UNIVERSITY OF BIRMINGHAM

AN APPLICATION OF DATA MINING TECHNIQUES TO DETERMINE THE EFFECTIVENESS OF MULTIPLE ROAD SAFETY COUNTERMEASURES

By

Atheer Naji Al-nuaimi

BEng, MEng

A Thesis submitted to the

University of Birmingham

for the degree of

Doctor of Philosophy

June 2019

ABSTRACT

Road safety is a major concern worldwide and requires appropriate management. This study focuses on the effectiveness of combining road safety countermeasures to reduce car occupants' fatalities in Low and Middle Income Countries (LMIC) using the star rating system in the international Road Assessment Programme (iRAP). Such countermeasures may be used individually or as a combination of multiple countermeasures. Yet it may be difficult for practising road safety engineers to make decisions about which combinations of countermeasures are the most effective to use when addressing road safety issues, especially with the availability of many treatments. Likely, errors in this decision-making will lead to inefficient use of limited road safety funding.

This study seeks to develop an innovative method whereby road infrastructure attributes may be combined and ranked in a new way, through processes of Knowledge Discovery in Databases (KDD). The computer environment Waikato Environment for Knowledge Analysis (WEKA) facilitated the above and enabled the extraction of knowledge in terms of decision trees and associated rules.

In addition, traditional statistical models were used to combine road infrastructure attributes in iRAP. Two statistical regression models were developed in this study, the Poisson and Negative Binomial. For the Poisson model, the results showed that overdispersion did exist as the overdispersion factor for the fatality rates data is higher than (1.0). Therefore, the Poisson model was omitted. For the Negative Binomial, the results showed that the independent variables have significant effect on the fatality numbers, as their confidence level is higher than 95%. Therefore, the Negative Binomial was shown to

be an appropriate statistical regression approach to correlate road traffic and geometric attributes to fatalities numbers for the database used in this study.

The knowledge discovery and data mining techniques were found to be an enhanced alternative to the statistical models, which finds the effectiveness of road safety multiple treatments by correlating the current road attributes associated with head-on and run-off crashes without the need for historical crash data. The data mining models developed in this study were verified by comparing the mined countermeasures combinations with the real ones from the iRAP model.

Using an iRAP database from India, the most effective combinations of countermeasures for each crash type and for each road safety level (star rating) were identified. For example, it was found that the most effective countermeasures combination for road sections with head-on and run-off crashes on single lane carriageways with traversable median types and gradient ≥0% to <7.5%, requires that the differential speed limits should not be presented, medium skid resistance, medium lane width (≥2.75m to <3.25m), straight or gentle curvature (radius greater than 900m), adequate quality of curve (advance warning signs of bend and roadside indicators to indicate the curvature), short roadside distance (1 to <5m), narrow paved shoulder (≥0m to <1.0m), and adequate delineation.

The above findings could lead practising road safety engineers to combine and rank road safety treatments more satisfactorily, and could also encourage countries to invest more in road safety to save more realistic estimated fatalities in the future, especially in LMICs.

ACKNOWLEDGMENTS

I have taken efforts in this research. However, it would not have been possible without the kind support and help of many individuals.

I would like first to express my very special gratitude for my supervisor Dr Harry Evdorides for his endless support, inspiring guidance, encouragement, and insightful comments that made this work possible.

Deserving of special mention are Dr Steve Lawson, James Bradford, and Luke Rogers – from the international Road Assessment Programme – for their support and the provision of the data needed in this research.

I sincerely thank Dr Michael Burrow (my second supervisor), Prof Miles tight, and Dr Fiona Raje – from the Department of Civil engineering – for their advice and support.

Many thanks go to Dr Sakdirat Kaewunruen and Dr Robert Akena for their suggestions, which helped in improving the presentation of this thesis.

A very special thanks for the person who is special to me in every way. Thank you for being my inspiration, partner, lover and best friend, my amazing wife, Maiss. A huge thank you for my lovely kids, Dode and Tofe, who their existence in my life was an essential motive for the completion of this project.

I would also like to express my deepest thanks and gratitude for my lovely parents, sister, and brother; Naji, Iman, Serar, and Ahmed for their support, encouragement, and prayer.

I am highly indebted to the Iraqi Ministry of Higher Education and Scientific Research (represented by Al-Mustansiriyah University) for their financial contribution to this study.

Thanks all for your support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ADT	Average Daily Traffic
ARFF	Attribute-Relation File Format
BTE	Bureau of Transport Economics
CART	Classification and Regression Tree
CMF	Crash Modification Factors
CSV	Comma-Separated Values
FN	False Negatives
FP	False Positives
FSI	Fatalities and Serious Injuries
ID3	Iterative Dichotomiser
IQR	Interquartile Range
iRAP	The international Road Assessment Programme
KDD	Knowledge Discovery in Databases
LMIC	Low and Middle Income Countries
MCCF	Multiple Countermeasure Correction Factor
NB	Negative Binomial
SR	Star Rating
SRS	Star Rating Scores

TN	True Negatives
TP	True Positives
USDOT	United States Department of Transportation
WEKA	Waikato Environment for Knowledge Analysis
WHO	World Health Organization

NOTATIONS

Y	*Dependent variable*
X1, X2……Xn	*Independent variables*
β0	*Regression constant*
β1, β2…….βn	*Regression coefficients*
F	*Expected reduction in the number of fatalities per km per year.*
Fe	*Existed number of fatalities*
Nol	*Number of lanes.*
Sres	*Skid resistance.*
Gr	*Gradient*
Ds	*Differential speed limits.*
Os	*Operating speed*
Mt	*Median type*
MCCFa	*The adjusted multiple countermeasures correction factor*
iRAPe	*The average difference between iRAP and observed fatalities reduction.*
NBe	*The average difference between NB and the observed fatalities reduction*
C4.5	*The decision tree based classification algorithm used in this study*

CHAPTER 1

INTRODUCTION

1.1 Global Assessment of Road Safety

Road safety is a major concern worldwide and requires appropriate actions. Every day, many people lose their lives or become disabled or injured in car crashes on the roads around the world. The highest percentage (90%) of the total number of causalities around the world occurs in low and middle income countries (LMIC) (WHO, 2015; World Bank, 2010).

The last report by the World Health Organization (WHO, 2015) indicated that:

1. About 3400 deaths on roads around the world daily;
2. Every year, tens of millions of persons are disabled or injured;
3. Vehicle occupant fatalities and injuries form the highest percentage of road users' casualties, as shown in Figure 1.1.

Improving road safety is an important issue because of the causalities. Its importance has come from humanity and economic reasons. In addition to the main reason, save individuals life, it is also essential to consider improving the economic figures for each country around the world. It is clear that even the cost of one death of car crashes is high. The cost of the entire number of causalities and damages from road accidents in one year forms a considerable part of countries economy. The joint report of WHO and the World Bank in 2004 indicated that, the cost of road accidents in low and middle income countries (LMIC), countries with Gross National Income £2967 or less (World Bank, 2017), is estimated as £49.5 billion every year (WHO, 2004).

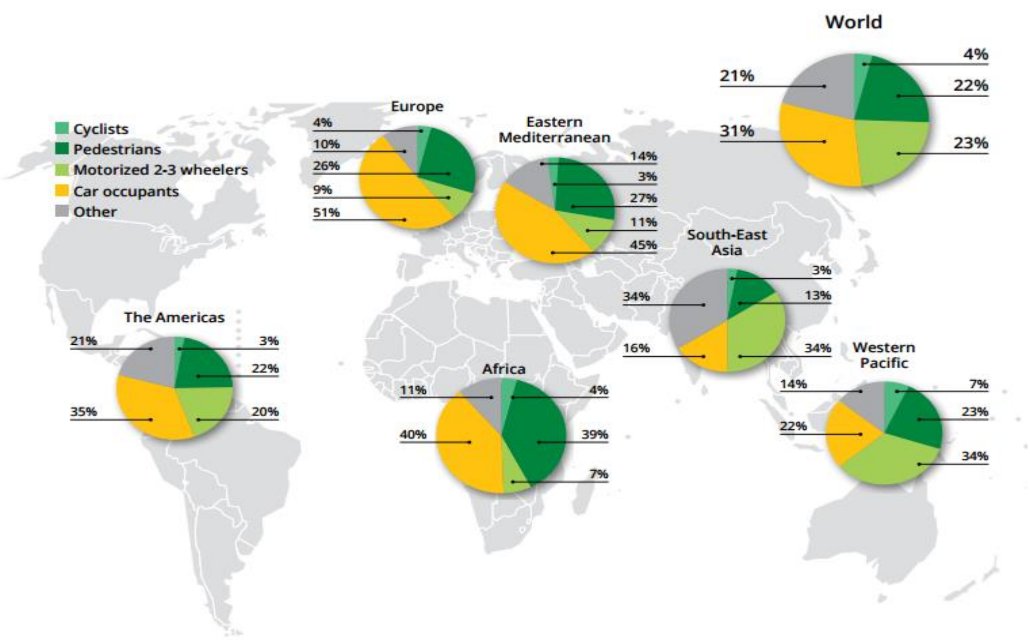
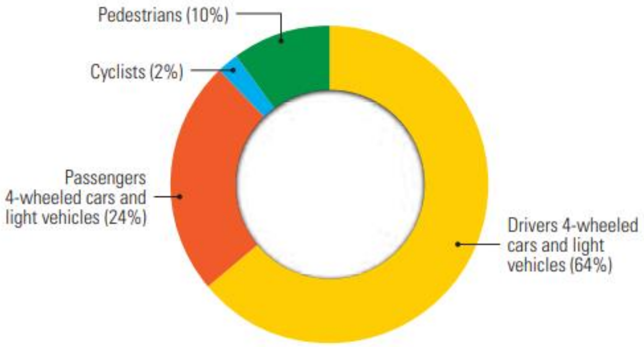


Figure 1.1 Percentages of fatalities and injuries on roads around the world (WHO, 2015)

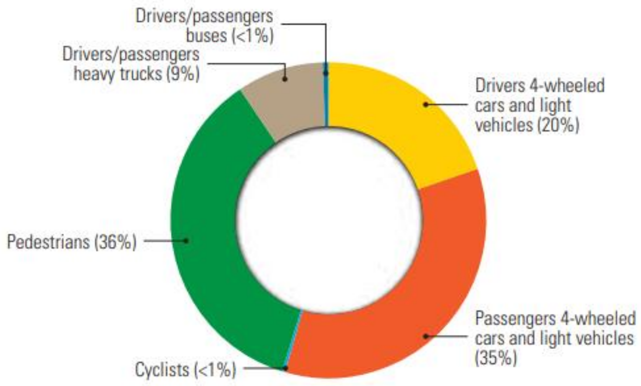
1.2 Road Safety in Low and Middle Income Countries (LMIC)

In LMICs, road safety is one of the main issues, which cause 1.13 million deaths and 45 million injuries for all road users every year (Perel et al, 2007). This is especially important for vehicle occupants' safety as this type of road users forms the highest percentage of road causalities. For example, in Argentina, the percentage of vehicle occupants' causalities (drivers and passengers) was 88% of the overall number of road accidents causalities in 2013, and in Armenia, they formed 55% for the same year as shown in Figure 1.2.

In January 2018, a study conducted by the World Bank showed that reducing road accidents fatalities and injuries could lead to substantial income gains, over the long-term, in LMICs. The study, also, found that countries with no investment in road safety could lose between 7 – 22% of their Gross Domestic Product (GDP) in 24 years period. This requires decision makers to prioritise investments in road safety (World Bank, 2018). These facts indicated that, there is an urgent need for immediate treatments to improve road safety, generally, and vehicle occupants' safety particularly.



Death by road user category in Argentina in 2013 (WHO, 2015)



Death by road user category in Armenia in 2013 (WHO, 2015)

Figure 1.2 Examples on the fatalities by road user in LMICs (WHO, 2015)

1.3 The Existing Situation

Road safety is an issue that is related to some factors such as educating people to behave more safely on the roadways, enforcement policies, which support more secure driving, and engineering of vehicles and roads as shown in Figure 1.3. Education means the transmission of the required skills and knowledge to those who lack these. In road safety, education has less to do with the individual potential development, implicated in wider use of the term "education," and typically refers to "driver education" and "public education" (Groeger, 2011). Although education and enforcement policies have an impact on road safety, this study focused on the civil engineering side (road infrastructure) as it plays an essential role in improving road safety (World Bank, 2010). For example, in the USA 53% of road fatalities and 38% of injuries are caused by accidents related to road infrastructure (Miller & Zaloshnja, 2009).

The infrastructure improvements can be carried out by implementing single or multiple countermeasures (two or more treatments). As multiple countermeasures are applied to improve road safety, there is a need to assess their combined effect on accident rates (Park, 2014). Many road infrastructure characteristics can have an effect on accident rates and vehicle occupants' safety. For example, practising road safety engineers in the international road assessment programme (iRAP) (a charity working to reduce road crashes) improved road

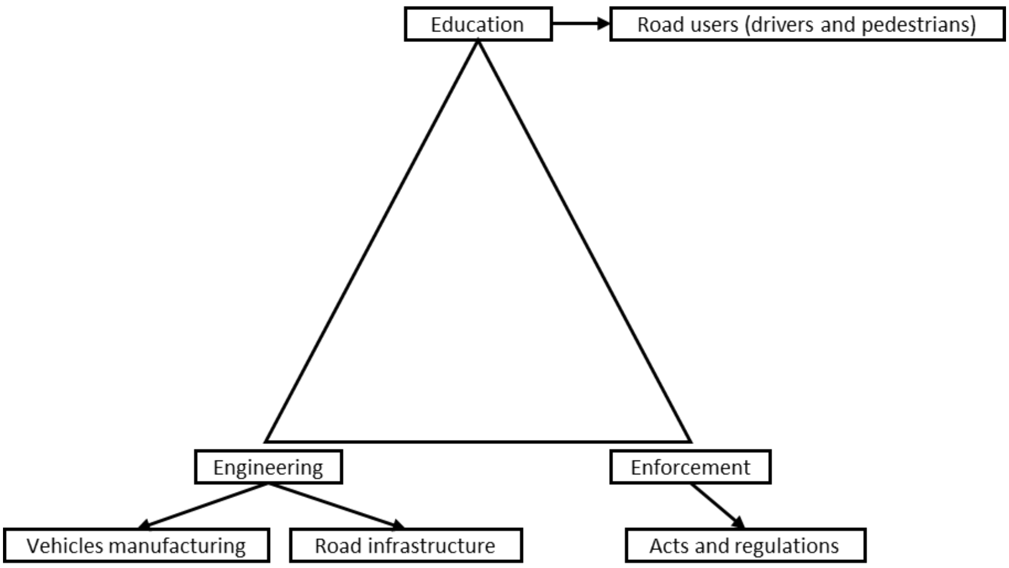


Figure 1.3 Road safety model and its components

safety level in Sao Paulo city (in Brazil) by reducing the risk associated with run-off crashes by providing or widening shoulders, and decreasing the severity of roadsides (removing or shielding fixed objects like trees or poles) (iRAP, 2014). Proposing the countermeasures, (i.e. the selection process of the appropriate sets of countermeasures) is one of the most essential steps to improve road safety. In addition, the selection from the possible array of countermeasures will affect the benefits that result from the implementation of those countermeasures, which may sometimes lead to overestimations in causalities reductions (iRAP, 2014).

The literature includes many previous studies, which are related to road attributes that affect road safety and crash reduction factors. Most of these studies focus on estimating the effect of road safety treatments by developing statistical models that consider crash modification factors CMFs as measures to reflect the effectiveness of these treatments (Abo-Qudais, 2001; Schneider et al, 2010; Daniello & Gabler, 2011; Nunn, 2011; Manan 2014). However, these models rely mainly on historical crash data, which have reliability issues in low and middle income countries (Phan, 2016), as explained in Chapter 2, and may suffer overestimations as in the iRAP model (iRAP, 2014).

The international road assessment programme (iRAP) is one of road safety assessment models that can deal with road safety in a cost-effective way in LMICs and focusing on improving road infrastructure. With regard to four road user categories, causalities may be divided into four categories: vehicle occupants, pedestrians, motorcyclists and cyclists’ fatalities. In the iRAP model, road safety is, ultimately, expressed by five star ratings from 1 star (least safe) to 5 star (most safe). These star ratings are based on star rating scores, which are calculated based on road geometrics and traffic conditions. After assessing the existing road safety condition, iRAP suggests countermeasures to improve the star rating, which, in turn, reduce crash rates and causalities numbers. These countermeasures may be used separately as a single countermeasure or as a mix of multiple countermeasures to reduce the risk of run-off, head-on, intersection, and property access crashes (iRAP, 2014).

It was felt necessary to identify the combinations of treatments with the highest effectiveness. This information could then be used to help estimate the expected road safety levels, as well as used in refining models for estimating the effectiveness of combined treatments. It is likely that practising road safety engineers may make incorrect decisions about which combinations of treatments to use when addressing road safety issues. Likely errors in this decision-making will lead to inefficient use of limited road safety funding (Turner, 2011).

To this end, this research developed a methodology and an automated framework for combining road safety treatments to reduce the number of three vehicle occupants (due to the available data) related crash types: head-on only, run-off only, and head-on together with run-off (non-intersected segment crashes). The outcomes of the developed models may be used to support practising road safety engineers in making decisions related to improve vehicle occupants' safety.

1.3.1 Problem Statement and Novelty

Road safety is a major concern worldwide and needs appropriate management. In the iRAP model, the concerned road safety programme in this study, road safety levels are classified into five levels using the star rating system from the lowest level (star rating 1) into the highest level (star rating 5). To improve road safety using the star rating approach in the iRAP model, a number of countermeasures can be used. Such countermeasures may be used individually or as a set of multiple countermeasures. Yet it may be difficult for practising road safety engineers to decide which combinations of countermeasures are the most effective to use when addressing road safety issues, especially with the availability of many treatments. Likely, errors in this decision-making will lead to inefficient use of limited road safety budgets. Therefore, the hypothesis of this study is that an innovative method based on processes of Knowledge Discovery in Databases (KDD) will provide an enhanced approach whereby road infrastructure attributes may be combined and ranked using a robust computational method. This automated data mining framework will also enable the extraction of the most effective combinations of countermeasures in terms of decision trees and associated rules. The model will use data from an iRAP database from Andhra Pradesh state in India.

1.4 Aim and Objectives

This study aimed to enhance vehicle occupants' safety by assessing the impact of road safety multiple countermeasures and providing an automated framework for combining and ranking sets of countermeasures to support the decision making process for improving road safety levels in high fatality traffic environments, such as those found in low and middle income countries (LMICs).

To achieve this aim, the below objectives were considered:

1. To develop a framework that could enhance iRAP model in the process of selecting road safety countermeasures to improve vehicle occupants' safety in low and middle income countries.
2. To examine the effect of combining road safety countermeasures, for three accident types, head-on, run-off, and head-on + run-off, on vehicle occupants' safety.
3. To develop road safety analysis models that rely on existing road infrastructure instead of historical crash data.
4. To identify the most effective sets of countermeasures combinations for the three concerned crash types.
5. To provide an automated framework to enhance the existing ViDA software used by the international road assessment programme (iRAP) for assessing road safety.

1.5 Thesis Structure

This thesis was structured in eleven Chapters as follows:

- Chapter 1 gives introduction about the importance and the urgent need to address road safety issues and the aim and objectives of this study, which may help in solving these issues.
- Chapter 2 presents the literature review around the subject of this research. It gives the review of two main approaches for modelling road safety analysis, the traditional statistical methods and the data mining modelling techniques. It also reviews the existing studies related to the use of these approaches to consider the effect of road infrastructure on road safety levels.
- Chapter 3 describes the methodology of this study that includes the considered modelling approaches (statistical and data mining). It presents the benefits of the approaches followed to determine the effect of road safety treatments combinations and the method used to select these combinations.
- Chapter 4 presents the engineering knowledge in the international road assessment programme (iRAP) and how the iRAP star rating system works. In addition, it presents the triggers that are used to examine the applicability of road safety treatments.
- Chapter 5 describes an application of the statistical regression approaches to refine equations for estimating the change in fatalities numbers because of the implementation of particular road safety treatments.

- Chapter 6 reviews the knowledge discovery approach. It gives an induction to the knowledge discovery and data mining approaches and the steps of this process that include data preparation, data mining, model building, model evaluation, and knowledge representation.
- Chapter 7 presents the knowledge discovery framework, which includes the database identification, the input data format, the data mining system used in this study, the model building process, and the model evaluation methods.
- Chapter 8 presents an application of the knowledge discovery and data mining techniques, described in Chapters 6 and 7, on the concerned database, which was offered by iRAP. It presents the developed data mining models for the three considered crash types and the most appropriate data mining technique for each type.
- Chapter 9 presents the knowledge extracted from the application of data mining on the iRAP database. This knowledge represents the extracted combinations of road safety treatments and the most appropriate sets for each crash type.
- Chapter 10 discusses the impact of the methodology followed in this study, which includes a comparison between the existing and developed models. It also presents the limitations of this study.
- Chapter 11 presents the conclusions and the recommendations of this study.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

This Chapter reviews the available literature on the models that deal with improving road safety by using single or multiple treatments that are associated with risk factors. These factors are numbers reflecting the effectiveness of road safety treatments. The Chapter is organized into four main sections. The first section presents two main modelling methods for road safety analysis. These are the statistical regression approach and the data mining approach. The second section reviews two main types of models that are used to assess and improve road safety. These are: 1) the crash frequency models, which are based on historical crash data, and 2) the star rating system model, which is based on existing data. The third section presents the factors that affect road safety. These include the human, vehicle, and infrastructure factors. Then, the fourth section, considers the effect of the traffic and infrastructure factors on road safety, as they are the concerned factors in this study.

2.2 Road Safety Analysis Approaches

Mainly, in assessing road safety, this study considers two main modelling methods for road safety analysis. The first approach is the traditional statistical modelling approach, which is the process of developing prediction models by using statistical regression methods such as the Negative Binomial and Poisson (Chengye and Ranjitkar, 2013). These models are developed based on historical accident records (Washington et al, 2010; Lord et al, 2005). The second approach is the data mining method, which is based on developing decision trees and tables from existing databases that include factors affecting road safety (Zhang and Fan, 2013).

2.2.1 The Traditional Statistical Approaches

These approaches evaluate road safety through risk factors (in some references are referred to as Crash Modification Factors CMFs). These factors represent the effect of implementing given road safety treatments at a specific site. To obtain these factors, a traditional statistical approach is used to build accident prediction models (statistical regression equations) to estimate the expected number of accidents on roads (HSM, 2009; Lee and Mannering, 2002). The statistical accident prediction models are mathematical equations that are used to correlate accidents frequency to road traffic and infrastructure conditions. They combine the crash modification factors and present them in the expression of mathematical formulas. Various prediction models for combining crash modification factors of traffic and geometric attributes have been developed to predict the combined effect of these attributes on crash frequency (Nunn, 2011).

For example, a study carried out by Manan (2014) estimated the change in the crash frequency because of the change in four traffic conditions: speed, passenger gender, driver behaviour, and helmet use. Daniello and Gabler (2011) developed a crash prediction model as a function of roadside object types. Another study by Abo-Qudais (2001) developed accidents prediction models as functions of the lane width, the number of lanes, the surface condition, the traffic volume, and the operating speed. They concluded that all these factors had an impact on the number of crashes. Moreover, they found that the number of lanes was the most effective factor on the crash frequency.

Schneider et al. (2010) developed a crash prediction model on rural roads as a function of some geometric factors such as the curve length, the curve radius, and the shoulder length. The data that were used in this study collected from rural two-lane highways in the state of Ohio for over six years from 2002 through the spring of 2008. The study showed that the considered factors significantly influenced the crash frequency.

2.2.1.1 The Limitations of the Statistical Approach

From the review of the available literature, it was clear that the statistical models had the attention of many previous studies in many regions (Wanvik, 2009; Haque, 2009; Harnen et al, 2006; Archer, 2004). These types of models focused on building statistical relationships between the crash frequency as the dependent variable and the traffic attributes as the explanatory variables. However, the key difficulty of such models, in low and middle income countries (LMICs), is the problem of obtaining reliable historical accident data that are needed in the process of developing these statistical models. Using unreliable crash records (because of underreporting road crashes and mismatch between police and hospital records (Kamaluddin et al, 2018)) to develop models for road safety analysis could result in inaccurate estimations for the reduction in crash numbers after the implementation of road safety treatments.

This may be related to some reasons. These are:

1. As road accidents happen randomly, there could be a fluctuation in the crash numbers on the same road section over time. In addition it is not easy to decide whether the change in the crash frequency is a result of the effectiveness of road safety treatments or by its fluctuation feature (HSM, 2009).
2. Because accidents are infrequent occasions, historical crash datasets need long periods of time to be considered sufficient for developing statistical models. Observations of 3 – 5 years may reflect the effect of crash modification factors of road safety treatments. Although there are new techniques, and mechanical and digital tools to collect data for long periods of time, there are some traffic and environmental characteristics, such as the traffic flow and the weather that may change over these periods of time. Therefore, it is difficult to correlate the change in such characteristics with the change in the crash numbers (Laureshyn, 2010).
3. There is interaction between different characteristics that may contribute in the accidents occurrence such as the contribution between the drivers' behaviour and some road geometric conditions, which may lead to crashes. For example, a hasty driver with no physical median may lead to a head-on accident (Garder, 2006; Kim et al, 2007;

Hosseinpour et al, 2014). Therefore, it is complicated to specify the contributing percentage of road safety treatments because the drivers' behaviour are not fully recorded in the accident reports. This, in turn, could lead to uncertainty in predicting the effect of the contribution of road safety treatments represented by the crash modification factors (HSM, 2009).

4. There is deficiency in reporting road accidents. This means that some crashes are not, or not accurately, recorded. Especially, those with property damages only or slight injuries. This could lead to unreliable estimations for the number of crashes and the causes behind them (Al-Dah, 2010).

As may be seen from the above-listed limitations, the use of historical accident records for building statistical models for road safety analysis has some issues that could affect its reliability. Therefore, this study concerns the use of an alternative method, i.e. data mining models to extract the useful knowledge from the databases that include road geometric and traffic conditions with their road safety levels. This knowledge could help to estimate road safety levels and relate them to the road traffic and geometric characteristics. This is because these models are based on the current road conditions rather than the dependence on historical crash data (Chang and Chen, 2005; Beshah and Hill, 2010; Zhang and Fan, 2013; Kwon et al, 2015; Mahajan, 2016).

2.2.2 The Data Mining Approaches

Data mining refers to extracting knowledge from a considerable amount of data. The most common data mining approaches that have been used in road safety aspects are decision trees and tables (Zhang and Fan, 2013). The decision trees and tables were initially presented by Ross Quinlan, a researcher in data mining, in the late seventies and early eighties. Thereafter, this has been expanded to the idea of data mining (Han et al, 2012).

Here, the decision trees and tables are methods to analyse road safety based on collecting the existing road traffic and geometric conditions without the need for historical accident records (Chang and Chen, 2005; Chang and Wang, 2006). These techniques are based on exploring the contribution of road traffic and geometric attributes to estimate the potential road safety level (Zhang and Fan, 2013).

Few studies considered the use of data mining in road safety. For example, a research was carried out by Chang and Chen (2005) to establish empirical relationships between road crashes and highway traffic, geometric, and environmental variables. They concluded that, the Classification and Regression Tree (CART), a data mining tree-based technique, is a good alternative for analysing freeway accidents frequency in comparison with the traditional statistical models. Zhang and Fan (2013) conducted a study that considered the effect of some major factors on traffic collisions (age, season, and gender) in Saskatchewan City, Canada. They found that the data mining technique produced decision trees with good accuracy (74 – 83%) to relate traffic collisions with these major factors.

2.3 Road Safety Assessment Models

2.3.1 Crash Frequency Models

As previously discussed in section 2.2, various crash frequency models were developed to consider road safety. These models have used risk factors to assess road safety with regard to the difference in the crash frequency before and after road safety treatments. The crash frequency models provide tools for evaluating the crash reduction because of the effectiveness of selected countermeasures. This is by identifying the road sections with the highest crash frequencies and determining the crash modification factors of the attributes that could affect the accidents frequency as potential treatments (HSM, 2009). To identify the road sections with the highest crash frequency, these models estimate the expected accident numbers for a road network by using prediction models. Each model reflects the impact of the independent variables that affect the crash frequency, which are subject to specific traffic and geometric site conditions (Schneider et al. 2010).

2.3.2 The International Road Assessment Programme

Another model, the international Road Assessment Programme (iRAP), considers risk factors to assess road safety with regard to a star rating system where the first star level represents the riskiest road sections and the five star level represents the safest road sections.

The international Road Assessment Programme (iRAP) is a comprehensive road safety model, which accounts for many cost-effective factors that affect road safety. These star ratings are based on star rating scores (SRS) which are calculated by an appropriate model that considers the risk factors of road attributes, traffic volumes and operating speeds. The outcome of this process are countermeasures that can be used to improve road safety and reduce fatalities and serious injuries (FSI) numbers. These countermeasures can be used separately as a single countermeasure or as a combination of multiple countermeasures for a particular location (iRAP, 2014).

The iRAP model could be useful as a proactive method that could be used, in situations where reliable accident records are not available, to reflect the effect of changing road attributes on road safety as will be described in Chapter 4. However, it should not be considered as a substitution to crash frequency models. There are some limitations related to this method. For example, the risk factors in iRAP were estimated based on studies carried out in high-income countries and adapted for application in LMICs. In addition, limited validation has been carried out in the traffic conditions of LMICs (Phan, 2016). It may be indicated, by comparison between crash frequency models and risk models, that the crash frequency models predict the change in crash numbers as a result of changing road traffic and infrastructure conditions. Whereas, the risk models expect the change in road safety levels (star ratings) as a result of the change in these conditions.

2.4 Factors Affecting Road Safety

There are some factors that affect road safety. As mentioned before, in Chapter 1, three main factors can have major impacts on road safety: (i) educating people, (ii) enforcing laws, and (iii) engineering of roads and vehicles (Charlton, 2011). In the context of road safety, education has a main role to transfer a knowledge body of information and skills to people who are already road users (Charlton, 2011). Enforcement is also a main factor that directly affect the number and severity of crashes. For example, speed limits could affect the severity of crashes as less severe can happen at lower speeds. The applicability of enforcement is related to the punishment certainty. If the punishment certainty is high, undesirable behaviours are less likely to be occurred (Li, et al., 2011).

The human factors, which are related to education and enforcement issues, and the vehicle factors, which are related to engineering and manufacturing issues, have an important effect on road safety and the number of crashes. However, this research will focus only on the effect of road infrastructures that can improve road safety and reduce accident rates because they have major impacts on road safety as explained in the sections below.

2.5 The Importance of Infrastructure on Road Safety

Infrastructure plays a crucial role in road safety. Well-designed roads can help people use roads safely and minimize the risk that a crash will occur. When a crash does happen, protective road infrastructure can mean the difference between life and death (World Bank, 2010). Although road accidents are mainly caused by human factors, the best-untapped possible solutions to prevent causalities on roads are through roads themselves (World Bank, 2010). Therefore, it is important to study the effect of infrastructure attributes on road safety (the number of fatalities and serious injuries), to treat the causes and reduce crash rates.

Below are the key infrastructure attributes that affect the non-intersected segment crashes according to the iRAP model (iRAP, 2014).

2.5.1 Lane width

Elvik and Vaa (2004) recommend accident rates decrease with increasing the lane width on rural roads, however, the rates may marginally increase on urban roads. Expanding lane widths, inside a standard design extent, may reduce crashes by up to 10%. For example expanding lanes by 0.5 m at intersections is anticipated to diminish crashes by 4-6% (Harnen et al., 2003). Lane widths of 3.2 m or more are anticipated to have 34% less crashes than narrower lanes as shown in Figure 2.1. However, the effect of lane width on road safety is less in low and middle income countries (iRAP, 2014).

Conversely, an examination of road asset and accident reports from Queensland, Australia, demonstrated that lane width has no connection with fatality and serious injuries crash results on

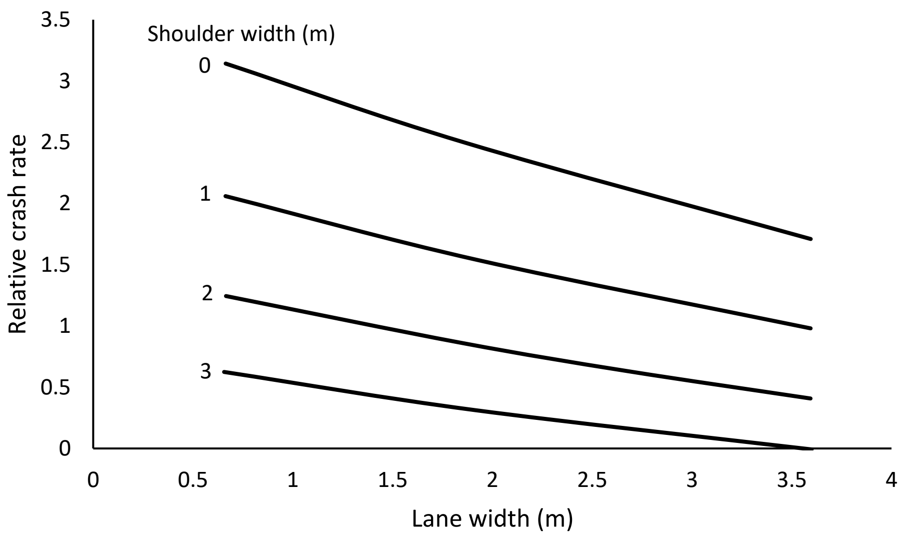


Figure 2.1 Effect of lane width on relative crash rates, adapted from (Zegeer, 1987)

two lane urban arterials with 60 km/h speed. There was, however, a connection for every other sort of urban and rural arterial roads with a speed of 80 km/h or more (Turner et al., 2009).

2.5.2 Curvature

In general, crash studies demonstrated that curves experience higher accident rates than straight roads areas, as shown in Figure 2.2, with rates fluctuating from 1.5 to 4 times more than the non-curve segments (Zegeer et al., 1992).

Wright and Robertson (cited in Mak and Sicking, 2003) investigated 300 fatal crashes of vehicles with fixed objects in Georgia trying to focus on crash rates at curves and on gradients by contrasting the features of the accidents' sites with control sections at distance of one mile upstream of the accident locations. They found that 70% of the crashes happened on curves. They also indicated that, downhill gradients of 2% or more were found to have some impact.

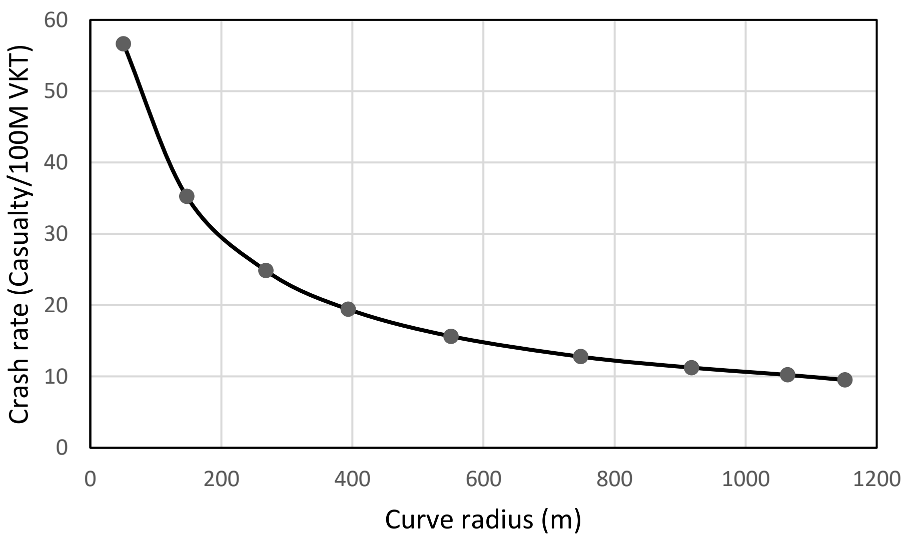


Figure 2.2 Crash rate by horizontal curvature radius, adapted from (Turner et al, 2010)

2.5.3 Quality of Curve

Turner et al. (2009) recommended researchers to study the effect of delineation on the quality of curve to the extent that delineation tools like signs and markings can advise the drivers to select the correct curvature.

Warning signs at curves can reduce accidents by (10 – 30 %), in addition, direction signs and painted guardrails at curves can decline crashes by (20 – 40 %). The curves that are close to each other, fewer than 0.5 curves per km, can be three times, more, potential sites for crashes in comparison with less frequent curves (Elvik and Vaa, 2004).

2.5.4 Delineation

It is assumed that if signs and markings were used in a right and standard way, they can reduce the number of crashes by 20%. A reduction of 30% and 35% is suggested by two studies, which refer to the reduction that can be achieved by using curve-warning signs in the appropriate way (iRAP, 2014). A more robust study conducted by Winnett and wheeler (2002) on the effect of using vehicle-activated signs showed that these signs may reduce crashes by 34% (Turner et al. 2009).

However, Sarin and Mital (1991), cited in (Chakrabarty et al., 2013) carried out a study on vehicles drivers in Delhi which showed that about 90% of the drivers ignore road signs, and traffic regulations and rules that control road users for safety. The study suggested that road signs and markings are less effective in developing countries because they are not selfenforcing. Aghabayk et al. (2012) conducted a study that revealed that the drivers' behaviour in developing countries, such as breaching traffic signals and stop signs, is due to poor attitudes towards road safety matters rather than lack in awareness or knowledge.

2.5.5 Roadside Object

A study by Elvik and Vaa (2004) showed that increasing the distance between roadway lanes and roadside objects from one meter to five meters can reduce injury accident rates by 22% and increasing this distance from five to nine meters could decline injuries in road crashes by 44%. They suggested that side slope may reduce the number of crashes by 42% if it was flatted from 1:3 to 1:4 and by 22% for flatting from 1:4 to 1:6.

2.5.6 Paved Shoulder

Turner et al. (2009) discussed the effect of paved shoulders on crash rates on undivided rural roads in Queensland, Australia with speed limits 100 km/hr. The study showed that there is a relationship between the width of paved shoulders and accident rates. Crash rates on sections with 0.5 m paved shoulder are 25% higher than crash rates on sections with 2.5 m width of shoulders. Figure 2.3 shows the relationship between the crash frequency and the sealed shoulder width.

Tuner et al. (2009) reviewed another study in Australia which concluded that road sections with 1.5 – 2.5 m paved shoulder had lowest crash rates in comparison with sections of less paved shoulder width. However, the crash frequency was higher in sections that have more than 3.0 m paved shoulder as some drivers may consider them as additional lanes.

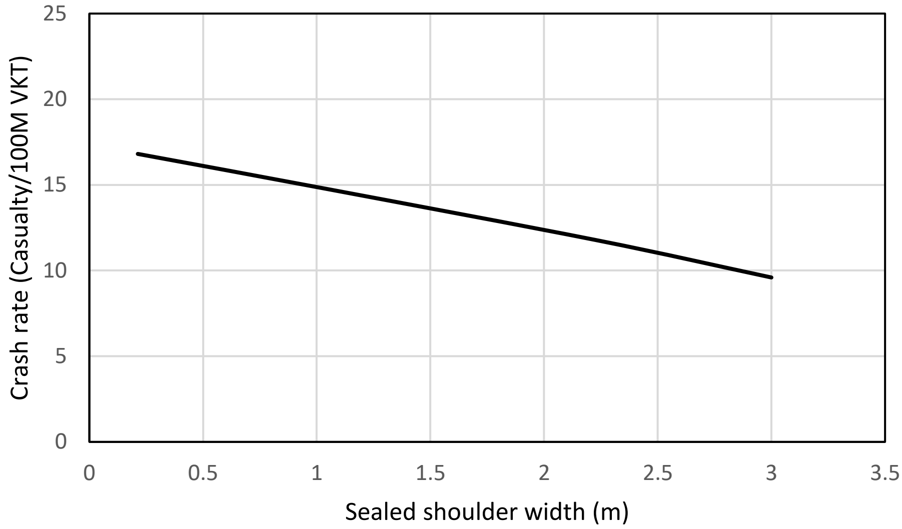


Figure 2.3 Effect of sealed shoulder width on crash rates on rural roads, adapted from (Jurewicz and Bennett, 2009 cited in Turner et al, 2009)

2.5.7 Number of Lanes

Adding travel lanes normally aims to provide more capacity. Although Some researchers argue that crash rates increase when the number of lanes increases, planners and engineers generally expect that reducing traffic congestion should improve road safety (Kononov et al, 2012). From the Highway Capacity Manual (HCM, 2000), it is known that increasing the number of lanes is associated with increasing the capacity because of higher free flow speed. A study by Council and Stewart (1999) stated a 40% to 60% reduction in accident rates because of converting two-lane sections into four lanes for the same traffic.

2.5.8 Median Type

Changing the median type could be a useful solution to reduce accident rates, especially for head-on crashes. The likelihood of severe accidents occurring is affected by the way that opposing flows are separated. The centreline alone as a separation between the two directions is the worst case. The movement of vehicles across median may be restricted by median barriers to reduce the chances for head-on crashes, as the errant vehicles' possibility to reach the opposite direction will be less (Elvik and Vaa, 2009). Also, the potential severity of an accident can be reduced by increasing the median width, whereas safety barriers may entirely remove that risk (iRAP, 2014).

2.5.9 Skid Resistance

Crash rates increase on roads that have low skid resistance, and unsealed roads have accidents frequency three times more than that of roads that are sealed adequately (iRAP, 2014). Turner and others found that a crash reduction factor of 35% for low skid resistance is recommended for urban environments, and this was similar to studies from the Bureau of Transport Economics (BTE) and Davies et al. (1997) (Turner et al, 2010).

2.5.10 Gradient

The gradient (downhill or uphill) has also a negative impact on the crash rates. A gradient of 7.5% - 10% could increase the number of crashes by 20% (Harwood, 2000). However, accident rates showed only slight raises for gradients of up to 6% (Choueiri et al, 1994). This was confirmed in a study by Elvik and Vaa (2004).

2.5.11 Differential speed limits

The differential speed limits (DSL) is a strategy that requires different speed limits for passenger cars and trucks. The effect of DSL on road safety has been investigated and some studies found that there is no difference after the introduction of a DSL. Other studies found that either the DSL or the uniform speed limit (USL) is better (FHWA, 2004). However, it seems more likely that using a DSL has negative effects on road safety as a result of the increase in the number of the car-truck overtake maneuvers, which increase the occurring possibility of head-on crashes (Ghods et al, 2012).

2.6 Summary

After reviewing the available literature, it was found that many studies have been carried out focusing on road safety. The majority of these studies developed traditional statistical prediction models to expect the average crash frequency for road networks. As stated in section 2.2.1.1, some limitations affect the certainty of these traditional models that are mainly related to the reliability of the historical accident records.

The literature also revealed that the data mining techniques have been tested and validated as a good alternative for considering road safety aspects. Few studies have been conducted focusing on the use of data mining techniques, represented by decision trees and tables, in correlating road traffic, geometric, and environmental factors to road safety. From that, it may be suggested that the data mining techniques may be considered to assess road safety as an alternative method to overcome the explained problems of the statistical models, which associated with the reliability of crash records (Beshah and Hill, 2010; Pakgohar et al, 2011). A study was conducted by Chang and Chen (2005) in Taiwan, to compare between using data mining and the Negative Binomial regression approach in assessing road safety, revealed that the data mining techniques resulted in higher accuracy than the Negative Binomial method for expecting the crash frequency.

It was also indicated, from the literature, that the previous studies focused on developing crash frequency models, using statistical methods, rather than the star rating system models. It may therefore be suggested that there is a lack in conducting research on assessing road safety using the star rating system to reflect the effect of changing road attributes on road safety in case of lack in reliable historical crash records. Therefore, there is a need to use the data mining as an alternative approach to overcome the limitations of the statistical methods and to develop a new methodology that may consider the use of the star rating system in the subject of road safety assessment.

CHAPTER 3

METHODOLOGY

3.1 Introduction

With limited resources and the responsibility to ensure public accountability in the use of budgets available for road safety programmes, it is important to provide decision makers with additional information to identify the countermeasures that could result in the greatest reductions of accidents and can be covered by the available fund. Thus, selecting the most effective set of countermeasures is a very important phase in the process of improving road safety (Preusser et al, 2008). Until this study is carried out, practising road safety engineers may have uncertainty in making decisions related to which sets of multiple countermeasures to be used to address road safety improvements.

Likely, errors in the decision making will lead to inefficient use of limited road safety funding (Turner, 2011) in addition to the possibility of the overestimation that can occur in the fatalities reduction rates which makes the real benefits less than the expected (David *et al*, 2008).

The outcome of iRAP methodology are the countermeasures that can be used to improve road safety and reduce fatalities and serious injuries (FSI) numbers. Each of the countermeasures can be used separately as a single countermeasure or together with others as a combination of multiple countermeasures for a location (iRAP, 2014).

The selection process of the appropriate countermeasures is one of the most effective steps to improve road safety. In addition, the selection from the possible array of countermeasures will affect the benefits that result from the implementation of those countermeasures such as reducing the number of fatalities, injuries, and property damages. However, there are some difficulties that can be overcome in the process of selecting countermeasures combinations.

3.2 Overestimation in the Current Crash Reduction Rates

The other issue relating to the use of multiple countermeasures is the overestimation in crash reductions (iRAP, 2014). There is general agreement that the adequacy of a countermeasure is liable to consistent losses when it is utilized as a part of mix with different countermeasures. That is, accident diminishment appraisals of individual countermeasures cannot be easily added together (Turner, 2011). The iRAP model philosophy makes utilization of a multiple countermeasure correction factor (MCCF) to predict diminishments in the effectiveness of road safety treatments when more than one countermeasure is chosen. However, the key restriction of this methodology incorporates a presumable overestimation in the predicted crash reduction (iRAP, 2014).

Therefore, one of the objectives of this research is to reduce this overestimation by developing models to calculate more accurately the effect of using multiple countermeasures to adjust the correction factor. Thus, this research could help in providing an added methodology for the iRAP model, which in turn could result in more effective road safety plans.

3.3 The Difficulties in the Current Countermeasures Selection Process

The iRAP is an assessment model that evaluates the existing road situations according to the star rating from 1 to 5 stars. After that, the model suggests countermeasures for each section or road. There are 94 countermeasures in the iRAP model that can be used to improve road safety. For each countermeasure, one or more sets of triggers has been defined. Any countermeasure cannot be considered unless all the triggers for this countermeasure are satisfied. For example, considering the “improve delineation” as a countermeasure depends on a trigger that the existing delineation situation is coded as “poor” in the iRAP coding (surveying) process (iRAP, 2014).

It is a challenge for practitioners to compare between multiple countermeasure combinations, especially, if there are many sets that can be applied for a section or an entire road. For example, if there are seven countermeasures that their triggers and specifications are satisfied with iRAP standards and can be considered to improve the star rating for a section or a road, the decision to choose from these countermeasures can be difficult as there are many options to choose for example two or three from the seven available. That means, practising engineers in iRAP may need to try every available set of treatments to find out and compare between the star rating, cost and benefit of each set, which consumes time and cost. Therefore, the main objective of this research is to develop a new methodology to compare between combinations of countermeasures according to their effectiveness, to help decision makers in the processes of optimization and selection of multiple countermeasures options.

3.4 Methodology Description

The methodology of this research was divided into three main parts as shown in Figure 3.1. The first part focuses on the current situation in iRAP methodology for the effectiveness of multiple countermeasures and their selection process to point out the areas that can be enhanced. As mentioned earlier in 3.2 and 3.3, the countermeasures selection process and the overestimation are multiple countermeasures issues that can be improved to obtain more accurate results.

In the second part, this study tries to present an enhancing methodology to the current situation. Although there are few models in the literature, as indicated in Chapter 2, that dealing with the effectiveness of multiple countermeasures, these models were built depending on data from jurisdictions where they developed and not from iRAP, which has special collecting and classifying data techniques. In addition, the available multiple countermeasure related models are not considering the star rating systems, which is the base system in iRAP model. Therefore, in this research, knowledge extraction techniques are needed to develop models by using datasets from iRAP itself to ensure that these models are compatible with iRAP data coding specifications and can consider the star rating system.

For the same reasons as above, statistical methods are used to address the overestimation in the fatalities reduction rates. The advantage of using the statistical methods is to develop fatalities estimation models from iRAP datasets as they are collected, classified and grouped according to iRAP standards. The statistical models could result in more accurate predicted fatalities numbers for each road or section as they are built by using data from the roads themselves. Although the process of developing a model for each road or section may consume more time, it could produce more accurate fatalities estimations by adjusting the multiple countermeasures correction factor (MCCF).

Finally, the third part discusses the validation process of the developed approaches. For the first issue, selecting the countermeasures, the knowledge extraction models are tested with the iRAP model to check the validity of the resulted countermeasures combinations. For the second issue, the overestimation, the statistical models that are built from iRAP datasets are tested with real data to see if these models can give fatality rates predictions closer to the real data than the equation that is used in iRAP. These testing processes will be discussed further in this Chapter.

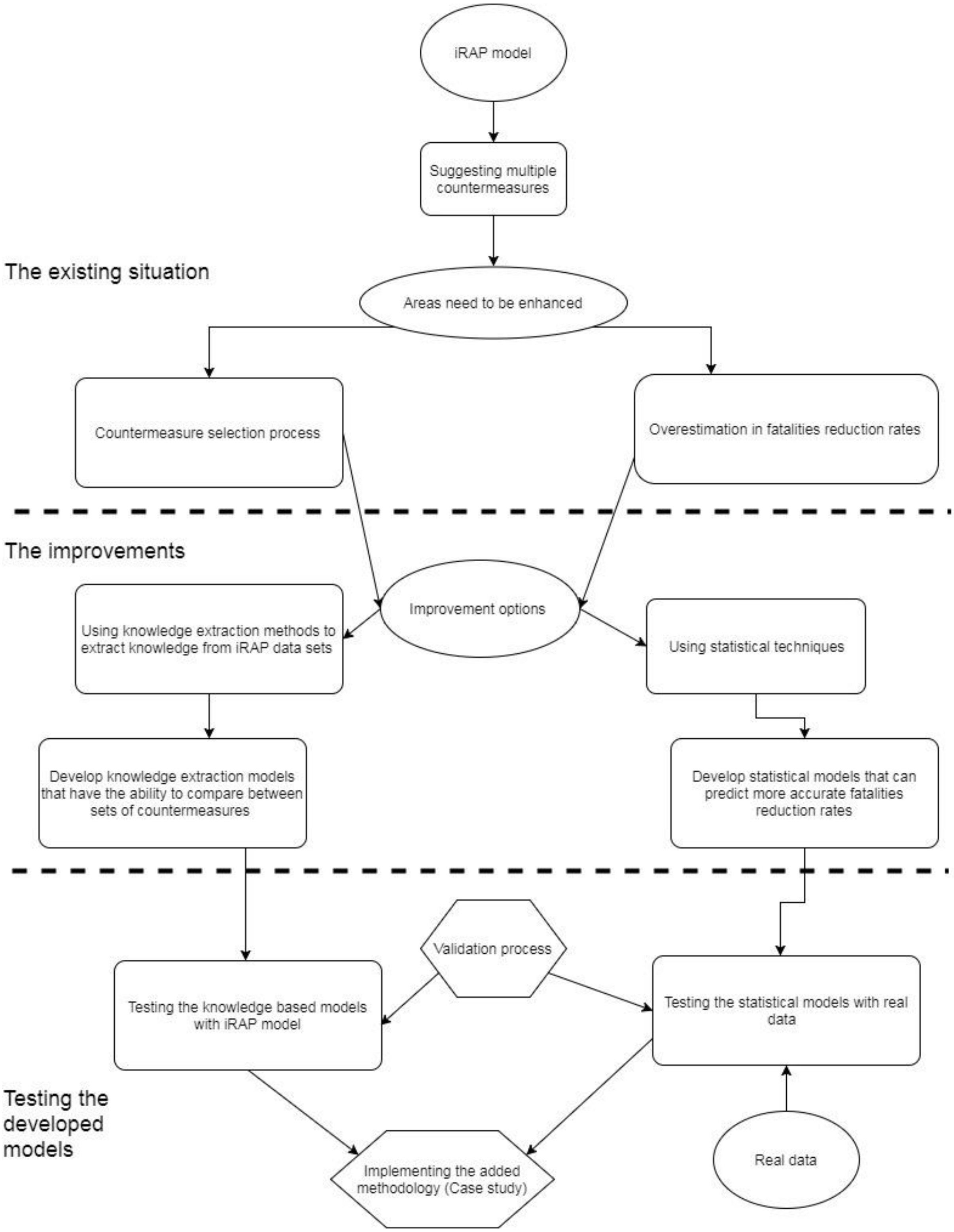


Figure 3.1. Methodology flow chart

3.5 Improvements for the Current iRAP Methodology

3.5.1 Improvements for the Multiple Countermeasure Correction Factor (MCCF)

3.5.1.1 Risk Factors

Crash modification factors (CMF), sometimes called risk factors, are used to relate road attributes and crash rates as follows: crash modification factors (CMFs) are multiplicative factors that are used to calculate the expected number of crashes after applying a given countermeasure at a specific section. For example, an intersection is experiencing 200 angle crashes per year. If a countermeasure that has a CMF of 0.7 is applied for angle crashes, then these crashes may be reduced to 140 per year following the implementation of the countermeasure (200 x 0.70 = 140) (iRAP, 2014).

