attributes Comparison of attributes	Saranyadevi et al. (2019)
	Association Rule; Multi-criteria aggregation (ELECTRE II)	Rules linking conditions and victims of crash Ranking of rules by importance	El Mazouri et al. (2019)
	K-modes; Association Rule	Clusters of various circumstances Rules for policy development	Dia et al. (2021)
	ANN, SVM, RF; Association Rule	Prediction of severity of accident Extraction of rules by importance	Kho et al. (2021)
Crash database, crowdsourcing data	PCA, K-means; Association Rule	Rules for strategy development Information for report	El Abdallaoui et al. (2019)

Table 7. Summary of statistical methods.

Data analysis			
Input	Procedure	Output	Reference
Crash database	Clustering analysis; NB; regression models; EB	Factors on accidents High-risk zones	Ng et al. (2002)
	Logistic regression; Discriminant analysis	Significance of variables Function for classification of accident type	Karacasu et al. (2014)
	Ordinal logistic regression	Significant predictors	Asare and Mensah (2020)

Table 8. Summary of multi-criteria analysis methods.

Resource allocation			
Input	Procedure	Output	Reference
Indicators	Fuzzy-AHP Model	Determination of weights and ranking score	Yu and Liu (2012)
Performance index	HDM-4; Utility-cost ratio; NPV-cost ratio	Ranking of projects	Odoki et al. (2015)

trade-offs so that many alternatives may be identified and compared with the criteria of preference. Two studies were identified and summaried in Table 8.

Yu and Liu (2012) proposed a multi-criteria model for project prioritisation in highway safety enhancement based on a novel analytical hierarchy process with fuzzy logic, which takes into account a number of indicators, concerning technical, economic and social criteria and weighs them properly for decision making. Fuzzy scaling was developed by employing two fuzzy functions to classify indicators into two types: "the-lower-the-better" and "the-higher-the-better" based on the characteristics of each indicator. Comparison matrix was constructed by using standard deviation to determine the relative importance of pair-wise indicators. Through a non-linear optimisation method to obtain the weights of criteria, the final ranking of each candidate project can be synthesized.

Odoki et al. (2015) also considered the economic, environmental, and social impacts of various investment alternatives. The utility/cost ratio technique was employed in this research in order to propose an MCA methodology to enable road agencies to prioritise and optimise road network investments under different budgets. This method expanded the application of high-level resource allocation and programming to the network level.

3.4. Data envelopment analysis (DEA)

DEA is a method for identifying the performance of decision-making units (DMUs). DEA has found application in road safety management to assess the performance of different regions or jurisdictions (summaried in Table 9). By comparing inputs and outputs, it helps identify best practices and areas requiring improvement. With the growing number of indicators being created, a composite road safety index was developed for evaluation. Shen et al. (2012) developed a DEA-based road safety model (DEA-RS) which aims at minimizing output levels given input levels. They defined input as indicators of exposure: the population size, the number of registered vehicles, distance travelled, and output as the number of road fatalities to assess the overall risk situation in terms of road safety. Road safety performance of 27 EU countries were ranked by the efficiency scores and it enabled to identify the benchmark for improving the operations of those underperforming countries.

Fancello et al. (2020) proposed a DEA-based decision support method to identify which roads most urgently needed safety improvements. The average number of conflict points at junctions and traffic flow were taken into account as inputs in this model, which used the social cost of crashes as its only output indicator. Two different DEA models constant returns to scale based and variable returns to scale based were carried out to validate the rankings of road sections that needed improvements immediately.

Dadashi and Mirbaha (2019) presented a ranking approach that combined DEA with Monte Carlo simulation to consider uncertainty to prioritise road safety improvement projects. Instead of running a DEA model only, a Monte Carlo simulation was used to provide stochastic values as the inputs and outputs and two ranking indicators were established based on the efficiency scores of each DMU.

3.5. Optimisation methods

Optimisation methods are used to identify solutions that maximise or minimise a certain objective function, which

Table 9. Summary of data envelopment analysis.

Input	Procedure	Output	Reference
Data analysis			
Indicators of exposure	DEA-RS	Risk ranking of each country	Shen et al. (2012)
AADT, conflict points, social cost	DEA	Score of road sections and urgent improvement need ranking of sections	Fancello et al. (2020)
Resource allocation			
Cost and crash frequency	DEA with Monte Carlo simulation	Ranking of projects	Dadashi and Mirbaha (2019)

Table 10. Summary of optimisation methods.

Input	Procedure	Output	Reference
Resource allocation			
Performance index	Weighted-objective Linear Optimisation	Maintenance priority list	Chassiakos et al. (2005)
Countermeasures, cost and CRF	Linear Programming	Best combination of safety improvements	Saha and Ksaibati (2016)
Countermeasures, cost and benefits	Mixed Integer Linear Optimisation	Prioritization of safety improvements	Byaruhanga and Evdorides (2022)

play a vital role in strategic road safety purposes. They assist in determining the optimal distribution of resources to maximize safety outcomes within budgetary constraints. Three studies were outlined in Table 10.

Regarding resource allocation, Chassiakos et al. (2005) proposed a computerised system to provide road agencies with valuable suggestions based on crashes data characteristics for setting improvement priorities and likely countermeasures. A priority list was developed by index of accident and fatality rates as well as repetitive occurrence of a particular accident type. A weighted-objective linear optimisation was employed to assist decision makers in justifying safety improvement alternatives.

Saha and Ksaibati (2016) developed a traffic safety management system (TSMS) that sought the ideal combination of safety enhancement initiatives that would reduce crashes and benefit society. In order to minimise the overall predicted crash, particularly fatal-and-injured crash, on the segments with high traffic volume, linear optimisation was used to determine the best combination of safety improvement projects based on the cost and crash reduction factor (CRF) of each countermeasure.

Byaruhanga and Evdorides (2022) applied mixed integer linear optimisation to determine the most reasonable expenditure between capital and maintenance works. It enabled to maximise the economic benefits created from the number of all crashes saved including property damage only (PDO) crashes under a constrained budget. They considered the effectiveness of countermeasure alternatives over their life span together with budget constraints to achieve high-level safety plan development.

3.6. Other methods

These studies outlined in Table 11 contribute to the advancement of road safety evaluation methodologies, involving diverse techniques and tools for assessing road safety performance, risk, and resource allocation.

Han et al. (2011) established a high-level index system and developed a road safety assessment method for regional road safety. They employed six representative indicators: fatality per 10 thousand vehicles (U1), traffic fatality (U2), fatality per 100 thousand population (U3), fatality per million passenger transport (U4), accident rate per hundred kilometres (U5), accident rate per million tons freight transport (U6) These indicators were merged into a comprehensive road safety index using a combination of analytic hierarchy process (AHP) and fuzzy logic theory.

Hogg and Mathie (2011) developed a safety risk model based on three types of data analysis - standard and bespoke outputs, bow-tie modelling and what-if analysis for segmentation and factor scaling, which were feasible to produce risk calculations and obtain insights and information for road safety decision making.

A supplementary tool called the beta-binomial test was applied to prioritisation of critical areas. Park and Young (2012) demonstrated the use of this tool to rank various jurisdictions facilitating strategic highway safety plans development.

Coll et al. (2013) put forward a non-linear aggregation method to evaluate road safety with a composite safety performance index, which could be used to identify and rank different policing areas based on road performance.

In terms of resource allocation, Hasan et al. (2021) proposed a computational toolkit including sorting, identification, treatment, ranking algorithms to parse crash database. It enabled to identify crash hotspots, suggest proper remedies, and estimate potential safety benefits. They used safety improvement indices (SII), a benefit-cost ratio, to compare project safety benefits to implementation costs.

4. Existing practical methods

In addition to academic research findings, there were five works created by road-related organisations for the better understanding of the current situation of road safety network management.

Herbel et al. (2010) developed a technical report to describe the Highway Safety Improvement Program (HSIP) and provided a roadway safety management process and strategic approach to improving highway safety under the support of the Federal Highway Administration (FHWA). The authors introduced benefit-cost analysis for countermeasure prioritisation by converting benefits to a monetary value. They advocated the adoption of the net present value (NPV) technique as well as the benefit-cost ratio (BCR) to assess the efficiency, effectiveness, and feasibility of potential countermeasures. According to the report, project prioritisation can be established by using ranking, incremental benefit-cost analysis and optimisation methods. Subsequently, in 2021, the FHWA produced a comprehensive guide titled 'Selecting Projects and Strategies to Maximize Highway

Safety Improvement Program Performance' (Gross et al., 2021) to explain in detail how to conduct BCR method. It represented an in-depth examination of the BCR in relation to fatal and severe-injury accidents exclusively, with the aim of methodically optimizing the safety impact of countermeasures in terms of mitigating loss of life and severe injuries. It introduced a countermeasure score to represents how much it costs to generate 1% of those potential benefits. Moreover, the guide clarified the limitation and advantages of diverse approaches—ranging from site-specific and systemic to systematic—in order to assist road safety personnel in selecting the most fitting approach for devising optimal strategies. Apart from the HSIP, the FHWA established a guidance for Strategic Highway Safety Plans (SHSP) and created an implementation process model (Federal Highway Administration, 2010). It is also a data-driven process, and the integration of all types of data is regarded as the fundamental element serving the identification of emphasis areas, crash types and strategy development. Performance management, strategic goals and objectives and emphasis areas are critical components of the SHSP process. It considered all aspects of road safety, including engineering, management, operation, education, enforcement and emergency service elements, by coordinating different road-related plans from the corresponding agencies, as illustrated in Figure 2. It assists state entities in identifying and prioritising their most urgent road safety needs and developing a programme or strategy that has the best chance of preventing fatalities and serious injuries.

Table 11. Summary of other methods.

Input	Procedure	Output	Reference
Data analysis			
Evaluation index	AHP Fuzzy theory	Weight of each evaluation index Road safety index	Han et al. (2011)
Crash database, Vehicle miles Working hours	Safety Risk Model	Standard report of various dimensions Risk analysis	Hogg and Mathie (2011)
Crash database	Overdispersion test; Beta-binomial test	Ranking of each jurisdiction	Park and Young (2012)
Crash database, indicators	Non-linear aggregation method	Composite road safety index Identification and ranking of policing areas	Coll et al. (2013)
Resource allocation			
Raw crash data, project limits, cost and treatment	Phase 1: treatment algorithm Phase 2: sorting, identification and ranking algorithms	Calculation of safety improvement index. List of projects recommended to improve	Hasan et al. (2021)

International Road Assessment Programme (iRAP) aims at mitigating the risk of road around the world, which is a world standard road safety amassment organisation. The programme introduced two methods, risk maps and star rating maps, which enables road safety agencies to develop investment strategies based on ranking risk level of road and identify high-potential risk area based on crash related data. Jane et al. (2011) verified the feasibility and suitability of the RAP in Manitoba. They obtained the international jurisdictional survey regarding the application of RAP and discussed different aspects of RAP implementation in terms of the effectiveness, strengths, weaknesses and so on. Risk mapping methodology was applied to Manitoba with the process of determination of time period of data, road classification, road segmentation and risk level development, and star rating mapping was used to realise overall risk ranking, which includes four steps: roadway data collection, road protection scores (RPS) calculation, conversion of RPSs and overall star rating. In addition, iRAP star rating and investment plan was updated by iRAP in 2019 and ViDA web software was developed for safety analyses (International Road Assessment Programme, 2019). A list of safety countermeasures that can improve star ratings and lower the level of risk is called a Safer Roads Investment Plan. It is a crash-based model to appraise safety countermeasure according to benefit cost ratio. More than 90 road treatment options are considered to produce feasible and economically sound investments, which is designed to provide a network-level assessment of risk and cost-effective countermeasures.

