Using Data Mining Techniques to Analyze Crash Patterns in Sri Lanka Road Accident Data

U L A S Perera
158768U

Faculty of Information Technology
University of Moratuwa
February 2019

Using Data Mining Techniques to Analyze Crash Patterns in Sri Lanka Road Accident Data

U L A S Perera
158768U

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Degree of Master of Science in Information Technology

February 2019

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information delivered form the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of the Student	Signature of the student
U. L. A. S. Perera	.....
	Date:

Supervised by

Name of Supervisor	Signature of Supervisor
S. C. Premaratne	.....
	Date:

Acknowledgements

I would like to express sincere gratitude to my project supervisor, Mr. S. C. Premaratne, Senior Lecturer in University of Moratuwa, who spent his valuable time for guiding this research to make it a success. I would also like to thank Prof. Asoka Karunananda and Dr. M. F. M. Firdhous who are lecturers in charge of literature review, research methodology and thesis writing subjects which were the basis for this research. Not only that my thanks should go to all the lecturers in MSc in Information Technology, all the facilitators and batch mates of MSc IT 2015/2016 batch.

Moreover a very special thanks should go to DIG Traffic Administration, Mr. Hemantha and his staff members of Sri Lanka Police for assisting me in collecting dataset and giving me their valuable comments on my research.

Finally I wish to thank my wife Kumudu for her understanding and guidance throughout this project and my two sons, Venura and Sandaru for their love and support.

Abstract

The road safety has been identified as a major factor that influences the sustainable development worldwide. This growing interest in road safety, is reflected by including it in Sustainable Development Goals of United Nations as "Halve the number of global deaths and injuries from road traffic accidents by 2020". According to road accident statistics published by Sri Lanka traffic police in 2015, every three and half hours a person is killed due to a road accident and two are seriously injured. This shows that travelling on local roads becoming more and more unsafe and risky. When improving the road safety conditions, it is necessary to identify the major factors contributing to road crash injuries and deaths, in order to take appropriate safety measures.

The Sri Lanka Police department uses MAAP (Microcomputer Accident Analysis Package) system for the storage and analysis of Road Traffic Accidents (RTA) data. However MAAP has its own limitations of analysis of accident data.

In the area of road traffic accident analysis, data mining technique has been recognize as a reliable technique which can be used beyond the conventional techniques. When analyzing road traffic accidents, different models were developed to identify factors affecting the severity of a traffic accident.

The objectives of this study are to explore the underlying factors influencing on injury severity, to identify the human, environment and vehicle factors influencing the road traffic accident severity and to identify crash proneness of road segments using available road and crash factors. In this study, data mining classification model is used to detect factors which influence on road accidents. We conducted an experiment with road accident data in 2015, provided by Sri Lanka Police.

In this research we proposed an accident severity model based on selected data mining techniques to identify influential factors for the severity of road traffic accidents. The solution model is developed using Weka software tool.