When more than one countermeasures are to be applied at a particular location, their combined effect is based on a multiplicative formula as presented in Chapter 5. The weakness in this methodology is that the MCCF can be subjected to some issues like:

1. Issue of Independence

The major issue is the possibility that the safety effects of the treatments may not be independent from each other. That means, the impact of the simultaneous or sequential application of treatments could not simply be the outcome of the CMFs for the individual treatments. Because of correlations among the CMFs, the real combined effect of multiple countermeasures may be greater than, equal, or less than the simple outcome. For example, a basic case in which there are two treatments represented by CMF1 and CMF2. The combined impact is the actual effect of the combination of these two individual treatments (CMF12) (Gross and Hamidi, 2011). As described in the HSM (2000), the current method is to simply multiply the two CMFs, if they are assumed to be independent. This assumption may lead to three different outcomes as illustrated in Equations 3-1, 3-2, and 3-3. The real safety impact of the combination of treatments may be overestimated (Equation 3-1), underestimated (Equation 3-2), or accurately estimated (Equation 3-3) (Gross and Hamidi, 2011).

$$CMF12 > CMF1 * CMF2 \qquad \text{(Eq. 3-1)}$$

$$CMF12 < CMF1 * CMF2 \qquad \text{(Eq. 3-2)}$$

$$CMF12 = CMF1 * CMF2 \qquad \text{(Eq. 3-3)}$$

2. Issue of applicability

CMFs may be associated with total accidents or particular accident types and/or severities. The accident type and severity related to a CMF defines the accidents for which the related treatments is targeted. It could not be appropriate to apply a CMF of a particular accident type or severity to other accident types and severities because a treatment may reduce particular accident types or severities while increasing other accident types and severities. Even if multiple countermeasures are independent, the respective CMFs may be associated with different accident types and/or severities. If this is the case, the CMF for a certain accident type and/or severity must only be used to the accidents expected for that accidents type and/or severity. This indicates that, practising road safety engineers should consider accident types and severities when combining the impact of CMFs because simply multiplying the CMFs together without doing so would likely lead to incorrect results (Gross and Hamidi, 2011).

To this end and to improve the accuracy of the expected reduction in fatalities numbers, this research seeks to develop a new methodology to adjust the multiple countermeasures modification factor (MCCF) by developing fatality rates estimation models using a data-driven methodology that will explore the relationship between fatality rates and the countermeasures implemented.

3.5.2 Improvements for the Countermeasures Selection Process

Countermeasures are selected by practising road engineers to improve road safety. The decision of choosing a single countermeasure from the available array for a specific section or entire road depends on the engineering knowledge, the triggers for this countermeasure, its applicability, and the need for it (iRAP, 2014).

Engineers in iRAP might use some countermeasures to improve the star rating into a higher level without knowing what the other available combinations of countermeasures are. This could be one of the reasons behind the overestimation in the expected crash reduction rates because the selected treatments could not be the most effective. This could result in higher numbers of fatalities than the expected after the application of these countermeasures (iRAP, 2014). Therefore, there is a need to have data-driven models such as decision making tools to help in the decision making process in case of comparison between multiple countermeasures as these models use data from roads themselves to reflect the real effect of road attributes on road safety for the roads under study. This is the reason behind using decision trees as a decision making tool in this research, which, to the best of the researcher knowledge, has not been utilised in the domain of multiple treatments comparisons for road safety issues related to vehicle occupants.

3.5.2.1 Benefits of Data Mining and Decision Trees

Knowledge extraction from databases (from iRAP in this study), also called data mining, is recognized as one of the fast growing areas in the field of computing (Holsheimer *et al.*, 1991; Pujari, 2001; Hegland, 2001; Ali et al, 2014). The reasons for it are numerous and can be summarized as follows (Canarelli, 1996):

- Because of the rapidly decreasing cost of storing and processing information, it has become easier and more convenient to build up databases.
- There is growing pressure toward exploiting the knowledge included in these databases, in particular for finding a business competitive advantage.
- There is a growing recognition that traditional statistical techniques are not able to extract as much useful information as possible, which is partly due to the recent awareness about systems' complexity.
- There is an increasing availability of both free/cheap computing power and improved algorithms related, in particular, to machine learning.

“Decision trees are a simple, but powerful form of multiple variable analysis. They provide unique capabilities to supplement, complement, and substitute for” (SAS, 2017):

- Traditional statistical models (such as Poisson regression).
- Other data mining techniques (e.g. neural networks).

The base element in building decision trees is the datasets that can be used to extract the useful knowledge. According to that and to all the above advantages, it was suggested that the iRAP datasets from Andhra Pradesh state in India, which include the road attributes and their star ratings could be used as input data for the data mining techniques to extract useful knowledge. The extracted knowledge could help in the decision making process by identifying the most accurate countermeasure combinations for the concerned roads as will be explained in Chapter 9.

3.5.2.2 Incorporation of Decision Trees in iRAP

Decision trees are used widely and could be useful for the purpose of improving iRAP model and providing a smooth way for comparison between road safety treatments combinations. To achieve this goal, in this study, iRAP datasets are used to build decision trees for all the countermeasures that are related to run-off and head-on crashes in iRAP. The outcome of these decision trees is the star rating for each combination of treatments that can be applied for a section or a road in the iRAP projects. This facility can provide a smooth way for practitioners to compare between sets of countermeasures according to the resulted star ratings, after checking their applicability and compatibility with the iRAP standards and specifications.

In addition to the decision trees, there are the decision rules (sets of countermeasures) that can be extracted from the trees. The benefit of these rules is that, they can be ranked by knowledge extraction tools according to the frequency of the same star rating for each set. That means, for example the set results in star rating three (SR3) for a hundred sections is more accurate than the set gives the same star rating for fifty sections. This is another benefit of using decision trees and rules in iRAP where combinations of countermeasures can be ranked for practitioners, according to their accuracy, so they can start checking the set that is ranked number one and then number two, if the first one is not applicable, descendingly.

3.5.2.3 The Process of Discovering Knowledge

Databases stored in iRAP contains all the traffic and geometric information for each 100 meter segment of roads. All road attributes, in these datasets, are coded according to the iRAP coding classifications. For example, lane width between (>0 - <2.7m) is classified as narrow and between (≥2.7 - <3.25) and (≥3.25) are classified as medium and wide respectively.

According to the above coding process, each attribute is categorized with a feature number that explains its description. For the above example, narrow, medium, and wide lane width are classified as features 1, 2, and 3 respectively. In addition, the datasets contain the star rating for each section, which is the result of all the traffic and the geometric characteristic for that section. These datasets can be useful if converted into engineering knowledge to support the decision making process about the most effective sets of countermeasures to improve road safety.

Data mining algorithms and tools could be found in the field of knowledge discovery and data mining. The term Knowledge Discovery in Databases (KDD) refers to the broad process of extracting knowledge from data, and emphasizes the application of specific data mining approaches. It is of interest to researchers in pattern recognition, machine learning, databases, artificial intelligence, statistics, knowledge provision for expert systems, and data visualization (Han *et al*., 2012).

The benefit of data mining is to develop computerized-based models that can support human judgments. They are statistics based techniques but differ from statistical models in that they have the ability to interpret non-numerical (nominal) data into structured and hierarchical models. Data miming systems use several well established tools and techniques, such as statistical approaches, neural networks, genetic algorithms, k-nearest neighbor algorithm, various techniques of using domain knowledge, Bayesian networks, case-based reasoning, and so on (Chang, 2001).

In this research, some data mining algorithms will be used to help in building the decision trees and rules that can be extracted from the iRAP datasets and used for comparison between road safety treatments.

3.6 Validation of the Statistical and Knowledge extraction Models

3.6.1 Validation of the Statistical Models

This section presents the validation procedure for the fatality rates estimation models. The testing method evaluates the accuracy of fatality rates predicted from the developed statistical models by plotting the differences between the observed and predicted values.

Validation of the models against other sections of the roads under study for the same years of accident data will be used to assess the models' ability to estimate fatality rates across the road. The statistical models will be developed by using data for three years (2010 – 2012). Fatality estimations will be calculated by using the models for the same period. Then, these numbers will be compared with the observed fatalities for these three years to assess the models' ability to expect fatality rates across other sections which have different road attributes but with the same human and vehicles factors on the roads.

3.6.2 Validation of the Knowledge Extraction Models

The knowledge extraction techniques will be used to develop added models to the iRAP countermeasures selection process. As these models are used to calculate the star rating and suggest the countermeasures that could improve it, it is important to check the validity of these models with iRAP. This testing process will be carried out by applying the countermeasures that are suggested by the knowledge extraction models to iRAP to see if the star rating from these models is the same as what resulted from iRAP.

Comparison between the star ratings of the suggested combinations and that determined by iRAP will be undertaken to filter the valid combinations.

3.7 Scope and Contents of iRAP Database

As iRAP is the concerned road safety programme in this study, iRAP database is chosen to develop the statistical and the data mining models in this study. Two sets of data from Andhra Pradesh state in India for a road network of about 3137 km length are used in the study: Multi project - Core Data, and Multi project – Fatality Rates. These datasets were provided by the international Road Assessment Programme (iRAP). The first set of data (Multi project - Core Data) provided the geometric characteristics of the roadway (e.g. number of lanes, median type, gradient, skid resistance, and differential speed limits). In addition, it covered the traffic conditions in terms of (operating speeds and traffic volumes). These data were collected and coded by iRAP using survey vehicles and a specific coding software as explained in section (3.5.2.3).

The second dataset (Multi project - Fatality Rates) provided the average number of fatalities for each homogenous section and over a three-year period (2010 – 2012). The number of fatalities for each homogenous section was divided on the section's length to determine the fatality rate per kilometer for analysis purposes.

Where available, existing records of the number of road fatalities occurring per year was used. However, detailed road crash data was not available on many of the road sections and therefore on these roads, estimates of the number of fatalities were made by iRAP based on typical fatality rates per billion vehicle kilometers travelled (vkt) (iRAP, 2014). An extracts from the Multi project - Core Data dataset and the Multi project – Fatality Rates dataset are presented in Appendix A.

3.8 Summary

According to the above issues related to the accuracy of the expected reduction in fatalities numbers, this research develops a statistical model to adjust the multiple countermeasures modification factor (MCCF). Adjusting this factor, with regard to real fatality data, for each road or section may lead to more accurate fatality estimations. In order to develop a new methodology for adjusting MCCF, statistical techniques are used to build models from the existing iRAP datasets as will be explained in Chapter 5. With regard to the main issue, combining and ranking road safety countermeasures, and to identify the most effective sets of multiple road safety treatments, this research develops data mining models as an appropriate automated framework to address, more accurately, the effectiveness of multiple road safety treatments, as will be shown in Chapter 8. The novelty of this research is to use these data mining models to combine and rank road safety countermeasures in the iRAP model as a new methodology that could help practising road safety engineers to address road safety issues.

CHAPTER 4

ENGINEERING KNOWLEDGE IN THE IRAP MODEL

4.1 Introduction

This Chapter outlines part of the engineering knowledge in the international Road Assessment Programme (iRAP) and the part of its methodology to calculate the effect of multiple road safety treatments on the reduction of road fatalities. Within the iRAP model, a number of trigger sets are applied for each project, which remove the inapplicable countermeasures. This Chapter presents the iRAP triggers for road safety treatments, which may affect the selection of the extracted sets of multiple countermeasures from the knowledge discovery process (mentioned in Chapter 3). This is because some countermeasures, for a specific project, may be overridden if they do not pass these triggers. This means that, any of the extracted sets may be omitted if one or more of its countermeasures were overridden by these triggers. These triggers include economic, application, and hierarchy factors as will be presented later in this Chapter.

4.2 Road Safety Protocols in iRAP

The iRAP was established to challenge the devastating impact of road safety on humanity and economy as the annual number of road fatalities around the world is expected to reach 2.4 million people in year 2030 if no intervention is undertaken. The iRAP has developed four protocols to evaluate and enhance road safety. These protocols are:

1. Risk Maps: show the number of fatalities and injuries on road networks.
2. Star Ratings: define the road safety into levels from one to five stars according to the roads' design and conditions.
3. Saver Road Investment Plans (SRIP): consider about 94 countermeasures to provide sounder infrastructure treatments to improve road safety.
4. Performance tracking: enables to track the performance of road safety on road networks by using risk maps and star ratings.

In this research, the focus will be only on protocols 2 and 3, as these protocols are directly related to the multiple countermeasures subject. These two protocols could be used as a part of a systematic approach to assess road infrastructure risks and make improvements based on where severe accidents can occur and how they can be prevented. The process of these two protocols is illustrated in Figure 4.1 (iRAP, 2014).

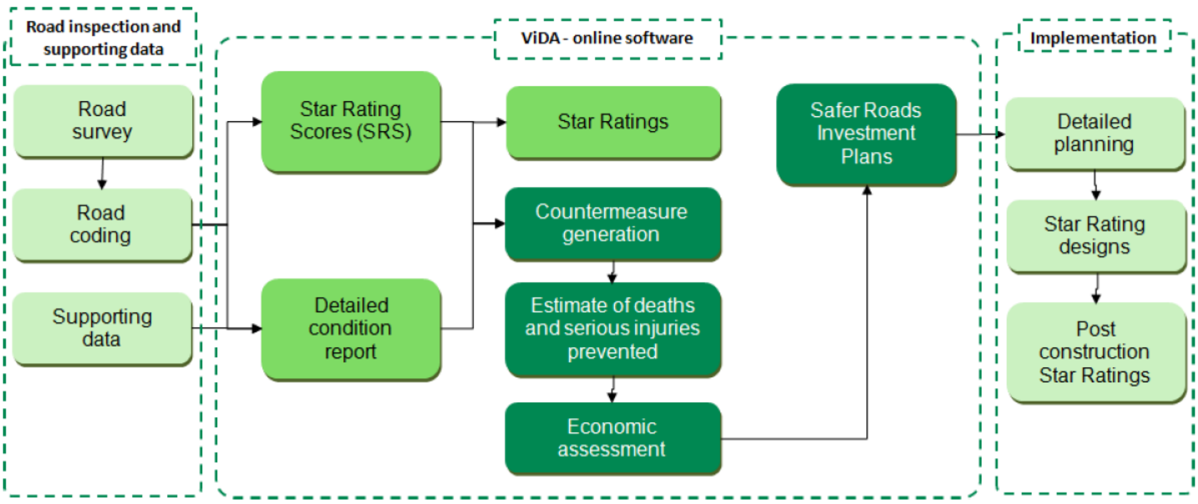


Figure 4.1 The iRAP Star Rating and Safer Roads Investment Plan process (iRAP, 2014)

4.2.1 The Star Ratings

Star ratings are simple measures of road safety levels according to inspection of road attributes that have effects on the likelihood of accidents and their severity. The star ratings are awarded between 1 and 5 stars depending on the safety level of a section or an entire road. The roads with star ratings 4 and 5 are the safest roads. These roads normally have road attributes, which are appropriate for the prevailing traffic conditions. Safer road attributes could include separated directions by a barrier or a wide median, adequate delineation, paved shoulders, wide lanes, and hazards free roadsides such as trees and poles. The riskiest roads (star ratings 1 and 2) do not have road attributes related to the prevailing traffic conditions. These might include frequent curves, unpaved shoulders, narrow lanes, poor delineation, roadside hazards such as unprotected tress and poles (iRAP, 2014).

4.2.1.1 Calculating the Star Rating Score (SRS)

There are four road users concerned in the iRAP model. These are vehicle occupants, motorcyclist, bicyclists, and Pedestrians (iRAP, 2014). In this research, the focus is on vehicles occupants as they form the highest number of fatalities in road accidents (WHO, 2015). According to the iRAP methodology, there are five factors that may be considered to calculate a Star Rating Score (SRS) for each 100 metre segment of road using equation 4.1. After that, each set of the consecutive 100 metre segments that reflect homogenous road characteristics (e.g. traffic volume, land use, carriageway type, and speed) are considered as a homogenous section. Then, the SRS for each homogenous section is calculated by averaging the SRS for all the 100 metre segments contained in that section (iRAP, 2014):

$$SRS = \Sigma\ \text{Crash Type Scores} = \text{run-off crash score} + \text{head-on crash score} + \text{intersection crash score} + \text{property access crash score} \quad \text{(Eq. 4-1)}$$

Where:

SRS = represents the relative risk of death and serious injury for an individual road user; and

$$\text{Crash Type Scores} = \text{Likelihood} \times \text{Severity} \times \text{Operating speed} \times \text{External flow influence} \times \text{Median traversability.} \quad \text{(Eq. 4-2)}$$

Where:

- Likelihood represents the chance a crash can happen according to risk factors of road attributes.
- Severity represents the severity of a crash that result from risk factors of road attributes.
- Operating speed is the changing in the risk according to speed change.
- External flow influence is the risk that a person can have to involve in crash relating to the existing of another person on the road.
- Median traversability is the potential risk that can be result from crossing the median by a vehicle (run-off and head-on accidents).

The value of the star rating score (SRS) for the vehicle occupants can be measured by multiplying the relative risk factors for all the road attributes that affect each accident type as shown in Figure 4.2. After calculating the star rating score for each homogenous section, the road sections are defined to star ratings from 1 to 5 using the ranges in Table 4.1. Then, a safer road investment plan is suggested to improve the sections with low star ratings by using single or multiple countermeasures (iRAP, 2014).

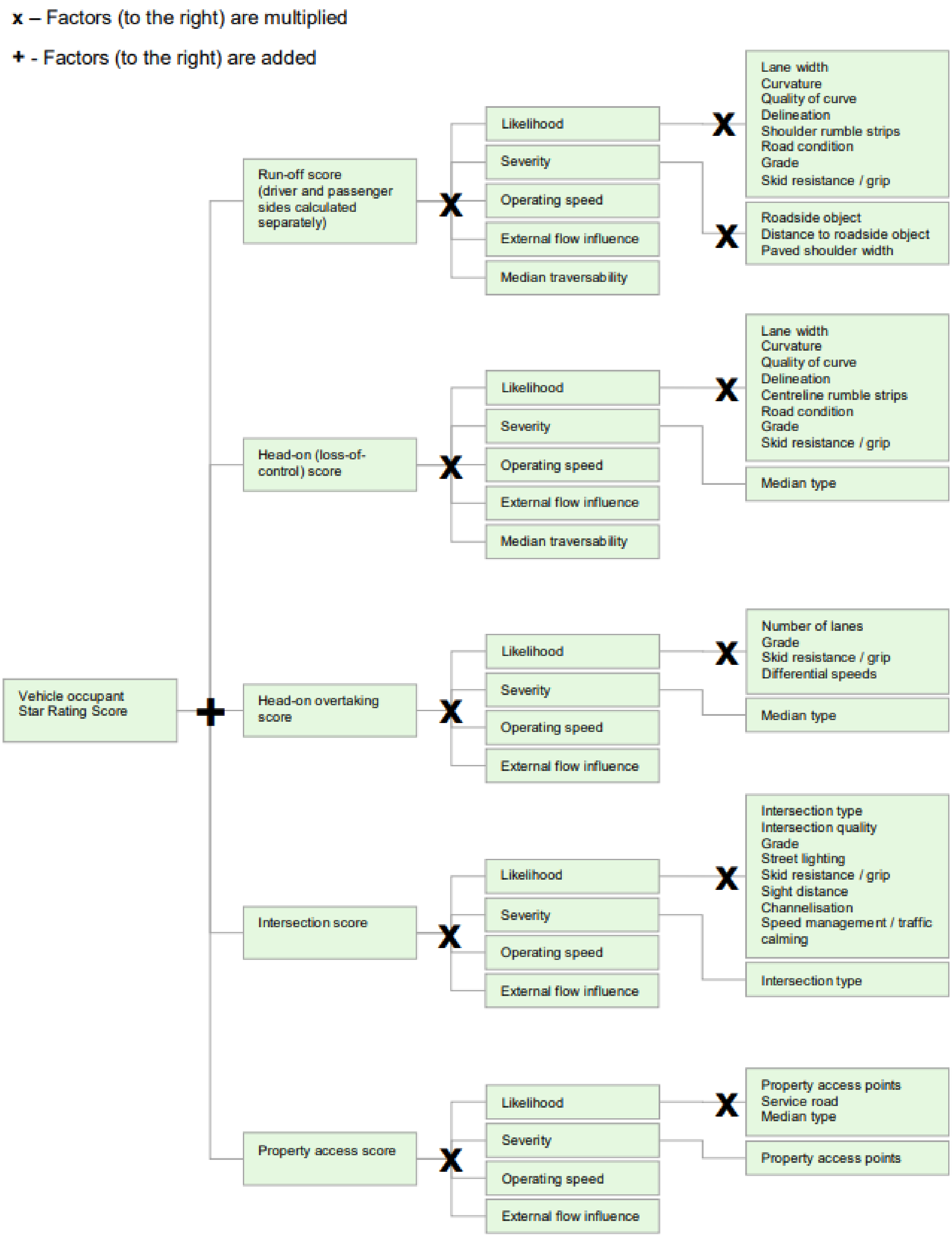


Figure 4.2 Calculating the star rating score for vehicle occupants (iRAP, 2014)

Table 4.1 Star Rating bands in the iRAP model (iRAP, 2014)

Star Rating	Star Rating Score
5	0 to < 2.5
4	2.5 to < 5
3	5 to < 12.5
2	12.5 to < 22.5
1	22.5 +

4.2.2 Safer Roads Investment Plans

Safer Roads Investment Plans (SRIP), based on the present value of the investment and the benefit cost ratio, are built from treatments that may improve the star ratings and reduce infrastructure-related risk in a cost-effective way. These plans are based on an economic analysis of a range of treatments, which is carried out by comparing the cost of applying and maintaining the treatments for 20 years (the design life for treatments in iRAP) with the reduction in crash costs that may result from this application over 20 years. The plans contain extensive planning and engineering information such as road characteristic records, treatment proposals, and economic evaluation for each 100 metre segment in a road network using the ViDA software.

To interpret the results, it is important to recognize that the iRAP model is designed to provide a network level assessment of risk and cost effective treatments. Because of that, any implementation of the proposals will ideally contain the following steps (iRAP, 2014):

1. Local examination of suggested treatments (including workshops, including all relevant stakeholders, where costs may be associated with benefits).
2. Preliminary scheme investigation studies, design and cost details for each proposal, final assessment and then construction.

4.3 The Economic Impact on Selecting Road safety treatments

Each identified treatment option is subjected to a benefit-cost ratio (BCR) analysis. Treatments that failed to achieve the BCR that meet a defined threshold for a given 100 meter segment is omitted from the analysis. The benefit of a treatment is determined by calculating the present value of fatalities and serious injuries that would be prevented over 20 years if the treatment was installed. The reduction in fatalities and serious injuries is calculated by replacing the Star Rating Score (SRS) used in the original estimation with a new, lower Star Rating Score (iRAP, 2014).

4.3.1 Countermeasures Identification

For each 100 metre segment of road, a series of treatments that feasibly may be applied are identified. This is achieved by considering each treatment's ability to minimize the risk (as measured by iRAP), application, and hierarchy rules. For example, a section of road with poor pedestrian star rating and high pedestrian activity is likely to benefit from the implementation of a pedestrian crossing, pedestrian refuge or signalised pedestrian crossing. Similarly, a road section with poor delineation and a high activity of car occupants is likely to benefit from better delineation (iRAP, 2014).

4.3.2 The Economic Benefit

An economic evaluation is undertaken in the iRAP methodology in order to identify the list of potential treatments for a limited budget. The economic analysis is performed twice during an iRAP assessment process. The first one during the treatment selection process, each treatment at each 100 metre road segment must produce benefit cost ratio (BCR) that passes a predefined threshold in order to be included in the Safer Roads Investment Plan (SRIP). At the end of the treatment selection process a second economic assessment of all the treatments, both individually and as a mix, is performed for inclusion in the SRIP (Lynam. 2012) cited in (iRAP, 2014).

The economic benefit is based on the number of casualties that can be prevented, the economic value of a human life (which is determined by considering the lifetime earnings), and the economic value of a serious injury. In case there is no reliable evidence of the economic value of a serious injury and the economic value of a human life, the iRAP model uses the rule of thumbs determined by McMahon and Dahdah (2008), which indicates that the value of a statistical life is about 70 times the GDP per capita for the country concerned (iRAP, 2014).

4.4 Countermeasures Selection Process

The method used to select treatments along a 100 metre segment of road is as follows (iRAP, 2014):

1. Each treatment's triggers are checked. Any treatment that does not pass its triggers is not considered further.
2. For each treatment, the number of fatalities and serious injuries that could be prevented if it was implemented by itself (that is, if no other treatments were applied at the same time) is calculated.
3. Each treatment is subject to an economic assessment, in which, its benefit-cost ratio (BCR) must exceed a predefined threshold set by the project team. Any treatment that does not satisfy the benefit-cost ratio threshold is not considered further.
4. Each treatment is subject to minimum length, minimum spacing and hierarchy rules (section 4.5.1). Any treatment that does not meet the rules is not considered further.
5. The effect of all viable treatments is calculated at the 100 metre level and the combined number of fatalities and serious injuries after all viable treatments have been implemented is determined for each accident type.
6. The total number of fatalities and serious injures prevented by the multiple treatments at that section is then adjusted, using a Multiple Countermeasures Correction Factor

(MCCF), to take into account the reduction in the effect of each countermeasure when it is combined with other countermeasures (iRAP, 2014).

7. At the end of the treatment selection process, a final economic assessment for all the treatments selected across all roads is carried out and a Safer Roads Investment Plan (SRIP) for 20 years is produced.

4.5 Triggers

In the iRAP model, for each treatment, a series of triggers (or prerequisite conditions) could be identified. A trigger must be satisfied before that treatment can be chosen for a 100 metre segment of road. The triggers are often a function of one or more of:

- Star Rating, which are based on a Star Rating Score (SRS).
- Road attribute, such as median type or quality of curvature.
- Vehicle (or road user) flow.

As a simple example, the trigger for the treatment ‘Improve Delineation’ is: if delineation is recorded as ‘poor’, then the treatment ‘improve delineation’ is applied, regardless of what the road’s Star Rating or vehicle flow. The countermeasure ‘signalise 3-leg intersection’, provides examples of more complex triggers. In the first trigger set, a 3-leg signalised intersection may be chosen if:

- There is an existing 3-leg intersection;

- Vehicle flow on the concerned road is higher than or equal to 5,000 vehicles per day; and
- The intersected road vehicle flow is higher than 1,000 vehicles per day.

In the second set of triggers, a 3-leg signalised intersection could be triggered if:

- The road is in a rural area;
- There is an existing 3-leg intersection;
- Vehicle flow on the concerned road is higher than or equal to 5,000 vehicles per day; and
- The intersected road vehicle flow is higher than 5,000 vehicles per day.

The base iRAP model includes more than 300 default triggers. As the traffic and geometric conditions may be different from a country to another, iRAP has developed different sets of triggers for worldwide use (iRAP, 2014).

4.5.1 Minimum length, minimum spacing and hierarchy rules

In addition to the triggers, there are a series of application rules that iRAP allows for certain treatments. Analysis teams could choose whether they wish to apply these additional rules. Treatments that satisfy these additional rules are considered for inclusion in the SRIP. These rules help to ensure that the treatment recommendations are logical and satisfy with established engineering practise. An example of a minimum length rule is:

Additional roadway lanes (such as overtaking lanes or 2+1 cross section) must be used for a minimum length of one kilometre within a homogenous section (that is defined during the data collection stage) of road before they are considered viable.

Examples of minimum spacing rules are:

- Grade-separated crossings for pedestrian must be at least one kilometre apart (that is, 10 consecutive 100 metre segments of road).

- New signalised crossings for pedestrian (non-intersection facilities) must be at least 600-metres apart (which is six consecutive 100 metre segments of road).

The hierarchy rules are only applied after the various economic, minimum length, and spacing rules above are applied. Therefore, these rules are based on the preliminary list of all viable treatments at a location. As an example of a hierarchy rule (sometimes called as an override), footpath treatments are listed below in order of their potential to reduce fatalities and serious injuries (from highest to lowest reduction) (iRAP, 2014):

- Provision of footpath on passenger side (with barrier).
- Provision of footpath on passenger side (>3m from road).
- Provision of footpath on passenger side (adjacent to road).
- Provision of footpath on passenger side (informal path >1m).

That is, all other things being equal, the implementation of ‘footpath on passenger side (with barrier)’ is preferable more than all the other footpath types. If all four forms of the footpath provision are viable then the best form of that treatment is selected for the final SRIP and all lower order countermeasures addressing the same issue are cancelled.

4.6 The Existing iRAP Methodology for Multiple Treatments

According to iRAP methodology, there are four steps to calculate the multiple countermeasures effectiveness for reducing the fatalities and serious injuries numbers (FSI) (iRAP, 2014):

1. The reductions of all the countermeasures are summed to calculate the Reduction ADDITIVE (R1).
2. A reduction is estimated by using a multiplicative method as in Equation 4-3 to calculate the Reduction MULTIPLIED (R2). When all of the proposed countermeasures are applied at the location, the new SRS and associated fatality estimation is based on a multiplicative formula:

$$\textit{Reduction MULTIPLIED} = FSI * [1 - (1 - CMF1)(1 - CMF2) * \ldots\ldots\ldots (1 - CMF_n) \qquad \textit{(Eq. 4-3)}$$

CMF = the crash modification factors (relative risk factors in the iRAP model).

3. A multiple countermeasure correction factor (MCCF) is calculated for each 100 metre segment of road and for each crash type. The factor is calculated as follows:

$$\textit{MCCF = Reduction MULTIPLIED (R2) / Reduction ADDITIVE (R1)} \qquad \textit{(Eq. 4-4)}$$

4. Reductions are estimated using the multiple countermeasure correction factor. The reduction in FSI for each individual countermeasure is then calculated as follows:

Reduction = Reduction SIMPLE (R3) x MCCF *(Eq. 4-5)*

R3 = the reduction for each countermeasure according to iRAP risk factors.

5. Total reduction = Sum of R3 for all the countermeasures used.

The weakness in the above methodology is that the multiplication of the crash modification factors in step 2 can be subjected to some Independence and applicability issues, which were explained in Chapter 3 (section 3.5.1.1).

4.7 Summary

The international Road Assessment Programme (iRAP) is a comprehensive road safety model, which accounts for many cost-effective factors that affect road safety. Four protocols have been developed by iRAP. These are risk maps, star ratings, safer roads investment plans (SRIP), and performance tracking. Only two protocols (star ratings and SRIP) have been considered in this study as they are directly related to the multiple countermeasures subject. The star ratings are simple measures of road safety levels that reflect the effect of road infrastructure on the likelihood of road crashes and their severity. In the iRAP model, road safety is, ultimately, expressed by five star ratings from 1 star (least safe) to 5 star (most safe). These star ratings are used in this study as the dependent variable to be associated to road traffic and geometric characteristics (as the independent variables) through developing data mining models to assess the effect of multiple countermeasures on road safety levels (star ratings) as will be explained in chapter 8). The safer roads investment plan represents the suggested countermeasures to improve road safety. In this study the countermeasures will be suggested by the data mining models as will be explained in Chapter 9.

CHAPTER 5

STATISTICAL MODELS

5.1 Introduction

Statistical techniques have been used to establish relationships between crash frequencies and road attributes (Chang, 2005). These include Multiple Linear regression, Poisson regression and Negative Binomial (NB) regression (Chin and Quddus, 2003). The results of any model mainly depend on the used regression technique. Earlier, studies on traffic crashes used Multiple Linear regression and created linear relationships between the dependent and the explanatory variables (Abdel-Aty and Radwan, 2000) cited in (Park, 2014). However, such multiple linear regression endures some undesirable statistical issues when used with crash analysis including high standard errors (the higher the standard error, the higher the deviation between the predicted and actual values). Therefore, linear regression seems not to be the most appropriate technique and should be used with caution (Abdel-Aty and Radwan, 2000; Jovanis and Chang, 1986; Chengye and Ranjitkar, 2013).

To overcome the statistical problems associated with the linear regression technique, Jovanis and Chang (1986) suggested the Poisson regression technique as an appropriate approach to model crash frequencies.

The Poisson model assumes that the mean and the variance of the accident related variable (i.e. the dependent variable) are equal. Yet, in most accident data, the variance of the accident frequency exceeds the mean and, in such cases, the data is considered overdispersed. However, overdispersion does not change the conclusion about the relationship between accidents and both the examined traffic and the highway geometric design variables. To overcome the problem of overdispersion, the Negative Binomial technique may be considered as the best option to analyze traffic accidents (Abdel-Aty and Radwan, 2000; Miaou, 1994; Poch and Mannering, 1996; Shankar et al, 1995; Lord et al, 2005; Montella, 2008). To this end, both the Poisson and Negative Binomial approaches will be explained in this research to develop models to describe the effect of the multiple countermeasures related to head-on overtaking crashes on fatality reduction rates in India. The effect of the multiple countermeasures related to run-off and head-on + run-off crashes was not presented in this study as the results were not statistically significant.

5.2 The Effect of Infrastructure on Road Safety

To develop the new model sought, there is a need to examine the effect of countermeasures on road safety, the effect of traffic on road safety, and the way in which these two aspects may be considered in a mathematical form.

5.2.1 The Effect of Countermeasures on Head-on Crashes

According to the iRAP model, there are five countermeasures can be used to improve road safety in relation to head-on crashes. These countermeasures are: adding travel lanes, changing median type, changing gradient, improving skid resistance, using differential speed limits, and using centreline rumble strips. The centreline rumble strips countermeasure was excluded from the analysis as it was not available in the concerned road network. It has been shown that these countermeasures, according to the available literature, have the ability to change the number of fatalities over time as presented in section 2.4.

5.2.2 Effect of Traffic Volume on Road Safety

In the last few decades, the need for vehicles has been increasing dramatically and will continue to do so. For example, the number of kilometers travelled in the Netherlands increased by 5% from 2000 to 2008 and the car use increased by 10% for the same period. The effect of traffic volume on road safety is not however obvious. Traffic jams may have a positive effect on road safety because of lower speeds. On the other hand, rear-end accidents could occur at the end of the queue, when traffic jam happens, because of the big difference between vehicles speeds (Marchesini and Weijermars, 2010). Because of the increase in the need for vehicles and kilometers travelled, it is important to consider the effect of flow rates on crash numbers, and to build a model that can give a more precise prediction of fatality numbers.

5.2.3 The Effect of Operating Speed on Road Safety

There is a solid evidence that speed has an effect on the likelihood and severity of crashes (iRAP, 2014). Speed is one of the most factors that affect road accidents. Higher speeds result in higher speeds of collision and serious injuries and fatalities. Also, when speeds are higher the time that is needed to process information and taking an action will be shorter, and the opposite for stopping distance that will be longer (Aarts and Van Schagen, 2006).

5.3 Choice for Models

Road safety is related to human, vehicle, and road infrastructure factors and exhibited by a number of crashes types (such as head-on, run-off, and intersection). An analysis to identify the optimum countermeasures to prevent all road users from any type of crashes is very challenging. Therefore, this study focused on head-on overtaking crashes for vehicle occupants as they usually lead to serious injuries and deaths (Kahane, 1994), and vehicle occupants' causalities have the highest percentage of the overall number of casualties (WHO, 2015). In addition, emphasis was put on infrastructure, which is a major accidents causation factor (Miller and Zaloshnja, 2009). However, the approach followed in this work can be generalized to include any other factors likely to contribute to road crashes.

To this end, the main objective of using statistical regression models (Poisson and Negative Binomial) in this study is to develop models to estimate vehicle occupants' fatality rates related to head-on crashes after the installation of combined treatments in low and middle income countries (LMICs). These models could be used to estimate a more accurate MCCF for the iRAP model to overcome the problem of the overestimation (explained in Chapter 3) in the predicted reduction of fatality rates after the implementation of multiple countermeasures.

5.4 Data Issues for Statistical Models

5.4.1 Data Used

A Sample of 350 road sections from the Multi project - Core Data dataset (explained in section 3.7) with their fatality rates (resulted from head-on overtaking crashes) from the Multi project – Fatality Rates dataset was used to develop the statistical model (300 of them were used to develop the model and 50 were used to evaluate it). As the lengths of the 350 concerned road sections were different, the number of fatalities for each homogenous section was divided on the section's length to determine the fatality rate per kilometer for analysis purposes. The data from the two datasets were integrated and tabulated as an Excel file. The file concerned seven road attributes (as independent variables). These are: five geometric characteristics of the roadway that affect head-on overtaking crashes according to iRAP (i.e. number of lanes, median type, gradient, skid resistance, and differential speed limits), and two traffic characteristics in terms of (operating speeds, and traffic volumes) and over a three-year period (2010 – 2012). Also, the file concerned the fatality rate as the dependent variable to be modeled with regard to the above independent variables. The fatality rates for the 350 road sections over a three-year period (2010 – 2012) were used to develop and evaluate the statistical model. The data selection process complied with the iRAP (iRAP, 2014) protocols so that a comparison could be made with an established road safety model.

5.4.2 Data Quality

The gathered data for each road section were checked for, data entry errors like missed entries, repeated entries, and wrong values. As data quality is an important factor that can affect the accuracy of the developed model, the sections with missing or repeated values were omitted. In addition, to ensure a reasonable quality of data, descriptive statistics techniques were used to clean the fatality rates data from outliers. One of the common descriptive statistics techniques is the Interquartile range (IQR). It is a measure of the dispersion in a range of data. In this method, the dispersion is measured based on identifying three quartiles Q1, Q2, and Q3 for a set of data (Shien, et al, 2007; Patton, 2005). The first quartile (Q1) is the median value of the first half of the data range, the second quartile (Q2) is the median of the data range, and the third quartile (Q3) is the median value of the second half (Doane and Seward, 2005) as may be seen in Figure 5.1. Interquartile range (IQR) is the difference between the third and the first quartiles (Q3 - Q1). In this technique, all the values that are smaller than (Q1 – 1.5 IQR) or greater than (Q3 + 1.5 IQR) are considered outliers and omitted from the datasets to ensure the quality of the data used (Zwillinger and Kokoska 1999). By using Microsoft Excel, the quartiles for the concerned dataset were calculated as follows: Q1 = 0.01, Q3 = 0.122, IQR = Q3 – Q1 = 0.112 as shown in Table 5.1.

Table 5.1 Descriptive statistics for the data concerned to develop the statistical model

Number of road sections	Minimum number of fatalities / km / year	Maximum number of fatalities / km / year	The first quartile (Q1)	The third quartile (Q3)	Interquartile range (IQR)
300	0.01	1.04	0.01	0.122	0.112

Therefore, the lower outlier bound is:

Q1 – (1.5 x IQR) = -0.15 fatality/km/year *(Eq. 5-1)*

and the upper outlier bound is:

Q3 + (1.5 x IQR) = 0.29 fatality/km/year. *(Eq. 5-2)*

Considering the above values, 24 road sections, with fatality rates higher than 0.29 were considered outliers and omitted from the analyses to ensure the quality of the data used.

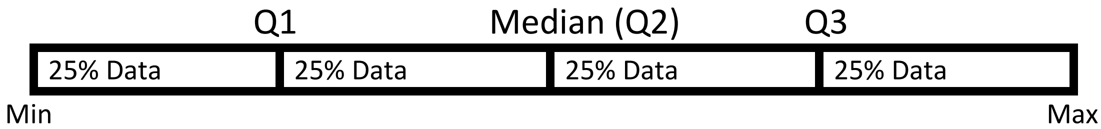


Figure 5.1 Data distribution and quartiles

5.5 Poisson and Negative Binomial Regression Models

Using the R statistical environment (R Core Team, 2016), Poisson and Negative Binomial regressions models were developed. All the seven geometric and traffic characteristics (independent variables) that affect head-on crashes according to the iRAP methodology were considered. The dependent variable was the fatality numbers (F). The general form for both the Poisson and Negative Binomial regression models is shown in Equation 5-3 (UCLA, 2018; Hilbe, 2011):

$$Y = exp\ (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots + (\beta_n X_n) \qquad (Eq.\ 5\text{-}3)$$

Where:

Y= Dependent variable

X_1, X_2……X_n= Independent variables

β_0 = Regression constant

β_0, β_1…… β_n= Regression coefficients

The two models were further tested to check the required statistical aspects (significance, overdispersion, residual deviance, and degrees of freedom).

5.6 Results of the Statistical Models

For the Poisson model, the results showed that some of the independent variables were insignificant because the correlation between the number of fatalities and these variables was less than 95%. The Poisson model was:

$$F = \exp(2.813 - 0.514Nol - 0.334Sres + 0.462Gr - 0.832Ds + 0.379Os + 0.000264AADT - 2.752Mt) \quad \text{(Eq. 5-4)}$$

Where:

F = Estimated number of fatalities per km per year

Nol = Number of lanes.

Sres = Skid resistance.

Gr = Gradient.

Ds = Differential speed limits.

Os = Operating speed

Mt = median type

AADT = annual average daily traffic.

In addition, the results showed that overdispersion did exist as the overdispersion factor for the fatality rates data is higher than (1.0) (R Core Team, 2016). This means that high standard errors will occur when using the Poisson model to estimate the fatality numbers after the implementation of multiple countermeasures (Abdel-Aty and Radwan, 2000). Therefore, the Poisson model was omitted and the Negative Binomial technique was then examined as the next option to develop the model sought. Table 5.2 shows the statistical parameters for the Negative Binomial model. The form of the Negative Binomial model was:

$$F = \exp(-3.22 - 0.703Nol - 0.244Sres + 0.502Gr - 0.573Ds + 0.534Os + 0.000136AADT - 1.763Mt) \quad \text{(Eq. 5-5)}$$

Where:

F = Estimated number of fatalities per km per year

Nol = Number of lanes.

Sres = Skid resistance.

Gr = Gradient.

Ds = Differential speed limits.

Os = Operating speed

Mt = median type

AADT = annual average daily traffic.

Table 5.2 The statistical parameters for the Negative Binomial model

Coefficients	**Estimate**	**Standard error**	**Probability**
Intercept	-3.22	2.814-e01	< 2e-16
Number of lanes	-0.703	7.459e-02	< 2e-16
Skid resistance	-0.244	3.419e-01	9.32e-10
Grade	0.502	3.509e-01	8.66e-09
Differential speed limit	-0.573	1.305e-01	0.000238
Operating speed	0.534	5.925e-02	< 2e-16
Vehicle flow (AADT)	0.000136	9.547e-02	< 2e-16
Median type	-1.763	1.051e-01	< 2e-16

The model was tested for the significance of each of the independent variables (i.e. the countermeasures and the traffic related conditions) on the number of fatalities in terms of the Negative Binomial Regression. In addition, overdispersion was examined. The results showed that the independent variables have significant effect on the fatality numbers, as their probability values are lower than 5%. To evaluate the ability of the NB model to relate the fatality rates with roads traffic and road geometric conditions without overestimation, the goodness of fit for the NB models was checked with regard to its residual deviance and degrees of freedom. The residual deviance shows how well the dependent variable (fatality rates) is predicted by the explanatory variables. The residual deviance is a measure of the goodness of fit for a model: the less the residual deviance the higher the goodness of fit (Khan, 2012). With regard to the term 'degrees of freedom', it is a statistical indicator that refers to the variability of the instances. Statistical models can be considered appropriate if the degrees of freedom are equal to the residual deviance. This means that the data used are not overdispersed (i.e. the model fits the data). Thus, to check that the data that were used to develop the NB model are not overdispersed, the ratio of the residual deviance to the degrees of freedom should be about 1 (R Core Team, 2016).

For the Negative Binomial model, the residual deviance was 321 and its ratio to the degrees of freedom was (0.93). This means that the NB model has a high goodness of fit in comparison with Poisson model, which its residual deviance was 2326 and its ratio to the degrees of freedom was (6.8). As the NB model had superior performance over the Poisson model, the NB model was considered appropriate for further tests and use.

5.7 Validation of the NB Model

To validate the fatality estimation model, fifty new road sections but from the same road network under study for the same period of analysis were examined with regard to the data quality criteria (section 5.4.2), and seven sections were identified as outliers. Therefore, only 43 sections were further considered to validate the model. The fatality rates calculated from the developed model using the road characteristics of each homogenous section were multiplied by the length of that section to determine the number of fatalities for each homogenous section. The results were plotted against the observed values for these 43 road sections and shown in Figure 5.2. It may be seen that there is a high converge between the predicted and actual data. Therefore, the NB model was considered appropriate to describe the effectiveness of multiple countermeasures reflected by the varying road attributes.

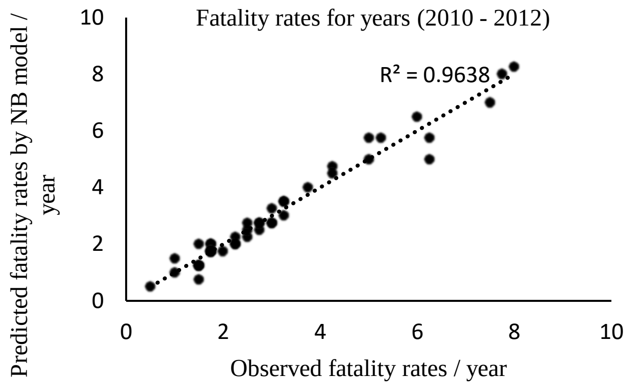


Figure 5.2 Validation of the Negative Binomial model

5.8 Application of Negative Binomial for the iRAP Model

5.8.1 Data Analysis

To test the effect of multiple countermeasures and their variations, appropriate data is required. These data should provide comparison between fatality reduction rates before and after implementing road safety multiple treatments. As such data is difficult to be available (Haghani et al, 2018), substitute data was gathered as follows. Eighty new sections from the iRAP database were divided into two groups of forty, where each group had the same road users, vehicles, traffic and geometric characteristics, except two or more of the five concerned road attributes were different to reflect the effect of the multiple countermeasures. The first group had high fatality rates and low star ratings and was considered as the 'test' group. The other forty sections with lower fatality rates and safer road attributes (higher star ratings) were considered as the 'control' group. Each section in the test group was compared with a section in the control group. The difference in fatality rates between the road sections in these two groups were considered as real fatality reductions in case that the road attributes of the sections in the control group were implemented on the sections in the test group.

Then, the iRAP formulas to calculate the reduction in the fatalities numbers (Equations 5-6 and 5-7) were applied on the forty sections in the test group to calculate the fatality reductions after improving the road attributes based on the infrastructure differences with the forty sections in the control group. After that, the results were multiplied by the length of each homogenous section to determine the number of fatalities for each homogenous section and compared with the real fatalities numbers of the sections in the control group.

In the iRAP model, the expected fatalities reduction (F) is given by:

$$F = Fe * MCCF \qquad \text{(Eq. 5-6)}$$

Where:

F = Expected reduction in the number of fatalities per km per year.

Fe = Existed number of fatalities.

and:

$$MCCF = [1 - (1 - CMF1)\,(1 - CMF2) * \ldots\ldots\ldots\ (1 - CMFn)] \qquad \text{(Eq. 5-7)}$$

Where:

MCCF = Multiple countermeasure correction factor.

CMF = Crash modification factor for treatments 1, 2,……. n.

Then, the same process was repeated but this time by applying the NB model (Equation 5-5) on the sections in the test group and considering the differences in the road attributes from the control group. And the resulted fatality rates were also compared with the real ones in the control group.

To clarify the above process, a road section (1 km) from the test group with one vehicle occupant fatality each year may be considered as an example. Its differences from a section in the control group were skid resistance improvement and the elimination of the differential speed limits as shown in Table 5.3.

Table 5.3 The crash modification factors for implementing the two suggested countermeasures according to the iRAP methodology

Countermeasures	***iRAP Risk factor (before)***	***iRAP Risk factor (after)***	***CMF (Percentage of the difference between before and after risk factors)***
Improving the skid resistance	1.4	1.0	29% ((1.4 – 1.0) / 1.4)
Eliminating the differential speed limits	1.2	1.0	17% ((1.2 – 1.0) / 1.2)

Using the iRAP method, the expected fatalities reduction is calculated using Eq. 5-7.

F = Fe * MCCF = 1 * 0.41 = 0.41 fatality/km/year

Where:

MCCF = [1 – (1 – 0.29) (1 – 0.17)] = 0.41

This reduction means that the expected number of fatalities, according to the iRAP methodology, will be reduced by 0.41 fatality each year. Therefore, the expected fatalities rate will be reduced from 1 into 0.59 fatality each year. By comparing this reduction with the real fatalities rate 0.68, from the control group, an overestimation of 13% [((0.68 – 0.59) / 0.68) * 100] exists in the iRAP estimations. Subsequently, the expected number of fatalities, after the implementation of countermeasures, was calculated using the NB model (Equation

5-5). The resulted number was 0.64 fatalities each year. When comparing this number with the observed fatalities number 0.68, an overestimation of 6% is found in the NB fatalities estimations.

The same process was followed to determine the differences between the real observed and iRAP estimated fatalities numbers for all the forty sections in the test group. The results may be seen in Figure 5.3, together with the trend line between the iRAP estimations and observed fatalities, which shows low converge between them (68%).

In addition, the differences between the observed fatalities and NB estimations for all the forty sections in the test group may be seen in Figure 5.4, which shows that the converge between the NB estimations and the observed fatalities numbers is better than the iRAP's and equal to 82%.

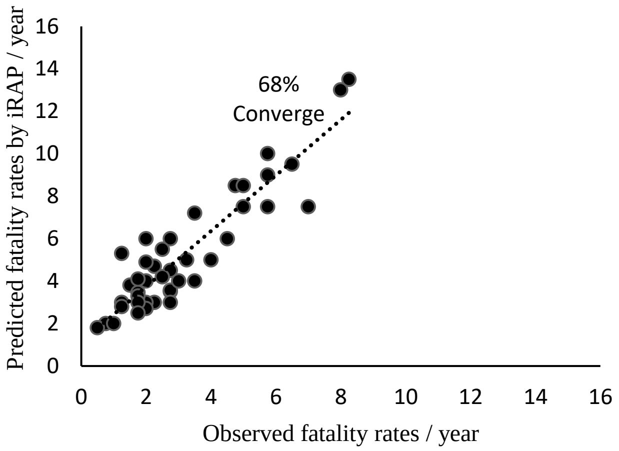


Figure 5.3 The iRAP estimated vs observed fatality rates

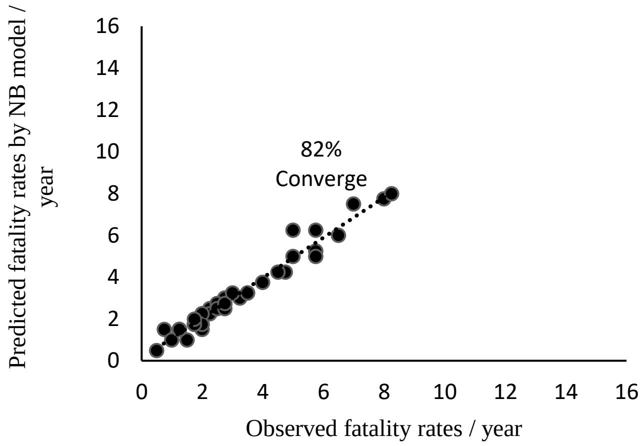


Figure 5.4 Negative Binomial model predicted vs observed fatality rates

5.8.2 The Contribution Method

The novelty of using statistical models in this study is not in developing the statistical models using statistical regression methods but the novelty is in the concept of contributing these models with the iRAP model.

A comparison between the iRAP and the observed data showed that an average overestimation of 32% was found in the concerned sections, whereas, an average over and underestimation of 18% appeared when the Negative Binomial model was used. This research proposes a concept that the difference between the two models (14% for the concerned dataset) may be used as a correction factor to adjust the MCCF to make the iRAP fatalities estimations closer to the observed data as follows:

$$MCCFa = MCCF\ (1 - (iRAPe - NBe)) \qquad \text{(Eq. 5-8)}$$

Where:

MCCFa = the adjusted multiple countermeasures correction factor

iRAPe = the average difference between iRAP and observed fatalities reduction.

NBe = the average difference between NB and the observed fatalities reduction.

An important note need to be made with regard to using Equation 5-8, is that, in case the difference between the iRAP fatalities estimations and real data is less than the difference between the NB model and the real data (i.e. NBe is greater than iRAPe), no adjustment is needed for the MCCF. Because the MCCF that is calculated by iRAP, in this case, is more reliable for that specific section or road.

Following the contribution process, the iRAP MCCFs were adjusted for the forty sections in the test group to calculate the expected fatalities numbers in iRAP. The results showed that the overestimation (i.e. the difference between observed and iRAP expected fatalities numbers) decreased from 32% to 17%. This means that, the iRAP estimations became closer to the real data after adjusting the MCCFs. Even, closer than the NB model estimations.