The New Zealand Transport Agency has introduced a novel One Network Framework, which integrates the movement function, coordinating diverse transportation

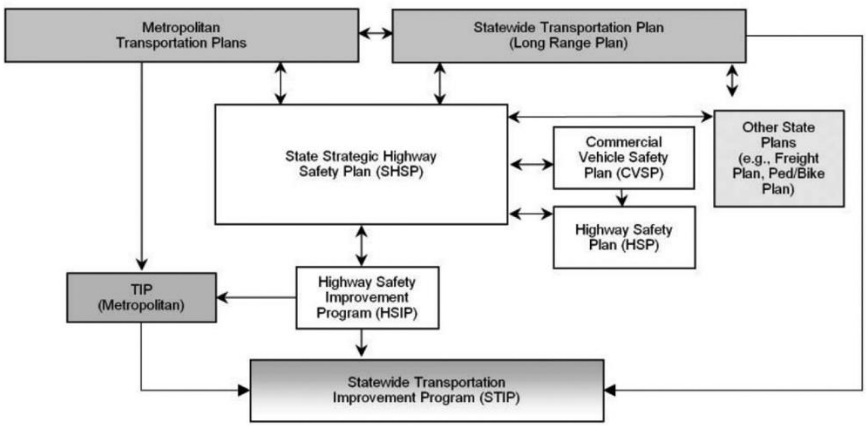

Figure 2. Coordinated transportation safety planning (Federal Highway Administration, 2010).

modes, and the place function, aimed at improving accessibility and travel efficiency in specific geographic areas. This framework serves as a comprehensive approach to facilitate strategic and well-informed decision-making (Waka Kotahi NZ Transport Agency, 2022). Furthermore, in the 2021–24 National Land Transport Programme, the agency utilized three investment criteria—government policy statement (GPS) on land transport alignment, scheduling, and efficiency—to prioritize proposed activities. It is an interesting and decisive transport funding allocation framework aligning with strategic objectives. In terms of its alignment with GPS priorities, it assesses how proposed activities correlate with strategic goals and underscores their potential contributions towards achieving these goals. The national land transport programme (Waka Kotahi NZ Transport Agency, 2021) outlined four strategic priorities—safety, enhanced travel options, improved freight connectivity, and addressing climate change. Additionally, scheduling within this framework denotes the importance and interconnectedness of proposed activities within a program or package. Efficiency indicates anticipated returns on investment, considering comprehensive cost-benefit analyses, with the benefit-cost ratio (BCR) serving as a key metric for evaluation. The prioritisation process involved the application of a matrix to rate the three factors explained above to determine the order of precedence for these activities and all details of rating criteria within three factors are elaborated by Waka Kotahi NZ Transport Agency.

5. Discussion

The main findings of the review describe various methods used for strategy development, categorizable into two groupings. One is data analysis considering different sources of data based on a big data background, and the other is allocation of resources, involving tasks such as project ranking, prioritization, and investment combinations.

5.1. Interpretation and limitations of evidence

Concerning the practical applications of methods in various countries aimed at enhancing strategic road safety, this review identified a few articles that employed data mining, statistical techniques, and risk modelling to analyse crash-related data. These methods aim to uncover hidden patterns of fatalities, injuries, crash occurrence to develop strategies for improving overall safety of road network.

Data mining has been widely applied to predict the severity of crashes, explore the influential factors of severity, and investigate the contributory factors of crash occurrence. However, the prediction model developed by Castro and Kim (2016) only considered road-related attributes to forecast severity, of which evidence was insufficient for strategic purposes, given the recognized relationship between crash severity, drivers, vehicles, and the environment. Jain et al. (2016) managed to classify different levels of crash-prone states and find their dominant crashes contributors. Nevertheless, from a strategic perspective, there exists a gap in comprehending the different safety performance of these states under analogous jurisdictional contexts, where Maji et al. (2018) offered an intriguing approach suggesting that an effective safety strategy and policy could be developed by exploring strategies of those benchmarking countries and validating these well-performance strategies in the context of one's own safety situation.

Among the data mining methodologies, association rule mining (ARM), an unsupervised technique, stands out by creating comprehensive rules based on data characteristics to bolster strategy development. The efficacy of ARM has been validated by Saranyadevi et al. (2019) and El Mazouri et al. (2019), yet a more exhaustive comprehension of the rules, avoiding biases induced by minimal support and confidence thresholds, remains a subject warranting further investigation. It holds significance that infrequent occurrences in crash databases, such as fatal incidents, which are a main concern for road administrators Moreover, obtaining a comprehensive knowledge of a multitude of rules becomes imperative within the road safety context, where integrating a clustering model into the analysis of rules could provide a systematically classified framework, enhancing comprehensive understanding of roadway crashes and enabling a more thorough grasp of their reality.

From a strategic management standpoint, the application of ARM could be extended to extract rules pertinent to fatal crashes, road class and regional areas for diverse strategic purposes. This potential research direction will enable authorities to formulate relevant policies and strategies by thoroughly comprehending the rules related to fatal crashes, road class and regions. This adaptability caters to the varying administrative requisites across different levels of governance.

In terms of resource allocation, various techniques such as multiple criteria analysis (MCA), data envelopment analysis (DEA), and optimization methods have been employed to prioritize safety enhancement projects. Yu and Liu (2012) and Odoki et al. (2015) utilized MCA, whereas Gross et al. (2021) and Herbel et al. (2010) employed benefit-cost ratio (BCR) to convert unaccountable benefits into monetary value, considering different impacts on projects. It is important to note that while BCR proves inadequate for comparing projects across multiple sites and identifying optimal project combinations, the multi-criteria method Odoki et al. (2015) and optimization approach (Saha and Ksaibati, 2016; Byaruhanga and Evdorides, 2022) are capable of accommodating project combinations. Only three articles used linear optimisation method to achieve resource allocation. To establish the accuracy and efficacy of the optimization method, there arises a necessity to formulate alternative optimization methods that facilitate the pursuit of the most optimal solutions and enable result comparison for various objectives.

Concerning the methodologies presently employed within the road safety sector, all of them are now considering data analysis on a road-segment level, and resources allocation is being based on benefit and cost analyses of individual countermeasures. Notably, while the US published a strategic highway safety plan (SHSP) with an implementation process model (Federal Highway Administration, 2010), it relies on the safety performance function (AASHTO, 2010) and highway safety improvement program (Herbel et al., 2010) for plan implementations. An intriguing and innovative approach is exemplified by iRAP, which initiates with segmentation analysis and culminates in a comprehensive risk map and a ranking of road networks. It developed a software named ViDA to provide a valuable countermeasures database for professionals engaged in road safety strategic management, aiding them in identifying the most cost-effective road safety upgrades. While Waka Kotahi NZ Transport Agency (2021) has formulated a comprehensive strategic management framework for the transportation system to align with governmental policies, there remains a need to explore methodologies that can assist decision-makers in the automatic development of user-defined investment plans for local authorities within budgetary constraints, utilizing diverse datasets such as crash data and road inventory data etc. Data-driven methods for road safety strategy development will provide a set of recommendations for road agencies to reduce human efforts and subjectivity. It is essential to develop a practical procedure or programme for high-level road safety management from data, thus there is a gap of using data mining methods to conduct strategic data analysis to enable a systematic and automated resource allocation approach that caters to the diverse objectives of a strategy.

5.2. Limitations of review process

However, there are limitations to every systematic review. Resource constraints may have limited the authors' ability to access a complete set of relevant articles, potentially influencing the overall scope of the synthesis. As mentioned in Figure 1, in this review there were 44 articles to which the authors have no access due to license constraint. In addition, the subjective nature of the synthesis process is a main concern with regards to reliability and objectivity of the synthesized information as the selective inclusion of studies may introduce inadvertent biases. In an effort to mitigate these limitations, the review diligently employed reasonable inclusion criteria, combined different tools, and followed search strategy to structure a transparent process. In terms of methodology, emerging tools like PRISMA and the critical appraisal skills program (CASP) hold promise for future applications in systematic reviews. These tools, aligning with the principles and theories elaborated by Gough et al. (2017), offer a more detailed and transparent approach in the selection and appraisal processes.

6. Conclusion

The objectives of this study were to perform a systematic review of the existing methods and proposed potential methods of improving road safety performance within the realm of strategic management and to obtain a wider and more in-depth understanding of their applications. This review set out to address the question:

- What is the evidence of supporting the different methods used for decision making in road safety management for strategic purposes?

In order to answer the question, 30 articles were identified, which demonstrated different methods applied to strategic road safety management, including:

- data mining
- statistical analysis
- multi-criteria analysis
- data envelope analysis
- optimisation methods

Additionally, the study identified practical guidance from road-related entities, which contributed to a better understanding of methodologies for road safety managers. Regarding the countermeasure selection, existing practical methods are restricted to project evaluation based on benefit cost ratio and cost effectiveness analysis. Nonetheless, it is imperative to acknowledge that strategic management involves a higher-level process in the endeavour to mitigate crash occurrences, in contrast to sectional analysis and crash prediction. Consequently, the development of effective strategies needs a stronger foundation of objective data-driven evidence and an automated decision-making tool to construct a safer road network system.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

AASHTO. (2010). *Highway safety manual*. AASHTO.

El Abdallaoui, H. E. A., El Fazziki, A., Ennaji, F. Z., & Sadgal, M. (2019). *A system for collecting and analyzing road accidents Big Data* [Paper presentation]. 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) (pp. 663–671).

Asare, I. O., & Mensah, A. C. (2020). Crash severity modelling using ordinal logistic regression approach. *International Journal of Injury Control and Safety Promotion*, *27*(4), 412–419. https://doi.org/10.1080/17457300.2020.1790615

Byaruhanga, C. B., & Evdorides, H. (2022). A budget optimisation model for road safety infrastructure countermeasures. *Cogent Engineering*, *9*(1), 2129363. https://doi.org/10.1080/23311916.2022.2129363

Castro, Y., & Kim, Y. J. (2016). Data mining on road safety: Factor assessment on vehicle accidents using classification models. *International Journal of Crashworthiness*, *21*(2), 104–111. https://doi.org/10.1080/13588265.2015.1122278

Chassiakos, A. P., Panagolia, C., & Theodorakopoulos, D. D. (2005). Development of decision-support system for managing highway safety. *Journal of Transportation Engineering*, *131*(5), 364–373. https://doi.org/10.1061/(ASCE)0733-947X(2005)131:5(364)

Coll, B., Moutari, S., & Marshall, A. H. (2013). Hotspots identification and ranking for road safety improvement: An alternative approach. *Accident; Analysis and Prevention*, *59*, 604–617. https://doi.org/10.1016/j.aap.2013.07.012

Counsell, C. (1997). Formulating Questions and Locating Primary Studies for Inclusion in Systematic Reviews. *Annals of Internal Medicine*, *127*(5), 380–387. https://doi.org/10.7326/0003-4819-127-5-199709010-00008

Dadashi, A., & Mirbaha, B. (2019). Prioritizing highway safety improvement projects: A Monte-Carlo based Data Envelopment Analysis approach. *Accident; Analysis and Prevention*, *123*, 387–395. https://doi.org/10.1016/j.aap.2018.11.003

Dia, Y., Seck, T. C., Diallo, S. M., Sall, O., Bousso, M., & Sarr E. N., & Mboup, L. M. (2021). *Senegalese Road Crash Data Analysis Based On Clustering And Association Rule Mining* [Paper presentation]. 2021 4th International Conference on Data Storage and Data Engineering (pp. 70–76). https://doi.org/10.1145/3456146.3456158

Elvik, R. (2010). *Assessment and applicability of road safety management evaluation tools: Current practice and state-of-the-art in Europe*. Institute of Transport Economics.