Table of Contents

DECLARATION **I**

ACKNOWLEDGEMENTS **II**

ABSTRACT **III**

TABLE OF CONTENTS **IV**

LIST OF FIGURES **VIII**

LIST OF TABLES **VIII**

CHAPTER 1 INTRODUCTION **1**

1.1 Introduction **1**

1.2 Background and Motivation **1**

1.3 Problem Definition **2**

1.4 Aim and Objectives **2**

1.5 Proposed Solution **2**

1.6 Structure of Dissertation **3**

1.7 Summary **3**

CHAPTER 2 DISCOVERING CRASH PATTERNS IN ROAD ACCIDENT DATA **4**

2.1 Introduction **4**

2.2 Road Safety and Traffic Accidents **4**

2.3 Road Accident Analysis **5**

2.4 Data Mining Techniques in Road Traffic Accident **6**

2.5 Summary of Challenges **7**

2.6 Problem Definition **9**

2.7 Aim and Objectives **9**

2.8 Summary **10**

CHAPTER 3 TECHNOLOGY ADAPTED **11**

3.1 Introduction **11**

3.2 Data Mining **11**

3.3 Steps of Knowledge Discovery Process **12**

3.4 Data Mining Models **12**

3.5 Major Applications of Data Mining **15**

3.6 Data Mining in Road Accident Analysis **15**

3.7 Tools used for Data Mining **16**

3.8 Data Mining Algorithms used **16**

3.8 Summary **17**

CHAPTER 4 A NOVEL APPROACH TO ANALYZE ROAD TRAFFIC ACCIDENT DATA **18**

4.1 Introduction **18**

4.2 Hypothesis **18**

4.3 Input **18**

4.5 Process **18**

4.5.1 The Proposed Model **19**

4.6 Users **20**

4.7 Summary **20**

CHAPTER 5 RESEARCH DESIGN FOR ANALYZING RTA DATA **21**

5.1 Introduction **21**

5.2 Research Design **21**

5.3 Summary **23**

CHAPTER 6 IMPLEMENTATION **24**

6.1 Introduction **24**

6.2 Solution for the Research Objective **24**

6.2.1 Data Pre-processing **25**

6.2.1.1 Data Pre-processing for the First Research Objective **25**

6.2.1.2 Data Pre-processing for the Second Research Objective **25**

6.2.2. Attribute/Feature Selection **25**

6.2.2.1 Attribute Selection for the First Research Objective **27**

6.2.2.2 Attribute Selection for Second Research Objective **29**

6.2.3. Measure the Variable Importance **32**

6.2.4. Classification Rule Extraction **34**

6.3 Summary **37**

CHAPTER 7 EVALUATION **38**

7.1 Introduction **38**

7.2 Evaluation for classification **38**

7.3 Evaluation of Injury severity **39**

7.4 Evaluation of Accident Severity **41**

7.5 Summary **43**

CHAPTER 8 CONCLUSION AND FURTHER WORKS **44**

8.1 Introduction **44**

8.2 Overview of the Research **44**

8.3 Key Findings **44**

8.4 Problems Encountered and Limitations **46**

8.5 Further Works **46**

8.6 Summary **46**

REFERENCES **47**

APPENDIX A: TABLE DETAILS OF ACCIDENT DATABASE **50**

APPENDIX B: INITIAL ATTRIBUTES SETS FOR SAMPLE SET 01 **51**

APPENDIX C: SELECTED ATTRIBUTES SETS FOR SAMPLE SET 01 **52**

APPENDIX D INITIAL ATTRIBUTES SETS FOR SAMPLE SET 02 **53**

APPENDIX E: SELECTED ATTRIBUTES SETS FOR SAMPLE SET 02 **54**

APPENDIX F: MEASURING ATTRIBUTE IMPORTANCE **55**

APPENDIX G: FURIA RULE EXTRACTION **57**

List of Figures

Figure 3-1 Steps of Knowledge Discovery Process 11
Figure 3-2 Data Mining Models 13
Figure 4-1 Proposed Model 20
Figure 4-2 Overall System design of the Proposed Solution 22
Figure 4-3 Weka GUI 24
Figure 4-4 Weka Attribute Selector Tool 26
Figure 4-6 Measure the attribute importance with RF 33

List of Tables

Table 2-1 Summary of literature 8
Table 3-1 Application of Data Mining Techniques 15
Table 4-1 Attributes selected after Wrapper Method 29
Table 4-2 Attributes selected after wrapper method 31
Table 4-3 VIM for Injury severity 33
Table 4-4 VIM for Accident Severity 34
Table 4-5 FURIA rule generation for injury severity 35
Table 4-6 FURIA rule generation for Accident severity 36
Table 4-7 Evaluation measures for classifiers 38
Table 4-8 J48 Tree Classifier evaluation summary for injury severity model 39
Table 4-9 The prediction accuracy of J48 by using the test data set 40
Table 4-10 Random Forest evaluation summary for injury severity model 40
Table 4-11 FURIA evaluation summary for injury severity model 40
Table 4-12 J48 Classifier Evaluation Summary for accident severity model 41
Table 4-13 The prediction accuracy of J48 by using test data set 41
Table 4-14 Random Forest evaluation summary of accident severity model 42
Table 4-15 FURIA evaluation summary of accident severity model 42
Table 4-16 Most influential factors of injury severity based on algorithm 45
Table 4-17 Most influential factors of accident severity based on algorithm 45