Although, 18% of the statistical model results differ from the actual fatality data, it may still be more accurate than the iRAP formula (before adjusting the MCCF). The reason behind this could be attributed to that the iRAP formula is a general model that could be used everywhere. Whereas, the developed model is specific to the studied road network and originated from the road data itself. In addition, the iRAP methodology for estimating the fatalities numbers depends on general crash modification factors. Whereas the NB model depends on the logarithmic weights of the independent variables (countermeasures) of a specific road network, and allows significant combination of the independent variables (UCLA, 2018). The limitation, however, of this approach is that a specific statistical model need to be developed for each road type in the network (in this study, for instance, the model concerned rural arterial roads). Therefore, the iRAP equation is still effective in terms of cost and time and its accuracy may be improved by using the adjusted MCCF as in Equation 5-8.

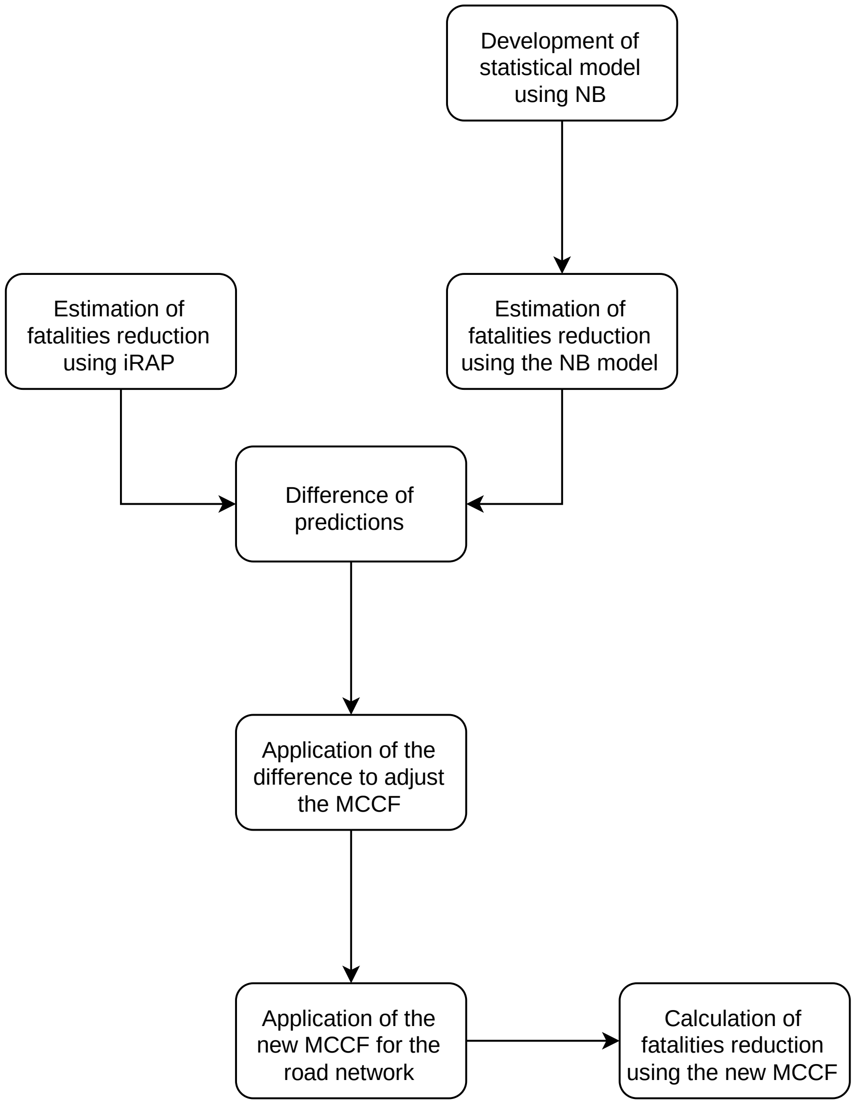


Figure 5.5 The process of contributing the NB model with the iRAP model

5.9 Summary

This part of the study developed a model to predict fatality rates on rural roads and contrasted it with the iRAP model. Poisson and Negative Binomial regression approaches were applied to a dataset, which covered traffic conditions, road geometrics and fatalities data for selected roads. The analyses showed that the Negative Binomial model developed is an appropriate and more accurate approach to predict road fatality rates for head-on crashes as a function of road design and traffic condition parameters such as number of lanes, median type, gradient, skid resistance, differential speed limits, operating speeds, and traffic volumes. The median type has the highest impact on the number of fatalities, the median type seeks to separate traffic directions to minimize the occurring of head-on crashes with non-traversable median types being the most effective ones. However, this is a high cost countermeasure and therefore, its applicability may be limited. The effect of the number of lanes on fatality rates, is the second most important factor and followed by the impact of the skid resistance, speed, and gradient. This is because, the provision of an additional lane in each direction, or at least one, will reduce the need for overtaking maneuvers which, in turn, will reduce the number of head-on crashes.

Although, the road fatality rates may be difficult to be predicted exactly, the developed model may be used to improve the iRAP methodology concerning the effectiveness of multiple countermeasures through an adjustment of the MCCF.

The improved MCCF enabled the iRAP estimations to converge better with the real data (the difference reduced from 32% to 17% in this study) and, as a result, reduced the overestimation of the effectiveness of the multiple countermeasures.

CHAPTER 6

METHODOLOGY FOR KNOWLEDGE DISCOVERY

6.1 Introduction

Following the development of the statistical models, this Chapter presents the knowledge discovery process and the data mining applications. In addition, it describes the generic data preparation issues and their effect on the process of building the knowledge discovery models. The Chapter is organised in five main parts. The first part presents a general description for the knowledge discovery and data mining. The second part describes the data preparation for knowledge discovery process. The third part demonstrates the data mining approach and its patterns. Then, the fourth part, the evaluation process for the data mining applications. Finally, the fifth part lists the representation methods for the outputs of the knowledge discovery process.

6.2 Knowledge Discovery and Data Mining

In recent years, the amount of data increases rapidly in many different domains because of the availability of the computers that provide an easy and inexpensive ways to save data. Thus, it is being difficult to understand and extract useful information from these large amounts of data. Therefore, knowledge discovery and data mining methods have been developed (Witten, 2017).

Data mining refers simply to extracting knowledge from a huge amount of data (Witten, 2017).

The term data mining, for many people, is a synonym for Knowledge Discovery. Whereas, for others is a fundamental step in the knowledge discovery process (Han et al, 2012). For the purpose of this study, data mining is treated as an essential step in the knowledge discovery process as shown in Figure 6.1.

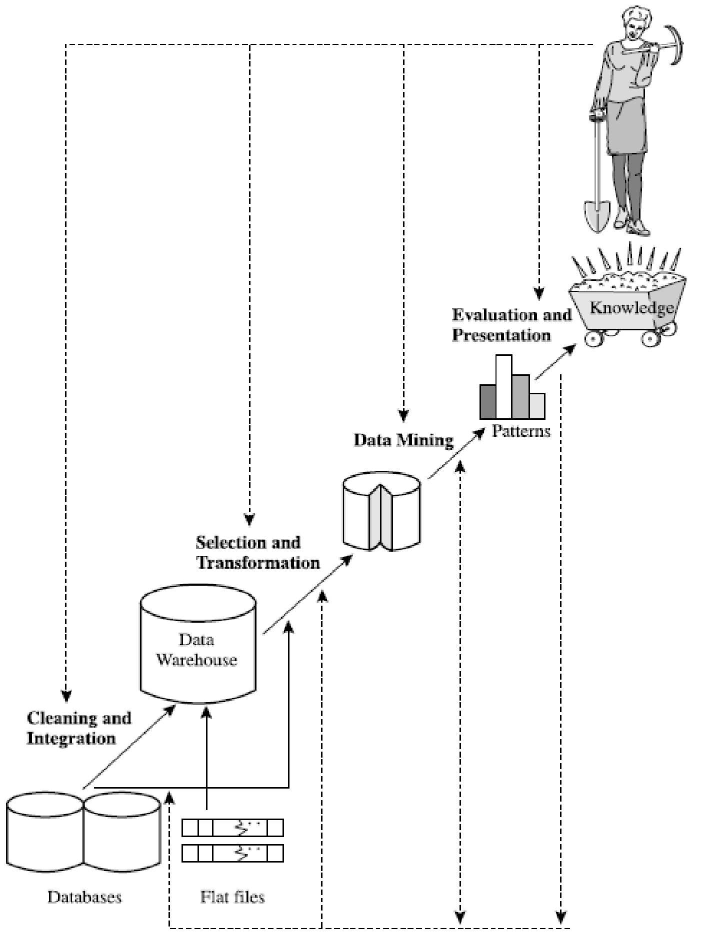


Figure 6.1 Data mining as a step in the knowledge discovery process (Han et al, 2012)

6.2.1 Knowledge Discovery in Databases

There is a difference between the two terms *Data* and *Database*. The term *Data*, in general, refers to a set of values that are collected together to be used as a reference or analysed for useful information (Beynon- Davies, 2002). In the knowledge discovery domain, the term *Data* refers to sets of attributes (independent variables) and their resulted classes (dependent variables). Whereas *Databases* are tables of data that are organized, by the computers, in a way that they can be easily managed. Each row in these tables represents an instance of the data (independent variables + dependent variable). Figure 6.2 shows a part of a general example from a database.

Databases can be useful if converted into useful mined information. Data mining learning algorithms and tools could be found in the field of knowledge Discovery in Databases. The term Knowledge Discovery in Databases (KDD) refers to the whole process of extracting knowledge from data, and emphasizes the application of particular data mining approaches. Whereas, data mining is a part of the KDD and it is the application of specific algorithms on these databases to extract the desired knowledge (Han et al, 2012).

Attributes (Independent variables)

Class (Dependent variable)

	Attribute 1	**Attribute 2**	**Attribute 3**	**Attribute 4**	**Attribute 5**	**Class**
Instance 1 →	Feature 1	Feature 3	Feature 1	Feature 2	Feature 2	Class 1
Instance 2 →	Feature 3	Feature 2	Feature 3	Feature 1	Feature 4	Class 2
Instance 3 →	Feature 3	Feature 3	Feature 1	Feature 2	Feature 1	Class 3

Figure 6.2 A general example from a database

6.2.2 Knowledge Discovery Patterns

The process of discovering knowledge includes defining data needs, preparation of datasets, developing models, assessment, and deployment (Rawlings 1999). The main two goals of the knowledge discovery process are verifying the user's hypothesis when the model is limited to verification and discovering when the model finds new information (extracted knowledge). Discovery is divided into description and prediction patterns. Description refers to the process of interpreting the data into descriptive patterns to describe the situation of the concerned variable that may result from the variables in the databases. While, prediction is the process of using explanatory variables in databases to predict the data patterns as values for the concerned variable (Fayyad et al 1996). In this research, only the first part (description patterns) has been considered as will be explained, why, later in Chapter 8.

6.2.3 The knowledge Discovery Process Steps

In order to use the knowledge discovery and data mining methods, a database is needed so that the user may extract some knowledge patterns from it, in addition to the need for a data mining application to extract the wanted knowledge from that database (Han et al, 2012). Figure 6.1 showed the data mining as a step in the KDD process. The database and the data mining application are used through the KDD steps. The knowledge discovery process consists of a series of steps that come in the following progression:

1. Data preparation
2. Data mining

3. Pattern evaluation
4. Knowledge representation

6.3 Data Preparation

Data may come from different sources and may be collected manually by people or automatically by equipment (Lin et al, 2007), depending on the requirements of the analysis and the available resources. Although manual methods may be slow, they can give higher level of details than the automated methods (McPherson, 2005).

There is high possibility that databases are susceptible to missing, inconsistent, and noisy data values because of their normally massive size. This could affect the quality of the mined results. Therefore, data need to be checked before processing them in knowledge discovery and data mining applications. There are a number of data preparation methods such as data cleaning, integration, and transformation (Witten et al, 2017).

6.3.1 Data Cleaning

It is unlikely that raw databases are perfectly clean (free of errors) (Brown, 2014). Therefore, data cleaning (sometimes called cleansing) techniques try to remove the errors, fill in the missing fields, and rectify the inconsistences in the databases. The next two sections discuss the methods of treating the missing values and removing the errors from the databases (Han et al, 2012).

6.3.1.1 Missing values

It is the case when some tuples or fields in the databases have no values for some attributes.

In most data mining systems, there is an implicit assumption that a missing value in an attribute has no practical significance on the data mining process. There are different cases that could be considered as missing values. For example, a zero in a field that could never typically be zero or a negative number in a field that can only be positive such as the number of causalities on a road which can never be negative. For the attributes that contain nominal values, the missing values could be blanks or indicted by dashes (Witten et al, 2017). These fields need to be inferred and cleaned by one of the following methods (Han et al, 2012):

1. **Ignoring the tuple:** this method is usually used when the class label (in case classification is the considered data mining pattern) is missing. The data mining tasks are described in section 6.4.1.
2. **Filling the missing field manually:** this approach could be unfeasible if many values are missed. Especially, in large databases as it is time consuming.
3. **Using a constant:** in this approach, all the missed values are replaced with same constant. The problem with this approach is that, if all the missed values are replaced by the same value then the data mining software could mistakenly consider them as an additional class, which is not one of the concerned classes in the database. Therefore, although it is a simple method, it could be unreliable.
4. **Using the mean:** in this approach, the mean of the attribute that contains missing values could be used to fill these fields. This approach could be used when the databases consist of numerical values only.

5. **Using the most probable value:** this approach depends on determining the missing values by using other statistical or mining methods such as regression or decision trees.

6.3.1.2 Noisy data

Noise is a random error that could happen in measuring any attribute value such as entering numerical data rather than nominal, repeating or deleting numbers. There are many possibilities why these errors can happen. There could be a fault in the instrument used to collect the data, computer or human errors that may happen during the data entry stage, or data transmission errors may also happen when transferring data between digital devices. The databases should be examined to detect the errors and changing them into the right format (for example, changing 0 into zero, or 2017/12/20 into 20/12/2017) (Han et al, 2012).

6.3.2 Data Integration and Transformation

There are other crucial operations that can markedly improve the results of the data mining models, which are the subject of this section. These processes engineer the input data into suitable format for the data mining application chosen (Witten et al, 2017):

1. Data Integration: the knowledge discovery process often requires data integration before using them to build the data mining models. Data integration is the process of merging data from different sources or databases into one dataset as a file that can be readable for data mining programmes.

2. Data Transformation: the data could also need to be transformed into a different format. In this process, the data are transformed into the appropriate format for the data mining applications (i.e. from numerical to nominal or the opposite). For example, transforming the numerical age range “(10 - 20)” years into the nominal form “Group1” (Han et al, 2012).

6.4 Data Mining

Data Mining may be defined as the operation of extracting patterns in data by using a specific data mining application. The patterns in data can be expressed as structures that can be tested and used to support future decisions in different knowledge fields. The expression (*structures*) comes from the ability of these patterns to explicitly describe the decision structure within the data mining domain.

The structure below is an example of a road engineer’s decision to choose among physical median, central hatching, or none median change.

```
If Head-on crashes = zero,
Then proposal = none,
Otherwise,
If Number of lanes = one,
Then suggestion = central hatching,
If Number of lanes = two,
Then suggestion = Physical median
```

The structural descriptions can be expressed as rules, as in the above example, or in another popular expression, which is decision trees as will be explained later in this Chapter. The topic data mining includes non-theoretical (i.e. practical) learning patterns, which describe and predict structural patterns in data. In this case, the input data are used as sets of examples for the data mining process. The output will be descriptions or predictions of the acquired knowledge.

6.4.1 Data Mining Patterns

There are, generally, four basic learning patterns in the applications of data mining. These are, classification, association, clustering, and numeric prediction (Berkhin, 2006). Classification can be defined as the machine learning scheme that produces a set of classified rules from which a classifying way for unseen data can be discovered. Association seeks for any relationships among all the features in a set of data, not just the features that describe or predict specific classification. Clustering learning seeks groups of examples, in data, that are related to each other. In numeric prediction, the output takes the form of a numeric value rather than a text.

The classification, in data mining, is sometimes given the name *supervised learning* (Caruana, and Niculescu-Mizil, 2004). This is because the scheme in this pattern is supervised by providing the actual outcome, for each instance in the data, to the machine. The outcome of each instance, in a set of data, is called *Class.* The success of the classification pattern can be tested by applying the knowledge that has been learned (mined) on an independent test set of data that its true classifications are known but were not provided to the machine. The classification success rate on the test set can be considered as an objective measure of how well the extracted knowledge has been described (Witten, 2017). This is the process that is followed in this study to build the data mining models for road safety levels (classes) as will be described in Chapter 8.

6.4.2 Data Mining Inputs

The instances in a dataset are the input data to the data mining systems (Hall et al, 2009). These instances are the examples that are expected to be classified, clustered, or associated (Frank et al, 2004). Each one of these instances is an independent and individual example of the knowledge to be learned in the data mining process (Witten et al, 2000). Each instance consists of a set of features that are called *attributes* that their values are predetermined. Each dataset contains a set of instances that are presented as a matrix of attributes versus classes (i.e. each set of attributes that produce a specific class in data) as showed in Figure 6.2 (Witten, 2017).

6.4.3 Types of Attributes

Each instance, as an input, in data mining provides a set of predefined attributes for the machine to learn (Hall and Holmes, 2003). Each instance is a row in the tables that form the datasets. The value of each attribute in a particular instance represents the quantity or the weight of this attribute as an impact on producing a specific class in the data. There are two types of attributes, numeric and nominal (text). Numeric attributes are measure numbers (either rational or integer values). A nominal attribute is a pre-specified value from a set of possibilities that are sometimes called categories (Witten, 2017). For example, one of the attributes that may affect road safety levels is the *Lane width.* Considering the lane width as an attribute may has possibilities of (narrow, medium, or wide). These are the categories of this attribute that can be used in the data mining models. The focus, in this study, is on the nominal attributes as will be described later in Chapter 8. The nominal values can be represented just as names or labels. Thus, they are called nominal, which is the Latin synonym for the word *name* (Han et al, 2012).

6.5 Pattern Evaluation

In data mining, evaluation would be considered the key to make progress. In case there is a large amount of data, the evaluation process would be easier by building the model using a large set of data and then applying it on another big test set. In some situations, the evaluation means checking the model's ability to classify the instances of the test set. In other cases, this could include probability predictions or numerical predictions instead of the nominal ones. As the classification is the concerned data mining pattern in this study (the reason explained in section 8.2), evaluating the ability of the data mining model as a classifier is considered. Three different methods of evaluation are used in this study. These are: Training set, Cross Validation, and Percentage Split as will be explained in Chapter 7 (Witten, 2017).

6.6 Knowledge Representation

As previously mentioned in this Chapter, data mining applications are used to extract structural patterns in data (Zhang et al, 2003). Before describing what are these applications and how they work, it is important to discuss how these patterns can be represented. There are different methods that can be considered for expressing the discovered patterns in data. These are (Witten, 2017):

1. Tables
2. Linear Models
3. Decision Trees
4. Rules

6.6.1 Tables

These are the simplest method of expressing the output from data mining applications. In this method, the extracted patterns are represented as rows in tables exactly as the input data as shown in Table 6.1. Tables can be used for both numeric and nominal attributes. These tables are sometimes referred to as regression or decision tables (Liu and Motoda, 2012).

Table 6.1 An example of the *Tables* representation method

	Attribute 1	**Attribute 2**	**Attribute 3**	**Attribute 4**	**Attribute 5**	**Class**
Decision 1 →	Feature 3	Feature 2	Feature 3	Feature 2	Feature 1	Class 1
Decision 2 →	Feature 1	Feature 2	Feature 3	Feature 1	Feature 3	Class 2
Decision 3 →	Feature 2	Feature 3	Feature 2	Feature 3	Feature 2	Class 3

6.6.2 Linear Models

This is another simple option for representing data mining outputs. In these models, the output is just the sum of the attributes values multiplied by their weights. The important thing here is to find out reliable values of the weights to make the model fits the desired output. Numerical values are the only type of attributes that can be used as inputs and outputs in this method of outputs representation. In statistics, the term regression is used for the process of predicting numeric values. Therefore, these models are sometimes called regression models.

Linear models are the simplest way to visualize the data mining outputs in two dimensions by drawing a fitting line through a set of points (Witten, 2017). Figure 6.3 shows an example of predicting the causalities rate of road crashes with only one attribute as input, which is the sealed shoulder width. In this example, the horizontal axis represents the *Class* (crash modification factors) and the vertical axis shows the *Attribute* (sealed shoulder width). As may be seen, the input and output values are both numerical. Here, the fitting line could be used to predict the class (crash modification factor) for unseen attribute (shoulder width) data on road sections that have the same traffic and geometric characteristics but the shoulder width.

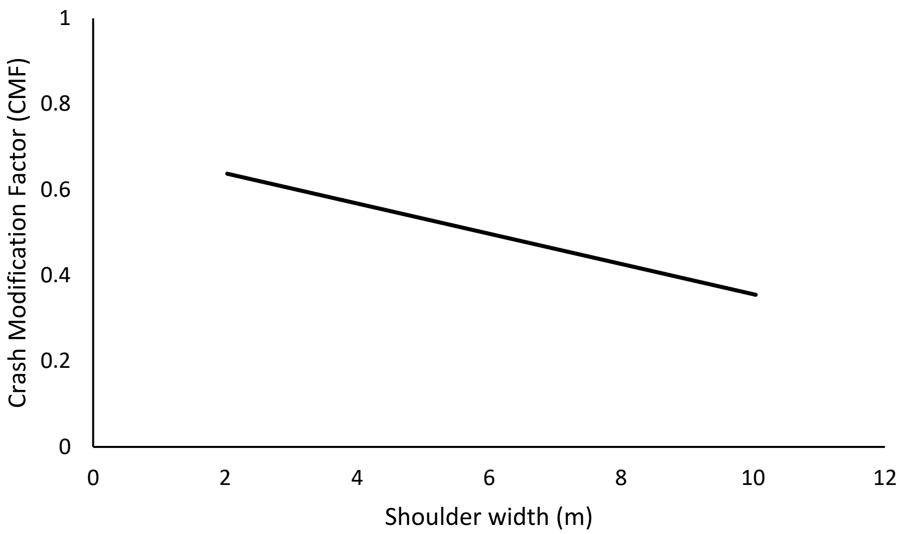


Figure 6.3 A relationship between the shoulder width and crash modification factors (Neudorff et al, 2016)

6.6.3 Decision Trees

These are another style for representing data mining outputs by learning from a set of instances. Decision trees are generally formed of a root node (the first node in a decision tree), internal nodes, branches, and leaf nodes. The leaf node is the last node in each path and the internal nodes are all the nodes between the first and last node in each path, as shown in Figure 6.4.

Testing the attributes is involved in the root and internal nodes. Whereas, the class of each set of attributes is represented in the leaf nodes (Han et al, 2012). Decision trees are flow charts that takes tree structure where each internal node indicates an attribute test, each path stands for an outcome from the test, and each leaf node represents a class label (Witten, 2017).

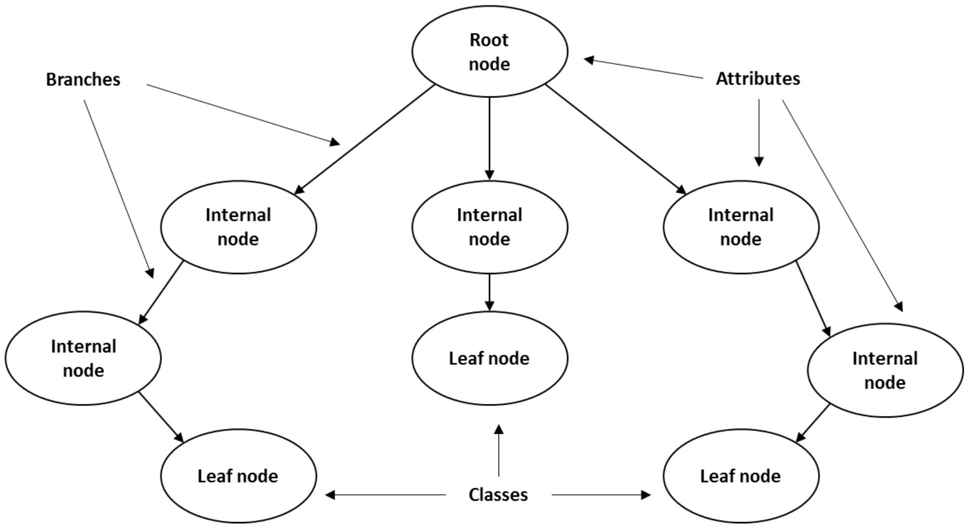


Figure 6.4 Decision tree terminology

6.6.3.1 Decision Trees Induction

During the late seventies and early eighties of the last century, Ross Quinlan, a researcher in data mining, succeeded in developing a decision tree algorithm which is ID3 (Iterative Dichotomiser). This has been expanded later to the concept of the data mining to produce the benchmark decision tree algorithm C4.5 as a successor of ID3. These algorithms adopt a non-backtracking method in which decision trees are structured in a top-down manner (Han et al, 2012).

The C4.5 is a decision tree based classification algorithm. It has progressed toward becoming from numerous points of view the business standard for most KDD classification algorithms with many programmes compared against it regarding its performance (Byrne et al, 2005). And it became quite popular after it was ranked number one in the top ten algorithms in data mining (Wu, 2008; Witten, 2017). This algorithm also has a very useful output visualization of decision trees that can provide a smooth way for the interpretation of the relationship between the dependent and explanatory variables (Witten et al, 2017). Therefore, it was selected to build the decision trees in this study as will be explained in Chapter 8.

The process of classifying an instance can be achieved by routing it down the tree according to the attributes values that were tested in a series of successive nodes until the instance reaches the leaf that is assigned for its class.

Decision trees can address the classification pattern in data mining as each leaf node denotes a class in the dataset. To classify an instance that its class label is unknown, the attributes values for that instance are tested against the decision tree beginning from the root node down to the leaf node. According to that, each path in a decision tree can be easily converted into a classification rule. The decision tree structure does not need any area experience. Thus, it is an appropriate scheme for the exploratory feature of knowledge discovery. In addition, the steps of the learning and classification processes in decision trees are simple, fast, and could be easily assimilated by people (Han et al, 2012). Therefore, they have been considered as one of the methods for representing the data mining outputs in this study.

6.6.3.2 Constructing Decision Trees

A recursive process can be utilized to build decision trees. This process starts by considering a partition of the data as a training set. Then the decision tree algorithm creates the first node (root node) by determining a value that is called the *splitting criterion*. The splitting criterion value indicates which attribute to be tested at each node by finding the best way to partition the given instances into individual classes. In addition, it tells which leaves to be grown from the root node. Then, the decision tree algorithm follows the same construction process to form the decision tree. This construction process stops only when one of the following conditions is exist (Han et al, 2012):

1. All the instances in the partitioned data give the same class.
2. The instances cannot be partitioned any further as there are no remaining attributes.
3. There are no instances for a given branch.

6.6.3.3 The Splitting Criterion

Constructing decision trees is a straightforward and simple process in a top-down manner except one problem, which is, which attribute to split on at each node. To solve this problem, a splitting value, which is called (the splitting criterion), is used at each node to determine the best attribute to split on at the first node and to follow this process down the tree. The measure that is used to determine the splitting criterion is called (Information). It is measured by units called (Bits). Therefore, choosing an attribute at the root node means that this attribute gains the most information that affect the classes in the concerned database. This, in turn, means that this attribute has the most impact on the classification process. Then, to go down from that point, a branch is created for each possibility of that attribute. This process will continue to determine the splitting criterion at each node until all the possible leaves are reached (Witten, 2017).

6.6.4 Rules

Decision rules are another common approach for representing data mining outcomes. It is simple to convert a decision tree to a set of classification rules. Each path in a tree can be treated as a rule. The preconditions of a rule are the attributes features in the nodes on the way from the root node to the leaf node of that path. The output of the rule is the class that is represented by the leaf node (Witten, 2017).

The preconditions of a rule are a sequence of tests, just like the tests in a decision tree node that give the class for the instances that belong to that rule. In most cases, the preconditions are generally treated as a package by using the conjunction (AND), which means all the tests should be passed before a set of attributes can be fired as a rule. In other cases of rule construction, the conjunction (OR) is used to indicate that if any one of the attribute tests is passed, then the class can be applied for that instance (Witten, 2017).

6.7 Summary

This Chapter presented the knowledge discovery models as interesting systems to extract the information that may provide useful knowledge in different domains. The steps of knowledge discovery and data mining to extract the engineering knowledge from the concerned database were considered. These steps are data preparation, data mining, evaluation, and knowledge representation.

The data preparation step included data cleaning, integration, and transformation. These was applied to clean the concerned database before using them in building the desired models as will be demonstrated in Chapter 8.

The data mining patterns classification, clustering, association, and prediction were explained. The classification pattern was considered appropriate and will be applied to develop the data mining models in this study.

The evaluation approaches for assessing the data mining models were presented. These are Training set, Cross Validation, and Percentage Split. All these assessment methods will be used, in Chapter 8, to evaluate the data mining models for road safety issues.

Also, the representation methods for the extracted knowledge were explained. Four methods were mentioned Tables, Linear models, Decision trees, and Rules. Two methods will be used to represent the data miming models in this study. The first method are *Tables* because it is the simplest method and the second one are Decision trees because they could provide simple and easily understandable representation approach by following the paths from the root node to the leaf node as rules that may provide useful knowledge. The Linear model method cannot be used in this study as it can be used only with numerical data, which is not the concerned data type in this research. Regarding the Rules method, they are already included in the Decision tree method as each path in the decision tree can be considered as a rule. All the above steps will be applied to develop data mining models for road safety levels.

CHAPTER 7

KNOWLEDGE DISCOVERY FRAMEWORK

7.1 Introduction

This Chapter presents the components of the knowledge discovery framework, that are considered in this study, and the available data mining systems to develop knowledge extraction models for road safety issues. The Chapter is organised in three main parts. The first part presents general description for the knowledge discovery framework. The second part describes the knowledge discovery and data mining systems including the one chosen for the purpose of this study and the reason why it has been chosen. The third part demonstrates the data miming implementations of the chosen system from the computing point of view.

7.2 Development of the KDD Framework

The development framework of the KDD process is illustrated in Figure 7.1. This process generally includes the following components (Chang, 2001):

1. Database Identification
2. Input data
3. Data mining system
4. Model building
5. Model evaluation

These may be elaborated as follows:

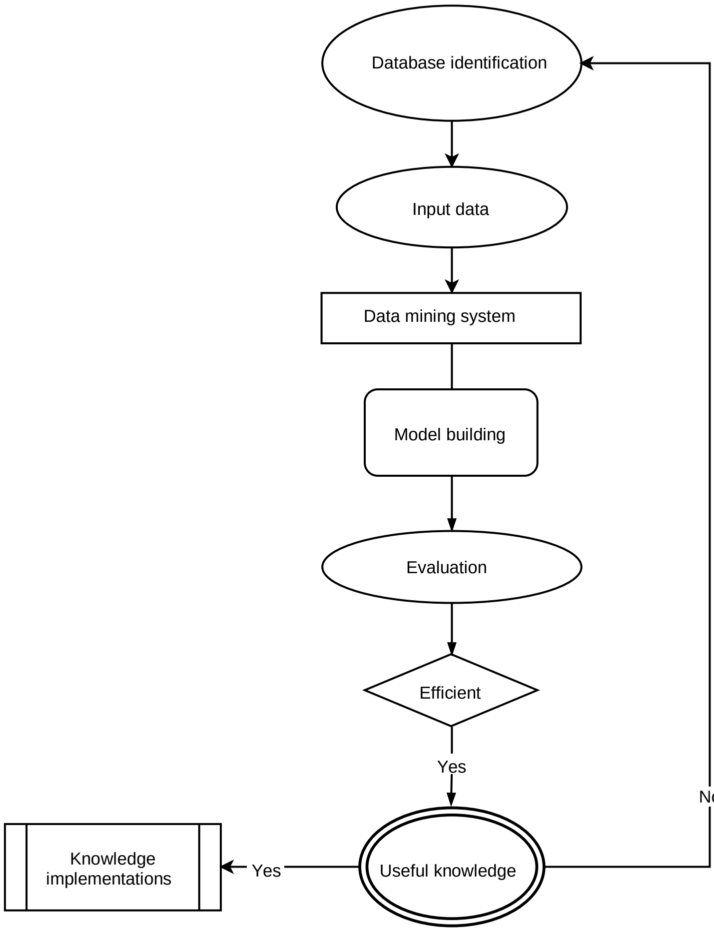


Figure 7.1 The development framework for the knowledge discovery in database

7.2.1 Database Identification

The first step, in the KDD, is the process of identifying the database to be used to develop the data mining models. In any data mining system, it is very important to know the inputs and outputs rather than the processes in between (Witten, 2017). Therefore, it is important to identify the databases that provide information about all the attributes that affect the concerned classes.

7.2.1.1 Databases Types

There are generally three types of databases. These are:

1. **Flat-file databases**: It is the simplest type of databases. In this type, the data are organized in files in the form of rows and columns. These databases may be used to store and retrieve the organized information as references or for analytical purposes. For example, using Microsoft Excel to store the first and last names for a number of students with their marks. Flat - file databases may be easily updated by removing or adding information (DOT, 2001).
2. **Relational databases**: These databases are more complex in comparison with the first type. They are organized by using specific programmes like Microsoft Access, Oracle, and SQL Server. They follow the same structure of the first type by using the rows and columns system (storing the data in tables). This type of databases, like the first one, allow the user to store, retrieve, and manage the data easily. However, it provides higher speed in managing the stored data by using statistical and mathematical functions that make relationships between the data in different tables in the same database (Oracle, 2018).
3. **Non-relational databases**: This is the most recent type of databases (Edlich, 2018). In these databases, instead of the tabular relations (rows and columns), the data are stored in clusters (Leavitt, 2010). This allows the user to track individual words in a specific cluster rather than searching all the stored data. This narrows down the

searching process. This, in turn, increases the speed of retrieving and managing the stored data (Mohan, 2013).

In this study, the concerned database is a flat-file database, which is created as tables by using Microsoft Excel software. The selected database provides information about the attributes that affect road safety levels in addition to their resulted classes. All the data in this database are numerical values.

7.2.2 Input Data

The input data form the most important part in the process of KDD (Rajagopalan et al, 2001). They need to be prepared before using them to build a data miming model. The preparation processes could involve data cleaning to remove errors and missing values, data integration to merge the data from different sources, and data transformation into the appropriate format (as described in Chapter 6, section 6.3). Improving the data quality is a crucial process in building any data mining model, as low data quality will lead to low quality mining results (Han et al, 2012). Therefore, the knowledge extraction process will not succeed without serious data preparation. After preparing the input data, they are entered as a table to a data mining system to build the required models.

7.2.3 Data Mining Systems

The word *systems* here refers to the available data mining programmes. A number of computer programmes may be used for the process of building the data mining models. These may include software such as Waikato Environment for Knowledge Analysis (WEKA), Statistical Analysis Software (SAS), RapidMiner Studio, and Datawatch Monarch software. The benefit of using these programmes in the domain of the KDD is to extract the useful knowledge from the databases and visualise it in an easy way to be understandable and interpretable for the users (Han et al, 2012).

Although data mining is a young domain when compared to statistics, many products of data mining applications are available in the market. Data mining systems are evolving, as new features and visualization methods are constantly added to the existing systems. The efforts to standardize the language for the data mining systems are still underway. Therefore, there are some differences between the available data mining systems (Han et al, 2012).

7.2.3.1 Choosing a Data Mining System

As many data mining systems are available, people may ask which type of data mining systems they should choose. Some people think that all the data mining systems share the same operations and the programming language and, therefore, they behave in a similar way on common functionalities like classification, clustering, association, and numerical prediction. Unfortunately, the reality is far from this assumption because some data mining systems have little consideration to some of these functionalities, could work with different datasets, or could even work on specific computer operating systems. Therefore, some features need to be considered in selecting a data mining system. These features are:

1. Data types
2. System issues
3. File format
4. Data mining functionalities
5. Visualization tools
6. Data mining language and graphical user interface

7.2.3.1.1 Data Types

Many of the available data mining systems can handle nominal (text), numerical (numbers), and symbolic values (like + or -). It is important to check that the data mining system can handle the concerned data before choosing it. Some data types require a particular kind of algorithms, and so, the chosen data mining system should be checked if or not it has these specific algorithms like decision tree algorithms (described in Chapter 6) (Han et al, 2012).

7.2.3.1.2 System Issues

A particular data mining system may work only on a specific operating system or on several. The most two popular operating systems that support the data mining environment are Microsoft Windows and Linux. There are data mining systems that run on other operating systems such as Macintosh (Han et al, 2012).

7.2.3.1.3 File Format

This is related to the input data file format that can operate with the data mining system. Some data mining systems can only operate with text file format. Whereas, many others can work with different types of data such as Microsoft Excel files format (Han et al, 2012).

7.2.3.1.4 Data Mining Functionalities

The core of a data mining system are the data mining functions. These includes functions such as classification, clustering, association, and prediction. Some data mining systems support only one functionality such as clustering. Whereas, others may provide two or more. For a particular function, the systems may provide one or many approaches for this function. Approaches may include neural networks, decision trees, genetic algorithms, and rules algorithms. Different problems may need different types of functions and different approaches may be more effective than others for different types of data (Han et al, 2012).

7.2.3.1.5 Visualization Tools

These tools could provide visualization forms, such as images and graphs that may be considered as an easily understandable method for the data mining models. The visualization in data mining, as a graphical function, can be divided into input visualization, output visualization. The quality and flexibility of a system's visualization tools may strongly affect its usability and attractiveness (Han et al, 2012).

7.2.3.1.6 Data Mining Language and Graphical User Interface

Features like high quality and ease of use for the graphical user interface in a data mining system are essential in promoting user guided and highly interactive data mining. Most of the available data mining systems include friendly graphical user interfaces. However, most of these systems do not share a common data mining programming language. This may make the standardization and the interoperability of the data mining products difficult (Han et al, 2012). As there is no, yet, a standardized data mining system, one need to be chosen for the purpose of this study. One of the systems that provides compatibility between all the above features and the concerned data needs to achieve the ultimate aim of this study is the Waikato Environment for Knowledge Analysis (WEKA) (Witten, 2017).

7.2.3.2 Waikato Environment for Knowledge Analysis

WEKA is a software that includes a collection of data mining algorithms. It provides, by its design, a flexible way to apply existing approaches on new data. It supports the process of mining useful information from databases. That includes, preparing the input data, using statistical techniques to evaluate the machine learning schemes, and visualizing the outputs of the data mining process. This programme was developed in the University of Waikato in New Zealand. The abbreviation WEKA stands for Waikato Environment for Knowledge Analysis. The software is written in java programming language. It can be used on almost any platform and has been checked on different computers' operating systems like Windows, Linux, and Mackintosh (Witten, 2017).

7.2.3.2.1 WEKA Implementations

WEKA includes approaches that consider the main problems in the domain of data mining. These are classification, association, regression, and clustering. There are many implementations that WEKA provides, which can be easily applied on datasets. In addition, WEKA provides some tools for transforming the data as a preparation step. WEKA does not need any programming codes as the programme provides an interactive user interface. One of the WEKA implementations is, applying data mining algorithms on datasets to analyze the output by providing statistical information in the WEKA output sheet about the attributes and the classes in a dataset. For example, information about which attribute has the highest impact on the concerned classes. Another implementation is to use existing models to get predictions on new data. A third one is to apply different models and compare their performance. The graphical user interface, which is called the *Explorer,* provides the easiest way to use the programme. It gives an easy access to all the system functions, for example, a decision tree can be quickly build for a dataset in a spreadsheet file (Witten, 2017).

The following points explain the compatibility of WEKA with the data needs in this study (Witten, 2017):

1. Regarding the data types, WEKA can handle nominal (text) data, which is the type of the concerned data in this study. In addition, it has the required algorithms, such as C4.5, for building decision trees, which are used as one of the representation methods for the extracted knowledge in this study.
2. Concerning the computers' operating systems, WEKA can be installed and work on different operating systems such as Windows, Mac OS X, and Linux. This provides flexibility when using different computers with different operating systems.

3. Also, for the data file format issues, WEKA can handle different file formats like Attribute-Relation File Format (ARFF), and Comma-Separated Values (CSV) (Frank et al, 2016). ARFF are text files that describe a list of instances that sharing a number of attributes. CSV are simple files that can store numerical or nominal data in a format that can be exported or imported into programmes that store data in tables such as Microsoft Excel (Computer Hope, 2018). In this study, the data were stored in CSV format as input data for WEKA system.
4. In respect to the data mining functionalities, WEKA is appropriate for the purpose of this study as it supports the functionality ‘Classification’ that is required for the concerned data (as will be explained in Chapter 8).
5. Regarding the visualization tools, WEKA has the ability to represent the data mining outputs in the required visualization tools. These are, in this research, Decision trees, and Tables.
6. In addition, WEKA is developed by using Java programming language, which provides flexibility for a large number of the developers to write a programme and run it on all the platforms that support Java without the need for the system to recompile it (Oracle, 2018). Therefore, many plugins are developed and can be used with WEKA to improve its performance. One of these plugins is ‘Prefuse Tree’ which is used in this study to improve the visualization of the decision trees by providing additional tools such as changing the

layout (portrait or landscape), changing the font size, and changing the nodes shapes (rectangular or circular).

In addition to all the features above, WEKA is an open source software, easily accessible, and free for all users (Frank et al, 2016).

7.2.4 Model Building

The model building is an iterative process to investigate different data mining models and identify the most accurate and efficient one for achieving the defined aims.

Knowledge extraction models are developed using data mining techniques when algorithms are provided with training sets of data that have pre-classified values of explanatory variables (road attributes) and dependant variable (the star rating). By using the instances in the training set, a provisional data mining model is developed. The next step is to find out how this model will perform when another set of data (test set) from the same database with hidden dependant variables is provided. In this step, the model performs classification according to the knowledge learnt from the training set. Then, the results of the dependent variable from the test set are compared with the actual data to adjust the provisional model. The final step is the model validation. In this step, a validation set is provided (where the values of the target variable are again hidden temporarily from the model) to reduce the error rate and develop the final model. The performance of the model when it is applied for future unseen data is then assessed by observing the percentage of the correctly classified instances as explained in the next section (Larose, 2005).

This process can be accomplished by utilizing a classification algorithm which separates the data into pre-characterized classes, that can be either quantitative (i.e. numbers) or categorical (i.e. text) which could then be used as a data mining model based on known data (Amado, 2000).

7.2.5 Model Evaluation

This part of the knowledge discovery process gives an evaluation of the models' performance and how efficiently the model is able to describe the data and group the attributes features that give each class in the concerned database. The evaluation of the model depends on the percentage of the instances that are correctly classified by the model. The higher the percentage of correctly classified instances, the higher the model accuracy (Król et al, 2017).

It is vital to evaluate the KDD model's accuracy before using it in real word projects. Using models that have low percentages of correctly classified instances, 50% percent or barely above, could lead to unwise decisions because low percentage of correctly classified instances means that the concerned data cannot be mined using this type of models (classification models) (Michael, 2010). The percentage of correctly classified instances should be checked for three different methods of test (Training set, Cross Validation, and Percentage Split) to identify the most accurate model (Witten, 2017).

7.2.5.1 Training set

In the data mining classification pattern, training data are used to build data mining models as classifiers. These data are called the training set. Measuring the performance of a classifier depends on the value of the *Error rate*. The concept of error rate is that all the instances in the test set are classified by a model that is developed from a training set of data. All the test instances that are classified correctly are considered as successes and all the instances that are classified incorrectly are considered as errors. Then, the error rate is determined as the proportion of the errors number over the total number of instances (Witten, 2017).

Calculating the error rate depending on test data from the same set that is used for training the machine and developing the model is a controversial issue. Because the model has been learned from very similar data, which could result in a biased error rate. Therefore, it is recommended to use a test set that has not played a part in building the initial model. This may give a reliable measure for the classifier’s performance. As a rule of thumb, the larger number of instances the better model’s performance. Likewise, the larger data in the test set the more accurate estimation of the error rate (Witten, 2017).

7.2.5.2 Cross Validation

In data mining, the *Holdout* approach is used to, randomly, partition the data between the training and testing sets. It is common to partition the data into three thirds, two for the training set and the third for the test set (Witten, 2017). The samples of the data that are used as training sets should be representative. Generally, it is difficult to tell whether a training set is representative. However, there is a sample checking method that may be beneficial. That is, representing, approximately, the right proportion of each class in the dataset in both the training and testing sets. Omitting a certain class from the training set or over representing it in the test set could affect the classification performance.

To ensure this unbiased representation for all the classes in the data, a statistical approach is utilized. This approach is called C*ross Validation.* In this approach, the data is divided into ten folds (partitions) in which all the classes are represented in about the same amount. Therefore, it is sometimes referred to as the *Tenfold* approach. Each fold of data is held out in turn and the machine is trained on the other nine folds. Thus, the data mining scheme is run ten times. Each run will give an error rate. Then, the average of these ten error rates will be determined to find the final error estimate. The reason behind choosing the number (ten) is that broad tests on huge number of different datasets have shown that the number (ten) is approximately the right number of data partitions that result in the best estimation of the error rate (Witten, 2017).

7.2.5.3 Percentage Split

In this evaluation method, the classifier is evaluated depending on how well a specific percentage of the instances, which is held out for the training set, represents the concerned data. The percentage of the instances that are held out for the training set depends on the user choice or the percentage that provides more accurate models (Witten, 2017). Therefore, in this research, different percentages of the instances for the training set were tested to see which of them gives more accurate data mining models for road safety issues as will be shown in Chapter 8.

All the above three evaluation methods will be used to evaluate the data mining models for road safety levels in this study.

7.2.5.4 The Evaluation Process in WEKA

In evaluating the accuracy of data mining models, WEKA is using a confusion matrix of two dimensions with a column and a row for each class in the dataset. Table 7.1 shows an example of a confusion matrix for a dataset of two classes. True Positives (TP) are the correctly classified instances for a particular class and the True Negatives (TN) are the correctly classified instances for the other class/s. The False Positives (FP) are those that are incorrectly classified positives for a particular class whereas their actual class/s are different. The False Negatives (FN) are the instances that are classified for different class/s while, in fact, they are related into another particular class (Tan et al, 2005).

Table 7.1 A typical confusion matrix of two dimensions

		Described Class	
		Group 1	Group 2
Actual Class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Each value in this matrix represents a number of instances from the dataset for which the described class is the column and the actual class is the row. The confusion matrix can be considered as a preliminary indication for the accuracy of the data mining models. Good accuracy for a data model corresponds to large numbers on the main diagonal of the matrix and small numbers, typically zero, off it (Witten, 2017). Figure 7.2. shows a confusion matrix form WEKA programme for a dataset that includes five classes a, b, c, d and e. It may be seen here, that the number (119), for example, represents the number of instances for which the actual class is the row (class b) and the described class is the column (class b). This means that all these (119) instances were correctly classified by WEKA as their described class is the same as their actual class. Whereas, the number (39), for example, represents the number of

```
=== Confusion Matrix ===

   a   b   c   d   e   <--
 184   0   0   0   0 |   a
   0 119   0  10   0 |   b
   0  39   0   0   0 |   c
   0  10   0  89   0 |   d
   0   0   0   0  84 |   e
```

Figure 7.2 An example of a confusion matrix instances for which the actual class is the row (class c) and the described class is the column (class b). This means that, all these (39) instances were incorrectly classified as (class b) whereas their actual class is (class c). In summary, all the matrix elements on the main diagonal are the correctly classified instances and all the elements off diagonal are the data instances that are classified incorrectly.

Although the confusion matrix gives the error rate as a measure of the overall successes for a model, there is possibility that some classifications may occur by chance. Therefore, and to overcome this problem, WEKA uses a statistical measure called (Kappa statistic), which considers this problem by deducting all the classifications that possibly occur by chance from the total number of successes (correct instances). The maximum percentage of Kappa statistic is 100%. The high the percentage of Kappa, the better the classification of a model (i.e. less classifications occur by chance) (Witten 2017).

7.3 Summary

The Chapter presented the general framework for the knowledge discovery process. The framework was presented as a flowchart that includes a set of steps that are applied in developing knowledge discovery models for road safety issues as will be described in Chapter 8. These are, database Identification, input data, data mining system, model building, and model evaluation.

The identified database in this study consists of sets of instances that include the road safety levels (as the dependent variable), and the affective attributes (as the independent variables). All the data in this database are nominal values. The input data will be prepared, using the techniques described in Chapter 6, before using them to build data mining models, as low quality of input data will lead to inaccurate data mining models. After preparing the input data, they will be entered to a data mining system to build the desired models. The importance of each of the above steps, in the knowledge discovery, was explained to be applied to build and evaluate the required models. In addition, this Chapter demonstrated the available data mining systems and the features or specifications that need to be considered in selecting a data mining system.

Because of the availability of many data mining systems in the market, there is a need to select one. For this research, the Waikato Environment for Knowledge Analysis (WEKA) is chosen because it has all the required features (presented in section 7.2.3.2.1) for building data mining models for road safety issues and the concerned database. For example, it can handle nominal data, it works on different computer operating systems like Windows, Linux, and Mac, and it has the classification algorithms that are required in this study. In addition, it has simple graphical user interface, which makes it easy to use the system, and provides the required representation tools (Tables and Decision trees).

CHAPTER 8

DATA MINING MODELS FOR ROAD SAFETY TREATMENTS

8.1 Introduction

Following the description of the fundamental steps in building data mining models in the previous Chapters, and their potential ability to help in the decision making process to select treatments to improve road safety levels as described in the methodology of this research, this Chapter discusses the process of building data mining models for the available road safety treatments (countermeasures) to improve road safety levels which are measured by the star rating system (described in Chapter 4). These models have the ability to optimise the sets of countermeasures by ranking them according to their effectiveness as exhibited in the field. This Chapter is organized in three main sections. The first one presents the specific description of applying the data mining systems on road safety issues. It describes the features of the practical problem of combining road safety treatments with regard to their effectiveness. It also gives the reason why classification is the appropriate data mining pattern for building models that may be able to make relationships between the countermeasures and road safety levels. The second section presents the process of building these models for the types of accidents considered in this study. Then, the third section shows the components of the data mining framework for this process, and the data mining model which is more accurate and efficient for estimating the combining effect of these countermeasures.

8.2 Data Mining Models for Road Safety

Before the process of building knowledge extraction and data mining models for selecting sets of countermeasures to improve road safety levels, it is necessary to understand the task required from the data mining models with regard to whether it is predictive or descriptive (Amado, 2000). In this research, the required task is descriptive because it seeks to describe the infrastructure conditions that give different levels of road safety in terms of the star rating system (described in Chapter 4).

The selection of the optimum data mining modelling approach should be based on the features of the practical problem to be solved. As the main aim of this study is to define more effective combinations of countermeasures for road sections that suffer low star ratings, the purpose of using data mining techniques in this study is to build models that may be able to classify the road safety treatments for each star rating class of the iRAP model. In other words, the computational problem in this research is how to define the combinations that are associated with each star rating class and then rank them according to their effectiveness. It seems that, the classification algorithm C4.5 that is a decision tree based classification algorithm (described in Chapter 6, section 6.6.3.1) in the data mining system (WEKA) is considered appropriate to address the above features. Defining and ranking the combinations of countermeasures for each star rating may be seen as an automated framework to specify appropriate combinations that result in high star ratings when implemented on sections with low star ratings, as will be explained later in Chapter 9.

8.3 Data Mining Models for Accident Types

To develop data mining models for road safety treatments, packages of countermeasures and their resulting star ratings as input data for these models are required. These packages may be selected with regard to the accident types because different types of countermeasures, which affect the star ratings, can be considered for different types of accidents in the iRAP model. There are four types of accidents in the iRAP model: head-on, run-off, intersection, and property access crashes. The first two types (head-on and run-off crashes) are referred to as non-intersected segment accidents. Whereas, the second two types (intersection and property access damage accidents) are related to intersected segment crashes. Due to data and time limitations, this study focused only on the countermeasures for the non-intersected segment crashes (head-on and run-off). Thus, three packages of countermeasures and their outcome (star ratings) were created as input data to build the models. These are, countermeasures related to head-on crashes, countermeasures related to run-off crashes, and countermeasures for both (head-on + run-off crashes) as shown Table 8.1.

A decision tree model for the specific countermeasures of each type of the non-intersected segment related crashes was considered as shown in Figure 8.1. Therefore, a decision tree was developed to show the treatments that may improve the star rating for head-on crashes and another decision tree was built to present the combinations that may improve the star rating for run-off crashes. After that, a third decision tree was developed for the whole package of the countermeasures that may improve the star rating for all non-intersected segment crashes (head-on + run-off) in iRAP. Each path in these decision trees represents a rule (set of road attributes) that results in a specific class (star rating). Figure 8.1 shows an example of the general form of a decision tree for road safety treatments. As may be seen in this figure, changing the features of road attributes (shown in Table 8.2) may change the resulted star rating. For example, changing the feature of ‘Road attribute 1’ from ‘Feature 3’ into ‘Feature 2’ may change the star rating from SR2 to SR3. If Road attribute 1, in this figure, is assumed ‘Lane width’. Then, changing it from, for example, ‘Narrow’ to ‘Wide’ may change the star rating from SR 2 to SR 3. The maximum number of levels for each of the above decision trees equals the number of the countermeasures considered for each accidents type in Table 8.1. Therefore, the maximum number of levels for head-on, run-off, and head-on + run-off decision trees are 6, 12, and 16 respectively.