Fancello, G., Carta, M., & Serra, P. (2020). Data envelopment analysis for the assessment of road safety in urban road networks: A comparative study using CCR and BCC models. *Case Studies on Transport Policy*, *8*(3), 736–744. https://doi.org/10.1016/j.cstp.2020.07.007

Federal Highway Administration. (2010). *The Essential Eight Fundamental Elements and Effective Steps for SHSP Implementation.*

Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22*(2), 213–228. https://doi.org/10.1080/02671520701296189

Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews*. Sage.

Gross, F., Harmon, T., Cynecki, M. J., Dittberner, R., & Chestnutt, C. (2021). *Selecting Projects and Strategies to Maximize Highway Safety Improvement Program Performance*.

Han, Y., Guan, H. Z., Zhao, K., & Wu, J.H. (2011). *Synthetic Evaluation Method of Regional Road Safety Level* [Paper presentation]. ICCTP 2011: Towards Sustainable Transportation Systems (pp. 1904–1913). https://doi.org/10.1061/41186(421)189

Hasan, M., Han, Z., & Machemehl, R. (2021). *Development of Crash Database Parsing Tools to Support the Highway Safety Improvement Program in Texas* [Paper presentation]. International Conference on Transportation and Development 2021 (pp. 297–308). https://doi.org/10.1061/9780784483534.026

Herbel, S., Laing, L., & McGovern, C. (2010). *Highway Safety Improvement Program (HSIP) Manual.*

Higgins, J. P. (2011). Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. www.cochrane-handbook.org.

Hogg, V., & Mathie, E. (2011). *Development of a safety risk model for the highways agency road network* [Paper presentation]. 6th IET International Conference on System Safety 2011. IET. https://doi.org/10.1049/cp.2011.0255

International Road Assessment Programme. (2019). iRAP Star Rating and Investment Plan Manual Version 1.3 (English).

Jain, A., Ahuja, G., Anuranjana., & Mehrotra, D. (2016). *Data mining approach to analyse the road accidents in India* [Paper presentation]. 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 175–179).

Jane, M., Henry, H., Craig, M., Thomas, B., Maryam, M., Keenan, P., Montufar, J., & Leanna, B. (2011). 2011 Conference and Exhibition of the Transportation Association of Canada - Transportation Successes: Let's Build on Them, Risk mapping and a physical-attribute-based star rating system for road safety network screening purposes on Canadian prairie region highways.

Karacasu, M., Ergül, B., & Altin Yavuz, A. (2014). Estimating the causes of traffic accidents using logistic regression and discriminant analysis. *International Journal of Injury Control and Safety Promotion*, *21*(4), 305–313. https://doi.org/10.1080/17457300.2013.815632

Kho, S. M., Pahlavani, P., & Bigdeli, B. (2021). Classification and association rule mining of road collisions for analyzing the fatal severity, a case study. *Journal of Transport & Health*, *23*, 101278. https://doi.org/10.1016/j.jth.2021.101278

Maji, A., Velaga, N. R., & Urie, Y. (2018). Hierarchical clustering analysis framework of mutually exclusive crash causation parameters for regional road safety strategies. *International Journal of Injury Control and Safety Promotion*, *25*(3), 257–271. https://doi.org/10.1080/17457300.2017.1416485

El Mazouri, F. Z., Abounaima, M. C., & Zenkouar, K. (2019). Data mining combined to the multicriteria decision analysis for the improvement of road safety: Case of France. *Journal of Big Data*, *6*(1), 1–30. https://doi.org/10.1186/s40537-018-0165-0

Ng, K. S., Hung, W. T., & Wong, W. G. (2002). An algorithm for assessing the risk of traffic accident. *Journal of Safety Research*, *33*(3), 387–410. https://doi.org/10.1016/s0022-4375(02)00033-6

Odoki, J. B., Di Graziano, A., & Akena, R. (2015). A multi-criteria methodology for optimising road investments. In *Proceedings of the Institution of Civil Engineers - Transport*, *168*2015(1), 34–47. https://doi.org/10.1680/tran.12.00053

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., …Moher, D. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews*, 10 (1), 89. https://doi.org/10.1186/s13643-021-01626-4

Park, P. Y., & Young, J. (2012). Investigation of a supplementary tool to assist in the prioritization of emphasis areas in North American strategic highway safety plans. *Accident; Analysis and Prevention*, *45*, 392–405. https://doi.org/10.1016/j.aap.2011.08.006

Saha, P., & Ksaibati, K. (2016). An optimization model for improving highway safety. *Journal of Traffic and Transportation Engineering (English Edition)*, *3*(6), 549–558. https://doi.org/10.1016/j.jtte.2016.01.004

Saloner, G., Shepard, A., & Podolny, J. (2005). *Strategic management*. John Wiley & Sons.

Saranyadevi, S., Murugeswari, R., & Bathrinath, S. (2019). *Fuzzy logic based recommendation system for road accident data analysis* [Paper presentation]. 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) (pp. 1–6). https://doi.org/10.1109/INCOS45849.2019.8951422

Shen, Y., Hermans, E., Brijs, T., Wets, G., & Vanhoof, K. (2012). Road safety risk evaluation and target setting using data envelopment analysis and its extensions. *Accident; Analysis and Prevention*, *48*, 430–441. https://doi.org/10.1016/j.aap.2012.02.020

Thomas, J., Brunton, J., & Graziosi, S. (2010). *EPPI-Reviewer 4.0: Software for research synthesis. EPPI-Centre Software*. Social Science Research Unit, Institute of Education.

Waka Kotahi NZ Transport Agency. (2022). *One Network Framework (ONF) Classification Guidance.*

Waka Kotahi NZ Transport Agency. (2021). *2021-24 NLTP Investment Prioritisation Method.*

World Health Organization. (2018). *Global status report on road safety 2018.*

World Health Organization. 2022. *High level meeting of the UN general assembly on global road safety.* https://www.who.int/news-room/events/detail/2022/06/30/default-calendar/high-level-meeting-of-the-un-general-assembly-on-global-road-safety.

Yu, J., & Liu, Y. (2012). Prioritizing highway safety improvement projects: A multi-criteria model and case study with SafetyAnalyst. *Safety Science*, *50*(4), 1085–1092. https://doi.org/10.1016/j.ssci.2011.11.018

Heliyon 10 (2024) e34293

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Advancing road safety strategy development: A data-driven multi-objective optimisation integrated approach

Bosong Jiao *, Harry Evdorides

School of Engineering, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

A R T I C L E I N F O

Keywords:
Road safety
Strategic management
Data-driven
Resource allocation

A B S T R A C T

Road traffic accidents pose a significant global health concern, with an alarming 1.19 million fatalities reported in 2021. Traditionally, strategies to address this challenge have relied on expert input and subjective evaluations. This study introduces an adaptive and novel two-stage model that minimizes expert interference by integrating multi-objective optimisation with association rule mining. This innovative approach provides a systematic framework to enhance the efficiency of decision-making and optimise resource allocation outcomes in road infrastructure management, facilitating adaptability to diverse objectives. A case study in Utrecht from iRAP validates the efficiency of the approach, demonstrating significant improvements across various objectives by using the non-dominant sorting genetic algorithm and enabling road local authorities to tailor investment plans to their specific road network characteristics by crash database mining. However, the methodology requires refinement, particularly in identifying risk levels considering the interactive effects of multiple road attributes. In conclusion, while representing a substantial advancement, further refinement is necessary to fully realize its potential in enhancing road safety.

1. Introduction

Road traffic collisions represent a significant global health concern, contributing substantially to both death and disability. As reported by the World Health Organization (2023), an estimated 1.19 million fatalities resulted from road traffic incidents in 2021, with a corresponding fatality rate of 15 deaths per 100,000 individuals. Road traffic injury remains the 12th leading cause of death when all ages are considered. These incidents not only result in immense human suffering but also impose significant economic costs, including medical expenses, lost productivity, and damage to infrastructure. This collective strain poses tough challenges for stakeholders ranging from policymakers and road authorities to the entire society. The core to address this challenge is the efforts to identify underlying problems and implement targeted interventions and countermeasures. Central to this endeavour is the strategic allocation of resources to initiatives aimed at improving road safety outcomes.

Several studies have demonstrated the ability of enhancing road safety through the implementation of structured safety management processes. For instance, the Highway Safety Improvement Program (HSIP) outlined a three-phase approach to allocate resources effectively for road safety, which involves identifying high-risk areas, proposing appropriate countermeasures, and allocating funds within budgetary constraints [1]. This systematic approach ensures an organised assessment of safety priorities and efficient resource allocation.

* Corresponding author.
E-mail addresses: bxj862@student.bham.ac.uk (B. Jiao), h.evdorides@bham.ac.uk (H. Evdorides).

https://doi.org/10.1016/j.heliyon.2024.e34293
Received 30 May 2024; Received in revised form 4 July 2024; Accepted 8 July 2024
Available online 9 July 2024

Similarly, international road assessment programme developed a suite of tools based on iRAP methodology to support road infrastructure safety management. It includes four protocols: risk mapping, star rating, safer road investment plans and performance tracking [2], which aligns with the three-step structure promoted by HSIP. The impressive impact of iRAP is its establishment of a structured framework beginning with section analysis and extending to road network analysis, facilitating the clear visualization of road risk and overall network performance. Concurrently, a systematic procedure is outlined for the identification of countermeasures, leveraging insights gained from risk mapping and star rating assessments.

In 2020, Waka Kotahi NZ Transport Agency [3] developed a definition of 'one network framework'(ONF) to put people, place and movement at the heart of planning and investment, which allows land use and transport planning to be integrated. The ONF aligns with strategic transport planning at all levels, where road to zero strategy is one of the objectives to achieve zero death on road. Meanwhile, they adopted an investment method that utilized three investment criteria—— government policy statement (GPS) on land transport alignment, scheduling, and efficiency—to prioritize proposed activities[4]. It is an interesting and decisive transport funding allocation framework aligning with strategic objectives.

In recent years, scholars and practitioners have investigated diverse methodologies and frameworks aimed at optimising the allocation of resources for road safety. These approaches involve traditional cost-benefit analyses to prioritize countermeasures, as well as more advanced optimisation techniques for selecting combinations of interventions. However, the search for an optimal investment plan in road safety strategic management is hindered by a variety of practical constraints. These constraints often conflict, leading to trade-offs between benefits, efficiency, and costs. Additionally, the unmeasurable impacts on road network operations exacerbate these challenges. These conflicting objectives and unaccountable effects emphasize the complexities and gaps in current road safety strategic management practices. Decision-making subjectivity and expert preferences often overlook those pertinent road network issues, highlighting the need for advanced and comprehensive optimisation methodologies that accommodate diverse objectives with less human intervention.

Given advancements in technology, data analytics, and computation, there is an increasing demand for data-driven methods in optimising road safety measures. Therefore, this study introduces a novel two-stage model that minimizes expert interference by integrating multi-objective optimisation with association rule mining. This innovative approach offers a systematic framework to enhance decision-making efficiency and optimise resource allocation in road infrastructure management. A case study in Utrecht, validated by iRAP, demonstrates the effectiveness of proposed model, achieving significant improvements across various objectives using the non-dominated sorting genetic algorithm. By mining crash databases, the model enables local road authorities to tailor investment plans to their specific network characteristics. While the methodology represents a substantial advancement in reducing reliance on expert judgment, further refinement is needed to account for the interactive effects of multiple road attributes in identifying risk levels.

This study exclusively examines investment planning for road safety, focusing on the third step of the Highway Safety Improvement Program (HSIP). The identification of high-risk areas and the proposal of countermeasures will not be addressed within the scope of this investigation. The remainder of this paper is organized as follows: Section 2 provides an overview of the optimisation related works developed for road safety improvement regarding resource allocation. Section 3 presents the methodology employed in this study, consisting of objectives identification, data collection, a brief introduction of definitions in multi-objective optimisation and integration of analytical techniques. Section 4 presents the findings of optimisation model and the result of the overall methodology based on information of Utrecht safer road investment plan, followed by a discussion in Section 6 to demonstrate the advancement and implications of the proposed methodology for decision making in practice. Finally, Section 6 offers concluding remarks and suggests directions for future research in the field of road safety strategic management.

2. Related work

The optimisation of road safety resource allocation is a critical endeavour in road infrastructure engineering, where the aim is to minimise crash or severity for improving overall safety outcomes within constrained budgets. Over the years, researchers have proposed various methodologies and frameworks to address this complex challenge.

The mixed integer linear programming (MILP) model was developed by Banihashemi and Dimaiuta [5] to maximize the benefits obtained by predefined alternatives for different segments of the highway which was achieved by minimizing the expected number of crashes. Their subsequent work [6] extended this model to incorporate travel time considerations, seeking to minimise both crash and delay costs within budget constraints.

Chowdhury et al. [7] introduced a multi-objective approach that emphasizes keeps the objectives in their respective units and provides a set of solutions rather than a single optimal. The core to their methodology is the introduction of disutility values, which estimated the relative severity of different crash levels. By identifying causal factors on road segments and suggesting appropriate countermeasures, the study aims to optimise resource allocation while minimizing expected loss of disutility at all road types.

Lambert et al. [8] proposed a multi-objective framework tailored for guardrail resource allocation, which facilitated interpretation of variety of benefits and costs in their own units. Their approach allows for subjective evaluations of different objectives, accommodating the varied needs and preferences of stakeholders.

Importance segregated multi-objective optimisation (ISMO) model was proposed by Yu et al. [9]. They separated objectives into two levels: decision and supporting level, so that fewer optimal solutions would be obtained from decision level and the process avoided overrating less important objectives.

Mishra et al. [10] presented optimal multiple resource allocation strategies to improve urban intersections under policy and budget constraints. focused on optimal resource allocation strategies for urban intersections, with a particular emphasis on equity in safety benefits and allocation within counties. Their approach considers policy and budget constraints to maximize safety benefits while ensuring fair distribution within communities.

Augeri et al. [11] proposed an interactive multi-objectives optimisation method using dominance-based rough set approach (DRSA) to provide an explicit structure for road safety decision makers. They applied DRSA to formulate decision rules after acquirement of pareto solutions, iteratively refining constraints until an optimal solution is reached.

In multi-objective optimisation, the preferences of the designer regarding the prioritization of objectives are captured [12]. The NSGA-II method is a widely used, fast, and elitist approach for deriving Pareto-optimal solutions across various domains.

Hojjati et al. [13]applied NSGA-II to optimise the operations of two reservoirs in the Ozan River catchment, aiming to maximize income from power generation and enhance flood control capacity. Their comparison with multi-objective particle swarm optimisation (MOPSO) revealed that NSGA-II performed better in optimising the reservoir system, offering improved coverage of the true Pareto front.

Rahimi et al. [14] conducted a comprehensive review and bibliometric analysis of application of NSGA-II in scheduling problems, highlighting its benefits and extensive use across different scheduling contexts. Their review underscored the flexibility and efficiency of NSGA-II in solving complex scheduling issues, emphasizing its adaptability in various scenarios.

Shaygan et al. [15] demonstrated the effectiveness of NSGA-II in spatial multi-objective optimisation for land use allocation, showing how the algorithm balances multiple conflicting objectives. This research illustrated its capability in handling spatial optimisation problems, making it a valuable tool for land use planning.

Ma et al. [16] discussed the concept of multi-objective optimisation and the foundation of NSGA-II, reviewing its family and modifications. They classified NSGA-II applications in engineering into categories such as manufacturing ([17,18]; H. [19]), transportation ([20,21]; C. [22]), supply chain [23,24], among others.

While the literature provides an overview of various methodologies for optimising road safety resource allocation and NSGA-II applications in different industries, there is a notable gap in discussing the integration of emerging data mining technologies with multi-objective optimisation (e.g., NSGA-II) in road safety strategic management. The use of artificial intelligence, machine learning, and big data analytics offers promising opportunities to enhance road safety initiatives. Hence, further exploration of holistic and data-driven optimisation procedures that consider broader socio-economic impacts of road safety interventions is needed.

3. Methodology

This section introduces a multi-objective optimisation methodology integrated with association rule mining for road safety resource allocation. The methodology aims to maximize the estimated net present value (NPV) of benefits, minimise the cost per KSI (Killed or Seriously Injured) saved, and account for the service life of countermeasures.

Since the objectives and the constraints involved are often heterogeneous and conflicting, a multi-objective optimisation methodology is proposed. Decision makers are trying to minimise the overall cost per KSI saved and minimise the construction impact to road network to maximize the overall benefits for road network and reduce KSI crash. This method is a data driven procedure composed of two stages to avoid subjectivity and bias in selection of countermeasures illustrated in Fig. 1. In the first stage, the Pareto optimal set generates a sample of solutions (e.g. Combination of countermeasures) where the non-dominated soring genetic algorithm

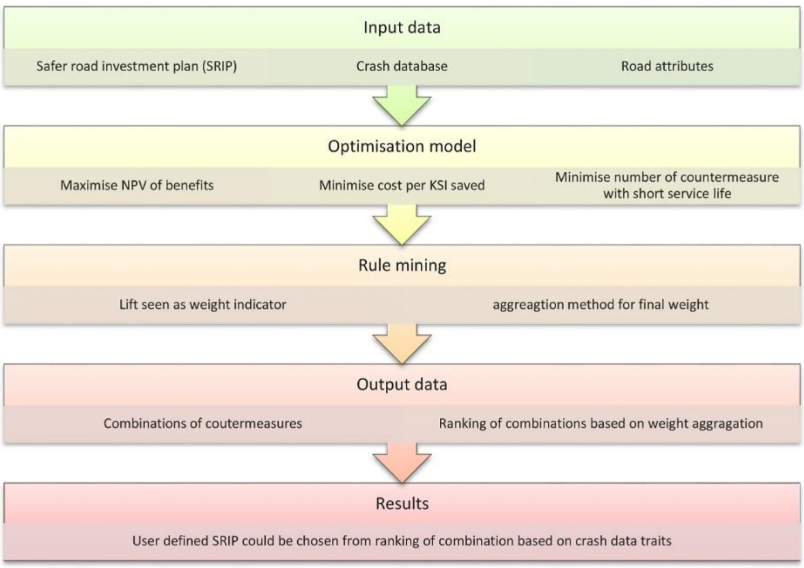

Fig. 1. Process of overall methodology.

(NSGA-II) was employed. In the second stage, the results of association rule mining are applied to weigh the optimal solutions through aggregation method, where a rule based model expressed in terms of easily understandable "if ..., then ..." decision rules is used.

3.1. Theoretical framework

Obviously, the multi-objective optimisation implies the exploration of optimal solutions, constituting a process essentially based on trade-off analysis. It is impossible to obtain a solution that is the best for all considered objectives. In comprehending the mechanics of multi-objective optimisation in facilitating trade-off analysis, it becomes imperative to understand several foundational definitions. As interpreted by Chang [12], these definitions serve as crucial conceptual foundations within multi-objective optimisation frameworks.

Definition 1. Nondominated and Dominated. A Pareto optimal point has no other point that improves at least one objective without detriment to another; that is, it is non-dominated, vice versa.

Definition 2. Pareto front. A set of solutions that are non-dominated to each other but are superior to the rest of solutions in the search space.

By elucidating these fundamental concepts, researchers gain a general understanding of the operational mechanisms governing trade-off analysis within the context of multi-objective optimisation. If a random data set is considered with two objectives aiming at maximization a plotted may be created as shown in Fig. 2.

As depicted in the above figure, the red dots are the set of non-dominated points called pareto front, each representing a unique trade-off between competing objectives. The Pareto front are deemed Pareto optimal (combination of countermeasures), as they offer the best achievable compromises across the objectives under consideration. Conversely, points located below or within the Pareto front as shown as blue dots are considered dominated, as they can be surpassed by other solutions in at least one objective without any corresponding improvement in other objectives.

3.2. Research methods

3.2.1. Stage 1: multi-objective optimisation

The first stage involves generating a Pareto optimal set of solutions using the Non-Dominated Sorting Genetic Algorithm II (NSGA II) [25], implemented in R Studio. NSGA II is chosen for its ability to provide a better spread of solutions and convergence compared to other multi-objective evolutionary algorithms.

It consists of two main steps: nondominated levels sorting and diversity preservation to search optimal solutions. The overall procedure of NSGA II is illustrated in Fig. 3 and demonstrated as followed. First, a random parent population of combinations of countermeasures (P_0) of size N is initialised and evaluated using the objective functions to rank their front tier level. Offspring (G_0) of size N is created by elitism election recombination, and mutation operators. The loop will begin once the first initial generation ends. The combined population of C_1 is composed of P_0 and G_0 with size of 2 N. Nondominated levels sorting is used to rank the front tier level (F_i) and search best solutions in the combined population of C_1. Next, in order to identify the new population (P_2) with size of N, which is composed of F_1 to F_i, crowding distance is calculated for each solution in ith front tier level to guarantee a good spread of solutions, where a partial operator was identified for binary tournament selection. Between two solutions, solution with a lower rank of front tier level is preferred; solution in a less crowded region if preferred if they are ranked as the same front tier level. The new population (P_1) of size N is now used for selection, crossover, and mutation to create an offspring (G_1) to make up the new combined population of C_2. The process will iterate until the maximum generation is achieved. It should be noted that more than one optimal

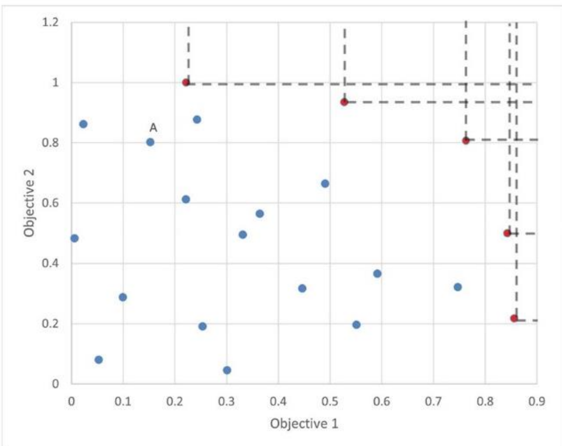

Fig. 2. Random data set for definition illustration.