Here, a question may be raised about the benefit of the first two models, as all the countermeasures were included in the third one. This is because a specific type of accidents

Table 8.1 Road safety treatments for each accident type in the iRAP model

Attribute	**Accident type**		
	Run-off	**Head-on overtaking**	**Total**
Differential speed limits		x	x
Median type		x	x
Number of lanes		x	x
Gradient	x	x	x
Skid resistance / grip	x	x	x
Lane width	x		x
Curvature	x		x
Quality of curve	x		x
Shoulder rumble strips *	x		x
Road condition	x		x
Roadside severity - passenger-side distance	x		x
Roadside severity - driver-side distance	x		x
Paved shoulder - driver-side	x		x
Paved shoulder - passenger-side	x		x
Delineation	x		x
Centreline rumble strips *		x	x
Total number	**12**	**6**	**16**

* These countermeasures were omitted as they were not present in the considered road network.

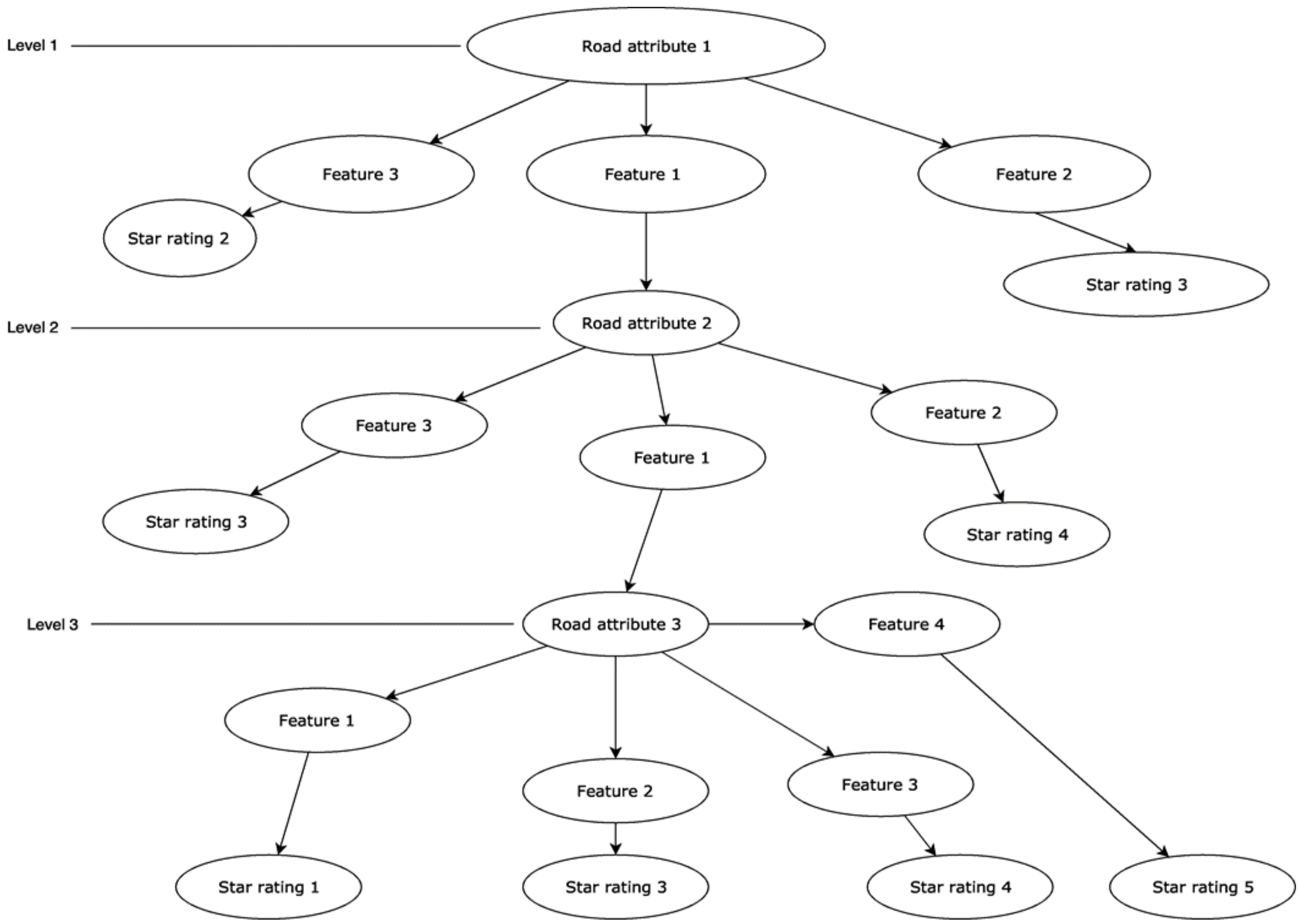


Figure 8.1 An example of the general form of a decision tree for road safety treatments

on non-intersected segments may be predominant. For example, on a one-way road there are no head-on crashes (just the run-off accidents can be considered).

Therefore, it is useful to provide models that can focus on fewer crash related countermeasures rather than considering just the full set of countermeasures, which makes the decision easier for engineers to compare between the available treatments for that particular type of accidents. In addition, developing data mining models for fewer number of road attributes (countermeasures) increases the accuracy of these models as will be shown later in this Chapter.

The countermeasures related to head-on overtaking crashes are adding travel lanes, changing median type (traversable or not), changing gradient, improving skid resistance, providing differential speed limits, and centreline rumple strips. The centreline rumble strips countermeasure was excluded from the analysis as it was not available in the concerned road network. These are all the countermeasures that are used by iRAP to assess and improve road safety issues that are resulted from head-on overtaking crashes.

Table 8.2 The features of the concerned road attributes (iRAP, 2014)

Road attribute	**Feature No.**	**Description**
Median type	1	Traversable
	2	non-traversable
Roadside severity - passenger side distance	1	0 to <1m
	2	1 to <5m
	3	5 to <10m
	4	>=10m
Roadside severity - drivers side distance	1	0 to <1m
	2	1 to <5m
	3	5 to <10m
	4	>=10m
Paved shoulder - drivers side	1	None
	2	Narrow (≥0m to <1.0m)
	3	Medium (≥1.0m to <2.4m)
	4	Wide (≥2.4m)
Paved shoulder - passenger side	1	None
	2	Narrow (≥0m to <1.0m)
	3	Medium (≥1.0m to <2.4m)
	4	Wide (≥2.4m)
Lane width	1	Narrow (≥0m to <2.75m)
	2	Medium (≥2.75m to <3.25m)
	3	Wide (≥3.25m)

Road attribute	Feature No.	Description
Curvature	1	Straight or gently curving
	2	Moderate
	3	Sharp
	4	Very sharp
Gradient	1	≥0% to <7.5%
	2	≥7.5% to <10%
	3	≥10%
Number of lanes	1	One
	2	Two
	3	Three
Differential speed limits	1	Not present
	2	Present
Skid resistance / grip	1	Sealed - adequate
	2	Sealed - medium
Skid resistance / grip	3	Sealed - poor
	4	Unsealed - adequate
	5	Unsealed - poor
Quality of curve	1	Adequate
	2	Poor
Shoulder rumble strips *	1	Not present
	2	present
Road condition	1	Good
	2	Medium
	3	Poor
Delineation	1	Adequate
	2	Poor
Centreline rumble strips *	1	Present
	2	Not present

* These countermeasures were omitted as they were not present in the considered road network.

8.4 Data Mining Framework for Road Safety Treatments

The framework of building data mining models, as a flowchart (shown in Figure 8.2), for road safety levels (the star ratings) in iRAP includes five main steps:

1. Identification of the database: the iRAP database (from India) is used to build the models. It includes the traffic and geometric road conditions (features), in addition, to the star ratings (class labels).
2. Preparation of the iRAP database: in this step the database is cleaned and transformed, as needed, to be used as input data for WEKA as explained previously in Chapter 6. Then, the data from different sets are integrated into one set to be used as an Excel file in WEKA.
3. Building the data mining models: three models are built for each type of accidents using three different methods in WEKA. These are, Training set, Cross Validation, and Percentage Split to compare between their accuracy and select the better.
4. Models evaluation: the percentages of the correctly classified instances are used as indicators to evaluate and compare between the three models, for each type of accidents, to select the most accurate one for extracting the countermeasures combinations for that type.
5. Evaluation of the extracted countermeasures: in this step, the engineering knowledge is needed to check the effectiveness of the mined sets before implementing them to improve the star rating. The overall data miming framework for road safety treatments is presented in Figure 8.2. The following sections explain, thoroughly, the above five steps.

8.4.1 iRAP Databases

For determining road safety levels in iRAP, data are needed to describe the infrastructure, its road safety performance, and its condition as follows. Infrastructure data (such as the lane width, number of lanes, and gradient) are related to road characteristics that remain constant over time whereas condition data (such as speeds and traffic volumes) refer to characteristics that change over time. Performance data (the star rating) represent whether or not road safety levels are improved by using sets of treatments. Infrastructure and condition data can be considered as the heart of the road safety assessment process in iRAP because the decision making process mainly depends on the quality and quantity of these data. Therefore, data were needed to build the relationship between the star ratings and the associated countermeasures. A database from a road network of a representative LMIC was provided, as an Excel file, by iRAP. It contains both the geometric and traffic conditions data for more than thirty-one thousands road segments. The datasets concerned the identification of the geometric characteristics of the roadway such as the number of lanes, the median type, the gradient, the skid resistance, the

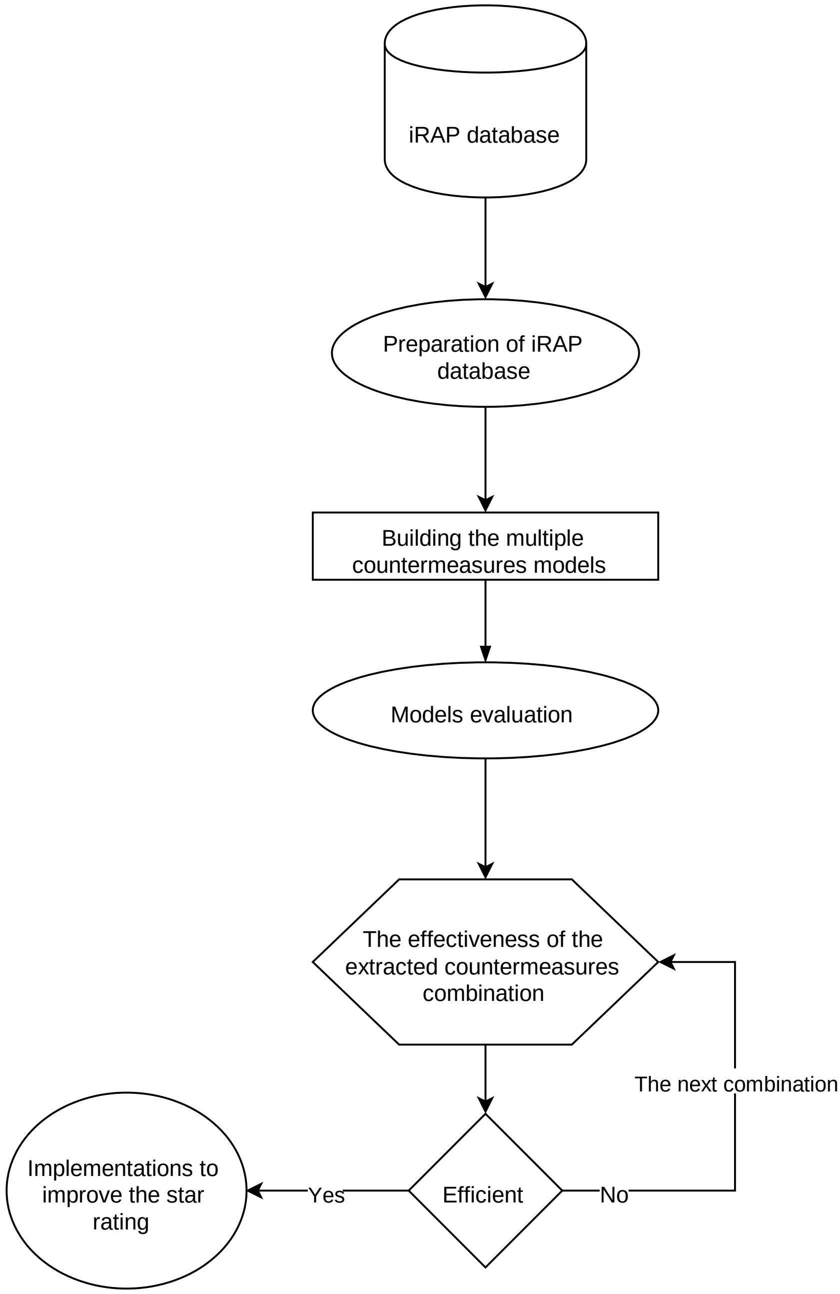


Figure 8.2 The data miming framework for road safety treatments

differential speed limits and the other road attributes presented in Table 8.3. In addition, these sets of data concerned the traffic conditions (operating speeds and traffic volumes) of the sections considered in this study.

These datasets were selected to build data mining models, which reflect the relationship between the star ratings and the use of the multiple road safety treatments in the iRAP model.

Three sets of data were adapted from the database, one for each type of accidents (run-off, head-on, and run-off + head-on) as shown in the presented sample (Table 8.3). Each set contains the geometric road attributes that affect the concerned type of accidents according to the iRAP specifications as shown in Table 8.1, in addition to the star ratings that are resulted from the combinations of these attributes for each section in the considered road network. The data were tabulated and prepared, as an Excel file, for the subsequent analyses. The following section describes the process of preparing an iRAP database before using it in the process of building the knowledge discovery models.

8.4.2 Preparation of the iRAP Database

The collected data, in iRAP, for each road section may include entry errors like missed entries, repeated entries, and wrong values. Therefore, and to consider data verification as an important factor that can affect the accuracy of the developed model, the data preparation techniques discussed in Chapter 6 (section 6.3) were applied to the iRAP database, before using it to build knowledge discovery models for the effect of the multiple countermeasures on road safety, as described below.

Table 8.3. A sample from the iRAP database showing all the data items for the considered accident types

Full set of countermeasures			
Road attributes	**Road section No.**		
	1	2	3
	Features		
Differential speed limits	1	1	2
Number of lanes	1	2	1
Gradient	1	1	2
Skid resistance / grip	5	1	2
Lane width	2	3	1
Curvature	3	2	1
Quality of curve	1	2	1
Road condition	3	3	1
Roadside severity - passenger-side distance	1	4	3
Roadside severity - driver-side distance	2	4	1
Paved shoulder - driver-side	3	2	4
Paved shoulder - passenger-side	2	4	1
Delineation	2	1	2
Median type	1	2	1
Resulted vehicle occupant Star Rating	SR1	SR1	SR2
Head-on crashes related countermeasures			
Differential speed limits	1	1	2
Number of lanes	1	2	1
Gradient	1	1	2
Skid resistance / grip	5	1	2
Median type	1	2	1
Resulted vehicle occupant Star Rating	SR1	SR1	SR2
Run-off crashes related countermeasures			
Gradient	1	1	2
Skid resistance / grip	5	1	2
Lane width	2	3	1
Curvature	3	2	1
Quality of curve	1	2	1
Road condition	3	3	1
Roadside severity - passenger-side distance	1	4	3
Roadside severity - driver-side distance	2	4	1
Paved shoulder - driver-side	3	2	4
Paved shoulder - passenger-side	2	4	1
Delineation	2	1	2
Resulted vehicle occupant Star Rating	SR1	SR1	SR2

8.4.2.1 iRAP Database Cleaning

Data cleaning techniques were used to treat the missing values and remove the noise from iRAP database. These considered the following techniques:

1. Missing Values

For the missing values, the 'Ignoring the tuple' method (described in Chapter 6, section 6.3.1.1) was used to consider the missing values. The reason behind choosing this approach is that the missing values in iRAP databases were noticed in the star rating classes for some instances (road sections). Therefore, and because the star rating is a class label, all the road sections that had missing values in the fields of star rating were omitted to avoid their impact on the models' accuracy as the data mining software could wrongly consider them as an additional star rating class.

2. Noisy Data

In order to remove the noise (errors) from the iRAP dataset, all the attributes values were checked by using the *Filter* feature provided in Microsoft Excel programme to inspect any inconsistences or inappropriate values that are incompatible with iRAP ranges for each attribute. For example, checking the attribute "Lane width" for any values other than 1, 2, or 3, which are the standard groups for this attribute according to iRAP. No errors were found in the concerned iRAP dataset as all the values within the iRAP ranges. This is because all the considered attributes' data were entered by using a specific coding programme, which allows iRAP observers to select one from specific options for each attribute. For example, the lane width in iRAP specifications has three groups to select from (1 - Wide ≥3.25m, 2 - Medium

≥2.75m to <3.25m, and 3 - Narrow 0m to <2.75m). Then, if the lane width for a section between 0 m and <2.75 m, the observer should select number "3" as a value for this section in the "Lane width" attribute in the concerned database.

8.4.2.2 Integrating and Transforming iRAP Databases

Data integration and transformation techniques were applied on the iRAP database to enhance the results of the knowledge discovery and data mining models. These considered the following techniques:

1. Data Integration

For the purpose of preparing one set of data that could be used as an input to WEKA programme, the data from different sets (different roads in the road network) were integrated into one Excel sheet to be readable as one package by the programme.

2. Data Transformation

Mixing the nominal with numerical data formats can influence the way that the data can be useful for the knowledge extraction process (Rajagopalan et al, 2001). All the raw data in the concerned iRAP dataset are numerical values. However, the classification algorithm C4.5 in WEKA uses nominal values. Therefore, all the numerical values of attributes in the iRAP dataset were transformed into nominal values. For example, the three values for the attribute "Lane width" (1, 2, and 3) were transformed into (Lw1, Lw2, and Lw3) respectively. Table 8.4 shows the ranges transformed for the countermeasures that affected the head-on crashes, as an example, which are in line with those applied in the iRAP model. In addition, it shows the iRAP's risk factors for these countermeasures. The lower the risk factor, the safer the road attribute. The ranges transformed for the other two accident types are presented in Appendix C.

8.4.3 The Computational Process of Building the Multiple Countermeasures Models

The data were taken from iRAP as Excel files. By using the filtration feature in the Excel programme, all the road sections that include intersection or property access accidents were omitted because this study focused only on the non-intersected types of accidents. Then, the data were cleaned and transformed into the appropriate format as described in section 8.4.2. After that, the Excel file was uploaded onto WEKA as the input data needed to build the decision tree models sought. There are three approaches for building and evaluating decision tree models in WEKA (described in Chapter 7). The three data mining modelling approaches (mentioned in section 8.4) were applied on head-on, run-off, and head-on + run-off crashes related countermeasures to build three decision trees models for each of the considered accident types and then selecting the most accurate model for each type as follows:

Table 8.4 Transformation ranges for the iRAP head-on related crashes countermeasures

Number of lanes label	**Description**	**iRAP Risk Factor**
NOL1	One lane in each direction	1.0
NOL2	Two lanes in each direction	0.02
NOL3	Three lanes in each direction	0.01
Skid resistance label	**Description**	**iRAP Risk Factor**
SkRes1	Sealed - adequate	1.0
SkRes2	Sealed - medium	1.4
SkRes3	Sealed - poor	2.0
SkRes4	Unsealed - adequate	3.0
SkRes5	Unsealed - poor	5.5
Gradient label	**Description**	**iRAP Risk Factor**
Gr1	≥0% to <7.5%	1.0
Gr2	≥7.5% to <10%	1.2
Gr3	≥10%	1.7
Median Type label	**Description**	**iRAP Risk Factor**
Traversable	Traversable median (only centre line)	1.0
Non-Traversable	Non-traversable median (safety barriers)	0.0
Differential speed label	**Description**	**iRAP Risk Factor**
Difspeed1	Speeds between cars and trucks exceed 20 km/h	1.2
DifSpeed2	Speeds between cars and trucks do not exceed 20 km/h	1.0

1. Training Set

Following the model building process that was described in Chapter 7 (section 7.2.4), a training set, a test set, and a validation set were needed to build the decision tree models. As previously mentioned, it is common to partition the data into three thirds, two for the training set and the third for the test set (Witten, 2017). Therefore, the prepared data package of each of the concerned accident types was divided into three thirds by Microsoft Excel. The first two thirds were considered as the training set for WEKA to learn. The third third was also divided into two equal sets. One was considered as the test set and the other as the validation set. After that, the training set was loaded to WEKA and a provisional model was created for each type of accidents. Then, the data mining scheme was run twice. The first run for the test set and the second one for the validation set to finalize the model. The finalized models showed low error rates, and high percentages of correctly classified instances (sets of countermeasures) 87.6%, 82.9%, and 79.1% for head-on, run-off, and both respectively. This means that, the data mining process to enable WEKA to combine road safety treatments for each star rating is successful because these percentages are highly above 50% percent (Michael, 2010).

2. Cross Validation

In this approach, the ten folds (partitions) Cross Validation method (described in Chapter 7, section 7.2.5.2) was followed by providing the full set of the prepared data for training and testing. Then the data mining scheme was run to build the decision tree model for road safety treatments. The developed model showed low error rates and good percentages of correctly classified instances 87.9%, 81.8%, and 77.8% for head-on, run-off, and both respectively, which indicate that this approach may be considered successful for building data mining models for road safety treatments.

3. Percentage Split

The data mining algorithm was also evaluated on how well it describes a particular percentage of the instances (road sections) which was partitioned to test the developed model. The percentage of the partitioned data depends on which value results in more accuracy. Therefore, all the percentages between 50 and 99% were tried as training sets. The researcher started the partitioning by 50% because the training set should be equal or more than the testing set (Witten, 2017). The results showed that 75% of the data as a training set gives the highest percentages of the correctly classified instances 89.5%, 82.1%, and 77.0% for head-on, run-off, and both respectively. This confirms that this approach, also, may be appropriate to describe road safety issues.

8.4.4 Evaluating the Countermeasures Models

As previously described in Chapter 7, evaluating a data mining model depends on its accuracy. The accuracy of a data mining model can be measured by observing the percentage of the correctly classified instances and selecting the model that gives the highest percentage. After building the data mining models by using the three different approaches in WEKA, as described in section 8.4.3, three data mining models for each type of accidents were developed for describing the infrastructure conditions. Therefore, a comparison between the three models of each accident type was then made depending on their accuracy values as shown in Table 8.5. This comparison showed that the models which were developed by using the first approach (Training set) are the most accurate ones for the run-off and head-on + run-off accident types. Whereas, the third approach (Percentage Split) is the most appropriate one for the head-on crashes with the highest percentage of correctly classified instances (the highest accuracy) as shown in Table 8.5. This means that, the two thirds of data was not the appropriate percentage that sufficiently represents all the star ratings in the head-on crashes dataset. Therefore, another percentage (75%) as a training set gave the highest accuracy for the decision tree model for head-on crashes.

The results also showed that the lower the number of countermeasures the higher the models accuracy. As may be seen in Table 8.5, the head-on accidents with the lowest number of related countermeasures (6 road attributes, Table 8.1) resulted in high percentages of correctly classified instances (85 – 90 range). Whereas, the full set of countermeasures (14 road attributes) resulted in lower ones (75 – 80 range). The percentages of the correctly classified instances could be found in the evaluation part in the output of WEKA decision tree models as presented in Appendix B.

Table 8.5 Percentage of correctly classified multiple countermeasures combinations for all the considered types of accidents

	Training Set	Cross Validation	Percentage Split (75%)
Head-on crashes models	87.6	87.9	**89.5**
Run-off crashes models	**82.9**	81.8	82.1
Head-on + Run-off crashes models	**79.1**	77.8	77.0

8.4.5 The Effectiveness of the Extracted Countermeasures

After the evaluation process of the three data mining models (Training set, Cross Validation, and Percentage Split) for each of the considered types of accidents, the most accurate model for each type of accidents is considered as a decision tree for the countermeasures of that type. Each path in these decision trees is considered as a rule (set of road attributes) that results in a particular star rating. In the WEKA model, to ensure that the most effective, assumed, countermeasures combinations are ranked in descending order, the "Best First algorithm" is used. It is an algorithm that searches the space of the attributes combinations and ranks them descendingly by using an evaluation function. This function evaluates the rules according to their frequency in the input dataset. Therefore, this algorithm puts the rule that matches the highest number of instances, in the dataset, in the beginning, as the first rule, and then ranks the other rules, descendingly, following the same evaluation approach (Russel et al. 2003). The extracted rules should be tested in the iRAP model before it can be considered as useful countermeasures combinations for improving the star raring as will be explained in Chapter 9.

8.5 Summary

This Chapter discussed the data mining methods for road safety issues. That is the data mining methods may overcome the problems of the interpretation of considerable amounts of data, and analysing the nominal (text) values. As the practical problem in this study is to define the road safety treatments combinations for each star rating class, the 'classification' was considered as the most appropriate data mining pattern. Thus, the classification algorithm C4.5, in WEKA, was used to build the decision tree models for road safety issues countermeasures.

Therefore, this Chapter presented the process of building data mining models, as a flowchart, for roads safety levels (the star ratings) in the iRAP model. This process included five main steps. These are: defining the needed database, preparing the selected datasets, building the data mining models sought, evaluating the accuracy of the developed models, and evaluating the effectiveness of the extracted sets of countermeasures. In the first step, the iRAP database was selected to reflect the road safety levels in terms of the star rating system used in the iRAP model. This database included all the required data for the countermeasures related to the accident types in focus (head-on, run-off, and head-on + run-off). Then, the datasets were prepared before using them to build the data mining models. The preparation process included cleaning the data from the missing values and transforming the data into the appropriate format for C4.5 algorithm (nominal values). For cleaning the dataset from the missing values, 'ignoring the tuple' method was used because the missing values were classes (star ratings). The third step included building the data mining models. In this step, three approaches were used (Training set, Cross Validation, and Percentage Split). The data were partitioned as needed for each of these approaches. In the fourth step, the accuracy of each approach was tested with regard to the percentages of the correctly classified instances. This is to define the most accurate model that can be used to build the decision tree for each type of accidents. The results showed that the 'Percentage Split' approach produced the most accurate data mining model for head-on crashes related countermeasures, and the 'Training set' method resulted in the most accurate for run-off and head-on + run-off crashes related countermeasures. This indicated that, the two thirds of data, which is the ideal percentage of data for training (Witten, 2017), was not the appropriate percentage to represent all the star ratings in the heads-on crashes dataset. The above approaches were considered to define the extracted combination of countermeasures. The fifth step included explanation about testing the effectiveness of the extracted sets in the iRAP model. This is to ensure that the extracted combinations are effective and applicable, with regard to the iRAP specifications, before implementing them to improve road safety levels as will be explained in Chapter 9.

CHAPTER 9

THE EXTRACTED KNOWLEDGE

9.1 Introduction

Following the process of developing decision trees and tables in Chapter 8, 2294 combinations of countermeasures for the concerned accident types (28, 605, and 1661 for head-on, run-off, and head-on + run-off respectively) were mined. The reliability of the extracted knowledge for each crash type should be tested before it can be considered as useful knowledge for road safety aspects. As proposed in the methodology of this research, the extracted combinations, for each star rating, were ranked descendingly and the first combinations of countermeasures for each star rating is assumed as the best set of countermeasures that produces that particular star rating. Therefore, the less the rule's number the more is the effectiveness of the combination. To check whether a set of countermeasures produces the actual star rating, the road attributes of that set should be provided into the iRAP model as input data to calculate the actual star rating. Then, this star rating should be compared with the star rating that was calculated by WEKA. If the actual and assumed star ratings are similar, then the concerned set of countermeasures may be considered to improve the star rating.

To cover the above, this Chapter is divided into three main sections. The first section presents the extracted knowledge for head-on crashes countermeasures and its effectiveness in improving road safety levels. The second section deals with the extracted knowledge for the run-off accidents type. Finally, the third section discusses the effectiveness of the extracted knowledge for the full set of the treatments that are related to the head-on + run-off crash type.

9.2 The Extracted Knowledge for Head-on Crashes

The knowledge (combinations of countermeasures) extracted from the data mining model for head-on crashes is illustrated in Table 9.1. To calculate the actual star ratings in the iRAP model, all the input data of the considered roads were entered to the ViDA software, which is the official iRAP online software to calculate the star ratings. The key requirements for this software are the infrastructure conditions of the concerned road network. These include all the traffic and geometric characteristics to calculate the star rating.

To validate the data mining model for the head-on crashes, each of the extracted combinations for head-on crashes countermeasures was applied on the road sections with low star ratings (SR1 or SR2) and only head-on crashes to check the star ratings, for these sections, after applying the road attributes of the extracted combinations. To ensure there were no run-off crashes, the sections were chosen with no roadside objects. In addition, to make sure that the crashes on these sections are head-on, the sections were chosen from single lane carriageways with traversable median types to ensure the need for the overtaking manoeuvres that caused the head-on crashes. The results of this test showed that the star ratings for 12 of the extracted combinations for head-on crashes countermeasures (rules 1 - 12 illustrated in Table 9.1) are in agreement with the actual star rating calculated by ViDA software. The agreement level between the extracted and actual star ratings reached 87% as shown in Table 9.1.

Table 9.2 shows an example of using the extracted sets to improve the star rating from SR1 to SR3 for a road section with regard to the head-on crashes related countermeasures. As may be seen in this table, the existing star rating of the considered section is SR1. The suggested sets of treatments were ranked by WEKA according to their effectiveness. Then, to check the validity of these suggested sets, the set ranked number 1, as the most accurate set according to the data mining process, was tested in the iRAP model. This is by applying the suggested road attributes of this rule onto iRAP model to calculate the actual star rating and compare it with the assumed. The test result showed that the actual star rating, from iRAP, was the same as the assumed from data mining. Therefore, this set could be considered appropriate to improve road safety (raising the star rating) for the concerned section by considering the differences between the existed and the suggested road attributes (set number 1). In this example, these differences were: eliminating the differential speed limit (changing DifSpeed2 to DifSpeed1) and improving the skid resistance feature (from SkRes3 to SkRes1). If the first set was not appropriate (i.e. the assumed and actual star ratings were not similar), the same process continues to check the effectiveness of the sets descendingly until the appropriate set is reached. The extracted combinations for the head-on crashes countermeasures are presented in Appendix D.

Table 9.1 The extracted knowledge for head-on crashes

Rule No.	Differential speed limits	Median type	Number of lanes	Gradient	Skid resistance / grip	Star rating	Agreement level %
1	DifSpeed1	Traversable	NOL1	Gr1	SkRes1	SR3	**87**
2	DifSpeed1	Traversable	NOL1	Gr1	SkRes2	SR3	81
3	DifSpeed2	Traversable	NOL1	Gr1	SkRes2	SR3	79
4	DifSpeed1	Traversable	NOL1	Gr2	SkRes1	SR3	72
5	DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes2	SR4	**85**
6	DifSpeed2	Non-Traversable	NOL2	Gr1	SkRes1	SR4	80
7	DifSpeed1	Non-Traversable	NOL2	Gr2	SkRes1	SR4	79
8	DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes1	SR5	**86**
9	DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes2	SR5	81
10	DifSpeed2	Non-Traversable	NOL2	Gr1	SkRes1	SR5	77
11	DifSpeed1	Non-Traversable	NOL2	Gr2	SkRes1	SR5	75
12	DifSpeed2	Non-Traversable	NOL2	Gr2	SkRes1	SR5	75

Table 9.2 An example of using data mining extracted rules to improve the star rating

Rule No.	Differential speed limits	Median type	Number of lanes	Gradient	Skid resistance / grip	Star rating
The existing situation is:						
	DifSpeed2	Traversable	NOL1	Gr1	SkRes3	SR1
The mined improvement rules are:						
1	DifSpeed1	Traversable	NOL1	Gr1	SkRes1	SR3
2	DifSpeed2	Traversable	NOL1	Gr1	SkRes2	SR3
3	DifSpeed1	Traversable	NOL1	Gr1	SkRes2	SR3
4	DifSpeed1	Traversable	NOL1	Gr2	SkRes1	SR3

To this end, it may be suggested that the most effective combination of countermeasures for head-on crashes on single lane carriageways (for the concerned road network) with traversable median types and gradient ≥0% to <7.5% that raises the star rating to SR3 is rule number 1 in Table 9.1 which requires that differential speed limits should not be presented and having adequate skid resistance. In addition, it may also be suggested that the road sections with higher gradient (7.5% to <10%) for the same conditions above could still have SR3 as shown in rule number 4 in Table 9.1.

And the most effective combination for head-on crashes on dual lane carriageways (for the concerned road network) with gradient ≥0% to <7.5% to raise the starting into SR4 is rule number 5 in Table 9.1 that requires: providing non-traversable median types, differential speed limits should not be presented, and medium skid resistance.

Also, the most effective combination to raise the star rating into SR5 is rule number 8 in Table 9.1, which requires the same above conditions for SR4 except improving the skid resistance from medium to adequate.

The most effective combination for each star rating (rules number 1, 5, and 8 for star ratings 3, 4, and 5 respectively) was applied on the low-star rating road sections (SR1 or SR2) that have only head-on crashes to calculate the expected reduction in fatality rates. The reductions percentages were calculated using the adjusted iRAP methodology for multiple countermeasures (explained in Chapter 5). The results showed that the reduction percentages are 13%, 19%, and 29% for star ratings 3, 4, and 5 respectively as shown in Table 9.3.

Table 9.3 The reduction in fatality rates caused by head-on crashes

	Star rating 3 (Rule No. 1)	**Star rating 4 (Rule No. 5)**	**Star rating 5 (Rule No. 8)**
Deaths (per year)			
Before countermeasures	56	56	56
After countermeasures	49	45	39
Prevented	7	11	16
Reduction %	13%	19%	29%
Estimated benefit form fatality reduction per year (£) *	735,980	1,156,540	1,682,240

* The value of a statistical life is about 70 times the Gross Domestic Product (GDP) per capita (McMahon and Dahdah, 2008). GDP for India £1502 (World Bank, 2017).

As may be seen from the above infrastructure characteristics for the three star ratings (3, 4, and 5), that the most effective countermeasures to improve the star rating from SR3 to SR4 are: changing the median type from traversable to non-traversable and adding another lane. This seems reasonable as providing non-traversable median such as median barriers and adding another lane for overtaking manoeuvres can eliminate the head-on crashes. However, these countermeasures may not be realistic in cases where the road safety improvements budget is limited. Furthermore, it was found that improving the skid resistance from medium to adequate could make the difference between SR4 and SR5.

The decision tree for the countermeasures that affect head-on crashes was chosen to be presented in this Chapter (Figure 9.1) as an example of road countermeasures and their impact on road safety levels in the iRAP model. The output of WEKA decision tree models for the other two accident types are presented in Appendix B.

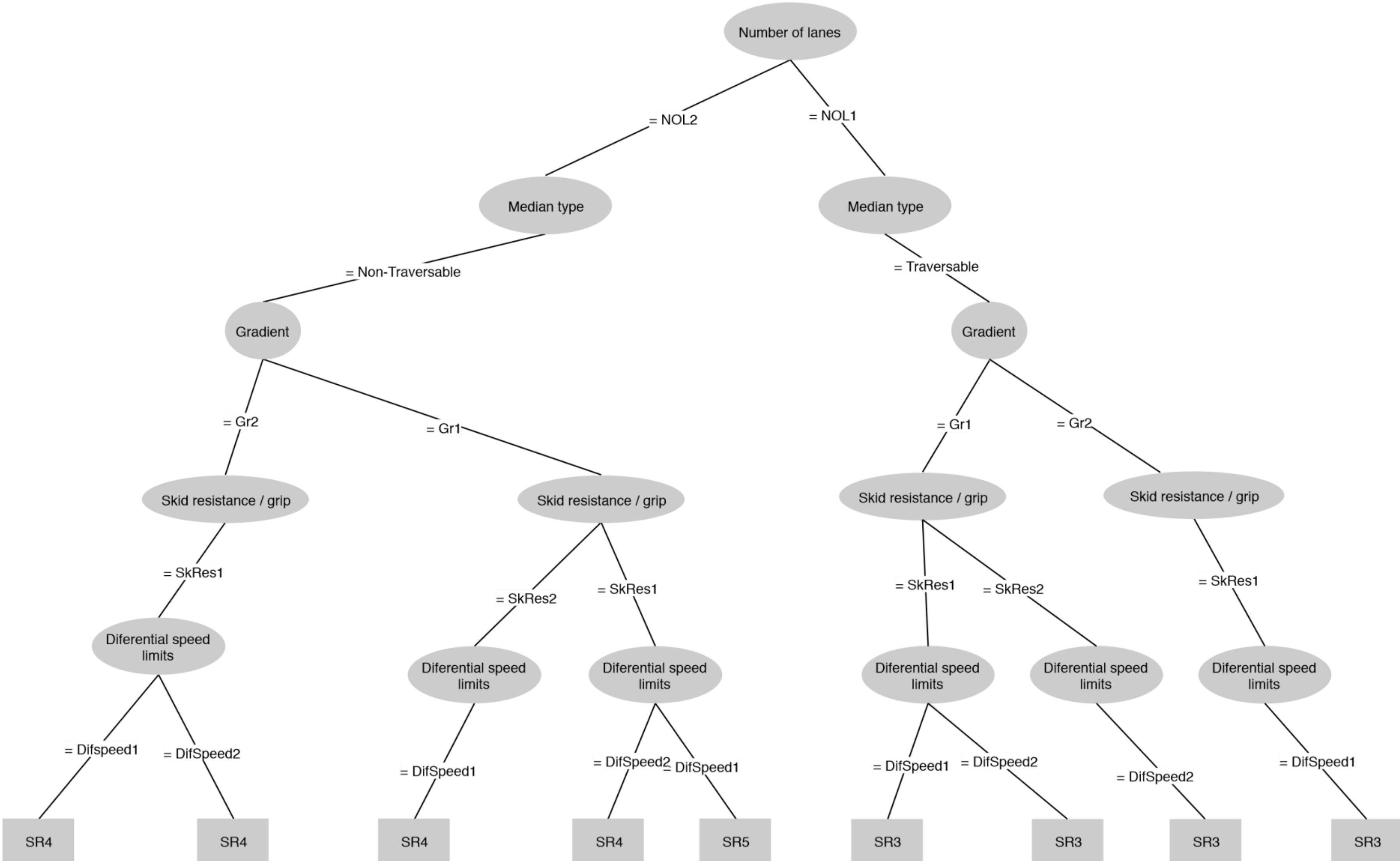


Figure 9.1 The decision tree for the countermeasures related to head-on crashes

9.3 The Extracted Knowledge for Run-off Crashes

The results showed that there are 605 possible treatment combinations of road attributes that are related to the run-off crashes. Table 9.4 shows the first five combinations for each star rating (3, 4, and 5) respectively, which their accuracy level no less than 63%. The full sets of the extracted treatments for run-off accidents are illustrated in Appendix E. Because of the big number of combinations and to test the effectiveness of the extracted countermeasure combinations, a sample of 150 combinations (the first fifty for each star rating (3, 4, and 5) were considered to calculate the actual star ratings and compare them with the expected from the data mining. To calculate the actual star ratings in the iRAP model, all the input data of the considered roads were entered to the iRAP ViDA software. These include all the traffic and geometric characteristics to calculate the star rating.

Each of the extracted countermeasures in this sample was applied on the road sections with low star ratings and only run-off accidents. To ensure there were no head-on crashes, the sections were chosen from dual carriageway roads (raised physical median or median barriers). The results showed that the expected star ratings were compatible with the actual star ratings. This reached 80% agreement level between the expected and the actual star ratings as shown in Table 9.4. It may be suggested that the most appropriate set of treatments for run-off accidents on single or dual lane carriageways (for the concerned road network) with gradient ≥0% to <7.5% to obtain star rating 3 (SR3) is rule number 1 in Table 9.4 with the highest agreement level. This rule requires medium skid resistance, medium lane width

(≥2.75m to <3.25m), straight or gentle curvature (radius greater than 900m), adequate quality of curve (advance warning signs of bend and roadside indicators to indicate the curvature), short roadside distance (1 to <5m), narrow paved shoulder (≥0m to <1.0m), and adequate delineation.

And the most effective combination to raise the starting into SR4 for run-off crashes on single or dual lane carriageways (for the concerned road network) with gradient ≥0% to <7.5% is rule number 6 in Table 9.4. It may be seen that improving the star rating from SR3 to SR4 requires adequate skid resistance, wide lane width (≥3.25m), adequate quality of curve, good road condition, medium roadside distance (5 to <10m), wide paved shoulder (≥2.4m), and adequate delineation.

In addition, it may be proposed that the most appropriate combination to raise the star rating into SR5 is rule number 11 in Table 9.4. It may be seen that there is a difference in one road condition between this rule and rule number 6 (SR4), this is by increasing the roadside distance from medium (5 to <10m) to long distance (>=10m).

It may be concluded by simple comparison between the infrastructure characteristics in rules (1, 6, and 11) in Table 9.4, for star ratings 3, 4, and 5, that the run-off crashes countermeasures that could raise the star rating from SR3 to SR4 are: improving the skid resistance, clearing more distance from hazardous objects on both roadsides, and providing wider shoulders. Also, it was found that increasing the roadside distance is the most effective run-off crashes countermeasure to improve the star rating from SR4 to SR5.

Table 9.4 Sample of the extracted knowledge for run-off crashes

Rule No.	Gradient	Skid resistance / grip	Lane width	Curvature	Quality of curve	Road condition	Roadside severity - passenger-side distance	Roadside severity - driver-side distance	Paved shoulder - driver-side	Paved shoulder - passenger-side	Delineation	Star Rating	Agreement level %
1	Gr1	SkRes2	LW2	Cu1	QoC1	RC2	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR3	**80**
2	Gr1	SkRes2	LW2	Cu3	QoC2	RC2	RSpsd2	RSdsd2	PSds2	PSps2	De2	SR3	78
3	Gr1	SkRes2	LW1	Cu3	QoC2	RC2	RSpsd2	RSdsd2	PSds2	PSps2	De2	SR3	76
4	Gr1	SkRes2	LW1	Cu1	QoC2	RC3	RSpsd2	RSdsd3	PSds3	PSps2	De2	SR3	75
5	Gr1	SkRes3	LW1	Cu1	QoC2	RC3	RSpsd3	RSdsd3	PSds3	PSps3	De2	SR3	75
6	Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps4	De1	SR4	**80**
7	Gr1	SkRes1	LW3	Cu2	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps4	De1	SR4	74
8	Gr1	SkRes1	LW3	Cu1	QoC2	RC2	RSpsd3	RSdsd2	PSds4	PSps4	De1	SR4	71
9	Gr1	SkRes2	LW2	Cu1	QoC2	RC2	RSpsd2	RSdsd2	PSds3	PSps3	De1	SR4	66
10	Gr1	SkRes2	LW2	Cu2	QoC2	RC2	RSpsd2	RSdsd2	PSds3	PSps3	De1	SR4	65
11	Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd4	RSdsd4	PSds4	PSps4	De1	SR5	**78**
12	Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps4	De1	SR5	72
13	Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps3	De1	SR5	68
14	Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps3	De1	SR5	64
15	Gr1	SkRes1	LW3	Cu1	QoC2	RC2	RSpsd3	RSdsd3	PSds3	PSps4	De1	SR5	63

The most effective combination for each star rating (rules number 1, 6, and 11 for star ratings 3, 4, and 5 respectively) was applied on the low-star rating road sections (SR1 or SR2) that have only run-off crashes to calculate the expected reduction in fatality rates. The results showed that the reduction percentages are 12%, 19%, and 31% for star ratings 3, 4, and 5 respectively as shown in Table 9.5.

Table 9.5 The reduction in fatality rates caused by run-off crashes

	Star rating 3 (Rule No. 1)	**Star rating 4 (Rule No. 6)**	**Star rating 5 (Rule No. 11)**
Deaths (per year)			
Before countermeasures	42	42	42
After countermeasures	37	34	29
Prevented	5	8	13
Reduction %	12%	19%	31%
Estimated benefit form fatality reduction per year (£) *	525,700	841,120	1,366,820

* The value of a statistical life is about 70 times the Gross Domestic Product (GDP) per capita (McMahon and Dahdah, 2008). GDP for India £1502 (World Bank, 2017).

9.4 The Extracted Knowledge for Head-on + Run-off Crashes

The results showed that there are 1661 possible countermeasures combinations of road attributes that are related to the (head-on + run-off) crashes (Appendix F). A test was carried out to determine the agreement level, between the suggested and actual star ratings, for the full set of countermeasures (head-on + run-off). As there are a huge number of combinations that need long time to be verified, a sample of 150 combinations (the first fifty for each star rating (3, 4, and 5) was considered to test the effectiveness of the mined knowledge and to verify the data mining model for the full set. Each of these combinations was applied on all the road sections that have low star ratings and (head-on + run-off) crashes. The results showed that the maximum agreement level reached between the suggested star ratings (from the data mining model) and the actual star ratings calculated directly from iRAP was 74% as shown in Table 9.6.

To this end, it may be proposed that the most effective combination of countermeasures, for head-on + run-off crashes on single lane carriageways with traversable median types and gradient ≥0% to <7.5%, that improves the star rating into SR3 is rule number 1 in Table 9.6. This requires that the differential speed limits should not be presented, medium skid resistance, medium lane width (≥2.75m to <3.25m), straight or gentle curvature (radius greater than 900m), adequate quality of curve (advance warning signs of bend and roadside indicators to indicate the curvature), short roadside distance (1 to <5m), narrow paved shoulder (≥0m to <1.0m), and adequate delineation.

And, to obtain star rating 4 (SR4), rule number 6 in Table 9.6 may be considered the most appropriate with the highest agreement level for this star rating. For head-on + run-off crashes on dual lane carriageways with non-traversable median types and gradient ≥0% to <7.5%, this rule requires that the differential speed limits should not be presented, adequate skid resistance, wide lane width (≥3.25m), adequate quality of curve, good road condition, medium roadside distance (5 to <10m), wide paved shoulder (≥2.4m), and adequate delineation.

Also, to reach star rating 5 (SR5) using the head-on + run-off crashes countermeasures, it may be seen that rule number 11 in Table 9.6 has the highest agreement level to achieve this star rating. This rule suggests one difference from rule number 6 (for SR4), that is longer roadside distance (>=10m) is required to achieve SR5.

It may be seen from the above countermeasures for star ratings 3, 4, and 5 that providing non-traversable median type, adding another lane, increasing the lane width, improving the road condition (repairing cracks and patches), increasing the roadside distance, and widening the shoulders were the head-on + run-off crashes countermeasures, for the concerned road network, that could raise the star rating from SR3 to SR4. However, and because of cost issues, it seems that the infrastructure attributes that could provide SR3 are more realistic. In addition, it was found that longer roadside distance is the countermeasure that makes the difference between SR4 and SR5 in considering the head-on + run-off crashes.

Table 9.6 Sample of the extracted knowledge for head-on + run-off crashes

Rule No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Differential speed limits	DifSpeed1	DifSpeed1	DifSpeed1	DifSpeed2	DifSpeed2	DifSpeed1	DifSpeed1	DifSpeed1	DifSpeed1	DifSpeed1	DifSpeed1	DifSpeed1	DifSpeed1	DifSpeed1	DifSpeed1
Median type	Traversable	Traversable	Traversable	Traversable	Traversable	Non-Traversable	Non-Traversable	Non-Traversable	Non-Traversable	Non-Traversable	Non-Traversable	Non-Traversable	Non-Traversable	Non-Traversable	Non-Traversable
Number of lanes	NOL1	NOL1	NOL1	NOL1	NOL1	NOL2	NOL2	NOL2	NOL2	NOL2	NOL2	NOL2	NOL2	NOL2	NOL2
Gradient	Gr1	Gr1	Gr1	Gr1	Gr1	Gr1	Gr1	Gr1	Gr1	Gr1	Gr1	Gr1	Gr1	Gr1	Gr1
Skid resistance / grip	SkRes2	SkRes2	SkRes2	SkRes2	SkRes3	SkRes1	SkRes1	SkRes1	SkRes2	SkRes2	SkRes1	SkRes1	SkRes1	SkRes1	SkRes1
Lane width	LW2	LW2	LW1	LW1	LW1	LW3	LW3	LW3	LW2	LW2	LW3	LW3	LW3	LW3	LW3
Curvature	Cu1	Cu3	Cu3	Cu1	Cu1	Cu1	Cu2	Cu1	Cu1	Cu2	Cu1	Cu1	Cu1	Cu1	Cu1
Quality of curve	QoC1	QoC2	QoC2	QoC2	QoC2	QoC1	QoC1	QoC2	QoC2	QoC2	QoC1	QoC1	QoC1	QoC1	QoC2
Road condition	RC2	RC2	RC2	RC3	RC3	RC1	RC1	RC2	RC2	RC2	RC1	RC1	RC1	RC1	RC2
Roadside severity - passenger-side distance	RSpsd2	RSpsd2	RSpsd2	RSpsd2	RSpsd3	RSpsd3	RSpsd3	RSpsd3	RSpsd2	RSpsd2	RSpsd4	RSpsd3	RSpsd3	RSpsd3	RSpsd3
Roadside severity - driver-side distance	RSdsd2	RSdsd2	RSdsd2	RSdsd3	RSdsd3	RSdsd3	RSdsd3	RSdsd2	RSdsd2	RSdsd2	RSdsd4	RSdsd3	RSdsd3	RSdsd3	RSdsd3
Paved shoulder - driver-side	PSds2	PSds2	PSds2	PSds3	PSds3	PSds4	PSds4	PSds4	PSds3	PSds3	PSds4	PSds4	PSds4	PSds4	PSds3
Paved shoulder - passenger-side	PSps2	PSps2	PSps2	PSps2	PSps3	PSps4	PSps4	PSps4	PSps3	PSps3	PSps4	PSps4	PSps3	PSps3	PSps4
Delineation	De1	De2	De2	De2	De2	De1	De1	De1	De1	De1	De1	De1	De1	De1	De1
Star Rating	SR3	SR3	SR3	SR3	SR3	SR4	SR4	SR4	SR4	SR4	SR5	SR5	SR5	SR5	SR5
Agreement level %	**74**	70	66	64	63	**71**	67	64	64	62	**72**	68	63	60	59

The most effective combination for each star rating (rules number 1, 6, and 11 for star ratings 3, 4, and 5 respectively) was applied on the low-star rating road sections (SR1 or SR2) that have head-on and run-off crashes to calculate the expected reduction in fatality rates. The results showed that the reduction percentages are 12%, 22%, and 30% for star ratings 3, 4, and 5 respectively as shown in Table 9.7. In addition, by considering that the value of a fatal accident is about 70 times the Gross Domestic Product (GDP) per capita (McMahon and Dahdah, 2008) and that the GDP per capita for India is £1502 (World Bank, 2017), the estimated benefit from the reduction of fatalities per year may be also seen in Table 9.7.

Table 9.7 The reduction in fatality rates caused by head-on and run-off crashes

	Star rating 3 (Rule No. 1)	**Star rating 4 (Rule No. 6)**	**Star rating 5 (Rule No. 11)**
Deaths (per year)			
Before countermeasures	180	180	180
After countermeasures	160	146	125
Prevented	22	40	54
Reduction %	12%	22%	30%
Estimated benefit form fatality reduction per year (£) *	2,313,080	4,205,600	5,677,560

9.5 Summary

Ranking the combinations of countermeasures is one of the essential aspects in road safety. The engineering facilities for improving road safety should be planned, designed, and maintained in consideration of the most effective combinations of countermeasures to ensure reduction of crashes and effective usage of the available infrastructure resources.

According to the results of this study, the developed data mining models extracted useful combinations of countermeasures and ranked them with regard to their frequency in the input data. For that, the 'Best first' algorithm in WEKA system was used as described in Chapter 8.

The agreement level between the suggested and the actual star ratings for all the extracted countermeasures combinations for head-on crashes, the 150 combinations sample of run-off crashes countermeasures, and the 150 combinations sample of the head-on + run-off crashes countermeasures were examined. The results showed high agreement levels between the star ratings suggested from the developed data mining models and the star ratings calculated by iRAP using the mined countermeasures combinations. The maximum agreement levels reached for the head-on, run-off, and head-on + run-off crashes were 87%, 80%, and 74% respectively.

A comparison between the above three agreement levels may show that the agreement level between the expected and actual star ratings for each single accidents type (head-on or run-off) is higher than the agreement level for the mixed combinations of the two accident types (head-on + run-off). This confirmed that the data mining models produced a reliable accuracy. This is because, if the accuracy increased with increasing the number of attributes, it may indicate that there is inter-correlation among the attributes, which could produce overfitting (meaningless accuracy) (Eaton, 2018).

The results, also, showed that the first combination in each star rating improved the existing road safety into the suggested levels for more road sections than the second combination. And the second combination raised the star ratings for more sections than the third extracted set. The same was found for the other combinations in a top-down manner. This verified that the data mining techniques have successfully extracted and ranked multiple countermeasures for the three concerned accident types and also confirmed the validity of the methodology used in this study for combining and ranking road safety treatments.

In conclusion, for road sections with only head-on crashes in the concerned road network, it may be suggested that adequate level of skid resistance (the highest level of this attribute in iRAP) and no differential speed limits could provide SR3 on single lane carriageways with gradients (0 – 10%). Moreover, changing the median into non-traversable and adding an additional lane may be the only possibility to achieve SR4 on these roads even with medium skid resistance. However, improving it to adequate could play a crucial role in gaining SR5 for these conditions.

For road sections with only run-off crashes in the concerned road network, it may be concluded that five attributes (medium skid resistance, medium lane width, medium road condition, at least narrow paved shoulder, and adequate delineation) could be the most effective countermeasures that may raise the star rating into the acceptable level (SR3). Four of these countermeasures could be improved (adequate skid resistance, wide lane width, good road condition, wide paved shoulder) in addition to medium roadside distance to raise the star rating on these roads into SR4. To get SR5, one of these conditions should be improved, that is through increasing the roadside distance into long (>=10m).

For road sections with both head-on and run-off crashes in the concerned road network, it may be suggested that providing medium skid resistance, no differential speed limits, medium lane width, medium road condition, narrow paved shoulder, and adequate delineation are the most effective conditions to raise the star rating into SR3. In addition, improving four of these conditions (adequate skid resistance, wide lane width, good road condition, wide paved shoulder) in addition to medium roadside distance could raise the star rating from 3 to 4. Furthermore, increasing the roadside distance into long with the previous conditions may make the difference between star ratings 4 and 5.

CHAPTER 10

DISCUSSION

10.1 Introduction

This research covered two issues in road safety. The first one is the problem of overestimation of multiple countermeasures' effectiveness in iRAP and the second is the question about which combinations of road crashes countermeasures are more effective to improve the star rating. With regard to the problem of overestimation of multiple countermeasures' effectiveness in the iRAP model, this study showed that statistical models may be developed using the Negative Binomial regression approach to adjust the MCCF in the iRAP model to estimate the future numbers of fatalities more accurately. However, these models rely on the availability of reliable historical crash data. Also, with regard to the second issue, this study showed that decision trees and tables may be developed to extract knowledge from road infrastructure databases with regard to the most effective combination of road safety treatments aimed at addressing head-on and run-off crashes. This is by using the data mining techniques as an appropriate alternative for the traditional statistical models in considering the effect of road traffic and geometric conditions on road safety levels.

This Chapter discusses the results of these models and the usefulness of the extracted knowledge for each accidents type. It also discusses the effectiveness of the extracted knowledge. In addition, it discusses the complexity of the computational process before and after the data mining process. This include preparing the databases, using the data mining system, and interpreting the extracted knowledge.

10.2 Value and Impact of the Data Mining Approach

For the purpose of developing models for ranking combinations of countermeasures, which are related to head-on, run-off, and head-on + run-off crashes, this research adopted the data mining technique. This is based on developing decision trees and tables to extract and compare road safety treatments combinations and rank them with regard to their effectiveness instead of relying only on the traditional statistical methods (described in Chapter 5), which develop prediction models from historical crash data. This is especially important in low and middle income countries as reliable history of accident records may not be available (Phan et al, 2016). The chosen data mining pattern in this study is Classification that is based on the statistical relationships between road attributes and road safety levels expressed in star ratings in iRAP. This is effected by correlating the star ratings with the recurrence of the features of road attributes, in other words, the features that result in each star rating. Therefore, by using the classification pattern in data mining, the combinations of the most frequent features of road attributes associated with each star rating may be determined.

As road safety forms an important aspect in road transport, suggesting improvements for this aspect should consider all the possible countermeasures and rank them to achieve plans that are more effective. Road safety may be measured by different indicators such as crash rates, crash severity, and star ratings. As the star rating system consider the road infrastructure, the data mining method may offer the decision makers the opportunity to compare between the existing and future situations after improving the road infrastructure.

Although the origin of data mining is statistics, these models have more flexibility over the statistical models. This is because they have the ability to develop models that combine the existing road infrastructure for each road safety level (star rating). The difference between road safety levels could be considered as a measure for the impact of changing road attributes instead of relying on historical crash data to reflect the effectiveness of this change, like in the statistical models, especially, in areas where reliable historical crash data are not available. The purpose of using star ratings of road infrastructure as a road safety measure is that, the improvement of road infrastructure can be strongly associated with the variation of road safety levels. That means, if a single or multiple road attributes were changed, the road safety level may be improved. This is directly related to the importance of using the data mining method to obtain an improved methodology to assess the variation in road safety levels as the data mining approach has been shown to be capable to convert the association between road safety levels and road safety treatments into decision trees and rules. It may be suggested that the significant correlation between road infrastructure and the star ratings, shown in this study, may ensure reliable estimations for the effectiveness of changing road attributes.

As stated in Chapter 8 (section 8.4), the correlation between the class labels (star ratings) and road attributes may be measured by the percentage of the correctly classified instances

(Michael, 2010). The ideal situation is 100% correctly classified instances when the data mining model fits all the instances, which in turn may produce the most effective combinations of countermeasures. Even when the ideal situation does not exist, strong relationships between the star ratings and road attributes, reached 89% in this study, could provide valuable information about the most effective sets of treatments. This comes in agreement with the work of Chang and Chen (2005), Chang and Wang (2006), and Zhang and Fan (2013), which indicated good effectiveness of using the data mining technique in the subject of road safety. The assumptions that were made in this research suggested that the data mining classification algorithm C4.5 may be capable to develop models that result in reliable combinations of countermeasures and rank them with regard to their effectiveness. These assumptions were verified, as shown in Chapter 9. Therefore, it may be suggested that the application of data mining in road safety may be appropriate to provide an automated framework for obtaining and ranking road safety treatments in low and middle income countries by taking into account the effect of changing road attributes on the road safety levels represented by the star ratings in the iRAP model.

10.2.1 Data Mining Usefulness

According to the proposed methodology and as described in Chapter 8 (section 8.2), the models were developed by using the classification data mining pattern and it was based on the theory of combining road attributes for each star rating. The countermeasures combinations were suggested from the frequency of specific road attributes with specific star ratings. The main advantage of the classification method is that it has the ability to provide the potential combinations from the available array of treatments and rank them according to their effectiveness. This may provide a useful approach for road designers to consider the combinations in a cost-effective way. Although the cost issues were beyond the scope of this study because of data limitations, the developed data mining models may provide a useful tool to give road designers more flexibility to choose another combination in case the cost of the first set is high. This means choosing another combination that improves the road safety level into the same star rating that could be achieved by using the first set but with lower cost. It is better to choose combinations with high agreement levels as sets of countermeasures with low agreement levels are likely result in star ratings other than the desired, as described in Chapter 9. Other advantages of the data mining approach are, it does not need historical crash data and it is an objective method that is not affected by the preferences of the practising road safety engineers.

The innovative feature of the data mining method is that the possible countermeasures can be determined, for each star rating, by using the existing road attributes instead of relying only on the engineering knowledge of the iRAP model (described in Chapter 4). Also, this methodology can be used to identify the most effective combinations of countermeasures instead of waiting years for sufficient historical crash data and carry out before-after studies to determine the most effective countermeasures for a specific area or road network.

10.3 Data Mining Complexity

10.3.1 Data Issues

As mentioned in Chapter 7, there are three types of databases Flat-file databases, Relational databases, and Non-relational databases. The database used in this study (iRAP database) is a Flat-file database. This database contains Tables (rows and columns format), which is the appropriate format for the data mining system. The data mining technique could be also applied on the Relational databases as data are stored in this type using the same format (rows and columns). However, the data mining could not be applied on the Non-relational databases as this type of databases stores the data in the format of clusters, which is not suitable for the available data mining systems (Witten, 2017).

To develop the data mining models in this study, a database that includes road infrastructure characteristics was needed. The database that was used in this study was offered by iRAP. As it may normally happen with any database, the considered iRAP database could contain missing values or noisy data such as repeated or wrong values (for example, numerical instead of nominal data) as described previously in Chapter 6. Therefore, the considered data were not used directly as they needed to pass through a preparation process to ensure the quality of the data used to avoid the impact of any errors on the developed data mining models. In addition to data cleaning and as the data mining algorithms used in this study can accept only nominal values, the data were transformed from numerical (its original format) into nominal values. In the process of preparing the data, it was found that there were some missing values in the iRAP database. Because all the missing values were star ratings, all the road sections that included missing values were omitted from the computational process (described in section 8.4.3) to avoid their effect on the accuracy of the developed models. It is essential to prepare the data before using them, this is to overcome the impact of data issues on developing accurate data mining models. This is because data issues can significantly affect the knowledge extraction process. For example, leaving some missing fields for the star rating (the class label) in the considered tables in the iRAP datasets can lead the data mining system (WEKA) to consider these empty fields as a sixth star rating class (Han et al, 2012). This, in turn, can affect the data mining process for combining the countermeasures for each star rating. With regard to the size of the database, it was found that the more the data the more the accuracy of the data mining model. This is because, the more the instances for the data mining training and testing processes (described in Chapter 7), the more the correctly classified instances. This means that using more instances for training and testing increased the ability of the data mining system to detect and correlate the frequency of road attributes with each star rating.

10.3.2 Complexity of Computations for Data Mining

The major complexity of the data mining approach is that, it is a time consuming method that needs trials for the three modelling approaches (Training set, Cross Validation, and Percentage Split). In addition, a data mining software and a well-trained researcher are required to prepare and analyse the data as described in the previous section, and build the decision trees and the rules' tables for the extracted knowledge. Also, the data mining systems cannot check the applicability of the extracted countermeasures. Therefore, the need for the engineering experience is an essential step to decide whether a specific combination can be implemented.

10.4 The Engineering Experience

Although the knowledge discovery process, in this study, has developed an automated framework for describing the effect of combining road safety treatments on the star rating, there is no doubt that it is essential to seek the engineering experience to check the applicability of any set of multiple countermeasures. For example, if an extracted rule suggests adding additional lane to improve the star rating for a road or section that has no enough space for this addition, this rule could not be applied. Another example, suggesting median barriers for a road that has one lane in each direction is not applicable because this will eliminate the overtaking manoeuvres to pass the slow or broken-down vehicles.

In addition, before considering any extracted set as appropriate, it should pass the iRAP countermeasures' triggers (described in Chapter 4). Therefore, the engineering experience is needed to check these triggers for the suggested countermeasures. Therefore, the data mining approach cannot be considered as a substitution for the engineering experience but rather they can be used together to obtain more accurate results.

10.5 Data Mining versus Statistics

As previously mentioned in Chapter 2, two main approaches can be considered for road safety analysis. These are:

1. The traditional statistical modelling approach.
2. Data mining modelling approach.

Both approaches were used in this study. The statistical modelling approach was used to develop a prediction model (mathematical equation) to adjust the multiple countermeasures correction factor (MCCF) to overcome the problem of the overestimation in the effectiveness of multiple countermeasures in the iRAP model as described in Chapter 5. It was found that the statistical approach resulted in significant relationship, between road infrastructure and fatalities numbers, for only one crash type. That is the head-on crashes, as the probability values for the countermeasures (the independent variables) were less than 0.05 (confidence level above 95%). Whereas, the probability values were above 0.05 for the other two crash types (run-off, and head-on + run-off) which are not statistically accepted (Chatterjee and Hadi, 2015). The reason behind that could be related to the number of the considered infrastructure attributes for each crash type, as larger numbers of independent variables increase the possibility of multi-collinearity (linear relationships between two or more independent variables). Meaning that, inaccurate estimations for the weights of the independent variables (road attributes), which resulted in confidence level less than 95% (Belsley, 2005).

The other approach, data mining, was used to correlate road infrastructure to the star ratings in the iRAP model. After evaluating the knowledge extracted from the developed data mining models in Chapter 9, it was found that the data mining technique has successfully correlated the road infrastructure to the star ratings for all the concerned crash types. The correlation percentages were 89.5%, 82.9%, and 79.1% for the head-on, run-off, and head-on + run-off crashes respectively.

It may be concluded, from the above, that the data mining technique has more flexibility over the statistical models in covering larger number of countermeasures. That is, the developed statistical models succeeded in covering only 5 countermeasures in this particular modelling approach, whereas, the data mining models covered 14 countermeasures without the need for historical crash data by using the star rating system as an alternative method to address road safety issues. This may entail that data mining techniques are shown to be powerful tools in combining high numbers of road safety treatments. This could enable practising road safety engineers in iRAP to examine the effect of all the road attributes, related to each crash type in the iRAP model, on road safety levels.

10.6 Application of the Data Mining for iRAP

This research proposed a new concept for the existing iRAP methodology to improve road safety levels for vehicle occupants through the star rating system. The enhanced approach is likely to propose more appropriate countermeasures of head-on, run-off, and head-on + run-off crashes for vehicles occupants which could lead to improve the star rating more satisfactorily. The crash countermeasures combinations associated with the classifying and ranking techniques used in the enhanced model cannot be reached by using the current iRAP ViDA software. Therefore, to combine the available countermeasures and rank the resulted combinations, it may be proposed that the current iRAP ViDA software could be upgraded with a data mining system and a ranking algorithm.

By exploring the current iRAP methodology, it was found that the iRAP methodology could suggest countermeasures with regard to the triggers that has been set by iRAP. The suggested countermeasures are used to produce safer road investment plans SRIP (described in Chapter 4) for three cost categories (low, medium, or high) without a specific functionality in ViDA software to consider all the available countermeasures combinations for each cost category. Therefore, it may be proposed that the data mining models developed in this research have the potential to contribute in the ViDA software to improve the star rating for vehicle occupants. However, further testing for the proposed methodology should be carried out before contributing it to the current iRAP methodology. For example, testing the data mining models on other states or countries, in addition, testing the effectiveness of these models on the intersection and property access accident types, which were not considered in this study. Moreover, similar data mining models could be developed for the other road users in the iRAP model (motorcyclists, cyclists, and pedestrians).

10.7 The Impact of the Research

The impact of this research comes from the success of the developed models in covering two main road safety issues: the overestimation in the reduction of road crash fatalities and the effectiveness of using multiple road safety treatments. The impact of this research could be summarised in the following achievements:

1. The developed statistical model provides more accurate estimations for the reduction in road crash fatalities after the implementation of road safety treatments, this model could be used to improve iRAP estimations as explained in Chapter 5. This could lead practising road safety engineers to consider additional treatments, and encourage countries to invest more in road safety to save more realistic estimated fatalities in the future.
2. The developed data mining models are capable of examining the effects of road safety treatments for three accident types (head-on, run-off, and head-on + run-off) to identify the most effective sets of countermeasures combinations for each crash type. This could provide more flexibility for decision makers to compare among the available sets of multiple treatments in a cost-effective way with limited budgets for road safety programmes. This means that, practising road safety engineers in iRAP could have the opportunity to choose the sets of countermeasures that result in the desired star ratings with lower cost. For example, in the situation mentioned in Table 9.2, rules

number 1 and 2 are both result in star rating 3 (SR3). However, rule number 1 requires two treatments (eliminating differential speed limits and improving skid resistance). Whereas rule number 2 requires one treatment (improving skid resistance), which means achieving the same star rating with lower cost by using rule number 2. This, in turn, could lead to improve road safety levels for more sections or roads (save more lives) within limited budget projects.

3. The developed data mining models provide an automated framework to enhance the existing ViDA software in the iRAP model.

Therefore, the impact of this research is that all the above achievements, if the iRAP methodology adopted the models developed in this study, could play a crucial role in improving road safety levels in India and other low and middle income countries with similar traffic conditions. The most effective combinations (extracted in this study) improved road safety for the road sections that have low star rating (SR1 or SR2), in the concerned road network in Andhra Pradesh state, with high agreement levels. These agreement levels reached 87%, 80%, and 74% for head-on, run-off, and head-on + run-off crashes respectively as explained in Chapter 9. That, in turn, could help in achieving significant reductions in hundreds of thousands of fatalities and injuries each year that cost billions of pounds.

This impact comes compatible with the urgent need for immediate treatments to improve road safety in low and middle income countries. With this regard, a study conducted by the

European commission indicated that road traffic fatalities are largely avoidable. Therefore, reaching combinations of countermeasures that are more effective could be considered as an essential part for more effective road safety management to prevent these fatalities and to achieve better road safety performance (European Commission, Directorate-General Transport and Energy, 2009).

10.8 Accuracy of models

As presented in Chapter 8, three data mining models were developed for each accident type using three different data mining modelling approaches, Training set, Cross Validation, and Percentage Split. The results of the validation part in Chapter 8 showed that the percentages of correctly classified instances, as the measure of the models accuracy, were high for all the crash types. The high percentages of the correctly classified instances, as previously shown in Chapter 8, indicated the significance of the statistical relationship between road attributes and the iRAP star ratings. This means that the accuracy of the developed models were high for all the data mining modelling approaches. However, there was a small difference between the three approaches, which made the Percentage spilt approach the most accurate for head-on crashes countermeasures. Whereas, the Training set approach was the most accurate for run-off and head-on + run-off crashes countermeasures. This revealed that, the two thirds of data, which is the ideal percentage of data for training (Witten, 2017), was not the appropriate percentage that sufficiently represent all the star ratings in the head-on crashes dataset.

Therefore, the Percentage Split approach gave higher accuracy with 75%. Meaning that three quarters of the data were needed to obtain the most accurate data mining model for head-on crashes countermeasures. This indicated the importance of exploring the three approaches without relying only on the ideal, Training set method. As this may not give the most accurate model as mentioned above.

10.9 Effectiveness of Countermeasures

After evaluating the effectiveness of the extracted sets of countermeasures in Chapter 9, it was found that, for the concerned case study, two countermeasures were the most effective to raise the star rating for head-on crashes into SR4 and SR5. As it was expected, the Median type change, and Adding travel lanes were the most effective countermeasures for this accident type. This is because changing the median type into non-traversable is directly related to the overtaking manoeuvres using the opposite direction that could be considered the main cause for the occurrence of head-on crashes. However, these countermeasures are considered high cost treatments (iRAP toolkit, 2010). Therefore, the countermeasures that raise the star rating into SR3 (eliminating the differential speed limit and improving the skid resistance) may be considered more appropriate for low budget projects.

For the run-off crashes, it was clear that providing more roadside distance and wider paved shoulders were the most effective countermeasures that may raise the star raring into SR4 and SR5. However, the countermeasures: improving the skid resistance, providing adequate delineation, and good road condition, which raise the star rating into SR3, seem more economical solutions.

In addition, it was found that the most effective countermeasures for head-on + run-off crashes were a collection of the head-on crashes countermeasures and the run-off crashes countermeasures, mentioned above. This may be an indication for the validity of the developed data mining models. These countermeasures, to get SR3, are, improving the skid resistance to adequate, providing adequate delineation, improving the road condition (fixing cracks and potholes), and eliminating the differential speed limit. The treatments that raise the star rating from SR3 to SR4 are, increasing the roadside distance, improving the skid resistance, changing the median type, increasing the shoulder width, and adding an additional lane. Moreover, to raise the star rating from SR4 to SR5, improving the skid resistance from medium to adequate, and increasing the roadside distance and the shoulder width are the appropriate treatments.

10.10 Study Limitations

The methodology developed in this study was based on road infrastructure databases. Due to the available data, it was only possible to develop data mining models based on data from only one of the LMICs (India). The infrastructure attributes were limited from some roads in the state of Andhra Pradesh in India. In addition, some of the concerned roads had similar road attributes, which reduced the ability to explore various combinations of countermeasures. Therefore, there is a need for more data from other states and countries to verify the data mining models in a wider range. However, that could need more resources and duration, which were beyond the scope of this research.

In addition, to develop the data mining models, data mining systems were needed. Because of the high cost and the time needed for training on such systems, only one data mining system (WEKA) was considered as previously described in Chapter 7 (section 7.2.3). Therefore, the accuracy of the other data mining systems in the market were not examined.

Due to the data limitations, the data mining models in this study were only developed for the non-intersected types of accidents. Therefore, the inclusion of the developed data mining models in the iRAP methodology could not be considered for road sections that include intersections or property access crashes.

Moreover, the decision trees that were developed for run-off and run-off + head-on crashes were large and therefore there was no possibility to present them on papers in the full size. However, the screen size was considered and presented to give an indication about the shape and the levels of the trees as shown in Appendix B. One exemption from that was the decision tree developed for the head-on crashes countermeasures because it considered the lowest number of attributes. Because of this limitation, the extracted combinations were presented in tables (Appendices D, E, and F). Each row in these tables represented a rule (set of countermeasures that resulted in a specific star rating).

10.11 Summary

This study, successfully, developed a new methodology that uses data mining techniques in combining road safety treatments. The extracted knowledge represented the frequency of the change in the star rating because of the change in one or more road attributes. Therefore, the data mining may be used to propose and rank road safety treatments for improving the star rating due to their approved relationships.

This research tried to focus the light on the usefulness of using these techniques in addressing vehicle occupants' safety. Although the findings of this research are elementary and need more testing as stated previously in this Chapter, they will be useful for helping the practising road safety engineers in making more accurate decisions for improving road safety. This is through considering the combinations with the highest accuracy levels, which could lead to save more people lives not just in India but also in the other low and middle income countries with similar traffic conditions.

In conclusion, the data mining approach developed in this study is capable of combining and ranking the road crashes countermeasures for the three considered crash types, which are head-on, run-off, head-on + run-off. This could significantly help in filling the gap of combining and ranking sets of road crashes treatments, which, to the best of the researcher knowledge, had not been considered in the domain of vehicles occupants' safety. The new concept could provide an invaluable automated framework to suggest treatments that may be missed by the practising road safety engineers when improving road safety levels because of the possibility of hundreds of combinations, especially with large numbers of available countermeasures. Therefore, it may potentially play a role in saving human lives that are lost yearly in India and countries with similar traffic and geometric road conditions.

CHAPTER 11

CONCLUSIONS AND FUTURE WORK

11.1 Conclusions

This study proposed a new methodology for estimating road safety levels and developed an automated framework to combine and rank road safety treatments. The innovation of this study is that the road infrastructure attributes were combined and ranked in a new way, which is the Knowledge Discovery in Databases (KDD). This is to assess the combined effect of these countermeasures on vehicle occupants' safety with regard to three types of accidents head-on, run-off, and (head-on + run-off). This methodology was developed to enhance the current iRAP methodology for combining treatments to improve the star rating into higher levels. The data mining developed in this study were build, evaluated, and applied on a database offered by iRAP. The new methodology developed models that capable to suggest treatments for road sections that suffer both head-on and run-off crashes. With regard to the objectives of this study (mentioned in Chapter 1), it may be concluded that:

11.1.1 Conclusions Related to the Statistical Models

1. Two statistical models developed in this study, one (the Negative Binomial model) provides a prediction model to adjust the MCCF to estimate the reduction in fatality rates, in iRAP, more accurately.
2. The Negative Binomial has been approved as the appropriate statistical regression approach to correlate road traffic and geometric attributes to road crashes fatalities for the database used in this study.

3. The road attributes included in the statistical model (number of lanes, median type, gradient, skid resistance, differential speed limits, operating speeds, and traffic volumes) have significant impact on the number of road crashes fatalities.
4. The statistical model developed in this study may be adopted by iRAP to adjust the MCCF to get more accurate estimations for the expected number of fatalities after the implementation of multiple countermeasures.
5. The statistical analysis for the considered countermeasures for head-on crashes showed that the attribute *Median type* has the highest impact on the fatality numbers caused by head-on crashes.

11.1.2 Conclusions Related to the Data Mining Models

1. The knowledge discovery and data mining techniques were found to be an appropriate approach to reflect the effectiveness of road safety multiple treatments by correlating the road attributes associated with head-on and run-off crashes to the star rating system in the iRAP model.
2. Developing the knowledge extraction models needed four steps, data preparation, data mining, model evaluation, and knowledge representations. In the data preparation step, the data were cleaned, integrated, and transformed.

a. It was found that the Ignoring the Tuple method (explained in Chapter 6) is the appropriate method to clean the data from the missing values. In addition, the data from different excel files were integrated to create one excel sheet as input data for the data mining system.

b. In the data mining step, the classification was considered as the appropriate data mining pattern to develop the data mining models. This is to classify the road attributes that may provide iRAP star ratings.

c. The Waikato Environment for Knowledge Analysis (WEKA) is an appropriate system with regard to the data used in this study.

d. In the evaluation step, three data mining modelling approaches were used, these are, Training set, Cross Validation, and Percentage Split. The Training set method is the appropriate method for head-on crashes and the Percentage Split is the appropriate method for run-off and (head-on + run-off) crashes.

e. For representing the extracted knowledge, it was found that the tables are considered as the simplest way to represent the data mining outputs (road safety treatments combinations and their star ratings in this study) and they provide easily understandable representation for these combinations.

3. The data mining models developed in this study provided a good automated framework to combine the head-on and run-off related countermeasures for each star rating without the need for historical crash data.

4. The *Best First* ranking algorithm in WEKA provided a good approach for ranking road safety treatments combinations.
5. The agreement levels between the suggested and actual star ratings for the combinations of road attributes related to each accident type separately were higher than those for the total number of road attributes related to both (head-on + run-off).
6. The lower the number of road attributes considered in the data mining model, the higher the accuracy of the extracted knowledge.
7. The newly developed data mining models of this study offer an innovative method in the subject of combining road attributes related to vehicle occupants' safety.
8. The extracted knowledge from the developed models showed that the head-on and run-off crashes are associated with the following attributes: Differential speed limits, Median type, Number of lanes, Gradient, Skid resistance, Lane width, Curvature, Quality of curve, Road condition, Roadside severity - passenger-side distance, Roadside severity - driver-side distance, Paved shoulder - driver-side, Paved shoulder - passenger-side, Delineation.
9. The developed data mining models can be used to compare between road safety levels (star ratings) before and after implementing a particular set of countermeasures.

10. The models may minimise the need for years to collect sufficient crash data for the statistical models, and may also increase the reliability of decision making based on such models.
11. The developed data mining models identified potential combinations of road safety treatments with high agreement levels as described in Chapter 10. These combinations (for the concerned road network) are:
    a. For head-on crashes:
        i. It was found that improving the skid resistance and eliminating the differential speed limit are the most effective countermeasures to improve the star rating from SR1 and SR2 into SR3.
        ii. Changing the median type and adding an additional lane are the countermeasures that could make the difference between SR3 and SR4.
        iii. Improving the skid resistance from sealed medium to sealed adequate could improve the star rating from SR4 to SR5.
    b. For run-off crashes:
        i. It was revealed that improving the skid resistance, widening the lanes into medium (2.75 – 3.25m), improving the road condition (fixing cracks and potholes), and providing adequate delineation are the most effective countermeasures to raise the star rating from SR1 and 2 into SR3.

ii. Increasing the roadside distance, increasing the shoulder width, and improving the skid resistance could raise the star rating from SR3 to SR4.

iii. Increasing the shoulder width and the roadside distance are the appropriate treatments to improve the star rating from SR4 to SR5.

c. For head-on + run-off crashes:

i. It was found that the countermeasures that improve the star rating from SR1 and 2 into SR3 are: improving the skid resistance, eliminating the differential speed limit, improving the road condition (fixing cracks and potholes), and providing adequate delineation.

ii. The countermeasures that raise the star rating from SR3 to SR4 are: changing the median type, adding an additional lane, increasing the roadside distance, increasing the shoulder width, and improving the skid resistance.

iii. To improve the star rating into SR5, the countermeasures that could make the difference are: improving the skid resistance from sealed medium to sealed adequate, increasing the shoulder width and the roadside distance.

11.2 Future Work

This study faced some limitations with regard to the available data and knowledge extraction systems.

Therefore, the following issues may be considered in the future work:

1. Developing data mining models for road safety treatments for road networks in other low and middle income countries with traffic conditions similar to India to enable a calibration process for the models developed in this study.
2. Only one data mining system, WEKA, was used in this study. Therefore, it is recommended to test the other data mining systems in the market to check the ability of these systems to develop such models for road safety issues on road networks.
3. This study focused only on non-intersected crashes (head-on and run-off). Therefore, it could be useful to develop data mining models that include also the intersected accidents in iRAP (intersection and property access accidents) on road networks. This is to apply the data mining models in a wider range and to develop comprehensive models that may cover all the accident types in the iRAP model.

References

Aarts, L. and Van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. Accident Analysis & Prevention, 38(2), pp.215-224.

Abdel-Aty, M.A., Radwan, A.E. (2000). Modelling traffic accident occurrence and involvement. *Accident Analysis and Prevention*, Vol. 32, pp. 633-642.

ABO-QUDAIS, S. (2001). URBAN ROADS ACCIDENTS PREDICTION MODELS. Routes/Roads, (309). p. 35-56.

Aghabayk, K., Sarvi, M. and Young, W. (2012), September. A macroscopic study of heavy vehicle traffic flow characteristics. In *ARRB Conference, 25th, 2012, Perth, Western Australia, Australia*.

Al-Dah, M.K. (2010). *Causes and consequences of road traffic crashes in Dubai, UAE and strategies for injury reduction.* Doctoral dissertation, Loughborough University, Loughborough, UK.

Ali, S.M. and Tuteja, M.R. (2014). Data Mining Techniques. *International Journal of Computer Science and Mobile Computing*, Vol.3 Issue.4, pp. 879-883.

Amado, V. (2000). Expanding the use of pavement management data. In *2000 MTC Transportation Scholars Conference, Ames, Iowa*, USA.

Archer, J. (2005). *Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling: A study of urban and suburban intersections.* Doctoral dissertation, Royal Institute of Technology, Stockholm, Sweden.

Belsley, D.A., Kuh, E. and Welsch, R.E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity* (Vol. 571). John Wiley & Sons.

Berkhin, P. (2006). A survey of clustering data mining techniques. *Grouping multidimensional data,* Springer, Berlin, Heidelberg (pp. 25-71).

Beshah, T. and Hill, S. (2010). Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors on Accident Severity in Ethiopia. In *AAAI Spring Symposium: Artificial Intelligence for Development*.

Beynon-Davies, P., 2002. *Information systems: An introduction to informatics in organisations*. Palgrave Macmillan.

Brown, M.S. (2014). Data mining for dummies. John Wiley & Son.

Byrne, M and Others (2005). Application of data mining in pavement performance modelling – a case study. *Road and transport research*. (14/4), 27-43.

Canarelli, P. (1996). *Knowledge Extraction from Data Using Genetic Algorithms*. European Commission.

Caruana, R. and Niculescu-Mizil, A. (2004). Data mining in metric space: an empirical analysis of supervised learning performance criteria. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining,* ACM, pp. 69-78.

Chakrabarty, N., Gupta, K. and Bhatnagar, A. (2013). A survey on awareness of traffic safety among drivers in Delhi, India. *The SIJ Transactions on Industrial, Financial and business management (IFBM)*, *1*(2), pp.106-109.

Chang, L.Y. (2005). Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. *Safety science*, *43*(8), pp.541-557.

Chang, L.Y. and Chen, W.C. (2005). Data mining of tree-based models to analyze freeway accident frequency. *Journal of safety research*, *36*(4), pp.365-375.

Chang, L.Y. and Wang, H.W. (2006). Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis & Prevention*, *38*(5), pp.1019-1027.

Chang, S.K. (2001). *Handbook of software engineering and knowledge engineering* (Vol. 1). World Scientific.

Chatterjee, S. and Hadi, A.S., 2015. *Regression analysis by example*. John Wiley & Sons.

Chengye, P. and Ranjitkar, P. (2013). Modelling motorway accidents using negative binomial regression. *Journal of the Eastern Asia Society for Transportation Studies*, 10, pp.1946-1963.

Chin, H.C. and Quddus, M.A. (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis & Prevention*, 35(2), pp.253-259.

Choueiri, E.M., Lamm, R., Kloeckner, J.H. and Mailaender, T. (1994). Safety aspects of individual design elements and their interactions on two-lane highways: international perspective. *Transportation Research Record*, (1445), pp. 34-46.

Computer Hope, (2018). How to create a CSV file? Available at: https://www.computerhope.com (Accessed 1 Apr. 2018).

Council, F. and Stewart, J. (1999). Safety effects of the conversion of rural two-lane to four-lane roadways based on cross-sectional models. *Transportation Research Record: Journal of the Transportation Research Board*, (1665), pp.35-43.

Daniello, A. and Gabler, H. (2011). Effect of barrier type on injury severity in motorcycle-to-barrier collisions in North Carolina, Texas, and New Jersey. *Transportation Research Record: Journal of the Transportation Research Board*, (2262), pp.144-151.

Davies, D. G., Taylor, M. C., Ryley, T. J., and Halliday, M. (1997) Cyclists at roundabouts – the effects of 'continental' design on predicted safety and capacity (TRL 285) Transport Research Laboratory, Crowthorne, UK.

Doane, D.P. and Seward, L.E. (2005). Applied statistics in business and economics. Irwin, USA.

Eaton, E. Classification: Decision Trees & Overfitting. Georgia Tech University. Available on https://www.cc.gatech.edu (Accesses 2 June 2018).

Edlich, P. (2018). NOSQL Databases. Available at: http://nosql-database.org (Accessed 6 Apr. 2018).

Elvik R. and Vaa T. (2004). *The Handbook of Road Safety Measures*. Elsevier.

Elvik, R., Vaa, T., Hoye, A. and Sorensen, M. eds. (2009). *The handbook of road safety measures*. Emerald Group Publishing, Bradford, UK.

European Commission (2009). Road Safety Management. Available at: http://ec.europa.eu (Accessed 16 September 2018).

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), p.37 - 54.

FHWA (Federal Highway Administration), (2004). The Safety Impacts of Differential Speed Limits on Rural Interstate Highways. FHWA Publication No. FHWA-HRT-04-156.

Frank, E., Hall, M., Trigg, L., Holmes, G. and Witten, I.H. (2004). Data mining in bioinformatics using Weka. *Bioinformatics*, *20*(15), pp.2479-2481.

Frank, E., Hall, M.A. and Witten, I.H., 2016. The WEKA workbench. *Data mining: Practical machine learning tools and techniques*, *4*. Morgan Kaufmann.

Garder, P. (2006). Segment characteristics and severity of head-on crashes on two-lane rural highways in Maine. *Accident Analysis & Prevention*, *38*(4), pp.652-661.

Ghods, A.H., Saccomanno, F. and Guido, G. (2012). Effect of Car/Truck differential speed limits on two-lane highways safety operation using microscopic simulation. *Procedia-Social and Behavioral Sciences*, *53*, pp. 833-840.

Groeger, J. (2011). How Many E's in Road Safety? Handbook of Traffic Psychology. 3-12.

Gross, F. and Hamidi, A. (2011). Investigation of existing and alternative methods for combining multiple CMFs. Highway Safety Improvement Program Technical Support. Task A, 9.

Haghani, M., Jalalkamali, R. and Berangi, M. (2018). Assigning crashes to road segments in developing countries. In *Proceedings of the Institution of Civil Engineers-Transport.* Thomas Telford, London, UK.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, *11*(1), pp.10-18.

Hall, M.A. and Holmes, G. (2003). Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data engineering*, *15*(6), pp.1437-1447.

Han, J., Pei, J. and Kamber, M. (2012). *Data mining: concepts and techniques*. Elsevier.

Haque, M.M., Chin, H.C. and Huang, H. (2009). Modeling fault among motorcyclists involved in crashes. *Accident Analysis & Prevention*, *41*(2), pp.327-335.

Harwood, D. W., Council, F. M., Hauer E., Hughes W E, and Vogt A. (2000). Prediction of the expected safety performance of rural two-lane highways, Report No. FWHA-RD-99-207, Federal Highway Administration, Washington, D.C., USA.

HCM (Highway Capacity Manual), (2000). Transportation Research Board, Washington, D.C., USA.

Hegland, M. (2001). Data mining techniques. *Acta numerica*, *10*, pp.313-355.

Hg.org. (2018). Head-On-Collisions - The Most Dangerous Type of Car Accident. [online] Available at: https://www.hg.org/legal-articles/head-on-collisions-the-most-dangerous-type-of-car-accident-29016 (Accessed 16 Sep. 2018).

Hilbe, J.M. (2011). *Negative binomial regression*. Cambridge University Press.

Holsheimer, M. and Siebes, A. (1991). *Data mining: The search for knowledge in databases*. Centrum voor Wiskunde en Informatica, Amsterdam, The Netherlands.

Hosseinpour, M., Yahaya, A.S. and Sadullah, A.F. (2014). Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from Malaysian Federal Roads. *Accident Analysis & Prevention*, *62*, pp.209-222.

HSM (Highway capacity manual) (2000). *Transportation Research Board National* Research Council. Washington, D.C., USA.

iRAP (international Road Assessment Programme), (2014). S*tar rating roads for safety: the iRAP methodology*. Available at: http://www.irap.org (Accessed 25 September 2017).

iRAP (International Road Assessment Programme) (2010) Road safety toolkit. Available at: http://toolkit.irap.org. (Accessed 03/11/2017).

Jovanis, P.P. and Chang, H.L. (1986). Modeling the relationship of accidents to miles traveled. *Transportation Research Record*, *1068*, pp.42-51.

Kahane, CJ. (1994). Correlation of NCAP Performance with Fatality Risk in Actual Head-On Collisions. National Highway Traffic Safety Administration, Washington, D.C.

Kamaluddin, N.A., Abd Rahman, M.F. and Várhelyi, A. (2018). Matching of police and hospital road crash casualty records–a data-linkage study in Malaysia. *International journal of injury control and safety promotion*, pp.1-8.

Khan, G., Bill, A.R., Chitturi, M. and Noyce, D.A. (2012). Horizontal curves, signs, and safety. *Transportation research record*, *2279*(1), pp.124-131.

Kim, D.G., Lee, Y., Washington, S. and Choi, K. (2007). Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. *Accident Analysis & Prevention*, *39*(1), pp.125-134.

Kononov, J., Reeves, D., Durso, C. and Allery, B. (2012). Relationship between freeway flow parameters and safety and its implication for adding lanes. *Transportation Research Record: Journal of the Transportation Research Board*, (2279), pp.118-123.

Król, D., Nguyen, N.T. and Shirai, K. (Eds.) (2017). Advanced Topics in Intelligent Information and Database Systems. Springer Science & Business Media.

Kwon, O.H., Rhee, W. and Yoon, Y. (2015). Application of classification algorithms for analysis of road safety risk factor dependencies. *Accident Analysis & Prevention*, *75*, pp.1-15.

Larose, D.T. (2005). *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons.

Laureshyn, A. (2010). *Application of automated video analysis to road user behaviour.* Doctoral Dissertation, Department of Technology and society Traffic Engineering, Lund University, Lund, Sweden.

Leavitt, N. (2010). Will NoSQL databases live up to their promise? *Computer*, *43*(2), pp. 12-14.

Lee, J. and Mannering, F. (2002). Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. *Accident Analysis & Prevention*, *34*(2), pp.149-161.

Lee, J. and Mannering, F. (2002). Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. *Accident Analysis & Prevention*, *34*(2), pp.149-161.

Lin, S., Gao, J., Koronios, A. and Chanana, V. (2007). Developing a data quality framework for asset management in engineering organisations. *International Journal of Information Quality*, *1*(1), pp.100-126.

Liu, H. and Motoda, H., 2012. *Feature selection for knowledge discovery and data mining*, Springer Science & Business Media, Vol. 454.

Lord, D., Manar, A. and Vizioli, A. (2005). Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. *Accident Analysis & Prevention*, *37*(1), pp.185-199.

Lynam, D. (2012) Development of Risk Models for the Road Assessment Programme, TRL, Crowthorne, UK.

Mahajan, N. and Kaur, B.P. (2016). Study the factors affecting road safety using decision tree Algorithms. *International Journal of Engineering Research and General Science*, 4(1), p. 168 – 173.

Mak, K. K., & Sicking D. L. (2003). *Roadside Safety Analysis Program (RSAP) - Engineer's manual* (NCHRP Report 492, TRB, ISBN 0–309–06812–6). Washington DC: Transportation Research Board.

Manan, M.M.A. (2014). Motorcycles entering from access points and merging with traffic on primary roads in Malaysia: Behavioral and road environment influence on the occurrence of traffic conflicts. *Accident Analysis & Prevention*, *70*, pp.301-313.

Marchesini, P. and Weijermars, W. (2010). The relationship between road safety and congestion on motorways. A literature review of potential effects, SWOV Institute for Road Safety Research, The Netherlands.

McPherson K. and Bennett C.R., 2005. Success Factors for Road Management Systems. The World Bank, Washington D.C., USA.

Miaou, S.P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis & Prevention*, *26*(4), pp.471-482.

Michael, A. (2010). *Classification and clustering*. [online] Ibm.com. Available at: https://www.ibm.com/developerworks/library/os-weka2/ (Accessed 17 Jan. 2018).

Miller, T. & Zaloshnja, E. (2009). On a crash course: The dangers and health costs of deficient roadways. Pacific Institute for Research & Evaluation. Available at: https://trid.trb.org (Accessed 16 September 2018).

Mohan, C. (2013). History repeats itself: sensible and Non sen SQL aspects of the NoSQL hoopla. In *Proceedings of the 16th International Conference on Extending Database Technology,* ACM, pp. 11-16.

Montella, A., Colantuoni, L. and Lamberti, R. (2008). Crash prediction models for rural motorways. *Transportation Research Record: Journal of the Transportation Research Board*, (2083), pp.180-189.

Neudorff, L., Jenior, P., Dowling, R. and Nevers, B., 2016. *Use of Narrow Lanes and Narrow Shoulders on Freeways: A Primer on Experiences, Current Practice, and Implementation Considerations* (No. FHWA HOP-16-060).

Nunn, S. (2011). Death by motorcycle: background, behavioural, and situational correlates of fatal motorcycle collisions. *Journal of forensic sciences*, *56*(2), pp.429-437.

Oracle, (2018). A Relational Database Overview. Available at: https://docs.oracle.com (Accessed 6 Apr. 2018).

Oracle, (2018). The Java Language Environment. Available at: http://www.oracle.com (Accessed 1 Apr. 2018).

Pakgohar, A., Tabrizi, R.S., Khalili, M. and Esmaeili, A. (2011). The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach. *Procedia Computer Science*, *3*, pp.764-769.

Park, J., Abdel-Aty, M. and Lee, C. (2014). Exploration and comparison of crash modification factors for multiple treatments on rural multilane roadways. *Accident Analysis & Prevention*, *70*, pp.167-177.

Patton, M.Q. (2005). Qualitative research. John Wiley & Sons, Ltd.

Perel, P., Ker, K., Ivers, R. and Blackhall, K. (2007). Road safety in low-and middle-income countries: a neglected research area. *Injury prevention*, *13*(4), pp.227-227.

Phan, V.L. (2016). *Crash risk models for motorcycle-dominated traffic environment of urban roads in developing countries.* Doctoral dissertation, University of Birmingham, UK.

Phan, V.L., Evdorides, H., Bradford, J. and Mumford, J. (2016). Motorcycle crash risk models for urban roads. In *Proceedings of the Institution of Civil Engineers-Transport.* Thomas Telford, London, UK, Vol. 169, pp. 397-407.

Poch, M. and Mannering, F. (1996). Negative Binomial Analysis of Intersection-Accident Frequencies. *Journal of Transportation Engineering*, 122(2), pp.105-113.

Preusser, D.F., Williams, A.F., Nichols, J.L., Tison, J. and Chaudhary, N.K. (2008). NCHRP report 622: Effectiveness of behavioral highway safety countermeasures. *Transportation Research Board of the National Academies,* Washington, DC.

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

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org (Accessed 03 March 2017).

Rajagopalan, B. and Isken, M.W. (2001). Exploiting data preparation to enhance mining and knowledge discovery. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *31*(4), pp.460-467.

Rawlings, Ian. (1999). Using data mining and warehousing for knowledge discovery. Computer Technology Review, 19(9), 20 - 22.

Russell, S.J., Norvig, P., Canny, J.F., Malik, J.M. and Edwards, D.D. (2003). Artificial Intelligence: A Modern Approach. Prentice Hall, Englewood Cliffs, New Jersey, USA.

Sarin, S.M. & Mittal, N. (1991), "Traffic Accidents in India", Internal Report Published at CRRI.

Schneider IV, W.H., Savolainen, P.T. and Moore, D.N. (2010). Effects of horizontal curvature on single-vehicle motorcycle crashes along rural two-lane highways. *Transportation Research Record*, 2194(1), pp.91-98.

Shankar, V., Mannering, F. and Barfield, W. (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis & Prevention*, *27*(3), pp.371-389.

Tan, P.N., Steinbach, M. and Kumar, V. (2006). *Introduction to Data Mining*. Addison-Wesley. Boston, USA.

Turner, B. (2011). Estimating the Safety Benefits when Using Multiple Road Engineering Treatments. *Road Safety Risk Reporter*, *11*. . ARRB Group, Melbourne, Australia.

Turner, B., Affum, J., Tziotis, M. and Jurewicz, C. (2009). Review of iRAP Risk Parameters. *ARRB Group Contract Report for iRAP*, Melbourne, Australia.

Turner, B., Imberger, K., Roper, P., Pyta, V. and McLean, J. (2010). Road Safety Engineering Risk Assessment Part 6: Crash Reduction Factors. Austroads (No. AP-T151/10).

USDOT (U.S. Department of Transportation), (2001). Data Integration Glossary. Available at: https://web.archive.org (Accessed 6 Apr. 2018).

Wanvik, P.O. (2009). Effects of road lighting: an analysis based on Dutch accident statistics 1987–2006. *Accident Analysis & Prevention*, *41*(1), pp.123-128.

Washington, S.P., Karlaftis, M.G. and Mannering, F. (2010). *Statistical and econometric methods for transportation data analysis*. Chapman and Hall/CRC Press.

WHO (World Health Organisation), (2004). World Report on road traffic injury prevention. WHO, Geneva, Switzerland.

WHO (World Health Organization), (2015). Road traffic injuries, Fact sheet No.358. Available at: http://www.who.int/mediacentre/factsheets/fs358/en/ (Accessed 01 May 2017).

Winnett, M. A. and Wheeler, A. H. (2002). Vehicle-Activated Signs – A Large Scale Evaluation, TRL report TRL548, TRL Limited, United Kingdom.

Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Witten, I.H., Frank, E., Trigg, L.E., Hall, M.A., Holmes, G. and Cunningham, S.J. (2000). Weka: Practical machine learning tools and techniques with Java implementations.

World Bank, (2010). Safe Roads for Development, A POLICY FRAMEWORK FOR SAFE INFRASTRUCTURE ON MAJOR ROAD TRANSPORT NETWORKS. Available at: https://www.irfnet.ch (Accessed 05 June 2016).

World Bank, (2017). World Bank Country and Lending Groups. Available at: https://datahelpdesk.worldbank.org (Accessed 01 October 2018).

World Bank, (2018). Road Deaths and Injuries Hold Back Economic Growth in Developing Countries. Available at: https://worldbank.org (Accessed 01 October 2018).

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Philip, S.Y. and Zhou, Z.H. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, *14*(1), pp.1-37.

Zegeer, C.V., Hummer, J., Herf, L., Reinfurt, D. and Hunter, W. (1987). *Safety Cost-Effectiveness of Incremental Changes in Cross-Section Design--Informational Guide* (No. FHWA/RD-87/094).

Zhang, S., Zhang, C. and Yang, Q. (2003). Data preparation for data mining. *Applied artificial intelligence*, *17*(5-6), pp.375-381.

Zhang, X.F. and Fan, L. (2013), May. A decision tree approach for traffic accident analysis of Saskatchewan highways. In *Electrical and Computer Engineering (CCECE), 2013 26th Annual IEEE Canadian Conference on* (pp. 1-4). IEEE.

Zwillinger, D. and Kokoska, S. (1999). CRC standard probability and statistics tables and formulae. CRC Press.

APPENDICES

Appendix A

A sample of the raw iRAP database used in developing the statistical and data mining models

A.1 Multi project - Core Data dataset

As the Multi project - Core Data dataset included more than a hundred columns, it was not possible to present them on papers. Therefore, this appendix presents some of the traffic and geometric characteristics as they were coded by iRAP practitioners and used in this research to develop the Negative Binomial and Poisson regression models as presented in Chapter 5, and also to develop the data miming models as explained in Chapter 8.