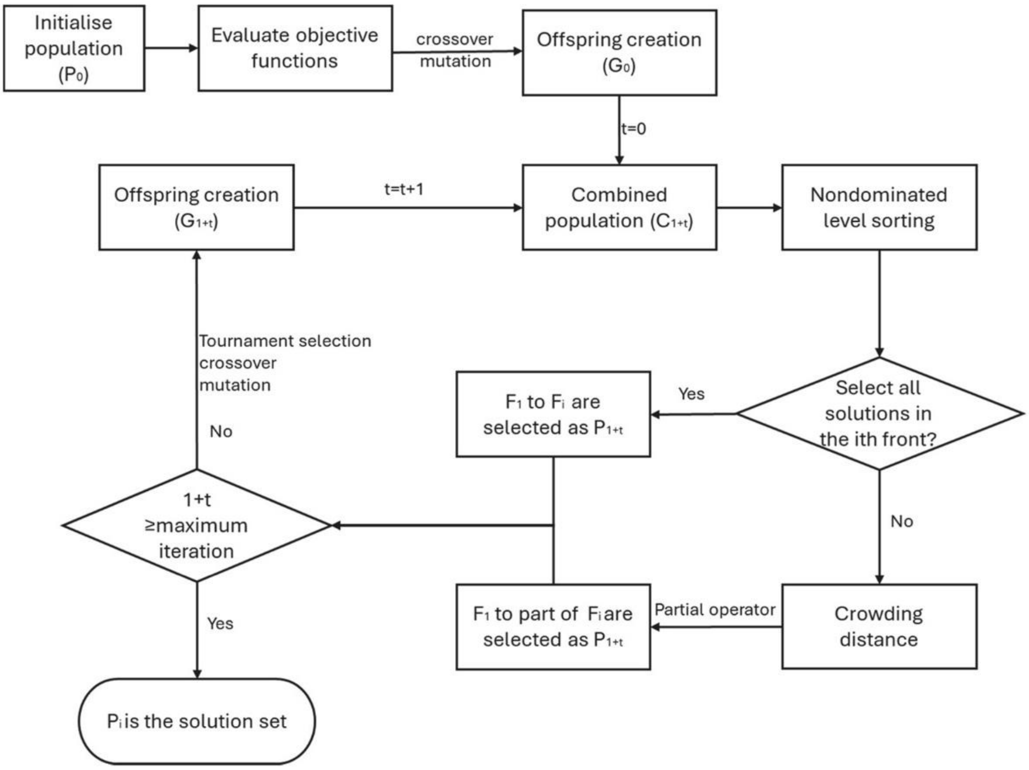

Fig. 3. Flow chart of NSGA II.

solution may be found.

3.2.2. Stage 2: association rule mining

In the second stage, association rule mining is applied to weigh the optimal solutions through an aggregation method. This process uses a rule-based model with "if …, then … " decision rules.

Once nondominated solutions are searched out, it is necessary to identify the most suitable solution based on different weight criteria. Some studies [9,11] tried to employ multi criteria and interactive method to select the best solution based on experts' preference and predefined criteria. Such methods are subjective resulting in bias and neglect of the importance of historical data. Thus, association rule mining is introduced to provide data-driven evidence in the context of road engineering.

Association rule mining is a widely used data mining method (Agrawal et al., 1994) to extract hidden information for road safety, which reflects the overall safety performance and potential problems of road network in the form of "if … then …". There is a crucial parameter named lift indicating the robustness of association between antecedent (e.g. road type) and consequent (e.g. fatalities) in a rule [26]. It quantifies the relationship between the observed support and the support that would be anticipated under the assumption of independence between the left-hand side (LHS) and right-hand side (RHS). Therefore, this study employs lift values as indicators to measure the strength of associations contributing to KSI crashes.

The value of lift implies how strong association of a rule, where a value of 1 indicates the items in rules are independent, a value of over 1 indicates items are more likely to occur than if items were independent to each other. A higher lift value in a rule signifies a greater likelihood of resulting in killed or serious injury outcomes. Therefore, road attributes data at all sites and road sections for proposed road improvements are required to carry out lift weight aggregation.

Given that the performance of investment plan is based on overall countermeasure effectiveness, resulting in various characteristics within the same road attribute category, the traits with higher lift values are used for lift aggregation. Ultimately, the lift aggregation of countermeasure is calculated by summing the lifts of all associated road attributes. Similarly, the lift aggregation of combinations is the sum of lift of all selected countermeasures, which stands for the risk level of roads where these selected countermeasures are required to enhance road safety.

$$\textit{Lift aggregation of combination} = \sum_{i=1}^{i} k_i \left(\sum_{n=1}^{n} l_{n,i} \right)$$

Where n indicates the nth road attributes associoated to fatal crash, i denotes the ith countermeasure proposed for road investment plan, $l_{n,i}$ is the lift value of the rule specified the strongness of association between nth attribute and KSI outcome for ith countermeasure option, and k_i is a binary parameter denoting that $k_i = 1$, if the ith countermeasure is selected in the combinations, otherwise, $k_i = 0$.

3.3. Modelling

3.3.1. Multi-objective identification

The optimisation model is structured to address multiple objectives and the objectives are considered from various perspectives. Benefit-cost analysis (BCA) and cost-effectiveness analysis (CEA) stand as two fundamental approaches for evaluating the efficacy of countermeasures. Within the iRAP methodology, the benefit-cost ratio (BCR) serves as a key metric for assessing countermeasures in economic terms. This involves computing the present value of safety benefits, considering the prevention of fatalities and serious injuries alongside the economic valuation of human life and severe injury. The costs associated with countermeasures are derived from construction expenses and data on service life. While BCR effectively quantifies the monetary implications of each countermeasure, decision-makers, operating under budgetary constraints, often aim to maximize the net present value of benefits, where the first objective is determined.

Behnood & Pino [27] proposed an optimisation framework grounded in cost-effectiveness analysis. Their procedure seeks to minimise the costs of countermeasures while simultaneously maximizing the reduction in the number of crashes. A critical aspect of this approach lies in understanding the resource allocation required for each additional unit of prevented fatalities and serious injuries. Accordingly, the second objective is identified as the minimization of the cost-effectiveness ratio (CER), which quantifies the resources expended per additional unit of prevented fatalities and serious injuries.

From the economical perspective, this model has chosen the service life as the third objective. Service life is a sensible measure in the way that it indicates the duration of the countermeasure after the implementation. Thus, the longer the service life, the more economical benefits (larger net present values) the countermeasures would yield for the same cost of implementation [9]. In addition, according to the study of Elvik [28] and Byaruhanga & Evdorides [29] the service life could affect the selection of countermeasures. In instances where the budget is sufficiently robust, prioritizing countermeasures characterized by long service life becomes instrumental in optimising economic advantages and ensuring enduring safety benefits. This strategic emphasis on countermeasures with continued effectiveness contributes to a reduction in treatment recurrence at specific locations, thereby minimizing disruptions to the operational efficiency of the road network.

Therefore, three objectives optimisation model is formulated focusing on: 1) maximize the net present value of benefits (see Eq. (1)); and 2) minimise the cost per KSI saved; (see Eq. (2)); and 3) minimise the total count of countermeasures characterized by a short service life (SSL) (see Eq. (3)).

$$MAX\ f_1 = \sum_{i=1}^{n} (B_i k) \quad \text{Equation 1}$$

$$MAX\ f_2 = \sum_{i=1}^{n} (E_i k) \quad \text{Equation 2}$$

$$MIN\ f_3 = \sum_{i=1}^{n} k_{i,SL=short} \quad \text{Equation 3}$$

In these equations, n represents the total number of recommended improvement activities outlined in the iRAP safer road investment plan and k is a binary parameter indicating the choice of selecting where $k = 1$ if the countermeasure is selected, otherwise $k = 0$. The parameter i signifies the i^{th} countermeasure proposed for road enhancement, where B_i denotes the NPV of benefits by implementing the i^{th} countermeasure, and E_i indicates the cost per KSI saved after implementing the i^{th} countermeasure. SL is an abbreviation for service life, representing the duration for which the countermeasure remains effective. The variable $k_{i,SL=short}$ signifies the decision to implement the i^{th} countermeasure with short service life (SSL). (where $k = 1$ if it is selected and it has short service lift, otherwise $k = 0$), and the main constraint of the model is the total budget (B) given by:

$$\sum_{i=1}^{n} C_i k \leq B \quad \text{Equation 4}$$

Where, C_i denotes the cost of i^{th} countermeasure.

3.3.2. NSGA-II parameters setting

In the context of using NSGA-II for optimising Pareto front solutions within R Studio, it is critical to determine appropriate parameters for the algorithm. A key initial consideration is the population size. According to empirical studies [30,31], a population size ranging from 30 to 100 is often sufficient to explore the search space effectively while maintaining computational efficiency. In this specific scenario, a population size of 50 is chosen, as it provides a balance that allows for robust exploration and enables decision makers to compare and select different countermeasure combinations.

The decision variables in this optimisation problem are binary, representing whether a countermeasure is selected or not. The fitness function for this scenario encompasses three objectives: maximizing the net present value of benefits, minimizing the cost per KSI saved, and minimizing the total number of countermeasures with a short service life (SSL).

The maximum iteration count is set to 1000, and the nBits parameter, which specifies the number of bits used in the binary encoding of the optimisation problem, corresponds to the number of proposed countermeasures in the iRAP investment plan for Utrecht.

The crossover rate, which determines how frequently the crossover operator is applied, is set to 0.8 by default. This rate strikes a balance by introducing new structures into the population at a rate that avoids the premature loss of high-performance structures while still promoting adequate exploration. Mutation, which increases population variability, is set to a default probability of 0.1. This low mutation rate prevents any single bit position from converging prematurely while avoiding the randomness associated with higher mutation rates [25].

3.3.3. Association rule mining parameters setting

For holistic road safety improvement, prioritizing the reduction of KSI incidents is a fundamental goal in road safety. There were an estimated 1.19 million road traffic deaths in 2021; this corresponds to a rate of 15 road traffic deaths per 100000 population [32]. In addition, EURoad Safety Policy Framework 2021–2030 aims at halving the number of fatalities and serious injuries on European roads by 2030, as a milestone on the way to 'Vision Zero' (European Commission, 2019), supporting the importance of understanding of KSI incidents to improve overall public safety and well-being. According to this, the consequent of the rule is set as KSI, so that all the rules extracted are related to KSI crashes on road network.

Minimum support and confidence are essential constraints for extracting useful information from data ([33]). To capture 99 % of the data corresponding to a specific consequent, the support value of that consequent is first calculated. The minimum support value is then set to 1 % of this support value. For instance, in 2014, there were 2799 crash records in Utrecht, 478 of which were KSI (Killed or Seriously Injured) crashes, resulting in a support value of 0.171. Consequently, the minimum support value is set to 0.002. Correspondingly, specifying confidence is less critical in this study since the primary parameter used is the lift value. Therefore, the minimum confidence is set to match the minimum support value, which is very low, reflecting patterns with weak association and frequency in the crash database.

Additionally, the number of items in each rule is set to two: one being the consequent of KSI (Killed or Seriously Injured) and the

Table 1
Example of safer road investment plan data.

Countermeasure	Length/ Sites	FSIs saved	PV of safety benefit	Estimated Cost	Net monetary benefit	Cost per KSI save	life
Additional lane (2 + 1 road with barrier)	15.10 km	80	25660031	20655000	5005031	258188	medium
Bicycle Lane (off-road)	3.50 km	3	1122681	392156	730525	130719	medium
Central hatching	5.70 km	0.6	184481	34173	150308	56955	medium
Central median barrier (1 + 1)	34.80 km	41	13286137	6288200	6997937	153371	medium
Central median barrier (no duplication)	0.70 km	1	365701	95182	270519	95182	medium
Centreline rumble strip/flexi-post	1.80 km	0.3	108423	19512	88911	65040	low
Clear roadside hazards (bike lane)	1.20 km	0.5	160633	216000	−55367	432000	low
Clear roadside hazards - driver side	0.10 km	0.1	34232	20000	14232	200000	low
Delineation and signing (intersection)	8 sites	0.5	149378	88582	60796	177164	low
Duplication with median barrier	1.20 km	26	8323159	6480000	1843159	249231	high
Footpath provision driver side (>3 m from road)	25.40 km	28	9060036	2922040	6137996	104359	low
Footpath provision driver side (adjacent to road)	27.90 km	35	11319681	4388280	6931401	125379	low
Footpath provision driver side (informal path >1 m)	4.70 km	1	307139	114747	192392	114747	low
Footpath provision passenger side (>3 m from road)	26.80 km	29	9267974	3085368	6182606	106392	low
Footpath provision passenger side (adjacent to road)	44.10 km	52	16843531	6939120	9904411	133445	low
Footpath provision passenger side (informal path >1 m)	5.80 km	1	406768	141451	265317	141451	low
Improve curve delineation	0.40 km	0.5	148419	7460	140959	14920	low
Improve Delineation	45.70 km	13	4190892	847446	3343446	65188	low
Lane widening (>0.5 m)	0.10 km	0.6	186272	152529	33743	254215	low
Lane widening (up to 0.5 m)	1.00 km	2	684333	656756	27577	328378	low
Overtaking lane	0.30 km	1	339394	405000	−65606	405000	medium
Parking improvements	1.50 km	0.3	99666	18900	80766	63000	low
Pedestrian fencing	27.20 km	9	2976955	179606	2797349	19956	medium
Protected turn lane (unsignalised 3 leg)	54 sites	74	23918473	7210571	16707902	97440	low
Protected turn lane (unsignalised 4 leg)	3 sites	14	4375323	535399	3839924	38243	low
Protected turn provision at existing signalised site (4-leg)	1 sites	0.6	204497	237955	−33458	396592	low
…							

other representing patterns in the crash database. This study assumes that all onsite attributes of road crashes can independently influence crash severity. Consequently, the lift aggregation method can be employed to assess the risk level of road locations or segments where road improvements are proposed by the iRAP investment plan.

3.4. Data collection

In this study, three different types of databases are considered. One is the national road crash registration (BRON), which is based in SWOV, the national scientific institute for road safety research in the Netherlands. The other two are safer roads investment plan (SRIP) and road attribute data stored in Vida web software embedded in iRAP methodology, which come from iRAP implementation in Utrecht, Netherlands.

The BRON database is a collection of all road traffic crashes in the Netherland. The database contains detailed information on crash circumstances, such as the year, day, date, road type, speed limit, mode of transport, and more. It serves as a crucial resource for road safety research, enabling the analysis of various factors related to roads, road users, vehicles, and traffic movements. In addition, the data is extensively used for guiding improvements in road safety infrastructure which is highly reliable as it records onsite crash situations, making it an indispensable source for identifying road infrastructure issues. For this study, crash data from 2014 were extracted to develop the proposed procedure.

SRIP are the results of iRAP embedded in the Vida application, a web online tool, which has been elaborated in the methodology of iRAP [2]. There are plenty of road related data required to develop SRIP in ViDA, which also includes a calibration of fatalities estimation. Brief process of SRIP development is provided consisting of 6 steps. Firstly, triggers, the prerequisite conditions to select countermeasures where iRAP has more than 300 triggers been defined, are tested for further consideration of investment plan. Secondly, the number of deaths and serious injuries prevented is calculated. Thirdly, benefit cost ratio (BCR) analysis is carried out for screening out countermeasures of which BCR is below the predefined threshold. Fourthly, minimum length, minimum spacing and hierarchy rules were considered by analysis team to define their preference for selecting countermeasures. Fifthly, the impact of all viable countermeasures is determined at the 100-m level and the combined deaths and serious injuries after all viable countermeasures have been applied is calculated for each crash type. Lastly, the total number of deaths and serious injures prevented for the individual countermeasure at that location is then adjusted to take account of multiple countermeasures that affect the same crash type. At the

Table 2
Example of road attribute data related to countermeasures.

Countermeasure	Length/ Sites	Urban/rural town or rural/ open area	Carriageway label	Street lighting	Speed limit (km/h)	Junction
Additional lane (2 + 1 road with barrier)	15.10 km	rural	Single car'way	partly	80	not
Bicycle Lane (off-road)	3.50 km	both	both	present	50/80	not
Central hatching	5.70 km	urban	Single car'way	present	50/60/80	not
Central median barrier (1 + 1)	34.80 km	rural	Single car'way	partly	60/70/80	not
Central median barrier (no duplication)	0.70 km	rural	Single car'way	present	70/80	not
Centreline rumble strip/flexi-post	1.80 km	rural	Single car'way	present	50/60/70/80	not
Clear roadside hazards (bike lane)	1.20 km	both	Single car'way	present	60/80	not
Clear roadside hazards - driver side	0.10 km	urban	Single car'way	present	50	not
Delineation and signing (intersection)	8 sites	both	Single car'way	present	50/60/80	at
Duplication with median barrier	1.20 km	rural	Single car'way	present	80	not
Footpath provision driver side (>3 m from road)	25.40 km	rural	Single car'way	partly	50/60/80	not
Footpath provision driver side (adjacent to road)	27.90 km	rural	both	partly	50/60/80	not
Footpath provision driver side (informal path >1 m)	4.70 km	rural	Single car'way	partly	50/60	not
Footpath provision passenger side (>3 m from road)	26.80 km	rural	both	partly	50/60/70/80	not
Footpath provision passenger side (adjacent to road)	44.10 km	rural	both	partly	50/60/70/80/ 100	not
Footpath provision passenger side (informal path >1 m)	5.80 km	rural	both	partly	50/60/70	not
Improve curve delineation	0.40 km	rural	both	present	80	not
Improve Delineation	45.70 km	both	both	partly	50/60/80/100	not
Lane widening (>0.5 m)	0.10 km	rural	Single car'way	present	80	not
Lane widening (up to 0.5 m)	1.00 km	rural	Single car'way	present	70/80	not
Overtaking lane	0.30 km	rural	Single car'way	present	70/80	not
Parking improvements	1.50 km	both	Single car'way	partly	60/50/80	not
Pedestrian fencing	27.20 km	both	both	partly	50/60/70/80/ 100	not
Protected turn lane (unsignalised 3 leg)	54 sites	rural	Single car'way	partly	50/60/80	at
Protected turn lane (unsignalised 4 leg)	3 sites	rural	Single car'way	present	80	at
Protected turn provision at existing signalised site (4-leg)	1 sites	rural	Dual car'way	present	80	at
…						

completion of the countermeasure selection process, a final economic analysis for all the countermeasures selected across all roads is undertaken and SRIP is produced. The SRIP data is highly detailed and structured, providing a thorough assessment of countermeasure effectiveness and economic viability based on road sections and location safety analysis. The example of data with performance of each countermeasure is summarised in Table 1.

Road attribute data, 78 attributes in total coded and stored in ViDA, are collected during road inspection by operating road survey and road coding. Road survey requires a detailed of geo-referenced image data of a road network and road coding records road attribute categories using survey images. The road attribute data is precise and systematically recorded, ensuring a comprehensive representation of road conditions for further development of SRIP, where attribute data are extracted based on locations or road segments of each proposed countermeasure. Compared road crash data in Netherland and ViDA embedded road attribute data, four distinct representative attributes are identified, which are recorded in both databases with expected significance to road infrastructure safety improvement. The road attributes data related to countermeasures is summarised in Table 2.

4. Result

4.1. Result of optimisation model

Budget constraint used in this study were identifies from the user-defined investment plan determined by the royal Dutch tourist

Table 3
Objective value of optimal solutions.

Solutions	1	2	3	4	5
Obj. 1: NPV of benefits	163745935	62355466	106094212	61624941	45647564
Obj. 2: cost per KSI saved	129629	69897	125084	69145	58020
Obj. 3: Number of SSL Countermeasures	21	2	2	2	1
Solutions	6	7	8	9	10
Obj. 1: NPV of benefits	62128373	57182693	62187724	102935572	58817531
Obj. 2: cost per KSI saved	153058	125139	153890	100203	50570
Obj. 3: Number of SSL Countermeasures	0	0	0	4	3
Solutions	11	12	13	14	15
Obj. 1: NPV of benefits	88288245	151958418	88295800	37919102	95286182
Obj. 2: cost per KSI saved	82046	123103	120504	25673	89228
Obj. 3: Number of SSL Countermeasures	8	14	1	1	8
Solutions	16	17	18	19	20
Obj. 1: NPV of benefits	49288692	93173174	89325513	21084157	141657702
Obj. 2: cost per KSI saved	118680	84663	121505	92654	108621
Obj. 3: Number of SSL Countermeasures	0	8	2	0	15
Solutions	21	22	23	24	25
Obj. 1: NPV of benefits	108866915	101958742	145938452	139524884	142835876
Obj. 2: cost per KSI saved	96799	98967	118656	120238	108972
Obj. 3: Number of SSL Countermeasures	5	3	9	7	16
Solutions	26	27	28	29	30
Obj. 1: NPV of benefits	131532448	129387034	155634348	106692602	101194832
Obj. 2: cost per KSI saved	98601	115371	126531	95753	143154
Obj. 3: Number of SSL Countermeasures	11	6	19	16	1
Solutions	31	32	33	34	35
Obj. 1: NPV of benefits	94390118	20143355	14086220	94960805	95066993
Obj. 2: cost per KSI saved	87614	84887	43457	93587	93634
Obj. 3: Number of SSL Countermeasures	8	0	0	3	3
Solutions	36	37	38	39	40
Obj. 1: NPV of benefits	105514428	13875943	95120643	65815468	38649627
Obj. 2: cost per KSI saved	95033	27584	87935	66995	28086
Obj. 3: Number of SSL Countermeasures	15	0	8	3	1
Solutions	41	42	43	44	45
Obj. 1: NPV of benefits	117578318	74322784	139007051	118748037	58905038
Obj. 2: cost per KSI saved	101829	72487	118310	116720	51984
Obj. 3: Number of SSL Countermeasures	7	6	8	4	3

association (ANWB). They ultimately chose 30 countermeasures out of 47 countermeasures recommanded by iRAP to improve the provincial road in Utrecht, which estimated saving 777 fatalities and serious injuries, 250 million euro, with benefit cost ratio of 2.2 by costing 113 million euro. Therefore, the budegt constraint was set to 113 million euro to validate the advancements of the methodology proposed in this study.

After eliminating duplicates and solutions exceeding the budget constraint, 45 optimal solutions emerged. These solutions were assessed based on their objective values, as detailed in Table 3, and the findings were visually presented in the accompanying Fig. 4. The net present value of benefits, as determined by ANWB, amounts to 135.7 million euros. Among the 45 solutions, 8 of them (designated as solutions 1, 12, 20, 23, 24, 25, 28, and 43) outperform ANWB's investment decision, specifically in maximizing the net present value of benefits. It costs 147,281 euros to save each KSI occupant in their investment plan, with only two solutions (designated as solutions 6 and 8) performing less favourably compared to ANWB's decision, while the majority of solutions prove to be more cost-effective. According to investment strategy determined by ANWB, which involves the selection of 19 countermeasures with short service life, it is noted that over half of the improvements necessitate reconstruction or rebuilding within a period of one to five years. This would result in periodic efforts and potentially disrupt the standard functioning of the road network significantly. Among the 45 solutions evaluated, only two investment combinations opt for at least 19 countermeasures with short service lives, while the remaining alternatives exhibit superior performance with respect to the third specified objective.