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
1	0.008	0.1	1	1	1	1	1	0	5039	3
1	0.109	0.1	1	1	1	1	1	0	5039	2
1	0.209	0.1	1	1	1	1	1	0	5039	2
1	0.309	0.1	1	1	1	1	1	0	5039	2
1	0.409	0.1	1	1	1	1	1	0	5039	2
1	0.508	0.1	1	1	1	1	1	0	5039	3
1	0.609	0.1	1	1	1	1	1	0	5039	3
1	0.709	0.1	1	2	1	1	1	0	5039	4
1	0.809	0.1	1	2	1	1	1	0	5039	4
1	0.909	0.1	1	2	1	1	1	0	5039	4
1	1.009	0.1	1	2	1	1	1	0	5039	4

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
1	1.109	0.1	1	2	1	1	1	0	5039	3
1	1.209	0.1	1	2	1	1	1	0	5039	4
1	1.309	0.1	1	2	1	1	1	0	5039	4
1	1.409	0.1	1	2	1	1	1	0	5039	4
1	1.507	0.1	1	2	1	1	1	0	5039	2
1	1.608	0.1	1	2	1	1	1	0	5039	1
1	1.708	0.1	1	1	2	1	1	0	5039	2
1	1.809	0.1	1	1	2	1	1	0	5039	1
1	1.908	0.1	1	1	2	1	1	0	5039	2
1	2.008	0.1	1	1	2	1	1	0	5039	2
1	2.108	0.1	1	1	2	1	1	0	5039	2
1	2.208	0.1	1	1	2	1	1	0	5039	1
1	2.308	0.1	1	1	2	1	1	0	5039	2
1	2.408	0.1	1	1	2	1	1	0	5039	2
1	2.508	0.1	1	1	1	1	1	0	5039	2
1	2.608	0.1	1	1	1	1	1	0	5039	2
1	2.708	0.1	1	1	1	1	1	0	5039	2
1	2.808	0.1	1	1	1	1	1	0	5039	1
1	2.908	0.1	1	1	1	1	1	0	5039	2

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
1	3.008	0.1	1	1	1	1	1	0	5039	2
1	3.108	0.1	1	1	2	1	1	0	5039	2
1	3.208	0.1	1	1	1	1	1	0	5039	2
1	3.308	0.1	1	1	1	1	1	0	5039	2
1	3.408	0.1	1	1	1	1	1	0	5039	2
1	3.508	0.1	1	1	1	1	1	0	5039	2
1	3.608	0.1	1	1	1	1	1	0	5039	1
1	3.708	0.1	1	1	1	1	1	0	5039	2
1	3.808	0.1	1	1	1	1	1	0	5039	2
1	3.907	0.1	1	1	1	1	1	0	5039	2
1	4.008	0.1	1	1	1	1	1	0	5039	1
1	4.108	0.1	1	1	1	1	1	0	5039	2
1	4.208	0.1	1	1	1	1	1	0	5039	2
1	4.308	0.1	1	1	1	1	1	0	5039	2
1	4.408	0.1	1	1	1	1	1	0	5039	1
1	4.508	0.1	1	1	1	1	1	0	5039	2
1	4.607	0.1	1	1	1	1	1	0	5039	2
1	4.707	0.1	1	1	1	1	1	0	5039	2
1	4.807	0.1	1	1	1	1	1	0	5039	2

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
1	4.907	0.1	1	1	1	1	1	0	5039	1
1	5.007	0.1	1	1	1	1	1	0	5039	2
1	5.107	0.1	1	1	1	1	1	0	5039	2
1	5.207	0.1	1	1	1	1	1	0	5039	2
1	5.307	0.1	1	1	1	1	1	0	5039	2
1	5.407	0.1	1	1	1	1	1	0	5039	2
1	5.507	0.1	1	1	1	1	1	0	5039	2
1	5.607	0.1	1	1	1	1	1	0	5039	2
1	5.707	0.1	1	1	1	1	1	0	5039	1
1	5.807	0.1	1	1	1	1	1	0	5039	2
1	5.907	0.1	1	1	1	1	1	0	5039	2
1	6.007	0.1	1	1	1	1	1	0	5039	2
1	6.107	0.1	1	1	2	1	1	0	5039	2
1	6.206	0.1	1	1	2	1	1	0	5039	2
1	6.307	0.1	1	1	1	1	1	0	5039	2
1	6.407	0.1	1	1	1	1	1	0	5039	1
1	6.506	0.1	1	1	1	1	1	0	5039	2
1	6.606	0.1	1	1	1	1	1	0	5039	2
1	6.706	0.1	1	1	1	1	1	0	5039	2

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
1	6.806	0.1	1	1	1	1	1	0	5039	1
1	6.906	0.1	1	1	2	1	1	0	5039	2
1	7.006	0.1	1	1	1	1	1	0	5039	1
1	7.106	0.1	1	1	1	1	1	0	5039	1
1	7.206	0.1	1	1	2	1	1	0	5039	2
1	7.306	0.1	1	1	1	1	1	0	5039	2
1	7.406	0.1	1	1	2	1	1	0	5039	2
1	7.506	0.1	1	1	2	1	1	0	5039	2
1	7.606	0.1	1	1	1	1	1	0	5039	2
1	7.706	0.1	1	1	2	1	1	0	5039	2
1	7.806	0.1	1	1	2	1	1	0	5039	3
1	7.906	0.1	1	1	2	1	1	0	5039	3
1	8.006	0.1	1	1	2	1	1	0	5039	2
1	8.106	0.1	1	1	1	1	1	0	5039	2
1	8.206	0.1	1	1	1	1	1	0	5039	2
1	8.306	0.1	1	1	1	1	1	0	5039	2
1	8.405	0.1	1	1	2	1	1	0	5039	1
1	8.505	0.1	1	1	2	1	1	0	5039	2
1	8.606	0.1	1	1	1	1	1	0	5039	2

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
1	8.706	0.1	1	1	1	1	1	0	5039	2
1	8.806	0.1	1	1	1	1	1	0	5039	2
1	8.905	0.1	1	1	1	1	1	0	5039	2
1	9.005	0.1	1	1	1	1	1	0	5039	2
1	9.105	0.1	1	1	1	1	1	0	5039	2
1	9.205	0.1	1	1	1	1	1	0	5039	2
1	9.305	0.1	1	1	1	1	1	0	5039	2
1	9.405	0.1	1	1	1	1	1	0	5039	2
1	9.505	0.1	1	1	1	1	1	0	5039	1
1	9.605	0.1	1	1	1	1	1	0	5039	1
1	9.705	0.1	1	1	2	1	1	0	5039	2
1	9.805	0.1	1	1	1	1	1	0	5039	2
1	9.905	0.1	1	1	1	1	1	0	5039	2
1	10.005	0.1	1	1	1	1	1	0	5039	2
1	10.105	0.1	1	1	2	1	1	0	5039	1
1	10.205	0.1	1	1	1	1	1	0	5039	2
1	10.305	0.1	1	1	1	1	1	0	5039	2
1	10.405	0.1	1	1	2	1	1	0	5039	2
1	10.505	0.1	1	1	2	1	1	0	5039	1

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
1	10.604	0.1	1	1	2	1	1	0	5039	2
1	10.705	0.1	1	1	1	1	1	0	5039	2
1	10.804	0.1	1	1	1	1	1	0	5039	2
1	10.905	0.1	1	1	1	1	1	0	5039	2
1	11.004	0.1	1	1	2	1	1	0	5039	1
1	11.104	0.1	1	1	2	1	1	0	5039	2
1	11.204	0.1	1	1	1	1	1	0	5039	2
1	11.304	0.1	1	1	1	1	1	0	5039	2
1	11.404	0.1	1	1	1	1	1	0	5039	2
1	11.504	0.1	1	1	1	1	1	0	5039	1
1	11.604	0.1	1	1	1	1	1	0	5039	1
1	11.704	0.1	1	1	1	1	1	0	5039	2
1	11.804	0.1	1	1	1	1	1	0	5039	2
1	11.904	0.1	1	1	1	1	1	0	5039	2
1	12.004	0.1	1	1	1	1	1	0	5039	2
1	12.104	0.1	1	1	1	1	1	0	5039	2
1	12.204	0.1	1	1	1	1	1	0	5039	2
1	12.304	0.1	1	1	2	1	1	0	5039	1
1	12.404	0.1	1	1	2	1	1	0	5039	2

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
1	12.504	0.1	1	1	2	1	1	0	5039	1
1	12.604	0.1	1	1	2	1	1	0	5039	2
1	12.704	0.1	1	1	2	1	1	0	5039	2
1	12.804	0.1	1	1	2	1	1	0	5039	1
1	12.904	0.1	1	1	2	1	1	0	5039	2
1	13.004	0.1	1	1	2	1	1	0	5039	2
1	13.104	0.1	1	1	2	1	1	0	5039	2
1	13.204	0.1	1	1	2	1	1	0	5039	3
1	13.304	0.1	1	1	2	1	1	0	5039	2
1	13.404	0.1	1	1	2	1	1	0	5039	2
1	13.504	0.1	1	1	2	1	1	0	5039	2
1	13.604	0.1	1	1	2	1	1	0	5039	2
1	13.704	0.1	1	1	1	1	1	0	5039	2
1	13.803	0.1	1	1	1	1	1	0	5039	2
1	13.903	0.1	1	1	1	1	1	0	5039	2
1	14.003	0.1	1	1	1	1	1	0	5039	3
1	14.103	0.1	1	1	1	1	1	0	5039	2
1	14.203	0.1	1	1	1	1	1	0	5039	2
1	14.303	0.1	1	1	1	1	1	0	5039	2

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
1	14.403	0.1	1	1	1	1	1	0	5039	2
1	14.503	0.1	1	1	1	1	1	0	5039	1
1	14.603	0.1	1	1	1	1	1	0	5039	2
1	14.702	0.1	1	1	1	1	1	0	5039	1
1	14.802	0.1	1	1	1	1	1	0	5039	1
1	14.902	0.1	1	1	1	1	1	0	5039	2
1	15.002	0.1	1	1	1	1	1	0	5039	1
1	15.102	0.1	1	1	1	1	1	0	5039	1
1	15.202	0.1	1	1	1	1	1	0	5039	2
1	15.302	0.1	1	1	1	1	1	0	5039	2
1	15.402	0.1	1	1	1	1	1	0	5039	2
1	15.502	0.1	1	1	1	1	1	0	5039	2
1	15.601	0.1	1	1	1	1	1	0	5039	2
1	15.702	0.1	1	1	1	1	1	0	5039	1
1	15.802	0.1	1	1	1	1	1	0	5039	3
1	15.902	0.1	1	1	1	1	1	0	5039	2
2	16.002	0.1	1	1	1	1	1	1	15000	3
2	16.102	0.1	1	1	1	1	1	1	15000	4
2	16.201	0.1	1	1	1	1	1	1	15000	3

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
2	16.301	0.1	1	1	1	1	1	1	15000	4
2	16.401	0.1	1	1	1	1	1	1	15000	4
2	16.501	0.1	1	2	1	1	1	1	15000	3
2	16.601	0.1	1	2	1	1	1	1	15000	4
2	16.701	0.1	1	2	1	1	1	1	15000	3
2	16.801	0.1	1	2	1	1	1	1	15000	3
2	16.901	0.1	1	2	1	1	1	1	15000	4
2	17.001	0.1	1	2	1	1	1	1	15000	4
2	17.101	0.1	1	2	1	1	1	1	15000	4
2	17.201	0.1	1	2	1	1	1	1	15000	4
2	17.301	0.1	1	2	1	1	1	1	15000	4
2	17.401	0.1	1	2	1	1	1	1	15000	4
2	17.501	0.1	1	2	1	1	1	1	15000	2
2	17.601	0.1	1	2	1	1	1	1	6459.41	2
2	17.701	0.1	1	2	1	1	1	1	6459.41	2
2	17.801	0.1	1	1	1	1	1	1	6459.41	2
2	17.901	0.1	1	1	2	1	1	1	6459.41	2
2	18	0.1	1	1	1	1	1	1	6459.41	2
2	18.101	0.1	1	1	1	1	1	1	6459.41	2

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
2	18.201	0.1	1	1	1	1	1	1	6459.41	2
2	18.301	0.1	1	1	1	1	1	1	6459.41	2
2	18.401	0.1	1	1	1	1	1	1	6459.41	2
2	18.501	0.1	1	1	2	1	1	1	6459.41	2
3	18.601	0.1	1	1	2	1	1	1	6459.41	2
3	18.701	0.1	1	1	2	1	1	1	6459.41	2
3	18.801	0.1	1	1	1	1	1	1	6459.41	2
3	18.901	0.1	1	1	1	1	1	1	6459.41	2
3	19.001	0.1	1	1	1	1	1	1	6459.41	2
3	19.101	0.1	1	1	1	1	1	1	6459.41	2
3	19.2	0.1	1	1	2	1	1	1	6459.41	2
3	19.3	0.1	1	1	2	1	1	1	6459.41	2
3	19.4	0.1	1	1	2	1	1	1	6459.41	2
3	19.5	0.1	1	1	1	1	1	1	6459.41	2
3	19.6	0.1	1	1	1	1	1	1	6459.41	2
3	19.701	0.1	1	1	1	1	1	1	6459.41	2
3	19.8	0.1	1	1	1	1	1	1	6459.41	1
3	19.9	0.1	1	1	1	1	1	1	6459.41	2
3	20	0.1	1	1	2	1	1	1	6459.41	2

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
3	20.1	0.1	1	1	1	1	1	1	6459.41	2
3	20.2	0.1	1	1	1	1	1	1	6459.41	2
3	20.3	0.1	1	1	1	1	1	1	6459.41	2
3	20.4	0.1	1	1	1	1	1	1	6459.41	2
3	20.5	0.1	1	1	1	1	1	1	6459.41	2
3	20.6	0.1	1	1	1	1	1	1	6459.41	1
3	20.7	0.1	1	1	1	1	1	1	6459.41	1
3	20.8	0.1	1	1	1	1	1	1	6459.41	3
3	20.9	0.1	1	1	1	1	1	1	6459.41	3
3	21	0.1	1	1	1	1	1	1	6459.41	2
3	21.1	0.1	1	1	1	1	1	1	6459.41	1
3	21.2	0.1	1	1	1	1	1	1	6459.41	2
3	21.3	0.1	1	1	1	1	1	1	6459.41	2
3	21.4	0.1	1	1	1	1	1	1	6459.41	2
3	21.5	0.1	1	1	1	1	1	1	6459.41	2
3	21.6	0.1	1	1	1	1	1	1	6459.41	2
3	21.7	0.1	1	1	1	1	1	1	6459.41	2
3	21.8	0.1	1	1	1	1	1	1	6459.41	2
3	21.9	0.1	1	1	1	1	1	1	6459.41	2

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
3	22	0.1	1	1	1	1	1	1	6459.41	2
3	22.1	0.1	1	1	1	1	1	1	6459.41	2
3	22.199	0.1	1	1	1	1	1	1	6459.41	2
3	22.299	0.1	1	1	1	1	1	1	6459.41	2
3	22.399	0.1	1	1	1	1	1	1	6459.41	1
3	22.499	0.1	1	1	1	1	1	1	6459.41	2
3	22.599	0.1	1	1	1	1	1	1	6459.41	2
3	22.699	0.1	1	1	1	1	1	1	6459.41	2
3	22.799	0.1	1	1	1	1	1	1	6459.41	2
3	22.899	0.1	1	1	1	1	1	1	6459.41	2
3	22.999	0.1	1	1	1	1	1	1	6459.41	2
3	23.099	0.1	1	1	1	1	1	1	6459.41	2
3	23.199	0.1	1	1	1	1	1	1	6459.41	2
3	23.299	0.1	1	1	1	1	1	1	6459.41	2
3	23.399	0.1	1	1	1	1	1	1	6459.41	2
3	23.499	0.1	1	1	1	1	1	1	6459.41	2
3	23.599	0.1	1	1	1	1	1	1	6459.41	2
3	23.699	0.1	1	1	1	1	1	1	6459.41	1
3	23.799	0.1	1	1	1	1	1	1	6459.41	2

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
3	23.899	0.1	1	1	1	1	1	1	6459.41	2
3	23.999	0.1	1	1	1	1	1	1	6459.41	2
3	24.099	0.1	1	1	1	1	1	1	6459.41	2
3	24.199	0.1	1	1	1	1	1	1	6459.41	2
3	24.299	0.1	1	1	1	1	1	1	6459.41	2
3	24.399	0.1	1	1	1	1	1	1	6459.41	2
3	24.499	0.1	1	1	1	1	1	1	6459.41	2
3	24.599	0.1	1	1	1	1	1	1	6459.41	1
3	24.698	0.1	1	1	1	1	1	1	6459.41	2
3	24.798	0.1	1	1	1	1	1	1	6459.41	2
3	24.898	0.1	1	1	1	1	1	1	6459.41	2
3	24.998	0.1	1	1	1	1	1	1	6459.41	2
3	25.098	0.1	1	1	1	1	1	1	6459.41	3
3	25.198	0.1	1	1	2	1	1	1	6459.41	3
3	25.297	0.1	1	1	1	1	1	1	6459.41	3
3	25.398	0.1	1	1	1	1	1	1	6459.41	3
3	25.498	0.1	1	1	1	1	1	1	6459.41	3
3	25.598	0.1	1	1	1	1	1	1	6459.41	3
3	25.698	0.1	1	1	1	1	1	1	6459.41	3

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
4	25.798	0.1	1	1	1	1	1	1	20000	3
4	25.898	0.1	1	1	1	1	1	1	20000	4
4	25.998	0.1	1	2	2	1	1	1	20000	4
4	26.098	0.1	1	2	1	1	1	1	20000	4
4	26.197	0.1	1	2	1	1	1	1	20000	4
4	26.297	0.1	1	2	1	1	1	1	20000	4
4	26.396	0.1	1	2	1	1	1	1	20000	2
4	26.497	0.1	1	2	1	1	1	1	20000	4
4	26.597	0.1	1	2	1	1	1	1	20000	4
4	26.697	0.1	1	2	1	1	1	1	20000	4
4	26.797	0.1	1	2	1	1	1	1	20000	4
4	26.897	0.1	1	2	2	1	1	1	20000	4
4	26.997	0.1	1	2	2	1	1	1	20000	3
4	27.096	0.1	1	1	2	1	1	1	20000	4
4	27.197	0.1	1	1	1	1	1	1	20000	3
4	27.296	0.1	1	1	2	1	1	1	20000	3
4	27.397	0.1	1	1	2	1	1	1	20000	3
4	27.497	0.1	1	1	2	1	1	1	20000	3
5	27.597	0.1	1	1	2	1	1	1	4014.12	2

Section	Distance	Length	Grade	Differential speed limits	Skid resistance	Number of lanes	Operating speed	Median type	Vehicle flow AADT	Star rating
5	27.697	0.1	1	1	2	1	1	1	4014.12	2
5	27.797	0.1	1	1	1	1	1	1	4014.12	2
5	27.897	0.1	1	1	1	1	1	1	4014.12	2
5	27.997	0.1	1	1	1	1	1	1	4014.12	2
5	28.097	0.1	1	1	1	1	1	1	4014.12	2
5	28.197	0.1	1	1	1	1	1	1	4014.12	2
5	28.297	0.1	1	1	1	1	1	1	4014.12	2

A.2 Multi project – Fatality Rates dataset

This section presents an extract from the raw fatality data for the road network considered in this study (a road network from Andhra Pradesh state, India). Where available, existing records of the number of road fatalities occurring per year was used. However, detailed road crash data was not available on many of the road sections and therefore on these roads, estimates of the number of fatalities were made by iRAP based on typical fatality rates per billion vehicle kilometres travelled (vkt) (iRAP, 2014).

Section	Distance	Vehicle Occupant Fatality Estimation Run-Off LOC Driver-side per km per year	Vehicle Occupant Fatality Estimation Run-Off LOC Passenger-side per km per year	Vehicle Occupant Fatality Estimation Head-On LOC per km per year	Vehicle Occupant Fatality Estimation Head-On Overtaking per km per year	Vehicle Occupant Fatality Estimation Intersection per km per year	Vehicle Occupant Fatality Estimation Property Access per km per year	Vehicle Occupant Fatality Estimation Total per km per year
1	0.008	0.008033	0.008345	0.003959	0.011832	0.253683	0.002477	0.288329
1	0.109	0.027501	0.028568	0.013554	0.048608	0	0.010175	0.128406
1	0.209	0.027501	0.028568	0.013554	0.048608	0	0.010175	0.128406
1	0.309	0.027501	0.028568	0.013554	0.048608	0	0.012026	0.130256
1	0.409	0.027501	0.028568	0.013554	0.048608	0	0.010175	0.128406
1	0.508	0.027501	0.028568	0	0	0	0.007123	0.063191
1	0.609	0.016117	0.016742	0	0	0	0.007123	0.039982
1	0.709	0.004797	0.004983	0	0	0	0.007123	0.016902
1	0.809	0.006907	0.007175	0	0	0	0.007123	0.021205
1	0.909	0.004797	0.004983	0	0	0	0.007123	0.016902
1	1.009	0.004797	0.004983	0	0	0	0.007123	0.016902
1	1.109	0.013431	0.013952	0	0	0	0.007123	0.034505
1	1.209	0.006907	0.007175	0	0	0	0.007123	0.021205
1	1.309	0.004797	0.004983	0	0	0	0.007123	0.016902
1	1.409	0.004797	0.004983	0	0	0	0.007123	0.016902
1	1.507	0.02878	0.029897	0	0	0	0.008418	0.067094
1	1.608	0.110882	0.115186	0.054649	0.048608	0	0.010175	0.339501
1	1.708	0.031681	0.03291	0.015614	0.048608	0	0.010175	0.138988
1	1.809	0.083162	0.086389	0.054649	0.048608	0	0.010175	0.282984
1	1.908	0.036301	0.041138	0.019517	0.048608	0	0.010175	0.15574
1	2.008	0.039601	0.008228	0.019517	0.048608	0	0.010175	0.126129
1	2.108	0.039601	0.008228	0.019517	0.048608	0	0.012026	0.127979
1	2.208	0.046201	0.035996	0.02277	0.048608	1.852723	0.010175	2.016474
1	2.308	0.046201	0.035996	0.02277	0.048608	0	0.010175	0.16375
1	2.408	0.028876	0.039995	0.018975	0.048608	0	0.010175	0.14663
1	2.508	0.033001	0.031425	0.016265	0.048608	0	0.010175	0.139474
1	2.608	0.02475	0.025711	0.016265	0.048608	0	0.010175	0.12551

Section	Distance	Vehicle Occupant Fatality Estimation Run-Off LOC Driver-side per km per year	Vehicle Occupant Fatality Estimation Run-Off LOC Passenger-side per km per year	Vehicle Occupant Fatality Estimation Head-On LOC per km per year	Vehicle Occupant Fatality Estimation Head-On Overtaking per km per year	Vehicle Occupant Fatality Estimation Intersection per km per year	Vehicle Occupant Fatality Estimation Property Access per km per year	Vehicle Occupant Fatality Estimation Total per km per year
1	2.708	0.033001	0.014998	0.016265	0.048608	0	0.010175	0.123047
1	2.808	0.158403	0.164551	0.07807	0.048608	0	0.012026	0.461658
1	2.908	0.033001	0.034282	0.016265	0.048608	0	0.010175	0.14233
1	3.008	0.029701	0.041138	0.019517	0.048608	0	0.012026	0.15099
1	3.108	0.02475	0.034282	0.016265	0.048608	0	0.012026	0.13593
1	3.208	0.033001	0.025711	0.016265	0.048608	0	0.010175	0.13376
1	3.308	0.033001	0.025711	0.016265	0.048608	0	0.012026	0.13561
1	3.408	0.033001	0.034282	0.016265	0.048608	0	0.012026	0.14418
1	3.508	0.02475	0.025711	0.016265	0.048608	0	0.010175	0.12551
1	3.608	0.118802	0.123414	0.07807	0.048608	0	0.010175	0.379069
1	3.708	0.038501	0.039995	0.018975	0.048608	0	0.012026	0.158105
1	3.808	0.02475	0.034282	0.016265	0.048608	0	0.010175	0.13408
1	3.907	0.038501	0.039995	0.018975	0.048608	0	0.012026	0.158105
1	4.008	0.033001	0.034282	0.016265	0.048608	1.354804	0.012026	1.498984
1	4.108	0.033001	0.034282	0.016265	0.048608	0	0.012026	0.14418
1	4.208	0.028876	0.039995	0.018975	0.048608	0	0.012026	0.14848
1	4.308	0.02475	0.025711	0.016265	0.048608	0	0.010175	0.12551
1	4.408	0.02475	0.034282	0.016265	0.048608	1.354804	0.012026	1.490734
1	4.508	0.038501	0.029996	0.018975	0.048608	0	0.012026	0.148106
1	4.607	0.02475	0.025711	0.016265	0.048608	0	0.010175	0.12551
1	4.707	0.02475	0.025711	0.016265	0.048608	0	0.010175	0.12551
1	4.807	0.038501	0.029996	0.018975	0.048608	0	0.010175	0.146256
1	4.907	0.158403	0.164551	0.07807	0.048608	0	0.012026	0.461658
1	5.007	0.038501	0.039995	0.018975	0.048608	0	0.012026	0.158105
1	5.107	0.033001	0.034282	0.016265	0.048608	0	0.010175	0.14233
1	5.207	0.033001	0.034282	0.016265	0.048608	0	0.012026	0.14418
1	5.307	0.014438	0.034282	0.016265	0.048608	0	0.012026	0.125618
1	5.407	0.038501	0.039995	0.018975	0.048608	0	0.012026	0.158105
1	5.507	0.038501	0.039995	0.018975	0.048608	0	0.010175	0.156255
1	5.607	0.033001	0.034282	0.016265	0.048608	0	0.010175	0.14233
1	5.707	0.033001	0.034282	0.016265	0.048608	1.354804	0.010175	1.497134
1	5.807	0.033001	0.014998	0.016265	0.048608	0	0.010175	0.123047
1	5.907	0.038501	0.039995	0.018975	0.048608	0	0.012026	0.158105
1	6.007	0.039601	0.041138	0.019517	0.048608	0	0.010175	0.15904
1	6.107	0.039601	0.041138	0.019517	0.048608	0	0.010175	0.15904
1	6.206	0.033001	0.034282	0.016265	0.048608	0	0.010175	0.14233

Section	Distance	Vehicle Occupant Fatality Estimation Run-Off LOC Driver-side per km per year	Vehicle Occupant Fatality Estimation Run-Off LOC Passenger-side per km per year	Vehicle Occupant Fatality Estimation Head-On LOC per km per year	Vehicle Occupant Fatality Estimation Head-On Overtaking per km per year	Vehicle Occupant Fatality Estimation Intersection per km per year	Vehicle Occupant Fatality Estimation Property Access per km per year	Vehicle Occupant Fatality Estimation Total per km per year
1	6.307	0.033001	0.034282	0.016265	0.048608	0	0.012026	0.14418
1	6.407	0.033001	0.034282	0.016265	0.048608	1.354804	0.012026	1.498984
1	6.506	0.014438	0.034282	0.016265	0.048608	0	0.012026	0.125618
1	6.606	0.033001	0.034282	0.016265	0.048608	0	0.010175	0.14233
1	6.706	0.033001	0.034282	0.016265	0.048608	0	0.010175	0.14233
1	6.806	0.110882	0.086389	0.054649	0.048608	0	0.010175	0.310704
1	6.906	0.033001	0.025711	0.016265	0.048608	0	0.010175	0.13376
1	7.006	0.033001	0.034282	0.016265	0.048608	1.354804	0.012026	1.498984
1	7.106	0.034651	0.035996	0.017078	0.048608	1.354804	0.010175	1.501312
1	7.206	0.033001	0.034282	0.016265	0.048608	0	0.010175	0.14233
1	7.306	0.039601	0.041138	0.019517	0.048608	0	0.010175	0.15904
1	7.406	0.039601	0.041138	0.019517	0.048608	0	0.010175	0.15904
1	7.506	0.033001	0.034282	0.016265	0.048608	0	0.010175	0.14233
1	7.606	0.029701	0.030853	0.019517	0.048608	0	0.010175	0.138855
1	7.706	0.029701	0.030853	0.019517	0.048608	0	0.010175	0.138855
1	7.806	0.017325	0.017998	0.019517	0.048608	0	0.012026	0.115474
1	7.906	0.017325	0.017998	0.019517	0.048608	0	0.010175	0.113624
1	8.006	0.028876	0.029996	0.018975	0.048608	0	0.010175	0.136631
1	8.106	0.038501	0.039995	0.018975	0.048608	0	0.012026	0.158105
1	8.206	0.038501	0.029996	0.018975	0.048608	0	0.010175	0.146256
1	8.306	0.039601	0.041138	0.019517	0.048608	0	0.010175	0.15904
1	8.405	0.046201	0.047994	0.02277	0.048608	0	0.012026	0.177599
1	8.505	0.038501	0.039995	0.018975	0.048608	0	0.010175	0.156255
1	8.606	0.038501	0.039995	0.018975	0.048608	0	0.012026	0.158105
1	8.706	0.038501	0.039995	0.018975	0.048608	0	0.010175	0.156255
1	8.806	0.038501	0.039995	0.018975	0.048608	0	0.012026	0.158105
1	8.905	0.038501	0.039995	0.018975	0.048608	0	0.010175	0.156255
1	9.005	0.038501	0.029996	0.018975	0.048608	0	0.012026	0.148106
1	9.105	0.038501	0.029996	0.018975	0.048608	0	0.010175	0.146256
1	9.205	0.028876	0.029996	0.018975	0.048608	0	0.012026	0.138481
1	9.305	0.038501	0.039995	0.018975	0.048608	0	0.012026	0.158105
1	9.405	0.038501	0.039995	0.018975	0.048608	0	0.010175	0.156255
1	9.505	0.038501	0.039995	0.018975	0.048608	1.354804	0.012026	1.512909
1	9.605	0.046201	0.047994	0.02277	0.048608	0	0.012026	0.177599
1	9.705	0.038501	0.039995	0.018975	0.048608	0	0.010175	0.156255
1	9.805	0.028876	0.039995	0.018975	0.048608	0	0.010175	0.14663

Section	Distance	Vehicle Occupant Fatality Estimation Run-Off LOC Driver-side per km per year	Vehicle Occupant Fatality Estimation Run-Off LOC Passenger-side per km per year	Vehicle Occupant Fatality Estimation Head-On LOC per km per year	Vehicle Occupant Fatality Estimation Head-On Overtaking per km per year	Vehicle Occupant Fatality Estimation Intersection per km per year	Vehicle Occupant Fatality Estimation Property Access per km per year	Vehicle Occupant Fatality Estimation Total per km per year
1	9.905	0.028876	0.039995	0.018975	0.048608	0	0.010175	0.14663
1	10.005	0.039601	0.041138	0.019517	0.048608	0	0.010175	0.15904
1	10.105	0.138603	0.191977	0.091082	0.048608	0	0.010175	0.480444
1	10.205	0.028876	0.039995	0.018975	0.048608	0	0.012026	0.14848
1	10.305	0.029701	0.030853	0.019517	0.048608	0	0.010175	0.138855
1	10.405	0.029701	0.030853	0.019517	0.048608	0	0.010175	0.138855
1	10.505	0.034651	0.035996	0.017078	0.048608	1.354804	0.010175	1.501312
1	10.604	0.033001	0.034282	0.016265	0.048608	0	0.010175	0.14233
1	10.705	0.033001	0.034282	0.016265	0.048608	0	0.010175	0.14233
1	10.804	0.033001	0.034282	0.016265	0.048608	0	0.010175	0.14233
1	10.905	0.039601	0.041138	0.019517	0.048608	0	0.010175	0.15904
1	11.004	0.046201	0.047994	0.02277	0.048608	0	0.010175	0.175749
1	11.104	0.033001	0.034282	0.016265	0.048608	0	0.010175	0.14233

Appendix B
The WEKA outputs and decision trees

B-1 Output information and the decision tree for run-off accidents

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: Run-off accidents
Instances: 13662
Attributes: 13
Grade
Skid resistance / grip
Lane width
Curvature
Quality of curve
Road condition
Roadside severity - passenger-side distance
Roadside severity - driver-side distance
Paved shoulder - driver-side
Paved shoulder - passenger-side
Delineation
Vehicle Star Rating
Test mode: split 75.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree

```
Curvature = Cu1
|   Delineation = De2
|   |   Roadside severity - driver-side distance = RSdsd2
|   |   |   Quality of curve = QoC2
|   |   |   |   Lane width = LW1
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   |   Grade = Gr1: SR2 (1985.0/283.0)
|   |   |   |   |   |   Grade = Gr3: SR2 (73.0/10.0)
|   |   |   |   |   |   Grade = Gr5: SR1 (4.0)
|   |   |   |   |   |   Grade = Gr3: SR2 (0.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   |   Road condition = RC1: SR3 (89.0/7.0)
|   |   |   |   |   |   Road condition = RC2
|   |   |   |   |   |   |   Paved shoulder - driver-side = PSds4: SR2 (54.0/14.0)
|   |   |   |   |   |   |   Paved shoulder - driver-side = PSds3: SR2 (9.0/1.0)
|   |   |   |   |   |   |   Paved shoulder - driver-side = PSds2: SR3 (17.0)
|   |   |   |   |   |   |   Paved shoulder - driver-side = PSds1: SR2 (0.0)
|   |   |   |   |   |   Road condition = RC3: SR2 (71.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4
|   |   |   |   |   |   Road condition = RC1: SR3 (14.0/3.0)
```

| | | | | | | Road condition = RC2: SR3 (7.0)
| | | | | | | Road condition = RC3: SR2 (6.0/1.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
| | | | | Lane width = LW2
| | | | | | Road condition = RC1: SR2 (52.0/13.0)
| | | | | | Road condition = RC2: SR2 (107.0/25.0)
| | | | | | Road condition = RC3
| | | | | | | Roadside severity - passenger-side distance = RSpsd2: SR1 (93.0/14.0)
| | | | | | | Roadside severity - passenger-side distance = RSpsd3: SR2 (35.0)
| | | | | | | Roadside severity - passenger-side distance = RSpsd4: SR2 (2.0)
| | | | | | | Roadside severity - passenger-side distance = RSpsd1: SR1 (0.0)
| | | | | Lane width = LW3: SR3 (1.0)
| | | | Quality of curve = QoC1: SR2 (0.0)
| | | | Quality of curve = QoC3
| | | | | Lane width = LW1
| | | | | | Road condition = RC1
| | | | | | | Roadside severity - passenger-side distance = RSpsd2
| | | | | | | | Skid resistance / grip = SkRes1
| | | | | | | | | Paved shoulder - passenger-side = PSps4: SR1 (124.0/27.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps3: SR3 (72.0/6.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps2: SR3 (44.0/2.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps1: SR3 (0.0)
| | | | | | | | Skid resistance / grip = SkRes2: SR3 (616.0/28.0)
| | | | | | | | Skid resistance / grip = SkRes5: SR3 (0.0)
| | | | | | | | Skid resistance / grip = SkRes3: SR3 (0.0)
| | | | | | | Roadside severity - passenger-side distance = RSpsd3: SR2 (17.0/5.0)
| | | | | | | Roadside severity - passenger-side distance = RSpsd4
| | | | | | | | Paved shoulder - driver-side = PSds4: SR2 (21.0/4.0)
| | | | | | | | Paved shoulder - driver-side = PSds3: SR3 (2.0)
| | | | | | | | Paved shoulder - driver-side = PSds2: SR3 (3.0)
| | | | | | | | Paved shoulder - driver-side = PSds1: SR2 (0.0)
| | | | | | | Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
| | | | | | Road condition = RC2
| | | | | | | Roadside severity - passenger-side distance = RSpsd2
| | | | | | | | Paved shoulder - driver-side = PSds4: SR1 (149.0/79.0)
| | | | | | | | Paved shoulder - driver-side = PSds3: SR3 (0.0)
| | | | | | | | Paved shoulder - driver-side = PSds2: SR3 (4.0/1.0)
| | | | | | | | Paved shoulder - driver-side = PSds1: SR2 (1.0)
| | | | | | | Roadside severity - passenger-side distance = RSpsd3: SR2 (3.0)
| | | | | | | Roadside severity - passenger-side distance = RSpsd4: SR2 (4.0/1.0)
| | | | | | | Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
| | | | | | Road condition = RC3
| | | | | | | Paved shoulder - driver-side = PSds4
| | | | | | | | Grade = Gr1: SR2 (234.0/89.0)
| | | | | | | | Grade = Gr3: SR2 (0.0)
| | | | | | | | Grade = Gr5: SR2 (0.0)
| | | | | | | | Grade = Gr3: SR3 (2.0)
| | | | | | | Paved shoulder - driver-side = PSds3: SR2 (3.0/1.0)
| | | | | | | Paved shoulder - driver-side = PSds2: SR3 (1.0)
| | | | | | | Paved shoulder - driver-side = PSds1: SR1 (12.0/4.0)
| | | | | Lane width = LW2

```
|   |   |   |   |   Skid resistance / grip = SkRes1: SR4 (5.0/1.0)
|   |   |   |   |   Skid resistance / grip = SkRes2
|   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR2 (740.0/248.0)
|   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR3 (7.0/2.0)
|   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR2 (0.0)
|   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
|   |   |   |   |   Skid resistance / grip = SkRes5: SR2 (0.0)
|   |   |   |   |   Skid resistance / grip = SkRes3: SR2 (1.0)
|   |   |   |   Lane width = LW3
|   |   |   |   |   Road condition = RC1: SR2 (115.0/7.0)
|   |   |   |   |   Road condition = RC2: SR2 (129.0/14.0)
|   |   |   |   |   Road condition = RC3
|   |   |   |   |   |   Skid resistance / grip = SkRes1: SR2 (0.0)
|   |   |   |   |   |   Skid resistance / grip = SkRes2
|   |   |   |   |   |   |   Grade = Gr1
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR2 (810.0/446.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR1 (8.0/4.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR3 (3.0/1.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
|   |   |   |   |   |   |   Grade = Gr3: SR2 (0.0)
|   |   |   |   |   |   |   Grade = Gr5: SR2 (0.0)
|   |   |   |   |   |   |   Grade = Gr3: SR1 (16.0/2.0)
|   |   |   |   |   |   Skid resistance / grip = SkRes5: SR1 (11.0)
|   |   |   |   |   |   Skid resistance / grip = SkRes3: SR1 (7.0/1.0)
|   |   Roadside severity - driver-side distance = RSdsd3
|   |   |   Lane width = LW1
|   |   |   |   Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   Paved shoulder - driver-side = PSds4
|   |   |   |   |   |   Road condition = RC1
|   |   |   |   |   |   |   Quality of curve = QoC2: SR3 (8.0)
|   |   |   |   |   |   |   Quality of curve = QoC1: SR3 (0.0)
|   |   |   |   |   |   |   Quality of curve = QoC3: SR2 (12.0/5.0)
|   |   |   |   |   |   Road condition = RC2: SR2 (26.0/9.0)
|   |   |   |   |   |   Road condition = RC3
|   |   |   |   |   |   |   Skid resistance / grip = SkRes1: SR2 (13.0/1.0)
|   |   |   |   |   |   |   Skid resistance / grip = SkRes2: SR1 (7.0/3.0)
|   |   |   |   |   |   |   Skid resistance / grip = SkRes5: SR2 (0.0)
|   |   |   |   |   |   |   Skid resistance / grip = SkRes3: SR2 (0.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds3
|   |   |   |   |   |   Road condition = RC1: SR3 (14.0/3.0)
|   |   |   |   |   |   Road condition = RC2: SR2 (1.0)
|   |   |   |   |   |   Road condition = RC3: SR2 (4.0/1.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds2
|   |   |   |   |   |   Road condition = RC1: SR3 (5.0)
|   |   |   |   |   |   Road condition = RC2: SR3 (20.0/2.0)
|   |   |   |   |   |   Road condition = RC3: SR2 (5.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds1: SR2 (0.0)
|   |   |   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   Road condition = RC1
|   |   |   |   |   |   Quality of curve = QoC2: SR3 (65.0/1.0)
|   |   |   |   |   |   Quality of curve = QoC1: SR3 (0.0)
```

| | | | | | Quality of curve = QoC3
| | | | | | | Paved shoulder - driver-side = PSds4: SR2 (3.0)
| | | | | | | Paved shoulder - driver-side = PSds3: SR3 (19.0)
| | | | | | | Paved shoulder - driver-side = PSds2: SR4 (3.0)
| | | | | | | Paved shoulder - driver-side = PSds1: SR3 (0.0)
| | | | | Road condition = RC2: SR3 (97.0/3.0)
| | | | | Road condition = RC3: SR3 (28.0/7.0)
| | | | Roadside severity - passenger-side distance = RSpsd4
| | | | | Quality of curve = QoC2: SR3 (21.0)
| | | | | Quality of curve = QoC1: SR3 (0.0)
| | | | | Quality of curve = QoC3
| | | | | | Paved shoulder - driver-side = PSds4: SR3 (3.0/1.0)
| | | | | | Paved shoulder - driver-side = PSds3: SR4 (3.0/1.0)
| | | | | | Paved shoulder - driver-side = PSds2: SR4 (2.0)
| | | | | | Paved shoulder - driver-side = PSds1: SR4 (0.0)
| | | | Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
| | | Lane width = LW2
| | | | Skid resistance / grip = SkRes1
| | | | | Roadside severity - passenger-side distance = RSpsd2: SR2 (21.0/1.0)
| | | | | Roadside severity - passenger-side distance = RSpsd3
| | | | | | Road condition = RC1: SR2 (0.0)
| | | | | | Road condition = RC2: SR3 (8.0/2.0)
| | | | | | Road condition = RC3: SR2 (18.0/3.0)
| | | | | Roadside severity - passenger-side distance = RSpsd4: SR3 (1.0)
| | | | | Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
| | | | Skid resistance / grip = SkRes2: SR3 (7.0)
| | | | Skid resistance / grip = SkRes5: SR2 (0.0)
| | | | Skid resistance / grip = SkRes3: SR2 (0.0)
| | | Lane width = LW3: SR3 (20.0/9.0)
| | Roadside severity - driver-side distance = RSdsd4
| | | Roadside severity - passenger-side distance = RSpsd2
| | | | Road condition = RC1
| | | | | Paved shoulder - driver-side = PSds4
| | | | | | Quality of curve = QoC2: SR3 (20.0)
| | | | | | Quality of curve = QoC1: SR3 (0.0)
| | | | | | Quality of curve = QoC3: SR2 (14.0/2.0)
| | | | | Paved shoulder - driver-side = PSds3: SR3 (53.0/11.0)
| | | | | Paved shoulder - driver-side = PSds2: SR3 (47.0/23.0)
| | | | | Paved shoulder - driver-side = PSds1: SR5 (3.0/1.0)
| | | | Road condition = RC2: SR3 (9.0/4.0)
| | | | Road condition = RC3
| | | | | Skid resistance / grip = SkRes1
| | | | | | Paved shoulder - driver-side = PSds4: SR2 (26.0)
| | | | | | Paved shoulder - driver-side = PSds3: SR2 (0.0)
| | | | | | Paved shoulder - driver-side = PSds2: SR3 (2.0)
| | | | | | Paved shoulder - driver-side = PSds1: SR2 (0.0)
| | | | | Skid resistance / grip = SkRes2: SR1 (4.0/2.0)
| | | | | Skid resistance / grip = SkRes5: SR2 (0.0)
| | | | | Skid resistance / grip = SkRes3: SR2 (0.0)
| | | Roadside severity - passenger-side distance = RSpsd3
| | | | Quality of curve = QoC2: SR3 (52.0/8.0)

```
|   |   |   |   Quality of curve = QoC1: SR3 (0.0)
|   |   |   |   Quality of curve = QoC3
|   |   |   |   |   Paved shoulder - driver-side = PSds4: SR2 (7.0/2.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds3: SR4 (8.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds2: SR4 (0.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds1: SR4 (0.0)
|   |   |   Roadside severity - passenger-side distance = RSpsd4
|   |   |   |   Lane width = LW1
|   |   |   |   |   Paved shoulder - driver-side = PSds4
|   |   |   |   |   |   Road condition = RC1: SR3 (10.0/2.0)
|   |   |   |   |   |   Road condition = RC2: SR4 (16.0/4.0)
|   |   |   |   |   |   Road condition = RC3: SR3 (30.0/14.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds3: SR4 (26.0/9.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds2
|   |   |   |   |   |   Road condition = RC1: SR3 (22.0/2.0)
|   |   |   |   |   |   Road condition = RC2: SR4 (7.0)
|   |   |   |   |   |   Road condition = RC3: SR3 (7.0/1.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds1: SR3 (0.0)
|   |   |   |   Lane width = LW2: SR4 (20.0/1.0)
|   |   |   |   Lane width = LW3: SR3 (5.0)
|   |   |   Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
|   |   Roadside severity - driver-side distance = RSdsd1: SR3 (1.0)
|   Delineation = De1
|   |   Road condition = RC1
|   |   |   Quality of curve = QoC2
|   |   |   |   Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   Grade = Gr1: SR3 (2191.0/175.0)
|   |   |   |   |   Grade = Gr3: SR3 (37.0/1.0)
|   |   |   |   |   Grade = Gr5: SR2 (8.0)
|   |   |   |   |   Grade = Gr3: SR3 (0.0)
|   |   |   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd2: SR3 (106.0/3.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd3: SR3 (660.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd4
|   |   |   |   |   |   Paved shoulder - driver-side = PSds4: SR3 (1.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds3: SR4 (29.0/11.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds2: SR3 (62.0/28.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds1: SR4 (1.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR3 (0.0)
|   |   |   |   Roadside severity - passenger-side distance = RSpsd4
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd2: SR3 (84.0/8.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd3
|   |   |   |   |   |   Paved shoulder - driver-side = PSds4: SR4 (0.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds3: SR3 (5.0/1.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds2: SR4 (44.0/1.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds1: SR4 (0.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd4: SR4 (428.0/169.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR4 (0.0)
|   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
|   |   |   Quality of curve = QoC1: SR3 (0.0)
|   |   |   Quality of curve = QoC3
```

```
|   |   |   |   Paved shoulder - driver-side = PSds4
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd2: SR3 (17.0/2.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd3: SR5 (4.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd4: SR4 (1.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR3 (0.0)
|   |   |   |   Paved shoulder - driver-side = PSds3
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   |   Skid resistance / grip = SkRes1: SR2 (82.0/13.0)
|   |   |   |   |   |   Skid resistance / grip = SkRes2: SR5 (3.0)
|   |   |   |   |   |   Skid resistance / grip = SkRes5: SR2 (0.0)
|   |   |   |   |   |   Skid resistance / grip = SkRes3: SR2 (0.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd2: SR2 (9.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd3: SR2 (12.0/6.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd4: SR3 (4.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd2: SR2 (12.0/4.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd3: SR3 (3.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd4: SR3 (9.0/1.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR3 (0.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR1 (1.0)
|   |   |   |   Paved shoulder - driver-side = PSds2: SR3 (17.0)
|   |   |   |   Paved shoulder - driver-side = PSds1: SR2 (0.0)
|   |   Road condition = RC2
|   |   |   Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   Roadside severity - driver-side distance = RSdsd2
|   |   |   |   |   Quality of curve = QoC2: SR2 (450.0/79.0)
|   |   |   |   |   Quality of curve = QoC1: SR2 (0.0)
|   |   |   |   |   Quality of curve = QoC3
|   |   |   |   |   |   Skid resistance / grip = SkRes1: SR4 (3.0/1.0)
|   |   |   |   |   |   Skid resistance / grip = SkRes2: SR3 (6.0)
|   |   |   |   |   |   Skid resistance / grip = SkRes5: SR3 (0.0)
|   |   |   |   |   |   Skid resistance / grip = SkRes3: SR3 (0.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd3: SR3 (30.0/5.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd4: SR3 (4.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
|   |   |   Roadside severity - passenger-side distance = RSpsd3: SR3 (53.0/6.0)
|   |   |   Roadside severity - passenger-side distance = RSpsd4
|   |   |   |   Roadside severity - driver-side distance = RSdsd2: SR3 (3.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd3: SR3 (1.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd4: SR4 (8.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR4 (0.0)
|   |   |   Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
|   |   Road condition = RC3
|   |   |   Roadside severity - passenger-side distance = RSpsd2: SR2 (74.0/11.0)
|   |   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   Paved shoulder - driver-side = PSds4: SR2 (3.0/1.0)
|   |   |   |   Paved shoulder - driver-side = PSds3: SR3 (2.0/1.0)
|   |   |   |   Paved shoulder - driver-side = PSds2: SR3 (2.0)
|   |   |   |   Paved shoulder - driver-side = PSds1: SR3 (0.0)
```

| | | Roadside severity - passenger-side distance = RSpsd4: SR4 (1.0)
| | | Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
Curvature = Cu3
| Roadside severity - driver-side distance = RSdsd2: SR1 (439.0/22.0)
| Roadside severity - driver-side distance = RSdsd3
| | Paved shoulder - driver-side = PSds4: SR1 (21.0/2.0)
| | Paved shoulder - driver-side = PSds3: SR1 (0.0)
| | Paved shoulder - driver-side = PSds2: SR2 (15.0/3.0)
| | Paved shoulder - driver-side = PSds1: SR1 (0.0)
| Roadside severity - driver-side distance = RSdsd4
| | Lane width = LW1
| | | Quality of curve = QoC2: SR1 (6.0/1.0)
| | | Quality of curve = QoC1: SR3 (7.0)
| | | Quality of curve = QoC3: SR3 (0.0)
| | Lane width = LW2: SR2 (2.0)
| | Lane width = LW3: SR1 (1.0)
| Roadside severity - driver-side distance = RSdsd1: SR1 (0.0)
Curvature = Cu2
| Skid resistance / grip = SkRes1
| | Roadside severity - passenger-side distance = RSpsd2
| | | Road condition = RC1
| | | | Roadside severity - driver-side distance = RSdsd2
| | | | | Grade = Gr1
| | | | | | Paved shoulder - driver-side = PSds4
| | | | | | | Lane width = LW1
| | | | | | | | Quality of curve = QoC2
| | | | | | | | | Delineation = De2: SR1 (16.0/5.0)
| | | | | | | | | Delineation = De1: SR2 (11.0/3.0)
| | | | | | | | Quality of curve = QoC1
| | | | | | | | | Paved shoulder - passenger-side = PSps4: SR2 (59.0/9.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps3: SR3 (2.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps2: SR2 (0.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps1: SR2 (0.0)
| | | | | | | | Quality of curve = QoC3: SR2 (0.0)
| | | | | | | Lane width = LW2: SR1 (10.0/2.0)
| | | | | | | Lane width = LW3: SR2 (0.0)
| | | | | | Paved shoulder - driver-side = PSds3
| | | | | | | Quality of curve = QoC2: SR2 (87.0/9.0)
| | | | | | | Quality of curve = QoC1
| | | | | | | | Delineation = De2: SR2 (37.0/9.0)
| | | | | | | | Delineation = De1: SR1 (7.0/1.0)
| | | | | | | Quality of curve = QoC3: SR2 (0.0)
| | | | | | Paved shoulder - driver-side = PSds2
| | | | | | | Delineation = De2
| | | | | | | | Quality of curve = QoC2: SR1 (3.0)
| | | | | | | | Quality of curve = QoC1: SR2 (21.0/2.0)
| | | | | | | | Quality of curve = QoC3: SR2 (0.0)
| | | | | | | Delineation = De1: SR2 (207.0/21.0)
| | | | | | Paved shoulder - driver-side = PSds1: SR2 (1.0)
| | | | | Grade = Gr3: SR2 (12.0/1.0)
| | | | | Grade = Gr5: SR1 (7.0)

```
|   |   |   |   |   Grade = Gr3: SR2 (0.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd3
|   |   |   |   |   Quality of curve = QoC2: SR3 (22.0/4.0)
|   |   |   |   |   Quality of curve = QoC1: SR2 (3.0/1.0)
|   |   |   |   |   Quality of curve = QoC3: SR3 (0.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd4
|   |   |   |   |   Paved shoulder - driver-side = PSds4
|   |   |   |   |   |   Quality of curve = QoC2
|   |   |   |   |   |   |   Delineation = De2: SR1 (6.0)
|   |   |   |   |   |   |   Delineation = De1: SR2 (2.0)
|   |   |   |   |   |   Quality of curve = QoC1: SR2 (7.0/1.0)
|   |   |   |   |   |   Quality of curve = QoC3: SR2 (0.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds3: SR2 (8.0/3.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds2: SR2 (42.0/9.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds1: SR5 (3.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
|   |   |   Road condition = RC2
|   |   |   |   Roadside severity - driver-side distance = RSdsd2
|   |   |   |   |   Paved shoulder - driver-side = PSds4: SR1 (99.0/26.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds3: SR1 (67.0/20.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds2
|   |   |   |   |   |   Quality of curve = QoC2: SR1 (38.0/12.0)
|   |   |   |   |   |   Quality of curve = QoC1: SR2 (16.0/4.0)
|   |   |   |   |   |   Quality of curve = QoC3: SR1 (0.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds1: SR1 (0.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd3
|   |   |   |   |   Delineation = De2
|   |   |   |   |   |   Lane width = LW1: SR1 (6.0/3.0)
|   |   |   |   |   |   Lane width = LW2: SR2 (3.0/1.0)
|   |   |   |   |   |   Lane width = LW3: SR1 (0.0)
|   |   |   |   |   Delineation = De1: SR2 (5.0/2.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd4: SR2 (4.0/1.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR1 (0.0)
|   |   |   Road condition = RC3
|   |   |   |   Roadside severity - driver-side distance = RSdsd2: SR1 (108.0/6.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd3: SR1 (22.0/4.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd4: SR2 (5.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR1 (0.0)
|   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   Road condition = RC1
|   |   |   |   Roadside severity - driver-side distance = RSdsd2
|   |   |   |   |   Paved shoulder - driver-side = PSds4
|   |   |   |   |   |   Quality of curve = QoC2
|   |   |   |   |   |   |   Delineation = De2: SR1 (3.0/1.0)
|   |   |   |   |   |   |   Delineation = De1: SR2 (3.0)
|   |   |   |   |   |   Quality of curve = QoC1: SR2 (6.0)
|   |   |   |   |   |   Quality of curve = QoC3: SR2 (0.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds3: SR2 (2.0/1.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds2: SR3 (23.0/9.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds1: SR2 (0.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd3
```

| | | | | Delineation = De2
| | | | | | Quality of curve = QoC2: SR2 (2.0)
| | | | | | Quality of curve = QoC1: SR3 (8.0/1.0)
| | | | | | Quality of curve = QoC3: SR3 (0.0)
| | | | | Delineation = De1: SR3 (54.0)
| | | | Roadside severity - driver-side distance = RSdsd4
| | | | | Paved shoulder - driver-side = PSds4: SR1 (1.0)
| | | | | Paved shoulder - driver-side = PSds3
| | | | | | Delineation = De2: SR3 (3.0)
| | | | | | Delineation = De1: SR2 (2.0)
| | | | | Paved shoulder - driver-side = PSds2: SR3 (14.0)
| | | | | Paved shoulder - driver-side = PSds1: SR3 (0.0)
| | | | Roadside severity - driver-side distance = RSdsd1: SR3 (0.0)
| | | Road condition = RC2
| | | | Roadside severity - driver-side distance = RSdsd2: SR2 (27.0/3.0)
| | | | Roadside severity - driver-side distance = RSdsd3: SR2 (11.0/5.0)
| | | | Roadside severity - driver-side distance = RSdsd4: SR3 (6.0)
| | | | Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
| | | Road condition = RC3
| | | | Paved shoulder - driver-side = PSds4
| | | | | Roadside severity - driver-side distance = RSdsd2: SR1 (31.0/4.0)
| | | | | Roadside severity - driver-side distance = RSdsd3: SR2 (14.0/4.0)
| | | | | Roadside severity - driver-side distance = RSdsd4: SR2 (7.0/2.0)
| | | | | Roadside severity - driver-side distance = RSdsd1: SR1 (0.0)
| | | | Paved shoulder - driver-side = PSds3: SR2 (28.0/1.0)
| | | | Paved shoulder - driver-side = PSds2: SR2 (2.0)
| | | | Paved shoulder - driver-side = PSds1: SR2 (0.0)
| | Roadside severity - passenger-side distance = RSpsd4
| | | Roadside severity - driver-side distance = RSdsd2
| | | | Paved shoulder - driver-side = PSds4: SR1 (5.0/2.0)
| | | | Paved shoulder - driver-side = PSds3: SR3 (4.0/1.0)
| | | | Paved shoulder - driver-side = PSds2: SR4 (10.0/3.0)
| | | | Paved shoulder - driver-side = PSds1: SR4 (0.0)
| | | Roadside severity - driver-side distance = RSdsd3: SR3 (7.0)
| | | Roadside severity - driver-side distance = RSdsd4
| | | | Paved shoulder - driver-side = PSds4
| | | | | Quality of curve = QoC2
| | | | | | Delineation = De2: SR2 (6.0)
| | | | | | Delineation = De1: SR3 (2.0)
| | | | | Quality of curve = QoC1: SR3 (35.0)
| | | | | Quality of curve = QoC3: SR3 (0.0)
| | | | Paved shoulder - driver-side = PSds3: SR4 (10.0/3.0)
| | | | Paved shoulder - driver-side = PSds2: SR3 (32.0/2.0)
| | | | Paved shoulder - driver-side = PSds1: SR3 (0.0)
| | | Roadside severity - driver-side distance = RSdsd1: SR3 (0.0)
| | Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
| Skid resistance / grip = SkRes2
| | Paved shoulder - driver-side = PSds4: SR1 (741.0/88.0)
| | Paved shoulder - driver-side = PSds3
| | | Quality of curve = QoC2: SR1 (5.0)
| | | Quality of curve = QoC1: SR4 (2.0)

| | | Quality of curve = QoC3: SR1 (0.0)
| | Paved shoulder - driver-side = PSds2: SR2 (3.0/1.0)
| | Paved shoulder - driver-side = PSds1: SR1 (0.0)
| Skid resistance / grip = SkRes5: SR1 (9.0)
| Skid resistance / grip = SkRes3: SR1 (1.0)
Curvature = Cu4: SR1 (30.0/1.0)

Number of Leaves : 311

Size of the tree : 435

Time taken to build model: 0.12 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