4.2. Result of integrated methodology

Once the optimal solutions are searched, all alternative combinations are employed to weight assignment process. As mentioned in methodology, the lift value is used as a weight metric in the form of rule to indicate the association between road attributes to KSI result, where association rule mining is applied to explore their association. Rules were extracted and demonstrated in Table 4.

According to the attributes derived from ViDA database and the result of crash databases rule mining analysis, four distinct risk attributes are used for lift aggregation method involving the road area, speed limit, situation of road, and street lightning. To highlight the urgency and significance of countermeasures recommended by ViDA across diverse locations and sections, the level of risk associated with roads will be signified by greater lifts of attributes among varying road segments. For instance, in the scenario where the construction of refuge islands is recommended for both three-leg and four-leg intersections, the lift value observed for the three-leg intersection will serve as a measure for urgency due to its comparatively higher lift. The top ten solutions ranked by lift aggregation with their objective value are demonstrated in Table 5.

5. Discussion

5.1. Model performance of case study

It is impossible to obtain a solution that is the best for all considered objectives, whereas it is imperative to discover the pareto front tier where a set of optimal solutions could be identified for decision makers. This approach enables an evidence-based assessment of prioritization concerning differential risk levels identified across distinct road infrastructure configurations, thereby informing the allocation of resources and interventions in road safety management. Contrasted with the investment plan outlined in the Utrecht report, which involved the selection of 33 countermeasures within a budgetary allocation of 114378840 euros, their strategy selected for 19 countermeasures characterized by a limited operational lifespan of fewer than 5 years. This selection was estimated to generate a net present value of benefits amounting to 135739970 euros, with an associated cost of 147282 euros per KSI saved. However, our model diverged in its selection.

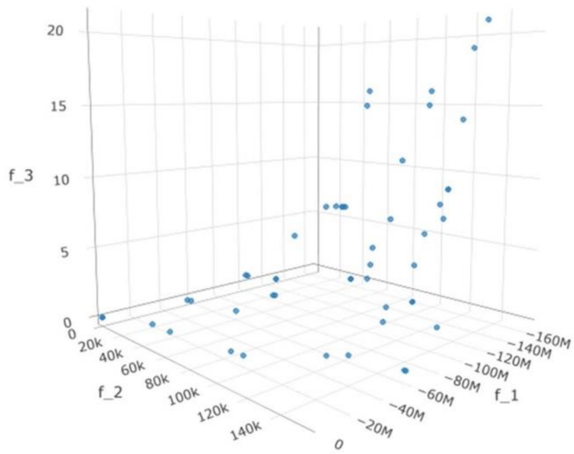

Fig. 4. Visualization of Pareto front.

Table 4
Decision rules associated to KSI crash.

LHS	RHS	support	confidence	lift
Inside outside built-up areas = Urban area	KSI	0.005359057	0.6	3.513389
Inside outside built-up areas = Rural area	KSI	0.005716327	0.484848485	2.839102
Speed limit road = 60 km/h	KSI	0.011432655	0.32	1.873808
Situation of road = Bend	KSI	0.016434441	0.238341969	1.395647
Situation of road = Intersection - 3 legs	KSI	0.019649875	0.224489796	1.314533
Road surface = Concrete	KSI	0.00785995	0.211538462	1.238695
Road surface = Porous asphalt	KSI	0.034297964	0.210065646	1.230071
Speed limit of road = 100 km/h	KSI	0.016791711	0.209821429	1.228641
Speed limit of road = 120 km/h	KSI	0.010360843	0.204225352	1.195872
Situation of road = Intersection - 4 legs	KSI	0.03965702	0.198214286	1.160673
Street lighting = Not burning	KSI	0.123258307	0.18935236	1.108781

Table 5
Ranking of investment plan.

Investment alternatives	Objective 1 (Maximize)	Objective 2 (Minimise)	Objective 3 (Minimise)	Lift aggregation	Ranking
1	163745935	129628.5	21	163.85	1
29	155634348	126531.3	19	145.06	2
26	142835876	108971.9	16	134.27	3
13	151958418	123103	14	132.94	4
30	106692602	95752.54	16	128.32	5
21	141657702	108620.5	15	119.24	6
37	105514428	95032.68	15	113.29	7
25	139524884	120238	7	96.42	8
24	145938452	118656	9	93.05	9
16	95286182	89227.85	8	89.47	10
Utrecht Plan	135739970	135739970	19	–	–

The top 10 alternatives are ranked by lift aggregation. In other word, these alternatives concern road locations and sections with high risky road attributes that are very likely to cause KSI crash. Each objective will be discussed separately to validate the advancement of our methodology. With regards to the objective 1 (Obj.1), aimed at maximizing the NPV of benefits, it is noteworthy that alternative 1 shows a visible enhancement of about 20 % when compared with the NPV estimation described in the Utrecht report. This improvement implies the advanced ability of our proposed methodology in searching the best solution for maximizing monetary

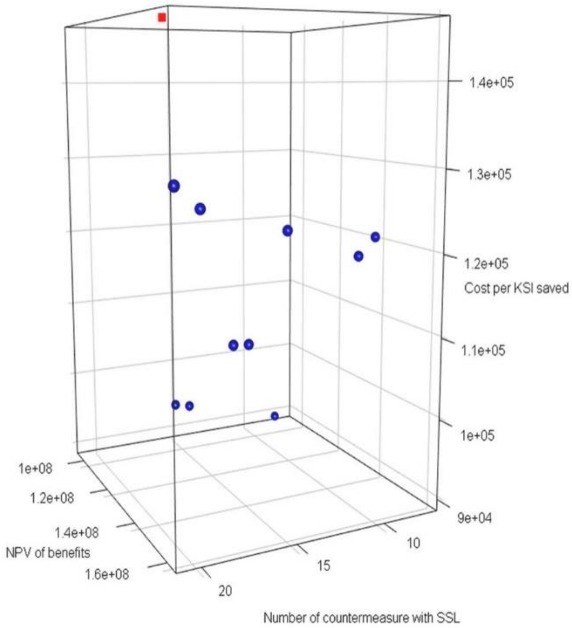

Fig. 5. 3D visualization of top 10 investment plan.

gains.

As for objective 2 (Obj.2), which strives to minimise the cost per KSI saved, alternative 16 emerges as a standout performer, with an average cost of 89,227.85 euros per KSI saved. This represents a significant improvement from the Utrecht plan, where the cost per KSI saved is approximately 147,281.5 euros, thereby signifying a substantial 39 % reduction in anticipated expenditures. This marked reduction underscores a valuable advancement by applying the proposed method in cost-effectiveness measures.

Regarding the objective 3 (Obj.3), trying to minimise the number of countermeasures with a short lifespan of less than 5 years selected, the Utrecht report identified a total of 19 such countermeasures. In contrast, Alternative 25 selected only seven short-lived road improvement interventions. This plan is anticipated to result in reduced frequency of construction activities and decreased disruption to the operation of the road network. This discrepancy also highlights the capacity to identify optimal solutions in quantitative terms. In addition, the proposed methodology facilitates the proficiency in simultaneously optimising various objectives across different metrics as it is a critical consideration given that real-world objectives typically extend beyond financial concerns to encompass diverse units of measurement.

Despite the presence of trade-offs among the three objectives, the model identifies top 10 optimal countermeasure combinations based on lift aggregation. Note that we aim to maximize NPV of benefits (Obj.1).while minimise Cost per KSI saved (Obj.2) and number of countermeasure with SSL (Obj.3).The 3D visualization depicts the 10 alternatives as blue circles, comparing with the investment plan of Utrecht represented by red squares as show in Fig. 5. Intriguingly, all solutions outperform in Obj.2, which suggests that these investment plans are particularly effective in reducing costs associated with preventing KSIs. Half of these solutions show superior performance across all three objectives compared to the investment plan formulated by ANWB. This indicates that these solutions not only excel in cost-effectiveness but also in achieving other objectives related to safety.

Due to their exceptional performance regarding the second objective, a 2D visualization is presented for a clearer understanding of the performance and compromises of 10 investment plans considering the first and third objectives as illustrated in Fig. 6. It interprets the comparative performance of various solutions in terms of objective 1 (NPV of benefits) and objective 3 (Number of countermeasures with SSL). Notably, alternatives 15, 29, and 36 exhibit a strategic inclination towards reducing the number of Short Service Lift (SSL) countermeasures, with sacrifice of the Net Present Value (NPV) of benefits. Conversely, combinations 1 and 28 prioritize the maximization of NPV of benefits, even employing an equal or greater number of SSL countermeasures. Remarkably, the methodology employed in this study has found five solutions (12, 20, 23, 24, and 25) which simultaneously optimise three objectives to varying extents.

This optimisation method is contributory in realizing substantial value within resource constraints, featuring the multiple considerations and diverse measurement metrics involved. This highlights the value of the proposed methodology in achieving a balance between competing priorities and underscores its potential to outperform conventional planning strategies in optimising outcomes across multiple dimensions.

In recent years, there has been a prevalence of single-objective optimisation methods in the road context, each focusing on different

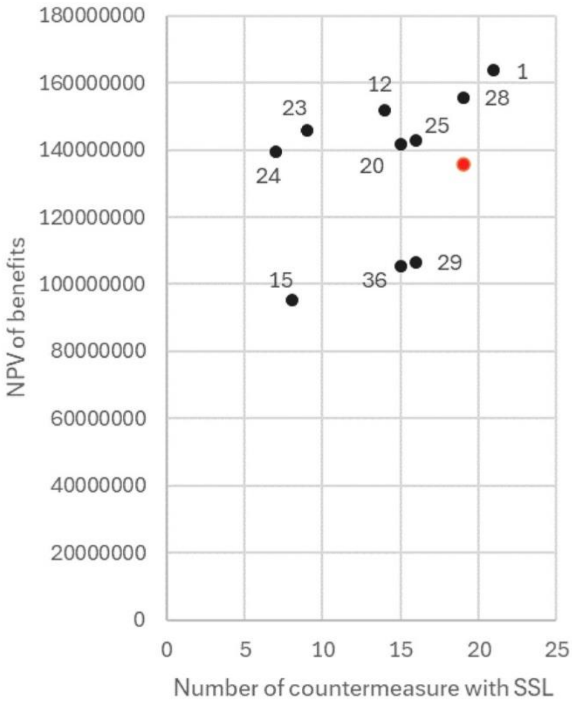

Fig. 6. Two objectives analysis.

aspects. These methods commonly revolve around optimising specific performance index Chowdhury et al. [7], minimizing the number of crashes and time delays [6], or considering the service life of countermeasures [29]. In addition, some multi-objective optimisation models have been developed, all rely heavily on expert input, often involving subjective evaluations of stakeholder needs and preferences [8], two-level optimisation processes by classifying decisions and supporting objectives [9], or the formulation of decision rules by experts [11].

In contrast, the proposed two-stage model not only considers diverse objectives, even when they differ in units, but also provides data-driven evidence of road context characteristics with minimal expert efforts to allocate resource. The contribution of this study is twofold. As demonstrated by the prototype developed in this study, the optimisation model effectively searches the Pareto frontier across various objectives within the constrained budget. Specifically, it underscores the practical applicability of the model in real-world decision-making scenarios and provides empirical evidence of the effectiveness of the non-dominant sorting genetic algorithm in identifying the Pareto frontier. Furthermore, this study highlights the potential of data-driven preferences to guide resource allocation and offers a framework for integrating data-driven insights with less expert efforts to enhance decision-making in complex road management scenarios.