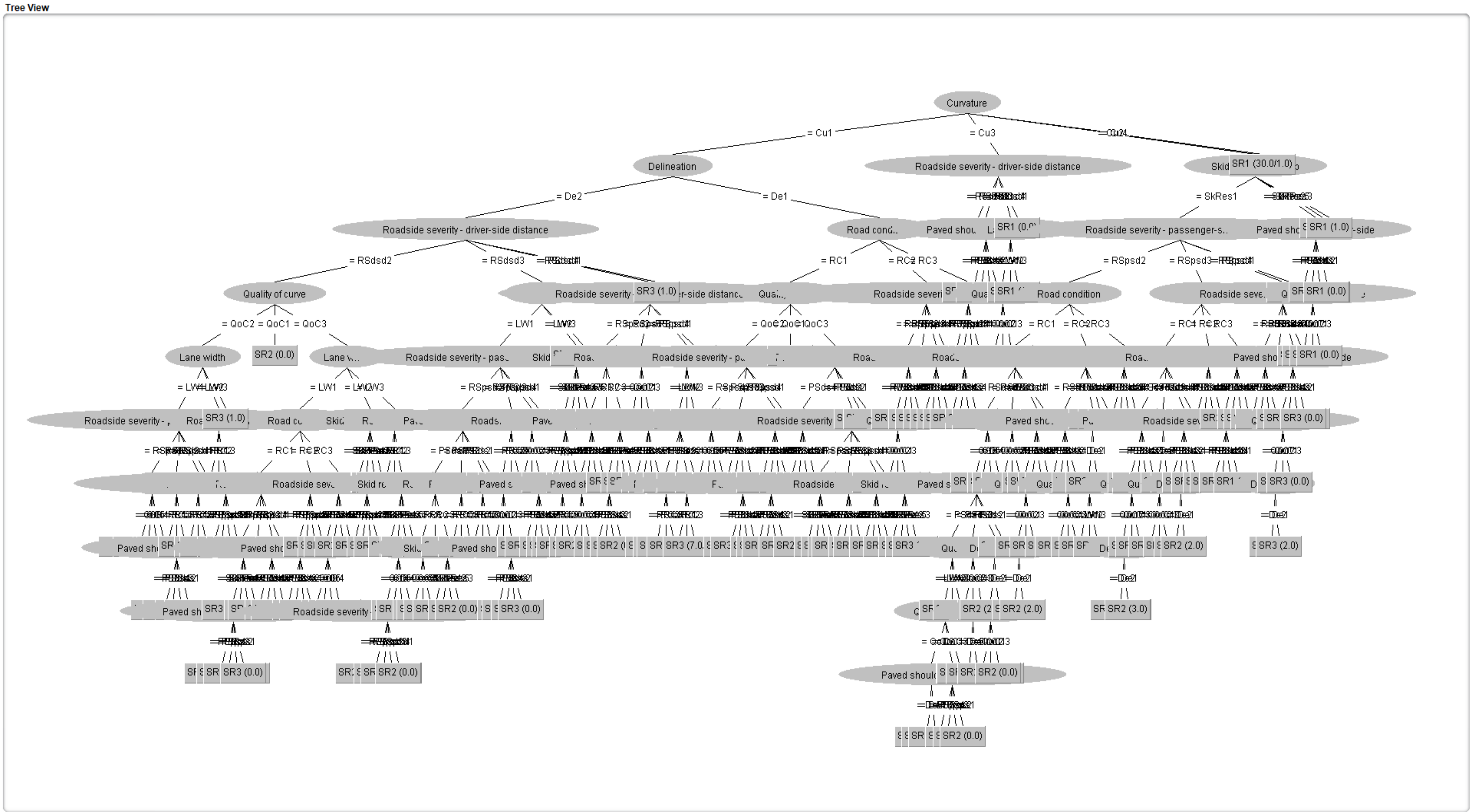
=== Summary ===

Correctly Classified Instances	2805	82.1376 %
Incorrectly Classified Instances	610	17.8624 %
Kappa statistic	0.7333	
Mean absolute error	0.1049	
Root mean squared error	0.2314	
Relative absolute error	39.02 %	
Root relative squared error	63.0697 %	
Total Number of Instances	3415	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.921	0.166	0.758	0.921	0.832	0.731	0.929	0.836	SR2
	0.791	0.050	0.917	0.791	0.849	0.763	0.927	0.903	SR3
	0.606	0.020	0.663	0.606	0.633	0.611	0.898	0.550	SR4
	0.791	0.029	0.840	0.791	0.815	0.781	0.975	0.880	SR1
	0.143	0.000	0.800	0.143	0.242	0.336	0.910	0.269	SR5
Weighted Avg.	0.821	0.086	0.831	0.821	0.819	0.742	0.933	0.848	

Figure B1. The decision tree for run-off accidents



B-2 Output information and the decision tree for (run-off + head-on) accidents

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     (run-off + head-on) accidents
Instances:    27888
Attributes:   17
              Differential speed limits
              Median type
              Number of lanes
              Grade
              Skid resistance / grip
              Lane width
              Curvature
              Quality of curve
              Road condition
              Roadside severity - passenger-side distance
              Roadside severity - driver-side distance
              Paved shoulder - driver-side
              Paved shoulder - passenger-side
              Delineation
              Vehicle Star Rating Raw
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------

Delineation = De2
|   Median type = MedType11
|   |   |   Quality of curve = QoC2
|   |   |   |   Roadside severity - driver-side distance = RSdsd2
|   |   |   |   |   Lane width = LW1
|   |   |   |   |   |   Curvature = Cu1
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   |   |   |   Grade = Gr1: SR2 (1842.0/186.0)
|   |   |   |   |   |   |   |   Grade = Gr3: SR2 (73.0/10.0)
|   |   |   |   |   |   |   |   Grade = Gr5: SR1 (4.0)
|   |   |   |   |   |   |   |   Grade = Gr3: SR2 (0.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   |   |   |   Road condition = RC1: SR3 (71.0/7.0)
|   |   |   |   |   |   |   |   Road condition = RC2
|   |   |   |   |   |   |   |   |   Paved shoulder - driver-side = PSds4: SR2 (49.0/14.0)
|   |   |   |   |   |   |   |   |   Paved shoulder - driver-side = PSds3: SR2 (7.0)
|   |   |   |   |   |   |   |   |   Paved shoulder - driver-side = PSds2: SR3 (17.0)
|   |   |   |   |   |   |   |   |   Paved shoulder - driver-side = PSds1: SR2 (0.0)
|   |   |   |   |   |   |   |   Road condition = RC3: SR2 (70.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4
|   |   |   |   |   |   |   |   Road condition = RC1: SR3 (6.0/1.0)
```

```
|   |   |   |   |   |   |   |  Road condition = RC2: SR3 (7.0)
|   |   |   |   |   |   |   |  Road condition = RC3: SR2 (6.0/1.0)
|   |   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
|   |   |   |   |   |  Curvature = Cu3: SR1 (5.0)
|   |   |   |   |   |  Curvature = Cu2: SR1 (13.0/3.0)
|   |   |   |   |   |  Curvature = Cu4: SR2 (0.0)
|   |   |   |   |  Lane width = LW2
|   |   |   |   |   |  Road condition = RC1: SR2 (46.0/13.0)
|   |   |   |   |   |  Road condition = RC2
|   |   |   |   |   |   |  Curvature = Cu1
|   |   |   |   |   |   |   |  Number of lanes = NOL1: SR2 (99.0/20.0)
|   |   |   |   |   |   |   |  Number of lanes = NOL2: SR4 (2.0)
|   |   |   |   |   |   |   |  Number of lanes = NOL3: SR2 (0.0)
|   |   |   |   |   |   |  Curvature = Cu3: SR1 (3.0)
|   |   |   |   |   |   |  Curvature = Cu2: SR2 (0.0)
|   |   |   |   |   |   |  Curvature = Cu4: SR2 (0.0)
|   |   |   |   |   |  Road condition = RC3
|   |   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd2: SR1 (93.0/14.0)
|   |   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd3: SR2 (35.0)
|   |   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd4: SR2 (2.0)
|   |   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd1: SR1 (0.0)
|   |   |   |   |  Lane width = LW3: SR3 (1.0)
|   |   |   |  Roadside severity - driver-side distance = RSdsd3
|   |   |   |   |  Lane width = LW1
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   |   |  Road condition = RC1: SR3 (18.0/1.0)
|   |   |   |   |   |   |  Road condition = RC2
|   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds4: SR2 (15.0/1.0)
|   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds3: SR2 (1.0)
|   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds2: SR3 (21.0/3.0)
|   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds1: SR3 (0.0)
|   |   |   |   |   |   |  Road condition = RC3: SR2 (21.0/2.0)
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   |   |  Curvature = Cu1: SR3 (182.0/11.0)
|   |   |   |   |   |   |  Curvature = Cu3: SR3 (0.0)
|   |   |   |   |   |   |  Curvature = Cu2: SR2 (2.0)
|   |   |   |   |   |   |  Curvature = Cu4: SR3 (0.0)
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd4: SR3 (21.0)
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
|   |   |   |   |  Lane width = LW2
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd2: SR2 (17.0)
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   |   |  Road condition = RC1: SR2 (0.0)
|   |   |   |   |   |   |  Road condition = RC2: SR3 (8.0/2.0)
|   |   |   |   |   |   |  Road condition = RC3: SR2 (18.0/3.0)
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd4: SR3 (1.0)
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
|   |   |   |   |  Lane width = LW3: SR3 (0.0)
|   |   |   |  Roadside severity - driver-side distance = RSdsd4
|   |   |   |   |  Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   |  Road condition = RC1: SR3 (24.0)
```

```
|   |   |   |   |   |   Road condition = RC2: SR3 (6.0/1.0)
|   |   |   |   |   |   Road condition = RC3
|   |   |   |   |   |   |   Paved shoulder - driver-side = PSds4: SR2 (25.0)
|   |   |   |   |   |   |   Paved shoulder - driver-side = PSds3: SR2 (0.0)
|   |   |   |   |   |   |   Paved shoulder - driver-side = PSds2: SR3 (2.0)
|   |   |   |   |   |   |   Paved shoulder - driver-side = PSds1: SR2 (0.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR3 (35.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4
|   |   |   |   |   |   Lane width = LW1
|   |   |   |   |   |   |   Road condition = RC1: SR4 (9.0/2.0)
|   |   |   |   |   |   |   Road condition = RC2: SR4 (25.0/4.0)
|   |   |   |   |   |   |   Road condition = RC3: SR3 (34.0/13.0)
|   |   |   |   |   |   Lane width = LW2: SR4 (20.0/2.0)
|   |   |   |   |   |   Lane width = LW3: SR4 (0.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
|   |   |   Quality of curve = QoC1
|   |   |   |   Roadside severity - driver-side distance = RSdsd2
|   |   |   |   |   Curvature = Cu1: SR1 (0.0)
|   |   |   |   |   Curvature = Cu3: SR1 (119.0/8.0)
|   |   |   |   |   Curvature = Cu2
|   |   |   |   |   |   Lane width = LW1
|   |   |   |   |   |   |   Road condition = RC1: SR2 (113.0/11.0)
|   |   |   |   |   |   |   Road condition = RC2
|   |   |   |   |   |   |   |   Paved shoulder - driver-side = PSds4
|   |   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   |   |   |   |   |   Grade = Gr1: SR1 (56.0/14.0)
|   |   |   |   |   |   |   |   |   |   Grade = Gr3: SR2 (3.0/1.0)
|   |   |   |   |   |   |   |   |   |   Grade = Gr5: SR1 (0.0)
|   |   |   |   |   |   |   |   |   |   Grade = Gr3: SR1 (0.0)
|   |   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR2 (5.0)
|   |   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR1 (0.0)
|   |   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR1 (0.0)
|   |   |   |   |   |   |   |   Paved shoulder - driver-side = PSds3
|   |   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR1 (14.0/3.0)
|   |   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR2 (3.0)
|   |   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR2 (1.0)
|   |   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR1 (0.0)
|   |   |   |   |   |   |   |   Paved shoulder - driver-side = PSds2: SR2 (13.0)
|   |   |   |   |   |   |   |   Paved shoulder - driver-side = PSds1: SR1 (0.0)
|   |   |   |   |   |   |   Road condition = RC3
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR1 (65.0/3.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   |   |   |   |   Paved shoulder - driver-side = PSds4: SR1 (12.0/2.0)
|   |   |   |   |   |   |   |   |   Paved shoulder - driver-side = PSds3: SR2 (27.0)
|   |   |   |   |   |   |   |   |   Paved shoulder - driver-side = PSds2: SR2 (2.0)
|   |   |   |   |   |   |   |   |   Paved shoulder - driver-side = PSds1: SR2 (0.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR1 (0.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR1 (0.0)
|   |   |   |   |   |   Lane width = LW2
|   |   |   |   |   |   |   Road condition = RC1: SR1 (7.0/2.0)
```

```
|   |   |   |   |   |   |   Road condition = RC2
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR1 (22.0/1.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR2 (9.0/2.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR1 (0.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR1 (0.0)
|   |   |   |   |   |   |   Road condition = RC3: SR1 (53.0/2.0)
|   |   |   |   |   |   Lane width = LW3: SR1 (0.0)
|   |   |   |   |   Curvature = Cu4: SR1 (12.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd3
|   |   |   |   |   Curvature = Cu1: SR1 (0.0)
|   |   |   |   |   Curvature = Cu3: SR1 (16.0/1.0)
|   |   |   |   |   Curvature = Cu2
|   |   |   |   |   |   Lane width = LW1
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR2 (9.0/4.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   |   |   |   Road condition = RC1: SR3 (8.0/1.0)
|   |   |   |   |   |   |   |   Road condition = RC2: SR2 (9.0/5.0)
|   |   |   |   |   |   |   |   Road condition = RC3: SR2 (2.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR3 (2.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
|   |   |   |   |   |   Lane width = LW2
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   |   |   |   Road condition = RC1: SR1 (0.0)
|   |   |   |   |   |   |   |   Road condition = RC2: SR2 (3.0/1.0)
|   |   |   |   |   |   |   |   Road condition = RC3: SR1 (17.0/1.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR2 (14.0/4.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR1 (0.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR1 (0.0)
|   |   |   |   |   |   Lane width = LW3: SR2 (0.0)
|   |   |   |   |   Curvature = Cu4: SR1 (2.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd4
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR2 (14.0/1.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   |   Road condition = RC1: SR3 (2.0)
|   |   |   |   |   |   Road condition = RC2: SR3 (5.0)
|   |   |   |   |   |   Road condition = RC3: SR2 (7.0/2.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR3 (47.0/3.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR1 (0.0)
|   |   |   Quality of curve = QoC3
|   |   |   |   Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   Road condition = RC1: SR3 (127.0/8.0)
|   |   |   |   |   Road condition = RC2: SR3 (4.0/2.0)
|   |   |   |   |   Road condition = RC3: SR4 (6.0/1.0)
|   |   |   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd2: SR2 (1.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd3
|   |   |   |   |   |   Paved shoulder - driver-side = PSds4: SR3 (0.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds3: SR3 (27.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds2: SR4 (3.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds1: SR3 (0.0)
```

| | | | | Roadside severity - driver-side distance = RSdsd4: SR4 (8.0)
| | | | | Roadside severity - driver-side distance = RSdsd1: SR3 (0.0)
| | | | Roadside severity - passenger-side distance = RSpsd4
| | | | | Roadside severity - driver-side distance = RSdsd2: SR3 (7.0)
| | | | | Roadside severity - driver-side distance = RSdsd3: SR4 (5.0/1.0)
| | | | | Roadside severity - driver-side distance = RSdsd4: SR4 (20.0/8.0)
| | | | | Roadside severity - driver-side distance = RSdsd1: SR4 (0.0)
| | | | Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
| | | Curvature = Cu1
| | | | Differential speed limits = DefSpeed1
| | | | | Roadside severity - driver-side distance = RSdsd2
| | | | | | Roadside severity - passenger-side distance = RSpsd2
| | | | | | | Paved shoulder - driver-side = PSds4
| | | | | | | | Paved shoulder - passenger-side = PSps4
| | | | | | | | | Road condition = RC1
| | | | | | | | | | Number of lanes = NOL1
| | | | | | | | | | | Skid resistance / grip = SkRes1: SR1 (3823.0/1064.0)
| | | | | | | | | | | Skid resistance / grip = SkRes2: SR2 (47.0/19.0)
| | | | | | | | | | | Skid resistance / grip = SkRes5: SR1 (0.0)
| | | | | | | | | | | Skid resistance / grip = SkRes3: SR1 (1.0)
| | | | | | | | | | | Skid resistance / grip = SkRes4: SR1 (2.0)
| | | | | | | | | | Number of lanes = NOL2: SR3 (9.0)
| | | | | | | | | | Number of lanes = NOL3: SR1 (0.0)
| | | | | | | | | Road condition = RC2: SR1 (1215.0/226.0)
| | | | | | | | | Road condition = RC3: SR1 (634.0/129.0)
| | | | | | | | Paved shoulder - passenger-side = PSps3
| | | | | | | | | Skid resistance / grip = SkRes1: SR3 (10.0/3.0)
| | | | | | | | | Skid resistance / grip = SkRes2: SR1 (2.0)
| | | | | | | | | Skid resistance / grip = SkRes5: SR3 (0.0)
| | | | | | | | | Skid resistance / grip = SkRes3: SR3 (0.0)
| | | | | | | | | Skid resistance / grip = SkRes4: SR3 (0.0)
| | | | | | | | Paved shoulder - passenger-side = PSps2: SR1 (35.0/11.0)
| | | | | | | | Paved shoulder - passenger-side = PSps1: SR3 (5.0)
| | | | | | | Paved shoulder - driver-side = PSds3
| | | | | | | | Road condition = RC1: SR1 (110.0/37.0)
| | | | | | | | Road condition = RC2
| | | | | | | | | Paved shoulder - passenger-side = PSps4: SR1 (0.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps3: SR1 (35.0/6.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps2: SR3 (34.0/15.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps1: SR1 (0.0)
| | | | | | | | Road condition = RC3: SR1 (2.0)
| | | | | | | Paved shoulder - driver-side = PSds2
| | | | | | | | Road condition = RC1
| | | | | | | | | Paved shoulder - passenger-side = PSps4: SR2 (5.0/3.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps3: SR3 (3.0/1.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps2
| | | | | | | | | | Lane width = LW1: SR2 (203.0/33.0)
| | | | | | | | | | Lane width = LW2: SR4 (4.0/1.0)
| | | | | | | | | | Lane width = LW3: SR2 (0.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps1: SR3 (1.0)
| | | | | | | | Road condition = RC2

| | | | | | | | | Skid resistance / grip = SkRes1: SR1 (65.0/20.0)
| | | | | | | | | Skid resistance / grip = SkRes2: SR3 (2.0)
| | | | | | | | | Skid resistance / grip = SkRes5: SR1 (0.0)
| | | | | | | | | Skid resistance / grip = SkRes3: SR1 (0.0)
| | | | | | | | | Skid resistance / grip = SkRes4: SR1 (0.0)
| | | | | | | | Road condition = RC3: SR1 (25.0/2.0)
| | | | | | | Paved shoulder - driver-side = PSds1
| | | | | | | | Road condition = RC1: SR2 (2.0)
| | | | | | | | Road condition = RC2: SR2 (1.0)
| | | | | | | | Road condition = RC3: SR1 (8.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd3
| | | | | | | Lane width = LW1
| | | | | | | | Road condition = RC1: SR2 (275.0/80.0)
| | | | | | | | Road condition = RC2
| | | | | | | | | Paved shoulder - driver-side = PSds4: SR2 (69.0/30.0)
| | | | | | | | | Paved shoulder - driver-side = PSds3: SR3 (3.0/1.0)
| | | | | | | | | Paved shoulder - driver-side = PSds2
| | | | | | | | | | Skid resistance / grip = SkRes1: SR2 (14.0/9.0)
| | | | | | | | | | Skid resistance / grip = SkRes2: SR3 (3.0)
| | | | | | | | | | Skid resistance / grip = SkRes5: SR3 (0.0)
| | | | | | | | | | Skid resistance / grip = SkRes3: SR3 (0.0)
| | | | | | | | | | Skid resistance / grip = SkRes4: SR3 (0.0)
| | | | | | | | | Paved shoulder - driver-side = PSds1: SR2 (0.0)
| | | | | | | | Road condition = RC3
| | | | | | | | | Paved shoulder - passenger-side = PSps4: SR1 (27.0/10.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps3: SR1 (0.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps2: SR2 (8.0/3.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps1: SR1 (0.0)
| | | | | | | Lane width = LW2: SR1 (210.0/85.0)
| | | | | | | Lane width = LW3: SR1 (127.0/33.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd4
| | | | | | | Paved shoulder - passenger-side = PSps4: SR2 (239.0/57.0)
| | | | | | | Paved shoulder - passenger-side = PSps3: SR2 (15.0/1.0)
| | | | | | | Paved shoulder - passenger-side = PSps2
| | | | | | | | Road condition = RC1: SR3 (4.0/1.0)
| | | | | | | | Road condition = RC2: SR1 (4.0/2.0)
| | | | | | | | Road condition = RC3: SR3 (0.0)
| | | | | | | Paved shoulder - passenger-side = PSps1: SR2 (0.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd1: SR1 (145.0/39.0)
| | | | | Roadside severity - driver-side distance = RSdsd3
| | | | | | Roadside severity - passenger-side distance = RSpsd2
| | | | | | | Lane width = LW1
| | | | | | | | Road condition = RC1: SR2 (219.0/73.0)
| | | | | | | | Road condition = RC2
| | | | | | | | | Paved shoulder - driver-side = PSds4: SR2 (53.0/25.0)
| | | | | | | | | Paved shoulder - driver-side = PSds3: SR3 (3.0/1.0)
| | | | | | | | | Paved shoulder - driver-side = PSds2: SR4 (15.0/8.0)
| | | | | | | | | Paved shoulder - driver-side = PSds1: SR2 (0.0)
| | | | | | | | Road condition = RC3
| | | | | | | | | Paved shoulder - passenger-side = PSps4: SR1 (23.0/10.0)
| | | | | | | | | Paved shoulder - passenger-side = PSps3: SR1 (0.0)

```
|   |   |   |   |   |   |   |   |  Paved shoulder - passenger-side = PSps2: SR2 (7.0/2.0)
|   |   |   |   |   |   |   |   |  Paved shoulder - passenger-side = PSps1: SR1 (0.0)
|   |   |   |   |   |   |  Lane width = LW2
|   |   |   |   |   |   |   |  Road condition = RC1: SR2 (89.0/47.0)
|   |   |   |   |   |   |   |  Road condition = RC2
|   |   |   |   |   |   |   |   |  Skid resistance / grip = SkRes1: SR1 (27.0/9.0)
|   |   |   |   |   |   |   |   |  Skid resistance / grip = SkRes2: SR3 (4.0/1.0)
|   |   |   |   |   |   |   |   |  Skid resistance / grip = SkRes5: SR1 (0.0)
|   |   |   |   |   |   |   |   |  Skid resistance / grip = SkRes3: SR1 (0.0)
|   |   |   |   |   |   |   |   |  Skid resistance / grip = SkRes4: SR1 (0.0)
|   |   |   |   |   |   |   |  Road condition = RC3: SR1 (17.0/5.0)
|   |   |   |   |   |   |  Lane width = LW3: SR1 (72.0/22.0)
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   |   |  Paved shoulder - passenger-side = PSps4: SR2 (315.0/120.0)
|   |   |   |   |   |   |  Paved shoulder - passenger-side = PSps3: SR4 (4.0/1.0)
|   |   |   |   |   |   |  Paved shoulder - passenger-side = PSps2: SR4 (23.0/8.0)
|   |   |   |   |   |   |  Paved shoulder - passenger-side = PSps1: SR2 (0.0)
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd4
|   |   |   |   |   |   |  Road condition = RC1: SR3 (116.0/25.0)
|   |   |   |   |   |   |  Road condition = RC2
|   |   |   |   |   |   |   |  Lane width = LW1
|   |   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds4: SR3 (2.0/1.0)
|   |   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds3: SR2 (0.0)
|   |   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds2: SR2 (3.0/1.0)
|   |   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds1: SR2 (0.0)
|   |   |   |   |   |   |   |  Lane width = LW2: SR3 (12.0/6.0)
|   |   |   |   |   |   |   |  Lane width = LW3: SR2 (2.0)
|   |   |   |   |   |   |  Road condition = RC3: SR2 (11.0/3.0)
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd1
|   |   |   |   |   |   |  Road condition = RC1: SR3 (10.0/6.0)
|   |   |   |   |   |   |  Road condition = RC2: SR1 (3.0)
|   |   |   |   |   |   |  Road condition = RC3: SR1 (1.0)
|   |   |   |   |  Roadside severity - driver-side distance = RSdsd4
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   |   |  Paved shoulder - passenger-side = PSps4: SR2 (247.0/53.0)
|   |   |   |   |   |   |  Paved shoulder - passenger-side = PSps3: SR2 (8.0/2.0)
|   |   |   |   |   |   |  Paved shoulder - passenger-side = PSps2: SR3 (16.0/5.0)
|   |   |   |   |   |   |  Paved shoulder - passenger-side = PSps1: SR2 (0.0)
|   |   |   |   |   |  Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   |   |  Road condition = RC1
|   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds4: SR3 (78.0/19.0)
|   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds3: SR3 (1.0)
|   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds2: SR5 (11.0/6.0)
|   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds1: SR3 (0.0)
|   |   |   |   |   |   |  Road condition = RC2
|   |   |   |   |   |   |   |  Lane width = LW1
|   |   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds4: SR3 (11.0/4.0)
|   |   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds3: SR4 (1.0)
|   |   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds2: SR4 (2.0/1.0)
|   |   |   |   |   |   |   |   |  Paved shoulder - driver-side = PSds1: SR3 (0.0)
|   |   |   |   |   |   |   |  Lane width = LW2: SR3 (5.0/2.0)
```

| | | | | | | | Lane width = LW3: SR2 (9.0/2.0)
| | | | | | | Road condition = RC3: SR2 (65.0/5.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd4
| | | | | | | Paved shoulder - passenger-side = PSps4: SR3 (246.0/18.0)
| | | | | | | Paved shoulder - passenger-side = PSps3: SR3 (2.0)
| | | | | | | Paved shoulder - passenger-side = PSps2: SR5 (7.0/2.0)
| | | | | | | Paved shoulder - passenger-side = PSps1: SR3 (0.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd1: SR2 (8.0/2.0)
| | | | | Roadside severity - driver-side distance = RSdsd1
| | | | | | Roadside severity - passenger-side distance = RSpsd2: SR1 (250.0/50.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd3
| | | | | | | Road condition = RC1: SR2 (9.0/5.0)
| | | | | | | Road condition = RC2: SR1 (2.0)
| | | | | | | Road condition = RC3: SR1 (0.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd4: SR2 (13.0/2.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd1: SR1 (444.0/127.0)
| | | | Differential speed limits = DefSpeed2
| | | | | Lane width = LW1
| | | | | | Road condition = RC1: SR3 (594.0/12.0)
| | | | | | Road condition = RC2: SR3 (66.0/17.0)
| | | | | | Road condition = RC3
| | | | | | | Roadside severity - driver-side distance = RSdsd2
| | | | | | | | Paved shoulder - driver-side = PSds4
| | | | | | | | | Grade = Gr1: SR2 (197.0/53.0)
| | | | | | | | | Grade = Gr3: SR2 (0.0)
| | | | | | | | | Grade = Gr5: SR2 (0.0)
| | | | | | | | | Grade = Gr3: SR3 (2.0)
| | | | | | | | Paved shoulder - driver-side = PSds3: SR2 (1.0)
| | | | | | | | Paved shoulder - driver-side = PSds2: SR3 (1.0)
| | | | | | | | Paved shoulder - driver-side = PSds1: SR3 (4.0)
| | | | | | | Roadside severity - driver-side distance = RSdsd3: SR3 (3.0)
| | | | | | | Roadside severity - driver-side distance = RSdsd4: SR4 (1.0)
| | | | | | | Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
| | | | | Lane width = LW2
| | | | | | Roadside severity - driver-side distance = RSdsd2
| | | | | | | Roadside severity - passenger-side distance = RSpsd2: SR2 (712.0/234.0)
| | | | | | | Roadside severity - passenger-side distance = RSpsd3: SR3 (6.0/1.0)
| | | | | | | Roadside severity - passenger-side distance = RSpsd4: SR2 (0.0)
| | | | | | | Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
| | | | | | Roadside severity - driver-side distance = RSdsd3: SR3 (3.0)
| | | | | | Roadside severity - driver-side distance = RSdsd4: SR3 (3.0/1.0)
| | | | | | Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
| | | | | Lane width = LW3
| | | | | | Road condition = RC1: SR2 (115.0/7.0)
| | | | | | Road condition = RC2: SR2 (128.0/13.0)
| | | | | | Road condition = RC3
| | | | | | | Skid resistance / grip = SkRes1: SR2 (0.0)
| | | | | | | Skid resistance / grip = SkRes2
| | | | | | | | Grade = Gr1
| | | | | | | | | Roadside severity - driver-side distance = RSdsd2: SR2 (654.0/289.0)
| | | | | | | | | Roadside severity - driver-side distance = RSdsd3: SR2 (0.0)

| | | | | | | | | Roadside severity - driver-side distance = RSdsd4: SR3 (2.0)
| | | | | | | | | Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
| | | | | | | | Grade = Gr3: SR2 (0.0)
| | | | | | | | Grade = Gr5: SR2 (0.0)
| | | | | | | | Grade = Gr3: SR1 (16.0/2.0)
| | | | | | | Skid resistance / grip = SkRes5: SR1 (11.0)
| | | | | | | Skid resistance / grip = SkRes3: SR1 (7.0/1.0)
| | | | | | | Skid resistance / grip = SkRes4: SR2 (0.0)
| | | Curvature = Cu3: SR1 (393.0/19.0)
| | | Curvature = Cu2
| | | | Roadside severity - driver-side distance = RSdsd2: SR1 (2009.0/317.0)
| | | | Roadside severity - driver-side distance = RSdsd3: SR1 (125.0/40.0)
| | | | Roadside severity - driver-side distance = RSdsd4
| | | | | Roadside severity - passenger-side distance = RSpsd2
| | | | | | Paved shoulder - driver-side = PSds4: SR1 (34.0/2.0)
| | | | | | Paved shoulder - driver-side = PSds3: SR1 (1.0)
| | | | | | Paved shoulder - driver-side = PSds2: SR2 (6.0/1.0)
| | | | | | Paved shoulder - driver-side = PSds1: SR1 (0.0)
| | | | | Roadside severity - passenger-side distance = RSpsd3
| | | | | | Road condition = RC1: SR1 (4.0/1.0)
| | | | | | Road condition = RC2: SR4 (3.0/1.0)
| | | | | | Road condition = RC3: SR1 (2.0)
| | | | | Roadside severity - passenger-side distance = RSpsd4: SR2 (32.0/7.0)
| | | | | Roadside severity - passenger-side distance = RSpsd1: SR1 (1.0)
| | | | Roadside severity - driver-side distance = RSdsd1
| | | | | Paved shoulder - driver-side = PSds4: SR1 (98.0/10.0)
| | | | | Paved shoulder - driver-side = PSds3: SR1 (0.0)
| | | | | Paved shoulder - driver-side = PSds2
| | | | | | Roadside severity - passenger-side distance = RSpsd2: SR3 (5.0/1.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd3: SR1 (0.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd4: SR1 (0.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd1: SR1 (7.0/1.0)
| | | | | Paved shoulder - driver-side = PSds1: SR1 (0.0)
| | | Curvature = Cu4: SR1 (41.0)
| Median type = MedType3
| | Road condition = RC1: SR4 (9.0/2.0)
| | Road condition = RC2: SR2 (2.0/1.0)
| | Road condition = RC3: SR4 (0.0)
| Median type = MedType7
| | Curvature = Cu1
| | | Roadside severity - driver-side distance = RSdsd2
| | | | Roadside severity - passenger-side distance = RSpsd2
| | | | | Road condition = RC1: SR2 (20.0/6.0)
| | | | | Road condition = RC2: SR4 (35.0/17.0)
| | | | | Road condition = RC3: SR1 (7.0/3.0)
| | | | Roadside severity - passenger-side distance = RSpsd3
| | | | | Quality of curve = QoC2: SR2 (8.0/1.0)
| | | | | Quality of curve = QoC1: SR2 (0.0)
| | | | | Quality of curve = QoC3: SR4 (3.0)
| | | | Roadside severity - passenger-side distance = RSpsd4: SR3 (3.0)
| | | | Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)

| | | Roadside severity - driver-side distance = RSdsd3
| | | | Quality of curve = QoC2: SR2 (7.0/1.0)
| | | | Quality of curve = QoC1: SR2 (0.0)
| | | | Quality of curve = QoC3
| | | | | Lane width = LW1: SR5 (2.0)
| | | | | Lane width = LW2: SR4 (4.0)
| | | | | Lane width = LW3: SR4 (0.0)
| | | Roadside severity - driver-side distance = RSdsd4
| | | | Quality of curve = QoC2
| | | | | Paved shoulder - driver-side = PSds4: SR2 (1.0)
| | | | | Paved shoulder - driver-side = PSds3: SR3 (30.0/10.0)
| | | | | Paved shoulder - driver-side = PSds2
| | | | | | Road condition = RC1: SR5 (10.0/1.0)
| | | | | | Road condition = RC2: SR2 (2.0)
| | | | | | Road condition = RC3: SR5 (0.0)
| | | | | Paved shoulder - driver-side = PSds1: SR3 (1.0)
| | | | Quality of curve = QoC1: SR4 (0.0)
| | | | Quality of curve = QoC3
| | | | | Road condition = RC1
| | | | | | Roadside severity - passenger-side distance = RSpsd2: SR4 (19.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd3: SR5 (7.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd4: SR5 (1.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd1: SR4 (5.0)
| | | | | Road condition = RC2
| | | | | | Skid resistance / grip = SkRes1: SR1 (3.0/1.0)
| | | | | | Skid resistance / grip = SkRes2: SR3 (2.0)
| | | | | | Skid resistance / grip = SkRes5: SR3 (0.0)
| | | | | | Skid resistance / grip = SkRes3: SR3 (0.0)
| | | | | | Skid resistance / grip = SkRes4: SR3 (0.0)
| | | | | Road condition = RC3: SR4 (3.0)
| | | Roadside severity - driver-side distance = RSdsd1
| | | | Roadside severity - passenger-side distance = RSpsd2: SR3 (12.0/2.0)
| | | | Roadside severity - passenger-side distance = RSpsd3: SR3 (1.0)
| | | | Roadside severity - passenger-side distance = RSpsd4: SR4 (3.0)
| | | | Roadside severity - passenger-side distance = RSpsd1: SR3 (5.0)
| | Curvature = Cu3: SR1 (3.0/1.0)
| | Curvature = Cu2
| | | Quality of curve = QoC2: SR3 (19.0/1.0)
| | | Quality of curve = QoC1
| | | | Roadside severity - driver-side distance = RSdsd2
| | | | | Paved shoulder - driver-side = PSds4: SR3 (7.0/4.0)
| | | | | Paved shoulder - driver-side = PSds3: SR1 (5.0)
| | | | | Paved shoulder - driver-side = PSds2: SR1 (0.0)
| | | | | Paved shoulder - driver-side = PSds1: SR1 (0.0)
| | | | Roadside severity - driver-side distance = RSdsd3: SR1 (1.0)
| | | | Roadside severity - driver-side distance = RSdsd4: SR2 (3.0)
| | | | Roadside severity - driver-side distance = RSdsd1: SR1 (0.0)
| | | Quality of curve = QoC3: SR3 (0.0)
| | Curvature = Cu4: SR1 (1.0)
| Median type = MedType4: SR1 (0.0)
| Median type = MedType6

```
|   |   Quality of curve = QoC2
|   |   |   Road condition = RC1
|   |   |   |   Number of lanes = NOL1: SR3 (14.0)
|   |   |   |   Number of lanes = NOL2: SR3 (177.0/47.0)
|   |   |   |   Number of lanes = NOL3: SR2 (7.0)
|   |   |   Road condition = RC2
|   |   |   |   Curvature = Cu1: SR4 (25.0/6.0)
|   |   |   |   Curvature = Cu3: SR4 (0.0)
|   |   |   |   Curvature = Cu2
|   |   |   |   |   Paved shoulder - driver-side = PSds4: SR1 (4.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds3: SR3 (5.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds2: SR3 (0.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds1: SR3 (0.0)
|   |   |   |   Curvature = Cu4: SR4 (0.0)
|   |   |   Road condition = RC3: SR3 (1.0)
|   |   Quality of curve = QoC1
|   |   |   Roadside severity - driver-side distance = RSdsd2
|   |   |   |   Paved shoulder - driver-side = PSds4
|   |   |   |   |   Number of lanes = NOL1: SR1 (0.0)
|   |   |   |   |   Number of lanes = NOL2: SR2 (3.0/1.0)
|   |   |   |   |   Number of lanes = NOL3: SR1 (3.0)
|   |   |   |   Paved shoulder - driver-side = PSds3: SR2 (3.0)
|   |   |   |   Paved shoulder - driver-side = PSds2
|   |   |   |   |   Road condition = RC1: SR1 (7.0/4.0)
|   |   |   |   |   Road condition = RC2: SR4 (4.0/2.0)
|   |   |   |   |   Road condition = RC3: SR1 (0.0)
|   |   |   |   Paved shoulder - driver-side = PSds1: SR1 (0.0)
|   |   |   Roadside severity - driver-side distance = RSdsd3: SR3 (2.0)
|   |   |   Roadside severity - driver-side distance = RSdsd4
|   |   |   |   Paved shoulder - driver-side = PSds4: SR3 (0.0)
|   |   |   |   Paved shoulder - driver-side = PSds3: SR3 (2.0)
|   |   |   |   Paved shoulder - driver-side = PSds2: SR3 (8.0)
|   |   |   |   Paved shoulder - driver-side = PSds1: SR5 (3.0)
|   |   |   Roadside severity - driver-side distance = RSdsd1: SR3 (0.0)
|   |   Quality of curve = QoC3
|   |   |   Road condition = RC1
|   |   |   |   Roadside severity - driver-side distance = RSdsd2
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR5 (1.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR4 (0.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR4 (3.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR3 (2.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd3
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR4 (5.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR4 (2.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR5 (13.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR4 (1.0)
|   |   |   |   Roadside severity - driver-side distance = RSdsd4
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   |   Paved shoulder - passenger-side = PSps4
|   |   |   |   |   |   |   Number of lanes = NOL1: SR4 (2.0)
|   |   |   |   |   |   |   Number of lanes = NOL2: SR2 (5.0/2.0)
```

| | | | | | | Number of lanes = NOL3: SR2 (0.0)
| | | | | | Paved shoulder - passenger-side = PSps3: SR4 (3.0)
| | | | | | Paved shoulder - passenger-side = PSps2: SR2 (2.0/1.0)
| | | | | | Paved shoulder - passenger-side = PSps1: SR4 (0.0)
| | | | | Roadside severity - passenger-side distance = RSpsd3: SR5 (9.0/1.0)
| | | | | Roadside severity - passenger-side distance = RSpsd4
| | | | | | Paved shoulder - passenger-side = PSps4: SR3 (6.0/1.0)
| | | | | | Paved shoulder - passenger-side = PSps3: SR5 (2.0)
| | | | | | Paved shoulder - passenger-side = PSps2: SR5 (7.0)
| | | | | | Paved shoulder - passenger-side = PSps1: SR5 (0.0)
| | | | | Roadside severity - passenger-side distance = RSpsd1: SR2 (1.0)
| | | | Roadside severity - driver-side distance = RSdsd1: SR4 (7.0)
| | | Road condition = RC2
| | | | Roadside severity - driver-side distance = RSdsd2
| | | | | Roadside severity - passenger-side distance = RSpsd2
| | | | | | Number of lanes = NOL1: SR3 (2.0)
| | | | | | Number of lanes = NOL2: SR1 (3.0)
| | | | | | Number of lanes = NOL3: SR1 (0.0)
| | | | | Roadside severity - passenger-side distance = RSpsd3: SR2 (4.0)
| | | | | Roadside severity - passenger-side distance = RSpsd4: SR2 (0.0)
| | | | | Roadside severity - passenger-side distance = RSpsd1: SR3 (1.0)
| | | | Roadside severity - driver-side distance = RSdsd3
| | | | | Roadside severity - passenger-side distance = RSpsd2: SR1 (9.0/3.0)
| | | | | Roadside severity - passenger-side distance = RSpsd3: SR4 (3.0)
| | | | | Roadside severity - passenger-side distance = RSpsd4: SR5 (3.0)
| | | | | Roadside severity - passenger-side distance = RSpsd1: SR4 (0.0)
| | | | Roadside severity - driver-side distance = RSdsd4
| | | | | Paved shoulder - passenger-side = PSps4
| | | | | | Roadside severity - passenger-side distance = RSpsd2: SR2 (5.0/1.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd3: SR3 (3.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd4: SR3 (5.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
| | | | | Paved shoulder - passenger-side = PSps3: SR4 (12.0/5.0)
| | | | | Paved shoulder - passenger-side = PSps2
| | | | | | Roadside severity - passenger-side distance = RSpsd2: SR4 (7.0/2.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd3: SR5 (9.0/3.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd4: SR2 (3.0/1.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd1: SR5 (0.0)
| | | | | Paved shoulder - passenger-side = PSps1: SR4 (0.0)
| | | | Roadside severity - driver-side distance = RSdsd1
| | | | | Roadside severity - passenger-side distance = RSpsd2: SR3 (4.0)
| | | | | Roadside severity - passenger-side distance = RSpsd3: SR4 (8.0)
| | | | | Roadside severity - passenger-side distance = RSpsd4: SR4 (2.0)
| | | | | Roadside severity - passenger-side distance = RSpsd1: SR3 (2.0)
| | | Road condition = RC3: SR3 (9.0/1.0)
| Median type = MedType5
| | Roadside severity - driver-side distance = RSdsd2: SR2 (2.0/1.0)
| | Roadside severity - driver-side distance = RSdsd3
| | | Curvature = Cu1: SR5 (6.0)
| | | Curvature = Cu3: SR3 (1.0)
| | | Curvature = Cu2: SR4 (2.0)

| | | Curvature = Cu4: SR5 (0.0)
| | Roadside severity - driver-side distance = RSdsd4
| | | Roadside severity - passenger-side distance = RSpsd2
| | | | Paved shoulder - passenger-side = PSps4: SR2 (3.0)
| | | | Paved shoulder - passenger-side = PSps3: SR3 (6.0/2.0)
| | | | Paved shoulder - passenger-side = PSps2: SR4 (2.0)
| | | | Paved shoulder - passenger-side = PSps1: SR2 (0.0)
| | | Roadside severity - passenger-side distance = RSpsd3: SR3 (4.0/1.0)
| | | Roadside severity - passenger-side distance = RSpsd4: SR4 (6.0/1.0)
| | | Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
| | Roadside severity - driver-side distance = RSdsd1: SR3 (1.0)
| Median type = MedType1: SR2 (4.0/1.0)
| Median type = MedType10: SR1 (0.0)
| Median type = MedType14
| | Roadside severity - passenger-side distance = RSpsd2
| | | Roadside severity - driver-side distance = RSdsd2: SR3 (17.0/4.0)
| | | Roadside severity - driver-side distance = RSdsd3: SR4 (2.0)
| | | Roadside severity - driver-side distance = RSdsd4: SR3 (0.0)
| | | Roadside severity - driver-side distance = RSdsd1: SR3 (2.0)
| | Roadside severity - passenger-side distance = RSpsd3: SR4 (6.0/1.0)
| | Roadside severity - passenger-side distance = RSpsd4: SR4 (4.0/1.0)
| | Roadside severity - passenger-side distance = RSpsd1: SR3 (5.0/1.0)
| Median type = MedType2: SR5 (2.0)
Delineation = De1
| Curvature = Cu1
| | Road condition = RC1
| | | | Roadside severity - passenger-side distance = RSpsd2
| | | | | Median type = MedType11
| | | | | | Grade = Gr1: SR3 (1943.0/109.0)
| | | | | | Grade = Gr3: SR3 (35.0)
| | | | | | Grade = Gr5: SR2 (8.0)
| | | | | | Grade = Gr3: SR3 (0.0)
| | | | | Median type = MedType3: SR3 (0.0)
| | | | | Median type = MedType7
| | | | | | Quality of curve = QoC2
| | | | | | | Paved shoulder - driver-side = PSds4: SR2 (11.0/4.0)
| | | | | | | Paved shoulder - driver-side = PSds3: SR2 (0.0)
| | | | | | | Paved shoulder - driver-side = PSds2: SR3 (5.0)
| | | | | | | Paved shoulder - driver-side = PSds1: SR2 (0.0)
| | | | | | Quality of curve = QoC1: SR2 (0.0)
| | | | | | Quality of curve = QoC3: SR5 (4.0)
| | | | | Median type = MedType4: SR3 (2.0)
| | | | | Median type = MedType6
| | | | | | Roadside severity - driver-side distance = RSdsd2
| | | | | | | Paved shoulder - driver-side = PSds4
| | | | | | | | Lane width = LW1: SR4 (19.0/8.0)
| | | | | | | | Lane width = LW2: SR3 (2.0)
| | | | | | | | Lane width = LW3: SR4 (0.0)
| | | | | | | Paved shoulder - driver-side = PSds3: SR2 (3.0)
| | | | | | | Paved shoulder - driver-side = PSds2: SR3 (47.0/26.0)
| | | | | | | Paved shoulder - driver-side = PSds1: SR3 (0.0)

```
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd3: SR3 (0.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd4: SR3 (167.0/7.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR3 (0.0)
|   |   |   |   |   Median type = MedType5
|   |   |   |   |   |   Paved shoulder - driver-side = PSds4: SR4 (0.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds3: SR3 (2.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds2: SR4 (10.0/1.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds1: SR4 (0.0)
|   |   |   |   |   Median type = MedType1: SR4 (1.0)
|   |   |   |   |   Median type = MedType10: SR3 (3.0)
|   |   |   |   |   Median type = MedType14: SR3 (0.0)
|   |   |   |   |   Median type = MedType2: SR3 (0.0)
|   |   |   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd2: SR3 (107.0/3.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd3
|   |   |   |   |   |   Quality of curve = QoC2: SR3 (660.0)
|   |   |   |   |   |   Quality of curve = QoC1: SR3 (0.0)
|   |   |   |   |   |   Quality of curve = QoC3: SR4 (6.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd4
|   |   |   |   |   |   Number of lanes = NOL1: SR4 (54.0/7.0)
|   |   |   |   |   |   Number of lanes = NOL2: SR3 (39.0)
|   |   |   |   |   |   Number of lanes = NOL3: SR4 (0.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR3 (0.0)
|   |   |   |   Roadside severity - passenger-side distance = RSpsd4
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd2: SR3 (87.0/8.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd3
|   |   |   |   |   |   Paved shoulder - driver-side = PSds4: SR4 (0.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds3: SR3 (5.0/1.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds2: SR4 (44.0/1.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds1: SR4 (0.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd4: SR4 (429.0/170.0)
|   |   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR4 (0.0)
|   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
|   |   |   |   Number of lanes = NOL1
|   |   |   |   |   Paved shoulder - driver-side = PSds4
|   |   |   |   |   |   Differential speed limits = DefSpeed1: SR2 (7.0/1.0)
|   |   |   |   |   |   Differential speed limits = DefSpeed2: SR3 (17.0/2.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds3
|   |   |   |   |   |   Skid resistance / grip = SkRes1
|   |   |   |   |   |   |   Paved shoulder - passenger-side = PSps4: SR2 (0.0)
|   |   |   |   |   |   |   Paved shoulder - passenger-side = PSps3
|   |   |   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd2
|   |   |   |   |   |   |   |   |   Grade = Gr1
|   |   |   |   |   |   |   |   |   |   Lane width = LW1
|   |   |   |   |   |   |   |   |   |   |   Median type = MedType11: SR2 (195.0/40.0)
|   |   |   |   |   |   |   |   |   |   |   Median type = MedType3: SR2 (0.0)
|   |   |   |   |   |   |   |   |   |   |   Median type = MedType7: SR2 (0.0)
|   |   |   |   |   |   |   |   |   |   |   Median type = MedType4: SR2 (0.0)
|   |   |   |   |   |   |   |   |   |   |   Median type = MedType6: SR2 (0.0)
|   |   |   |   |   |   |   |   |   |   |   Median type = MedType5: SR2 (0.0)
|   |   |   |   |   |   |   |   |   |   |   Median type = MedType1: SR2 (0.0)
```

| | | | | | | | | | | Median type = MedType10: SR4 (2.0)
| | | | | | | | | | | Median type = MedType14: SR2 (0.0)
| | | | | | | | | | | Median type = MedType2: SR2 (0.0)
| | | | | | | | | | Lane width = LW2: SR4 (3.0/1.0)
| | | | | | | | | | Lane width = LW3: SR2 (0.0)
| | | | | | | | | Grade = Gr3: SR2 (0.0)
| | | | | | | | | Grade = Gr5: SR2 (0.0)
| | | | | | | | | Grade = Gr3: SR1 (5.0)
| | | | | | | | Roadside severity - driver-side distance = RSdsd3
| | | | | | | | | Roadside severity - passenger-side distance = RSpsd2: SR2 (18.0)
| | | | | | | | | Roadside severity - passenger-side distance = RSpsd3: SR2 (11.0/5.0)
| | | | | | | | | Roadside severity - passenger-side distance = RSpsd4: SR3 (8.0)
| | | | | | | | | Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
| | | | | | | | Roadside severity - driver-side distance = RSdsd4
| | | | | | | | | Roadside severity - passenger-side distance = RSpsd2: SR2 (19.0/8.0)
| | | | | | | | | Roadside severity - passenger-side distance = RSpsd3: SR3 (6.0/1.0)
| | | | | | | | | Roadside severity - passenger-side distance = RSpsd4: SR3 (9.0)
| | | | | | | | | Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
| | | | | | | | Roadside severity - driver-side distance = RSdsd1: SR1 (25.0/8.0)
| | | | | | | Paved shoulder - passenger-side = PSps2
| | | | | | | | Roadside severity - driver-side distance = RSdsd2: SR1 (44.0/17.0)
| | | | | | | | Roadside severity - driver-side distance = RSdsd3: SR1 (0.0)
| | | | | | | | Roadside severity - driver-side distance = RSdsd4: SR1 (0.0)
| | | | | | | | Roadside severity - driver-side distance = RSdsd1: SR2 (5.0/2.0)
| | | | | | | Paved shoulder - passenger-side = PSps1: SR2 (0.0)
| | | | | | Skid resistance / grip = SkRes2
| | | | | | | Differential speed limits = DefSpeed1
| | | | | | | | Roadside severity - driver-side distance = RSdsd2
| | | | | | | | | Roadside severity - passenger-side distance = RSpsd2: SR1 (49.0/7.0)
| | | | | | | | | Roadside severity - passenger-side distance = RSpsd3: SR2 (4.0)
| | | | | | | | | Roadside severity - passenger-side distance = RSpsd4: SR2 (1.0)
| | | | | | | | | Roadside severity - passenger-side distance = RSpsd1: SR1 (0.0)
| | | | | | | | Roadside severity - driver-side distance = RSdsd3: SR2 (3.0)
| | | | | | | | Roadside severity - driver-side distance = RSdsd4: SR2 (3.0)
| | | | | | | | Roadside severity - driver-side distance = RSdsd1: SR1 (0.0)
| | | | | | | Differential speed limits = DefSpeed2: SR5 (3.0)
| | | | | | Skid resistance / grip = SkRes5: SR2 (0.0)
| | | | | | Skid resistance / grip = SkRes3: SR1 (1.0)
| | | | | | Skid resistance / grip = SkRes4: SR2 (0.0)
| | | | | Paved shoulder - driver-side = PSds2
| | | | | | Roadside severity - driver-side distance = RSdsd2
| | | | | | | Skid resistance / grip = SkRes1
| | | | | | | | Paved shoulder - passenger-side = PSps4: SR2 (0.0)
| | | | | | | | Paved shoulder - passenger-side = PSps3: SR3 (8.0/2.0)
| | | | | | | | Paved shoulder - passenger-side = PSps2: SR2 (737.0/398.0)
| | | | | | | | Paved shoulder - passenger-side = PSps1: SR2 (0.0)
| | | | | | | Skid resistance / grip = SkRes2
| | | | | | | | Roadside severity - passenger-side distance = RSpsd2: SR1 (47.0/27.0)
| | | | | | | | Roadside severity - passenger-side distance = RSpsd3: SR2 (16.0/8.0)
| | | | | | | | Roadside severity - passenger-side distance = RSpsd4: SR2 (2.0/1.0)
| | | | | | | | Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)

| | | | | | | Skid resistance / grip = SkRes5: SR2 (0.0)
| | | | | | | Skid resistance / grip = SkRes3: SR1 (4.0)
| | | | | | | Skid resistance / grip = SkRes4: SR2 (0.0)
| | | | | | Roadside severity - driver-side distance = RSdsd3: SR3 (343.0/154.0)
| | | | | | Roadside severity - driver-side distance = RSdsd4: SR3 (144.0/42.0)
| | | | | | Roadside severity - driver-side distance = RSdsd1: SR2 (16.0/6.0)
| | | | | Paved shoulder - driver-side = PSds1: SR5 (1.0)
| | | | Number of lanes = NOL2
| | | | | Roadside severity - driver-side distance = RSdsd2
| | | | | | Median type = MedType11: SR4 (1.0)
| | | | | | Median type = MedType3: SR4 (0.0)
| | | | | | Median type = MedType7
| | | | | | | Skid resistance / grip = SkRes1: SR4 (24.0/5.0)
| | | | | | | Skid resistance / grip = SkRes2: SR3 (4.0)
| | | | | | | Skid resistance / grip = SkRes5: SR4 (0.0)
| | | | | | | Skid resistance / grip = SkRes3: SR4 (0.0)
| | | | | | | Skid resistance / grip = SkRes4: SR4 (0.0)
| | | | | | Median type = MedType4: SR4 (0.0)
| | | | | | Median type = MedType6: SR3 (10.0/1.0)
| | | | | | Median type = MedType5: SR3 (11.0/2.0)
| | | | | | Median type = MedType1: SR4 (0.0)
| | | | | | Median type = MedType10: SR4 (0.0)
| | | | | | Median type = MedType14: SR4 (1.0)
| | | | | | Median type = MedType2: SR4 (0.0)
| | | | | Roadside severity - driver-side distance = RSdsd3: SR5 (2.0)
| | | | | Roadside severity - driver-side distance = RSdsd4
| | | | | | Roadside severity - passenger-side distance = RSpsd2
| | | | | | | Median type = MedType11: SR5 (0.0)
| | | | | | | Median type = MedType3: SR5 (0.0)
| | | | | | | Median type = MedType7
| | | | | | | | Paved shoulder - passenger-side = PSps4: SR4 (0.0)
| | | | | | | | Paved shoulder - passenger-side = PSps3: SR2 (38.0/21.0)
| | | | | | | | Paved shoulder - passenger-side = PSps2: SR4 (31.0/16.0)
| | | | | | | | Paved shoulder - passenger-side = PSps1: SR5 (1.0)
| | | | | | | Median type = MedType4: SR5 (0.0)
| | | | | | | Median type = MedType6
| | | | | | | | Paved shoulder - passenger-side = PSps4: SR2 (0.0)
| | | | | | | | Paved shoulder - passenger-side = PSps3: SR5 (40.0/25.0)
| | | | | | | | Paved shoulder - passenger-side = PSps2: SR2 (10.0)
| | | | | | | | Paved shoulder - passenger-side = PSps1: SR2 (0.0)
| | | | | | | Median type = MedType5: SR3 (27.0/9.0)
| | | | | | | Median type = MedType1: SR5 (0.0)
| | | | | | | Median type = MedType10: SR5 (0.0)
| | | | | | | Median type = MedType14: SR5 (0.0)
| | | | | | | Median type = MedType2: SR5 (0.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd3
| | | | | | | Median type = MedType11: SR3 (0.0)
| | | | | | | Median type = MedType3: SR3 (0.0)
| | | | | | | Median type = MedType7: SR5 (17.0)
| | | | | | | Median type = MedType4: SR3 (0.0)
| | | | | | | Median type = MedType6: SR3 (39.0/10.0)

| | | | | | | Median type = MedType5: SR4 (10.0/1.0)
| | | | | | | Median type = MedType1: SR3 (0.0)
| | | | | | | Median type = MedType10: SR3 (0.0)
| | | | | | | Median type = MedType14: SR3 (0.0)
| | | | | | | Median type = MedType2: SR3 (0.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd4: SR5 (33.0/6.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd1: SR2 (2.0/1.0)
| | | | | Roadside severity - driver-side distance = RSdsd1
| | | | | | Median type = MedType11: SR4 (2.0)
| | | | | | Median type = MedType3: SR4 (0.0)
| | | | | | Median type = MedType7: SR4 (17.0)
| | | | | | Median type = MedType4: SR4 (0.0)
| | | | | | Median type = MedType6: SR4 (14.0/3.0)
| | | | | | Median type = MedType5: SR3 (5.0/1.0)
| | | | | | Median type = MedType1: SR4 (0.0)
| | | | | | Median type = MedType10: SR4 (0.0)
| | | | | | Median type = MedType14: SR4 (0.0)
| | | | | | Median type = MedType2: SR4 (0.0)
| | | | Number of lanes = NOL3: SR3 (0.0)
| | Road condition = RC2
| | | Number of lanes = NOL1
| | | | Quality of curve = QoC2
| | | | | Roadside severity - driver-side distance = RSdsd2
| | | | | | Roadside severity - passenger-side distance = RSpsd2: SR2 (438.0/72.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd3: SR3 (8.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd4: SR3 (3.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
| | | | | Roadside severity - driver-side distance = RSdsd3: SR3 (43.0/5.0)
| | | | | Roadside severity - driver-side distance = RSdsd4
| | | | | | Roadside severity - passenger-side distance = RSpsd2: SR3 (3.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd3: SR3 (1.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd4: SR4 (8.0)
| | | | | | Roadside severity - passenger-side distance = RSpsd1: SR4 (0.0)
| | | | | Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
| | | | Quality of curve = QoC1: SR2 (0.0)
| | | | Quality of curve = QoC3
| | | | | Paved shoulder - driver-side = PSds4: SR3 (8.0/1.0)
| | | | | Paved shoulder - driver-side = PSds3
| | | | | | Lane width = LW1
| | | | | | | Roadside severity - driver-side distance = RSdsd2
| | | | | | | | Roadside severity - passenger-side distance = RSpsd2: SR1 (61.0/7.0)
| | | | | | | | Roadside severity - passenger-side distance = RSpsd3
| | | | | | | | | Skid resistance / grip = SkRes1: SR2 (6.0)
| | | | | | | | | Skid resistance / grip = SkRes2: SR2 (5.0/1.0)
| | | | | | | | | Skid resistance / grip = SkRes5: SR2 (0.0)
| | | | | | | | | Skid resistance / grip = SkRes3: SR1 (5.0)
| | | | | | | | | Skid resistance / grip = SkRes4: SR2 (0.0)
| | | | | | | | Roadside severity - passenger-side distance = RSpsd4: SR1 (3.0/1.0)
| | | | | | | | Roadside severity - passenger-side distance = RSpsd1: SR1 (0.0)
| | | | | | | Roadside severity - driver-side distance = RSdsd3
| | | | | | | | Roadside severity - passenger-side distance = RSpsd2: SR2 (4.0/1.0)

```
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR3 (3.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR2 (0.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
|   |   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd4
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR2 (4.0/1.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR2 (0.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR4 (2.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
|   |   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR4 (2.0/1.0)
|   |   |   |   |   |   Lane width = LW2: SR3 (6.0/1.0)
|   |   |   |   |   |   Lane width = LW3: SR1 (0.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds2
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd2
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes1: SR2 (114.0/55.0)
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes2: SR1 (7.0/3.0)
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes5: SR2 (0.0)
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes3: SR2 (3.0)
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes4: SR2 (0.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR2 (43.0/6.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes1: SR2 (18.0/1.0)
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes2: SR1 (4.0)
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes5: SR2 (0.0)
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes3: SR2 (0.0)
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes4: SR2 (0.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd3
|   |   |   |   |   |   |   Skid resistance / grip = SkRes1
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR2 (15.0/6.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR2 (17.0/5.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR3 (11.0/4.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
|   |   |   |   |   |   |   Skid resistance / grip = SkRes2: SR2 (11.0/2.0)
|   |   |   |   |   |   |   Skid resistance / grip = SkRes5: SR2 (0.0)
|   |   |   |   |   |   |   Skid resistance / grip = SkRes3
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR1 (5.0/1.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR3 (3.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR2 (1.0)
|   |   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR1 (0.0)
|   |   |   |   |   |   |   Skid resistance / grip = SkRes4: SR2 (0.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd4
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2: SR2 (8.0/1.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3: SR3 (10.0/3.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd4: SR3 (11.0/2.0)
|   |   |   |   |   |   |   Roadside severity - passenger-side distance = RSpsd1: SR3 (0.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
|   |   |   |   |   Paved shoulder - driver-side = PSds1: SR2 (0.0)
|   |   |   Number of lanes = NOL2
|   |   |   |   Quality of curve = QoC2
|   |   |   |   |   Median type = MedType11: SR3 (1.0)
```

```
|   |   |   |   |   Median type = MedType3: SR3 (0.0)
|   |   |   |   |   Median type = MedType7: SR2 (9.0/2.0)
|   |   |   |   |   Median type = MedType4: SR3 (0.0)
|   |   |   |   |   Median type = MedType6
|   |   |   |   |   |   Paved shoulder - driver-side = PSds4: SR3 (29.0/1.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds3: SR5 (5.0/2.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds2: SR3 (0.0)
|   |   |   |   |   |   Paved shoulder - driver-side = PSds1: SR3 (0.0)
|   |   |   |   |   Median type = MedType5: SR3 (0.0)
|   |   |   |   |   Median type = MedType1: SR3 (0.0)
|   |   |   |   |   Median type = MedType10: SR3 (0.0)
|   |   |   |   |   Median type = MedType14: SR3 (0.0)
|   |   |   |   |   Median type = MedType2: SR3 (0.0)
|   |   |   |   Quality of curve = QoC1: SR3 (0.0)
|   |   |   |   Quality of curve = QoC3
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd2
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd2
|   |   |   |   |   |   |   Median type = MedType11: SR5 (0.0)
|   |   |   |   |   |   |   Median type = MedType3: SR5 (0.0)
|   |   |   |   |   |   |   Median type = MedType7: SR3 (2.0/1.0)
|   |   |   |   |   |   |   Median type = MedType4: SR5 (0.0)
|   |   |   |   |   |   |   Median type = MedType6: SR5 (0.0)
|   |   |   |   |   |   |   Median type = MedType5: SR5 (0.0)
|   |   |   |   |   |   |   Median type = MedType1: SR5 (2.0)
|   |   |   |   |   |   |   Median type = MedType10: SR5 (0.0)
|   |   |   |   |   |   |   Median type = MedType14: SR5 (0.0)
|   |   |   |   |   |   |   Median type = MedType2: SR5 (0.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd3: SR4 (1.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd4
|   |   |   |   |   |   |   Median type = MedType11: SR2 (0.0)
|   |   |   |   |   |   |   Median type = MedType3: SR2 (0.0)
|   |   |   |   |   |   |   Median type = MedType7: SR4 (8.0/3.0)
|   |   |   |   |   |   |   Median type = MedType4: SR2 (0.0)
|   |   |   |   |   |   |   Median type = MedType6
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes1: SR2 (8.0/2.0)
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes2: SR4 (2.0)
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes5: SR2 (0.0)
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes3: SR2 (0.0)
|   |   |   |   |   |   |   |   Skid resistance / grip = SkRes4: SR2 (0.0)
|   |   |   |   |   |   |   Median type = MedType5: SR3 (3.0/1.0)
|   |   |   |   |   |   |   Median type = MedType1: SR2 (0.0)
|   |   |   |   |   |   |   Median type = MedType10: SR2 (0.0)
|   |   |   |   |   |   |   Median type = MedType14: SR2 (0.0)
|   |   |   |   |   |   |   Median type = MedType2: SR2 (0.0)
|   |   |   |   |   |   Roadside severity - driver-side distance = RSdsd1: SR3 (3.0/1.0)
|   |   |   |   |   Roadside severity - passenger-side distance = RSpsd3
|   |   |   |   |   |   Median type = MedType11: SR2 (0.0)
|   |   |   |   |   |   Median type = MedType3: SR2 (0.0)
|   |   |   |   |   |   Median type = MedType7: SR4 (2.0)
|   |   |   |   |   |   Median type = MedType4: SR2 (0.0)
|   |   |   |   |   |   Median type = MedType6: SR2 (5.0/2.0)
```

| | | | | | Median type = MedType5: SR2 (0.0)
| | | | | | Median type = MedType1: SR2 (0.0)
| | | | | | Median type = MedType10: SR2 (0.0)
| | | | | | Median type = MedType14: SR2 (0.0)
| | | | | | Median type = MedType2: SR2 (0.0)
| | | | | Roadside severity - passenger-side distance = RSpsd4: SR5 (7.0)
| | | | | Roadside severity - passenger-side distance = RSpsd1: SR4 (1.0)
| | | Number of lanes = NOL3: SR2 (0.0)
| | Road condition = RC3
| | | Median type = MedType11
| | | | Roadside severity - driver-side distance = RSdsd2: SR2 (86.0/17.0)
| | | | Roadside severity - driver-side distance = RSdsd3: SR2 (3.0/1.0)
| | | | Roadside severity - driver-side distance = RSdsd4: SR4 (3.0)
| | | | Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
| | | Median type = MedType3: SR2 (0.0)
| | | Median type = MedType7
| | | | Roadside severity - driver-side distance = RSdsd2: SR1 (3.0/1.0)
| | | | Roadside severity - driver-side distance = RSdsd3: SR2 (0.0)
| | | | Roadside severity - driver-side distance = RSdsd4: SR2 (2.0)
| | | | Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
| | | Median type = MedType4: SR2 (0.0)
| | | Median type = MedType6: SR3 (2.0/1.0)
| | | Median type = MedType5: SR2 (0.0)
| | | Median type = MedType1: SR2 (0.0)
| | | Median type = MedType10: SR2 (0.0)
| | | Median type = MedType14: SR2 (0.0)
| | | Median type = MedType2: SR2 (0.0)
| Curvature = Cu3
| | Roadside severity - driver-side distance = RSdsd2: SR1 (160.0/15.0)
| | Roadside severity - driver-side distance = RSdsd3
| | | Paved shoulder - driver-side = PSds4: SR1 (4.0)
| | | Paved shoulder - driver-side = PSds2
| | | | Roadside severity - passenger-side distance = RSpsd2: SR1 (10.0/5.0)
| | | | Roadside severity - passenger-side distance = RSpsd3: SR2 (9.0)
| | | | Roadside severity - passenger-side distance = RSpsd4: SR1 (1.0)
| | | | Roadside severity - passenger-side distance = RSpsd1: SR2 (0.0)
| | | Paved shoulder - driver-side = PSds1: SR2 (0.0)
| | Roadside severity - driver-side distance = RSdsd4
| | | Road condition = RC1: SR1 (11.0/3.0)
| | | Road condition = RC2: SR3 (3.0)
| | | Road condition = RC3: SR1 (0.0)
| | Roadside severity - driver-side distance = RSdsd1: SR2 (3.0/1.0)
| Curvature = Cu2
| | Roadside severity - passenger-side distance = RSpsd2
| | | | Road condition = RC1
| | | | | Grade = Gr1
| | | | | | Roadside severity - driver-side distance = RSdsd2: SR2 (302.0/28.0)
| | | | | | Roadside severity - driver-side distance = RSdsd3: SR3 (22.0/4.0)
| | | | | | Roadside severity - driver-side distance = RSdsd4
| | | | | | | Median type = MedType11: SR2 (6.0/2.0)
| | | | | | | Median type = MedType3: SR2 (0.0)

| | | | | | | Median type = MedType7: SR2 (0.0)
| | | | | | | Median type = MedType4: SR2 (0.0)
| | | | | | | Median type = MedType6: SR2 (31.0/2.0)
| | | | | | | Median type = MedType5: SR3 (6.0/1.0)
| | | | | | | Median type = MedType1: SR2 (0.0)
| | | | | | | Median type = MedType10: SR2 (1.0)
| | | | | | | Median type = MedType14: SR2 (0.0)
| | | | | | | Median type = MedType2: SR2 (0.0)
| | | | | | Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
| | | | | Grade = Gr3: SR2 (12.0/1.0)
| | | | | Grade = Gr5: SR1 (7.0)
| | | | | Grade = Gr3: SR2 (0.0)
| | | | Road condition = RC2
| | | | | Roadside severity - driver-side distance = RSdsd2
| | | | | | Quality of curve = QoC2: SR1 (87.0/26.0)
| | | | | | Quality of curve = QoC1: SR2 (2.0/1.0)
| | | | | | Quality of curve = QoC3: SR1 (0.0)
| | | | | Roadside severity - driver-side distance = RSdsd3: SR2 (5.0/2.0)
| | | | | Roadside severity - driver-side distance = RSdsd4: SR2 (2.0)
| | | | | Roadside severity - driver-side distance = RSdsd1: SR1 (0.0)
| | | | Road condition = RC3: SR1 (9.0/2.0)
| | | | Median type = MedType11: SR1 (219.0/88.0)
| | | | Median type = MedType3: SR1 (0.0)
| | | | Median type = MedType7
| | | | | Skid resistance / grip = SkRes1
| | | | | | Lane width = LW1: SR3 (9.0/3.0)
| | | | | | Lane width = LW2: SR2 (2.0)
| | | | | | Lane width = LW3: SR3 (0.0)
| | | | | Skid resistance / grip = SkRes2: SR2 (3.0)
| | | | | Skid resistance / grip = SkRes5: SR2 (0.0)
| | | | | Skid resistance / grip = SkRes3: SR2 (0.0)
| | | | | Skid resistance / grip = SkRes4: SR2 (0.0)
| | | | Median type = MedType4: SR1 (0.0)
| | | | Median type = MedType6: SR3 (9.0/2.0)
| | | | Median type = MedType5
| | | | | Quality of curve = QoC2: SR4 (2.0)
| | | | | Quality of curve = QoC1: SR2 (3.0/1.0)
| | | | | Quality of curve = QoC3: SR2 (0.0)
| | | | Median type = MedType1: SR5 (2.0)
| | | | Median type = MedType10: SR1 (0.0)
| | | | Median type = MedType14: SR1 (0.0)
| | | | Median type = MedType2: SR1 (0.0)
| | Roadside severity - passenger-side distance = RSpsd3
| | | | Road condition = RC1: SR3 (89.0/13.0)
| | | | Road condition = RC2: SR2 (7.0/1.0)
| | | | Road condition = RC3: SR2 (2.0/1.0)
| | | | Number of lanes = NOL1
| | | | | Road condition = RC1
| | | | | | Roadside severity - driver-side distance = RSdsd2: SR3 (18.0/11.0)
| | | | | | Roadside severity - driver-side distance = RSdsd3
| | | | | | | Quality of curve = QoC2: SR1 (17.0/8.0)

| | | | | | | Quality of curve = QoC1: SR2 (6.0/1.0)
| | | | | | | Quality of curve = QoC3: SR1 (0.0)
| | | | | | Roadside severity - driver-side distance = RSdsd4: SR2 (9.0/3.0)
| | | | | | Roadside severity - driver-side distance = RSdsd1: SR2 (0.0)
| | | | | Road condition = RC2: SR1 (12.0/1.0)
| | | | | Road condition = RC3: SR1 (0.0)
| | | | Number of lanes = NOL2
| | | | | Quality of curve = QoC2: SR3 (8.0)
| | | | | Quality of curve = QoC1: SR4 (5.0)
| | | | | Quality of curve = QoC3: SR3 (0.0)
| | | | Number of lanes = NOL3: SR1 (0.0)
| | Roadside severity - passenger-side distance = RSpsd4
| | | | Roadside severity - driver-side distance = RSdsd2
| | | | | Median type = MedType11: SR3 (3.0)
| | | | | Median type = MedType3: SR4 (0.0)
| | | | | Median type = MedType7: SR4 (0.0)
| | | | | Median type = MedType4: SR4 (0.0)
| | | | | Median type = MedType6: SR4 (8.0/1.0)
| | | | | Median type = MedType5: SR4 (0.0)
| | | | | Median type = MedType1: SR4 (0.0)
| | | | | Median type = MedType10: SR4 (0.0)
| | | | | Median type = MedType14: SR4 (0.0)
| | | | | Median type = MedType2: SR4 (0.0)
| | | | Roadside severity - driver-side distance = RSdsd3: SR3 (5.0)
| | | | Roadside severity - driver-side distance = RSdsd4
| | | | | Paved shoulder - driver-side = PSds4: SR3 (2.0)
| | | | | Paved shoulder - driver-side = PSds3: SR4 (6.0/1.0)
| | | | | Paved shoulder - driver-side = PSds2: SR3 (31.0/2.0)
| | | | | Paved shoulder - driver-side = PSds1: SR3 (0.0)
| | | | Roadside severity - driver-side distance = RSdsd1: SR3 (0.0)
| | | | Number of lanes = NOL1
| | | | | Skid resistance / grip = SkRes1
| | | | | | Road condition = RC1: SR2 (25.0/13.0)
| | | | | | Road condition = RC2
| | | | | | | Roadside severity - driver-side distance = RSdsd2: SR1 (4.0)
| | | | | | | Roadside severity - driver-side distance = RSdsd3: SR1 (3.0)
| | | | | | | Roadside severity - driver-side distance = RSdsd4: SR2 (5.0/1.0)
| | | | | | | Roadside severity - driver-side distance = RSdsd1: SR1 (0.0)
| | | | | | Road condition = RC3: SR2 (0.0)
| | | | | Skid resistance / grip = SkRes2: SR1 (4.0)
| | | | | Skid resistance / grip = SkRes5: SR2 (0.0)
| | | | | Skid resistance / grip = SkRes3: SR2 (0.0)
| | | | | Skid resistance / grip = SkRes4: SR2 (0.0)
| | | | Number of lanes = NOL2
| | | | | Paved shoulder - passenger-side = PSps4: SR4 (0.0)
| | | | | Paved shoulder - passenger-side = PSps3: SR4 (3.0)
| | | | | Paved shoulder - passenger-side = PSps2: SR5 (3.0)
| | | | | Paved shoulder - passenger-side = PSps1: SR4 (0.0)
| | | | Number of lanes = NOL3: SR2 (0.0)
| | Roadside severity - passenger-side distance = RSpsd1: SR1 (8.0/3.0)
| Curvature = Cu4: SR1 (4.0/1.0)

Number of Leaves : 830

Size of the tree : 1104

Time taken to build the model: 0.32 seconds

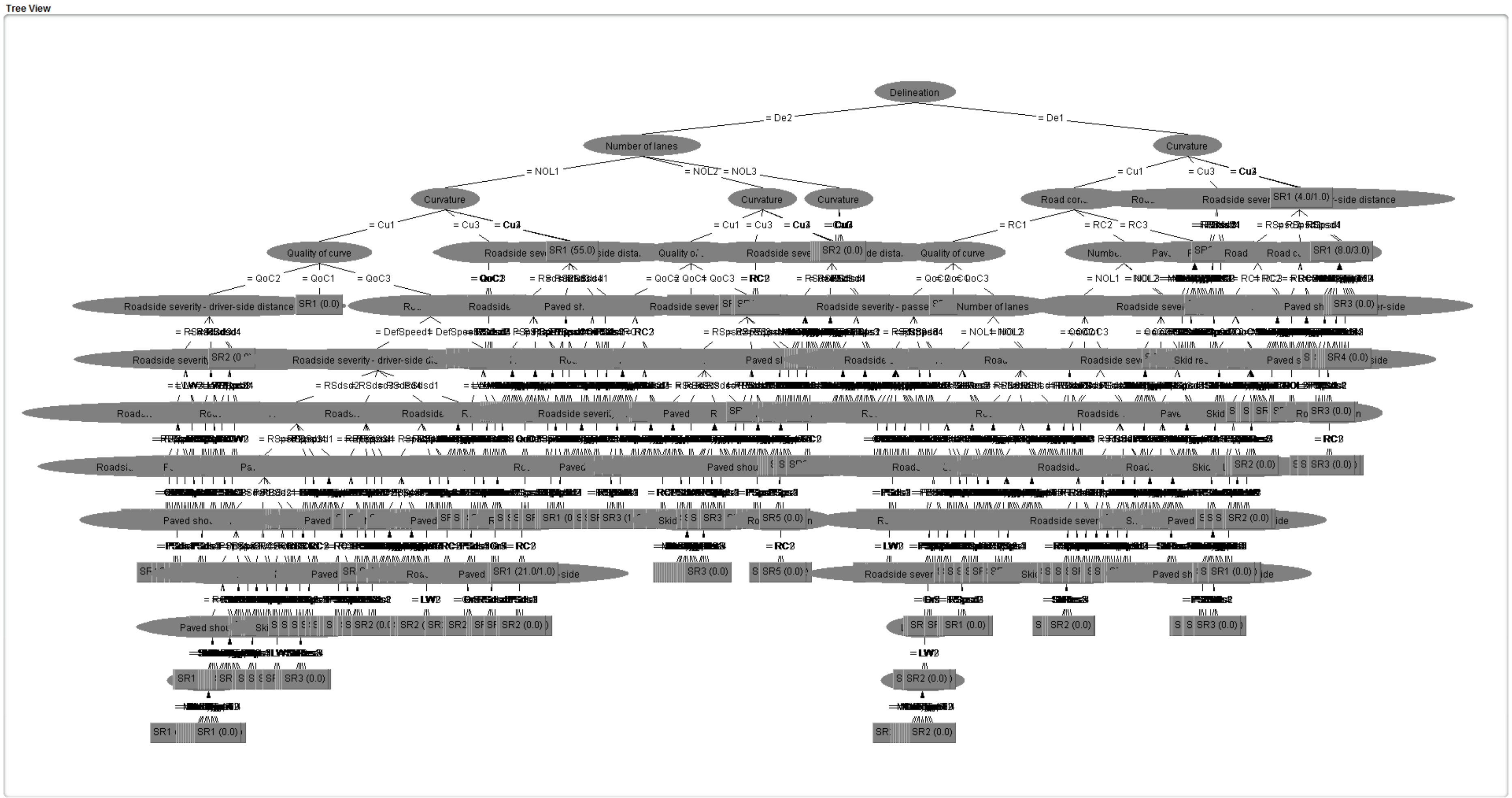
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	21701	77.8148 %
Incorrectly Classified Instances	6187	22.1852 %
Kappa statistic	0.6833	
Mean absolute error	0.1326	
Root mean squared error	0.2614	
Relative absolute error	46.4956 %	
Root relative squared error	69.2304 %	
Total Number of Instances	27888	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.811	0.114	0.738	0.811	0.773	0.678	0.905	0.797	SR2
	0.683	0.044	0.868	0.683	0.765	0.691	0.885	0.832	SR3
	0.330	0.012	0.676	0.330	0.444	0.448	0.861	0.482	SR4
	0.941	0.144	0.768	0.941	0.846	0.765	0.934	0.831	SR1
	0.397	0.003	0.667	0.397	0.497	0.509	0.938	0.528	SR5
Weighted Avg.	0.778	0.095	0.782	0.778	0.769	0.693	0.906	0.793	

Figure B2. The decision tree for (run-off + head-on) accidents



Appendix C
Transformation ranges for the iRAP countermeasures

This appendix presents the transformation labels for the iRAP countermeasures ranges that were used to transform the numerical data to nominal as the appropriate format for WEKA system. Then, these data were used as input data to develop the knowledge discovery models for run-off and (head-on + run-off) accidents.

Table C1. Transformation ranges for the iRAP run-on related crashes countermeasures

Roadside severity - passenger side distance label	**Description**	**iRAP Risk Factor**
RSpsd1	0 to <1m	1.0
RSpsd2	1 to <5m	0.8
RSpsd3	5 to <10m	0.35
RSpsd4	>=10m	0.1
Roadside severity - drivers side distance label	**Description**	**iRAP Risk Factor**
RSdsd1	0 to <1m	1.0
RSdsd2	1 to <5m	0.8
RSdsd3	5 to <10m	0.35
RSdsd4	>=10m	0.1
Paved shoulder - drivers side label	**Description**	**iRAP Risk Factor**
PSds1	None	0.77
PSds2	Narrow (≥ 0m to < 1.0m)	0.83
PSds3	Medium (≥ 1.0m to < 2.4m)	0.95
PSds4	Wide (≥ 2.4m)	1.0
Paved shoulder - passenger side label	**Description**	**iRAP Risk Factor**
PSps1	None	0.77
PSps2	Narrow (≥ 0m to < 1.0m)	0.83
PSps3	Medium (≥ 1.0m to < 2.4m)	0.95
PSps4	Wide (≥ 2.4m)	1.0
Lane width label	**Description**	**iRAP Risk Factor**
LW1	Narrow (≥ 0m to < 2.75m)	1.0
LW2	Medium (≥ 2.75m to < 3.25m)	1.2
LW3	Wide (≥ 3.25m)	1.5

Curvature label	Description	iRAP Risk Factor
Cu1	Straight or gently curving	1.0
Cu2	Moderate	1.8
Cu3	Sharp	3.5
Cu4	Very sharp	6.0
Quality of curve label	**Description**	**iRAP Risk Factor**
QoC1	Adequate	1.0
QoC2	Poor	1.25
Road condition label	**Description**	**iRAP Risk Factor**
RC1	Good	1.0
RC2	Medium	1.2
RC3	Poor	1.4
Delineation label	**Description**	**iRAP Risk Factor**
De1	Adequate	1.0
De2	Poor	1.2
Gradient label	**Description**	**iRAP Risk Factor**
Gr1	≥ 0% to <4%	1.0
Gr2	≥ 4% to <5%	1.2
Gr3	≥ 5% to <7.5%	1.7
Skid resistance label	**Description**	**iRAP Risk Factor**
SkRes1	Sealed - adequate	1.0
SkRes2	Sealed - medium	1.4
SkRes3	Sealed - poor	2.0
SkRes4	Unsealed - adequate	3.0
SkRes5	Unsealed - poor	5.5

Table C2. Transformation ranges for the iRAP run-on related crashes countermeasures

Roadside severity - passenger side distance label	Description	iRAP Risk Factor
RSpsd1	0 to <1m	1.0
RSpsd2	1 to <5m	0.8
RSpsd3	5 to <10m	0.35
RSpsd4	>=10m	0.1
Roadside severity - drivers side distance label	**Description**	**iRAP Risk Factor**
RSdsd1	0 to <1m	1.0
RSdsd2	1 to <5m	0.8
RSdsd3	5 to <10m	0.35
RSdsd4	>=10m	0.1
Paved shoulder - drivers side label	**Description**	**iRAP Risk Factor**
PSds1	None	0.77
PSds2	Narrow (≥ 0m to < 1.0m)	0.83
PSds3	Medium (≥ 1.0m to < 2.4m)	0.95
PSds4	Wide (≥ 2.4m)	1.0
Paved shoulder - passenger side label	**Description**	**iRAP Risk Factor**
PSps1	None	0.77
PSps2	Narrow (≥ 0m to < 1.0m)	0.83
PSps3	Medium (≥ 1.0m to < 2.4m)	0.95
PSps4	Wide (≥ 2.4m)	1.0
Lane width label	**Description**	**iRAP Risk Factor**
LW1	Narrow (≥ 0m to < 2.75m)	1.0
LW2	Medium (≥ 2.75m to < 3.25m)	1.2
LW3	Wide (≥ 3.25m)	1.5
Curvature label	**Description**	**iRAP Risk Factor**
Cu1	Straight or gently curving	1.0
Cu2	Moderate	1.8
Cu3	Sharp	3.5
Cu4	Very sharp	6.0
Quality of curve label	**Description**	**iRAP Risk Factor**
QoC1	Adequate	1.0
QoC2	Poor	1.25
Road condition label	**Description**	**iRAP Risk Factor**
RC1	Good	1.0
RC2	Medium	1.2
RC3	Poor	1.4

Delineation label	Description	iRAP Risk Factor
De1	Adequate	1.0
De2	Poor	1.2
Gradient label	**Description**	**iRAP Risk Factor**
Gr1	≥ 0% to <4%	1.0
Gr2	≥ 4% to <5%	1.2
Gr3	≥ 5% to <7.5%	1.7
Skid resistance label	**Description**	**iRAP Risk Factor**
SkRes1	Sealed - adequate	1.0
SkRes2	Sealed - medium	1.4
SkRes3	Sealed - poor	2.0
SkRes4	Unsealed - adequate	3.0
SkRes5	Unsealed - poor	5.5
Number of lanes label	**Description**	**iRAP Risk Factor**
NOL1	One lane in each direction	1.0
NOL2	Two lanes in each direction	0.02
NOL3	Three lanes in each direction	0.01
Median Type label	**Description**	**iRAP Risk Factor**
Traversable	Traversable median (only centre line)	1.00
Non-Traversable	Non-traversable median (safety barriers)	0.0
Differential speed label	**Description**	**iRAP Risk Factor**
Difspeed1	Speeds between cars and trucks exceed 20km/h	1.2
DifSpeed2	Speeds between cars and trucks do not exceed 20 km/h	1.0

Appendix D

The rules extracted for the head-on crashes related countermeasures

Differential speed limits	**Median type**	**Number of lanes**	**Gradient**	**Skid resistance / grip**	**Star rating**
DifSpeed2	Traversable	NOL1	Gr3	SkRes3	SR1
DifSpeed2	Traversable	NOL1	Gr1	SkRes5	SR1
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes1	SR5
DifSpeed2	Traversable	NOL1	Gr3	SkRes3	SR2
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes2	SR4
DifSpeed1	Traversable	NOL1	Gr1	SkRes1	SR3
DifSpeed1	Traversable	NOL1	Gr1	SkRes2	SR3
DifSpeed2	Traversable	NOL1	Gr3	SkRes2	SR2
DifSpeed1	Traversable	NOL1	Gr1	SkRes4	SR1
DifSpeed2	Traversable	NOL1	Gr1	SkRes2	SR3
DifSpeed1	Traversable	NOL1	Gr1	SkRes5	SR1
DifSpeed2	Non-Traversable	NOL2	Gr1	SkRes1	SR4
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes2	SR5
DifSpeed1	Traversable	NOL1	Gr2	SkRes1	SR3
DifSpeed2	Traversable	NOL1	Gr2	SkRes3	SR2
DifSpeed2	Traversable	NOL1	Gr3	SkRes2	SR2
DifSpeed2	Non-Traversable	NOL2	Gr1	SkRes1	SR5
DifSpeed2	Traversable	NOL1	Gr1	SkRes5	SR1
DifSpeed1	Non-Traversable	NOL2	Gr2	SkRes1	SR4
DifSpeed2	Traversable	NOL1	Gr3	SkRes3	SR2
DifSpeed1	Traversable	NOL1	Gr1	SkRes4	SR1
DifSpeed2	Traversable	NOL1	Gr2	SkRes3	SR2

Differential speed limits	Median type	Number of lanes	Gradient	Skid resistance / grip	Star rating
DifSpeed1	Non-Traversable	NOL2	Gr2	SkRes1	SR5
DifSpeed1	Traversable	NOL1	Gr2	SkRes2	SR3
DifSpeed2	Non-Traversable	NOL2	Gr2	SkRes1	SR5
DifSpeed1	Traversable	NOL1	Gr1	SkRes3	SR1
DifSpeed1	Traversable	NOL1	Gr1	SkRes1	SR4
DifSpeed2	Traversable	NOL2	Gr1	SkRes1	SR5

Appendix E

A sample of the rules extracted for the run-off cashes related countermeasures

Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
Gr1	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd3	RSdsd4	PSds1	PSps1	De1	SR2
Gr1	SkRes1	LW1	Cu1	QoC3	RC3	RSpsd4	RSdsd4	PSds2	PSps2	De2	SR2
Gr5	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd4	RSdsd4	PSds2	PSps2	De1	SR2
Gr1	SkRes2	LW2	Cu1	QoC1	RC2	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR3
Gr1	SkRes2	LW2	Cu3	QoC2	RC2	RSpsd2	RSdsd2	PSds2	PSps2	De2	SR3
Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps4	De1	SR4
Gr1	SkRes2	LW1	Cu3	QoC2	RC2	RSpsd2	RSdsd2	PSds2	PSps2	De2	SR3
Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd4	RSdsd4	PSds4	PSps4	De1	SR5
Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd1	RSdsd1	PSds3	PSps3	De1	SR2
Gr1	SkRes1	LW1	Cu2	QoC2	RC2	RSpsd4	RSdsd3	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu3	QoC2	RC1	RSpsd4	RSdsd4	PSds2	PSps2	De1	SR1
Gr1	SkRes1	LW1	Cu1	QoC3	RC3	RSpsd4	RSdsd4	PSds3	PSps3	De1	SR2

Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
Gr5	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd4	RSdsd4	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd4	RSdsd4	PSds2	PSps2	De2	SR2
Gr1	SkRes2	LW1	Cu1	QoC2	RC3	RSpsd2	RSdsd3	PSds3	PSps2	De2	SR3
Gr3	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd4	RSdsd4	PSds2	PSps2	De1	SR2
Gr1	SkRes3	LW1	Cu1	QoC2	RC3	RSpsd3	RSdsd3	PSds3	PSps3	De2	SR3
Gr1	SkRes1	LW3	Cu2	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps4	De1	SR4
Gr1	SkRes2	LW1	Cu1	QoC3	RC3	RSpsd2	RSdsd2	PSds1	PSps1	De2	SR1
Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd3	RSdsd4	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd4	RSdsd3	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW1	Cu2	QoC2	RC2	RSpsd2	RSdsd4	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd4	RSdsd4	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC2	RSpsd4	RSdsd3	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu3	QoC2	RC1	RSpsd3	RSdsd4	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd2	RSdsd4	PSds1	PSps1	De2	SR3
Gr1	SkRes1	LW3	Cu1	QoC2	RC2	RSpsd3	RSdsd2	PSds4	PSps4	De1	SR4

Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
Gr1	SkRes2	LW1	Cu1	QoC3	RC2	RSpsd2	RSdsd2	PSds1	PSps1	De2	SR2
Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd3	RSdsd4	PSds2	PSps2	De2	SR3
Gr1	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd3	RSdsd4	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd2	RSdsd4	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd2	RSdsd4	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC3	RSpsd4	RSdsd3	PSds2	PSps2	De2	SR2
Gr1	SkRes2	LW2	Cu1	QoC2	RC2	RSpsd2	RSdsd2	PSds3	PSps3	De1	SR4
Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW1	Cu1	QoC2	RC2	RSpsd2	RSdsd4	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu3	QoC1	RC2	RSpsd2	RSdsd3	PSds2	PSps2	De1	SR2
Gr1	SkRes2	LW2	Cu2	QoC2	RC2	RSpsd2	RSdsd2	PSds3	PSps3	De1	SR4
Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd3	RSdsd3	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC2	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW1	Cu2	QoC2	RC2	RSpsd2	RSdsd3	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu2	QoC2	RC3	RSpsd4	RSdsd2	PSds2	PSps2	De2	SR2

Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
Gr1	SkRes1	LW1	Cu3	QoC2	RC1	RSpsd2	RSdsd4	PSds2	PSps2	De1	SR1
Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd4	RSdsd4	PSds3	PSps3	De1	SR3
Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd4	RSdsd3	PSds2	PSps2	De2	SR4
Gr1	SkRes1	LW1	Cu1	QoC2	RC2	RSpsd4	RSdsd2	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd2	RSdsd2	PSds1	PSps1	De2	SR2
Gr1	SkRes1	LW1	Cu2	QoC1	RC3	RSpsd2	RSdsd3	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu2	QoC2	RC2	RSpsd3	RSdsd2	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd4	RSdsd4	PSds2	PSps2	De2	SR3
Gr1	SkRes2	LW1	Cu1	QoC3	RC1	RSpsd2	RSdsd2	PSds1	PSps1	De2	SR2
Gr1	SkRes1	LW1	Cu2	QoC1	RC3	RSpsd3	RSdsd2	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu2	QoC2	RC3	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC3	RSpsd3	RSdsd2	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW1	Cu3	QoC2	RC1	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu3	QoC2	RC1	RSpsd4	RSdsd2	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd2	RSdsd3	PSds2	PSps2	De1	SR3

Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd4	RSdsd4	PSds3	PSps3	De1	SR4
Gr1	SkRes1	LW1	Cu1	QoC2	RC2	RSpsd3	RSdsd4	PSds2	PSps2	De2	SR3
Gr1	SkRes1	LW1	Cu1	QoC2	RC3	RSpsd2	RSdsd4	PSds2	PSps2	De2	SR3
Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd2	RSdsd4	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd3	RSdsd2	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC2	RSpsd4	RSdsd3	PSds2	PSps2	De2	SR3
Gr1	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd2	RSdsd4	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd4	RSdsd4	PSds3	PSps3	De2	SR2
Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps4	De1	SR5
Gr5	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR2
Gr3	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW1	Cu1	QoC2	RC3	RSpsd3	RSdsd3	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd2	RSdsd3	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR3

Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
Gr1	SkRes1	LW1	Cu3	QoC2	RC3	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR1
Gr1	SkRes1	LW1	Cu1	QoC2	RC3	RSpsd4	RSdsd2	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd4	RSdsd2	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd3	RSdsd2	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW1	Cu1	QoC2	RC2	RSpsd2	RSdsd3	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps3	De1	SR5
Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd3	RSdsd4	PSds3	PSps3	De1	SR3
Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd3	RSdsd3	PSds2	PSps2	De2	SR4
Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd3	RSdsd2	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu2	QoC2	RC2	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR1
Gr1	SkRes1	LW1	Cu1	QoC2	RC2	RSpsd3	RSdsd2	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd4	RSdsd3	PSds3	PSps3	De1	SR3
Gr1	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd3	RSdsd4	PSds2	PSps2	De2	SR3
Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd4	RSdsd2	PSds2	PSps2	De2	SR3
Gr1	SkRes1	LW1	Cu3	QoC2	RC1	RSpsd2	RSdsd3	PSds2	PSps2	De1	SR2

Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd2	RSdsd4	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu2	QoC1	RC3	RSpsd2	RSdsd2	PSds2	PSps2	De2	SR1
Gr3	SkRes1	LW1	Cu1	QoC2	RC3	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC3	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd3	RSdsd4	PSds3	PSps3	De1	SR2
Gr3	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd4	RSdsd4	PSds3	PSps3	De1	SR2
Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd4	RSdsd4	PSds3	PSps3	De2	SR4
Gr1	SkRes2	LW1	Cu1	QoC3	RC3	RSpsd2	RSdsd2	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC2	RSpsd2	RSdsd4	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd4	RSdsd4	PSds3	PSps3	De1	SR4
Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps3	De1	SR5
Gr1	SkRes1	LW1	Cu1	QoC2	RC2	RSpsd3	RSdsd3	PSds2	PSps2	De2	SR3
Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR3
Gr1	SkRes1	LW1	Cu2	QoC2	RC2	RSpsd2	RSdsd3	PSds2	PSps2	De2	SR2

Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
Gr1	SkRes1	LW1	Cu3	QoC2	RC2	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR1
Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd2	RSdsd4	PSds3	PSps3	De1	SR2
Gr1	SkRes1	LW1	Cu3	QoC1	RC1	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR2
Gr3	SkRes1	LW1	Cu3	QoC1	RC1	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC3	RSpsd2	RSdsd3	PSds2	PSps2	De2	SR2
Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd4	RSdsd4	PSds3	PSps3	De2	SR4
Gr1	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd2	RSdsd2	PSds1	PSps1	De2	SR2
Gr1	SkRes1	LW1	Cu1	QoC2	RC2	RSpsd4	RSdsd2	PSds2	PSps2	De2	SR3
Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd4	RSdsd3	PSds3	PSps3	De2	SR2
Gr1	SkRes1	LW3	Cu1	QoC2	RC2	RSpsd3	RSdsd3	PSds3	PSps4	De1	SR5
Gr1	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd2	RSdsd3	PSds2	PSps2	De1	SR3
Gr5	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR1

Appendix F

A sample of the rules extracted for the (head-on + run-off) cashes related countermeasures

Differential speed limits	**Median type**	**Number of lanes**	**Grade**	**Skid resistance**	**Lane width**	**Curvature**	**Quality of curve**	**Road condition**	**Road side severity passenger-side distance**	**Roadside severity driver-Side distance**	**Paved shoulder driver-side**	**Paved shoulder passenger-side**	**Delineation**	**Vehicle Star Rating**
DifSpeed1	Traversable	NOL1	Gr1	SkRes2	LW2	Cu1	QoC1	RC2	RSpsd2	RSdsd2	PSds2	PSps2	De1	SR3
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC3	RSpsd1	RSdsd1	PSds2	PSps2	De2	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC3	RC2	RSpsd4	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC2	RC3	RSpsd1	RSdsd1	PSds2	PSps2	De2	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd1	RSdsd1	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd1	RSdsd1	PSds2	PSps2	De2	SR1
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps4	De1	SR4
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes1	LW3	Cu2	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps4	De1	SR4
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd1	RSdsd1	PSds3	PSps3	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes3	LW1	Cu1	QoC3	RC2	RSpsd4	RSdsd3	PSds2	PSps2	De1	SR2
DifSpeed1	Traversable	NOL1	Gr1	SkRes2	LW2	Cu3	QoC2	RC2	RSpsd2	RSdsd2	PSds2	PSps2	De2	SR3
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC3	RSpsd3	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC2	RC2	RSpsd4	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd1	RSdsd1	PSds2	PSps2	De2	SR1
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd3	RSdsd3	PSds1	PSps1	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd4	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW2	Cu2	QoC1	RC1	RSpsd1	RSdsd1	PSds3	PSps3	De1	SR2

Differential speed limits	Median type	Number of lanes	Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd1	RSdsd1	PSds2	PSps2	De2	SR1
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd3	RSdsd1	PSds3	PSps1	De1	SR2
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd1	RSdsd1	PSds3	PSps3	De1	SR2
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes1	LW3	Cu1	QoC2	RC2	RSpsd3	RSdsd2	PSds4	PSps4	De1	SR4
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd4	RSdsd4	PSds3	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd4	RSdsd3	PSds2	PSps2	De1	SR2
DifSpeed1	Traversable	NOL1	Gr1	SkRes2	LW1	Cu3	QoC2	RC2	RSpsd2	RSdsd2	PSds2	PSps2	De2	SR3
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd4	RSdsd1	PSds4	PSps1	De1	SR4
DefSpeed1	Non Traversable	NOL2	Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd3	RSdsd1	PSds3	PSps1	De1	SR2
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes2	LW1	Cu1	QoC3	RC2	RSpsd4	RSdsd3	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd1	RSdsd1	PSds3	PSps3	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW2	Cu1	QoC3	RC1	RSpsd1	RSdsd1	PSds3	PSps3	De1	SR1
DifSpeed2	Traversable	NOL1	Gr1	SkRes2	LW1	Cu1	QoC2	RC3	RSpsd2	RSdsd3	PSds3	PSps2	De2	SR3
DifSpeed2	Traversable	NOL1	Gr1	SkRes3	LW1	Cu1	QoC2	RC3	RSpsd3	RSdsd3	PSds3	PSps3	De2	SR3
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd3	RSdsd4	PSds2	PSps2	De1	SR3
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd4	RSdsd4	PSds4	PSps4	De1	SR5
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps4	De1	SR5
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd4	RSdsd3	PSds2	PSps2	De1	SR3

Differential speed limits	Median type	Number of lanes	Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd4	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd2	RSdsd1	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC3	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR2
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes2	LW2	Cu1	QoC2	RC2	RSpsd2	RSdsd2	PSds3	PSps3	De1	SR4
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd1	RSdsd1	PSds4	PSps1	De2	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd3	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC2	RC2	RSpsd4	RSdsd3	PSds2	PSps2	De1	SR1
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd1	RSdsd4	PSds4	PSps2	De1	SR2
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes2	LW1	Cu2	QoC1	RC2	RSpsd4	RSdsd2	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu2	QoC2	RC3	RSpsd2	RSdsd1	PSds3	PSps2	De1	SR2
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd1	RSdsd1	PSds3	PSps3	De1	SR2
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd2	RSdsd1	PSds3	PSps1	De1	SR2
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes2	LW2	Cu2	QoC2	RC2	RSpsd2	RSdsd2	PSds3	PSps3	De1	SR4
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd2	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC3	RSpsd2	RSdsd1	PSds2	PSps2	De2	SR1
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd4	RSdsd1	PSds3	PSps2	De2	SR2
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd1	RSdsd1	PSds3	PSps2	De2	SR1
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd1	RSdsd1	PSds3	PSps3	De1	SR2

Differential speed limits	Median type	Number of lanes	Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu2	QoC2	RC2	RSpsd1	RSdsd4	PSds4	PSps2	De1	SR2
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes2	LW1	Cu1	QoC3	RC1	RSpsd4	RSdsd3	PSds2	PSps2	De1	SR4
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC3	RSpsd4	RSdsd4	PSds3	PSps3	De1	SR2
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd2	RSdsd1	PSds2	PSps2	De1	SR2
DefSpeed1	Non Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd3	RSdsd1	PSds4	PSps1	De1	SR2
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes2	LW1	Cu1	QoC3	RC2	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes2	LW1	Cu1	QoC3	RC2	RSpsd4	RSdsd2	PSds2	PSps2	De1	SR1
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd3	RSdsd4	PSds2	PSps2	De1	SR3
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd3	RSdsd1	PSds3	PSps2	De1	SR2
DefSpeed1	Non Traversable	NOL2	Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd1	RSdsd4	PSds3	PSps3	De1	SR2
DefSpeed1	Non Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd1	RSdsd1	PSds3	PSps3	De2	SR3
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd1	RSdsd1	PSds4	PSps2	De2	SR2
DefSpeed2	Traversable	NOL1	Gr1	SkRes2	LW1	Cu1	QoC3	RC3	RSpsd2	RSdsd2	PSds1	PSps1	De2	SR3
DefSpeed1	Traversable	NOL1	Gr1	SkRes2	LW1	Cu1	QoC3	RC3	RSpsd2	RSdsd2	PSds1	PSps1	De2	SR1
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd1	RSdsd1	PSds3	PSps3	De2	SR2
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd4	RSdsd4	PSds3	PSps2	De1	SR5

Differential speed limits	Median type	Number of lanes	Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
DefSpeed1	Non Traversable	NOL2	Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd3	RSdsd1	PSds4	PSps1	De1	SR3
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd2	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes2	LW1	Cu2	QoC2	RC2	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes3	LW1	Cu1	QoC3	RC2	RSpsd2	RSdsd3	PSds2	PSps2	De1	SR1
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd4	RSdsd3	PSds2	PSps2	De1	SR3
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd1	RSdsd2	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd3	RSdsd1	PSds3	PSps2	De1	SR2
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes1	LW1	Cu3	QoC1	RC1	RSpsd4	RSdsd3	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu3	QoC1	RC2	RSpsd2	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes3	LW1	Cu1	QoC3	RC2	RSpsd3	RSdsd2	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes2	LW1	Cu2	QoC1	RC1	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes2	LW1	Cu2	QoC2	RC2	RSpsd4	RSdsd2	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd3	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd3	RSdsd4	PSds3	PSps2	De1	SR2
DefSpeed1	Non Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd1	RSdsd4	PSds4	PSps2	De1	SR2
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd4	RSdsd2	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd4	RSdsd3	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC3	RSpsd2	RSdsd3	PSds2	PSps2	De1	SR2

Differential speed limits	Median type	Number of lanes	Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu2	QoC2	RC2	RSpsd2	RSdsd1	PSds3	PSps2	De1	SR2
DefSpeed1	Non Traversable	NOL1	Gr5	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd4	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC3	RSpsd4	RSdsd4	PSds2	PSps2	De2	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC2	RC2	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes3	LW1	Cu1	QoC3	RC2	RSpsd4	RSdsd2	PSds3	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd2	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC2	RC1	RSpsd1	RSdsd2	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC3	RSpsd3	RSdsd2	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW2	Cu1	QoC3	RC1	RSpsd2	RSdsd1	PSds3	PSps2	De1	SR4
DefSpeed1	Non Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd1	RSdsd4	PSds3	PSps3	De1	SR2
DefSpeed1	Non Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd4	RSdsd1	PSds3	PSps2	De2	SR4
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps3	De1	SR5
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes1	LW3	Cu1	QoC1	RC1	RSpsd3	RSdsd3	PSds4	PSps3	De1	SR5
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC2	RC1	RSpsd3	RSdsd4	PSds1	PSps1	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes2	LW1	Cu1	QoC3	RC1	RSpsd2	RSdsd4	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC1	RC1	RSpsd3	RSdsd3	PSds2	PSps2	De1	SR2
DefSpeed1	Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC2	RC2	RSpsd4	RSdsd2	PSds2	PSps2	De1	SR1
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd4	RSdsd4	PSds3	PSps3	De1	SR2

Differential speed limits	Median type	Number of lanes	Grade	Skid resistance	Lane width	Curvature	Quality of curve	Road condition	Road side severity passenger-side distance	Roadside severity driver-Side distance	Paved shoulder driver-side	Paved shoulder passenger-side	Delineation	Vehicle Star Rating
DefSpeed1	Non Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd3	RSdsd1	PSds3	PSps2	De2	SR4
DefSpeed1	Non Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd3	RSdsd1	PSds3	PSps2	De2	SR4
DifSpeed1	Non-Traversable	NOL2	Gr1	SkRes1	LW3	Cu1	QoC2	RC2	RSpsd3	RSdsd3	PSds3	PSps4	De1	SR5
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu1	QoC3	RC1	RSpsd2	RSdsd4	PSds3	PSps1	De1	SR2
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd2	RSdsd3	PSds2	PSps2	De1	SR3
DefSpeed1	Non Traversable	NOL1	Gr1	SkRes1	LW1	Cu1	QoC3	RC2	RSpsd4	RSdsd4	PSds3	PSps3	De1	SR4
DefSpeed1	Traversable	NOL2	Gr1	SkRes1	LW1	Cu2	QoC1	RC2	RSpsd2	RSdsd4	PSds3	PSps2	De1	SR2