This model was developed by leveraging the outcomes of an investment plan derived from the iRAP methodology. This approach empowers local authorities to devise user defined investment plans tailored to the specific characteristics of their local road network environment. While not advocating for complete elimination of expert control, this model empowers experts with greater insights from natural data pertaining to diverse road contexts, thereby reducing the pressure associated with searching for optimal solutions.

5.2. *Challenges and limitations*

The proposed methodology, while demonstrating substantial potential in optimising resource allocation and enhancing road safety, presents several challenges and limitations that must be acknowledged.

A significant challenge lies in the extensive data requirements of the methodology. High-quality and comprehensive historical crash data, alongside detailed road attribute data, are essential for analysis. However, acquiring such data can be problematic, especially in regions with limited data collection infrastructure or inconsistent data recording practices. The variability in data quality and completeness can adversely affect the reliability of the results.

The study is conducted based on the overall performance of specific countermeasures derived from iRAP investment plan. However, different road sections or junctions may exhibit varied road attributes within the same countermeasure category, leading to a potential oversimplification in the analysis. Enhanced specificity could be attained by formulating investment plans based on homogeneous sections and junctions, each characterized by distinct countermeasure performance metrics. This would allow for more precise and effective countermeasure selection, tailored to the unique characteristics of specific road segments.

The current approach relies on the lift value derived from association rule mining to determine the weight of risk levels. This method assumes that identified road attributes, or risk factors, independently signify the urgency for countermeasures. However, this assumption overlooks potential interactive relationships among these attributes. Consequently, there is a risk of overestimating the risk levels associated with specific road sections or junctions, potentially leading to an exaggerated sense of urgency and misallocation of resources. A more detailed risk assessment approach that accounts for the interdependencies among road attributes would enhance the accuracy and effectiveness of the proposed methodology.

In conclusion, although the proposed methodology shows significant promise in optimising resource allocation and improving road safety outcomes, several challenges and limitations need to be addressed. The extensive data requirements, the need for a correlated risk identification are critical areas that require further attention. Future research should consider incorporating comprehensive datasets from different regions and contexts to thoroughly evaluate the performance of proposed model. Addressing these limitations will enhance the reliability and practicality of the methodology, making it a more robust tool for road safety management.

6. Conclusion

This study highlights the originality and importance of its methodology in addressing the complexities of multi-objective optimisation for road safety strategy development. By overcoming challenges associated with simultaneous optimisation of multiple objectives, this approach offers a novel solution that promises to improve decision-making processes and optimise resource allocation in road infrastructure management. Additionally, the study emphasizes the practical implications of the methodology by providing a systematic framework that integrates diverse objectives and decision rule mining. This equips road agencies and stakeholders with a valuable tool for more efficient resource allocation and road safety improvement, requiring less human intervention.

However, it is crucial to acknowledge the challenges associated with the quantity and quality of data, particularly in the context of developing countries. The performance of our methodology relies heavily on the availability and reliability of data, highlighting the need for ongoing efforts to improve data collection processes and enhance data quality assurance measures. Furthermore, while the methodology effectively addresses the complexities of multi-objective optimisation, there remains a need for refinement of identification of risk level considering interactive effects of multiple road attributes. Future research should focus on formulation of investment plans centred on homogeneous road sections and junctions and developing more sophisticated models that can capture these interactions more comprehensively, thereby enhancing the effectiveness of the methodology. In conclusion, while the methodology represents a significant advancement in the field of road safety strategy development, there is still much work to be done to fully realize its potential.

Data availability statement

Sharing research data enables other researchers to evaluate findings, build upon previous work, and enhance trust in published articles. Authors make their data as publicly accessible as reasonably possible. Below is a summary of the data availability for this study, which will be published alongside the article.

The data associated with this study has been deposited in a publicly accessible repository. Data utilized in this study were obtained from the National Road Crash Registration (BRON) database, managed by SWOV, the Safer Roads Investment Plan (SRIP) data from the ViDA application based on the iRAP methodology, and the road attribute data also stored in ViDA. Reproduction of this data is authorized, provided the source is acknowledged. Interested parties can access the data through the following link:

SWOV. (2014). National Road Crash Registration (BRON) database. SWOV, The Netherlands. Available at: https://swov.nl/en/data/crashes.

iRAP. (2014). Safer Roads Investment Plan (SRIP) data from ViDA application. iRAP, Utrecht, The Netherlands. Available at: https://vida.irap.org/en-gb/dashboard.

iRAP. (2014). Road attribute data from ViDA application. iRAP, Utrecht, The Netherlands. Available at: https://vida.irap.org/en-gb/dashboard.

CRediT authorship contribution statement

Bosong Jiao: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Harry Evdorides:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Herbel, L. Laing, C. McGovern, Highway Safety Improvement Program (HSIP) Manual, 2010.
[2] International Road Assessment Programme, iRAP Star Rating and Investment Plan Manual, 2019. Version 1.3 (English).
[3] Waka Kotahi NZ Transport Agency, One Network Framework (ONF) Classification Guidance, 2022.
[4] Waka Kotahi NZ Transport Agency, 2021-24 NLTP investment prioritisation methodWaka Kotahi NZ Transport Agency, 2021. https://www.nzta.govt.nz/planning-and-investment/planning-and-investment-knowledge-base/202124-nltp/2021-24-nltp-investment-prioritisation-method/.
[5] M. Banihashemi, M. Dimaiuta, Maximizing safety improvement benefits in crash prediction models with accident modification factors, Transport. Res. Rec. 1908 (1) (2005) 9–18.
[6] M. Banihashemi, Optimization of highway safety and operation by using crash prediction models with accident modification factors, Transport. Res. Rec. 2019 (1) (2007) 108–118, https://doi.org/10.3141/2019-14.
[7] M.A. Chowdhury, N.J. Garber, D. Li, Multiobjective methodology for highway safety resource allocation, J. Infrastruct. Syst. 6 (4) (2000) 138–144, https://doi.org/10.1061/(ASCE)1076-0342(2000)6:4(138.
[8] J.H. Lambert, J.A. Baker, K.D. Peterson, Decision aid for allocation of transportation funds to guardrails, Accid. Anal. Prev. 35 (1) (2003) 47–57, https://doi.org/10.1016/S0001-4575(01)00085-9.
[9] J. Yu, Y. Liu, J. Mao, An importance segregation based multi-objective optimization model for traffic safety improvement countermeasure selection, International Journal of Transportation Science and Technology 1 (2) (2012) 205–218.
[10] S. Mishra, M.M. Golias, S. Sharma, S.D. Boyles, Optimal funding allocation strategies for safety improvements on urban intersections, Transport. Res. Pol. Pract. 75 (2015) 113–133, https://doi.org/10.1016/j.tra.2015.03.001.
[11] M. Augeri, S. Greco, N. Distefano, S. Leonardi, Road safety resource allocation using interactive multiobjective optimization, European Transport\Trasporti Europei 84 (2021) 1–16.
[12] K.-H. Chang, Chapter 19 - multiobjective optimization and advanced topics, in: K.-H. Chang (Ed.), E-Design, Academic Press, 2015, pp. 1105–1173, https://doi.org/10.1016/B978-0-12-382038-9.00019-3.
[13] A. Hojjati, M. Monadi, A. Faridhosseini, M. Mohammadi, Application and comparison of NSGA-II and MOPSO in multi-objective optimization of water resources systems, Journal of Hydrology and Hydromechanics 66 (3) (2018) 323–329.
[14] I. Rahimi, A.H. Gandomi, K. Deb, F. Chen, M.R. Nikoo, Scheduling by NSGA-II: review and bibliometric analysis, Processes 10 (1) (2022), https://doi.org/10.3390/pr10010098.
[15] M. Shaygan, A. Alimohammadi, A. Mansourian, Z.S. Govara, S.M. Kalami, Spatial multi-objective optimization approach for land use allocation using NSGA-II, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (3) (2014) 906–916, https://doi.org/10.1109/JSTARS.2013.2280697.
[16] H. Ma, Y. Zhang, S. Sun, T. Liu, Y. Shan, A comprehensive survey on NSGA-II for multi-objective optimization and applications, Artificial Intelligence Review 56 (12) (2023) 15217–15270, https://doi.org/10.1007/s10462-023-10526-z.
[17] Z. Babajamali, F. Aghadavoudi, F. Farhatnia, S.A. Eftekhari, D. Toghraie, Pareto multi-objective optimization of tandem cold rolling settings for reductions and inter stand tensions using NSGA-II, ISA Transactions 130 (2022) 399–408.
[18] P. Mohapatra, A. Nayak, S.K. Kumar, M.K. Tiwari, Multi-objective process planning and scheduling using controlled elitist non-dominated sorting genetic algorithm, International Journal of Production Research 53 (6) (2015) 1712–1735.
[19] H. Wang, Y. Fu, M. Huang, G.Q. Huang, J. Wang, A NSGA-II based memetic algorithm for multiobjective parallel flowshop scheduling problem, Computers & Industrial Engineering 113 (2017) 185–194.
[20] A. Khavandi Khiavi, H. Mohammadi, Multiobjective optimization in pavement management system using NSGA-II method, Journal of Transportation Engineering 144 (2) (2018) 04018016. Part B: Pavements.
[21] A. Martınez-Puras, J. Pacheco, MOAMP-Tabu search and NSGA-II for a real bi-objective scheduling-routing problem, Knowledge-Based Systems 112 (2016) 92–104.
[22] C. Wang, H. Shi, X. Zuo, A multi-objective genetic algorithm based approach for dynamical bus vehicles scheduling under traffic congestion, Swarm and Evolutionary Computation 54 (2020) 100667.
[23] S. Bandyopadhyay, R. Bhattacharya, Applying modified NSGA-II for bi-objective supply chain problem, Journal of Intelligent Manufacturing 24 (2013) 707–716.

[24] J. Sadeghi, S. Sadeghi, S.T.A. Niaki, A hybrid vendor managed inventory and redundancy allocation optimization problem in supply chain management: an NSGA-II with tuned parameters, Computers & Operations Research 41 (2014) 53–64.
[25] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation 6 (2) (2002) 182–197.
[26] M.J.A. Berry, G.S. Linoff, Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management, John Wiley & Sons, 2004.
[27] H.R. Behnood, S. Pino, Road safety resource allocation using the cost-effectiveness approach; Case study in Iran, Proceedings of the Institution of Civil Engineers-Transport 176 (4) (2020) 213–223.
[28] R. Elvik, Problems in determining the optimal use of road safety measures, Research in Transportation Economics 47 (2014) 27–36.
[29] C.B. Byaruhanga, H. Evdorides, A budget optimisation model for road safety infrastructure countermeasures, Cogent Engineering 9 (1) (2022) 2129363.
[30] J.J. Grefenstette, Optimization of control parameters for genetic algorithms, IEEE Transactions on Systems, Man, and Cybernetics 16 (1) (1986) 122–128, https://doi.org/10.1109/TSMC.1986.289288.
[31] C. Reeves, J.E. Rowe, Genetic Algorithms: Principles and Perspectives: a Guide to GA Theory, vol. 20, Springer Science & Business Media, 2002.
[32] World Health Organisation, Global Status Report on Road Safety 2023, 2023.
[33] R. Agrawal, R. Srikant, others, Fast algorithms for mining association rules, Proc. 20th Int. Conf. Very Large Data Bases, VLDB 1215 (1994) 487–499.