

A Decision Support System to Predict Highway Accident Alerts in Sri Lanka

K. A. D. S. H. Rodrigo
179476A

Faculty of Information Technology
University of Moratuwa
2020

A Decision Support System to Predict Highway Accident Alerts in Sri Lanka

K. A. D. S. H. Rodrigo
179476A

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

2020

Declaration

I declare that this 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 Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student

K. A. D. S. H. Rodrigo

Signature of Student

.....

Date:

Supervised By

Name of Supervisor

S. C. Premarathne

Signature of Supervisor

.....

Date:

Acknowledgement

Firstly, I would like to express my sincere gratitude to my supervisor Mr. Saminda Premarathne for the continuous support of my dissertation study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this dissertation. I could not have imagined having a better advisor and mentor for my research study.

Besides my supervisor, I would like to thank Dr. Mohamed Firdhous who taught Research Methodology and Literature Review and Thesis Writing modules which helped me to widen my research from various perspectives.

I would not forget to remember the Sri Lanka Police for their timely support by providing all the accident records that have been required for the success of this research.

In addition, I would like to thank all the staff of Paragon Software Lanka for their insightful comments and encouragement which helped me lot to enhance my knowledge from various perspectives.

Furthermore, I would also like to thank all the batch mates of the M.Sc. in IT degree program who gave their valuable feedbacks to improve the results of the research, my family for the support they provided me throughout my entire life and in particular.

Abstract

Road traffic and rapidly increasing road accidents become a vast problem not only for Sri Lankans but for all human beings who are living in this planet. According to the Sri Lanka police highway reports, there have been a number of deaths recorded due to fatal accidents in highways and also affected significant amount of people for non-fatal accidents as well. Several key factors vastly contributed directly to cause an accident, such as environmental factors, human factors and road condition factors., etc. In this research, machine learning techniques and methods have been applied to the southern highway accident records retrieved by Police Highway Control Center for the period of 2015 to June 2019 in order to establish a model which enables to forecast reason for the accidents that will be occurring in the future. The python scikit-learn package has been used on top of the anaconda framework to discover hidden patterns of data with the help of decision tree, support vector machine and logistic regression powered by a one-vs-rest classifier. The two well-known ensembles; random forest and gradient boost classifiers have also performed in this dataset in order to enhance the accuracy. The performance of each classifier has compared critically based on their results. The results obtained by performing various experiments show that the ensemble's work well when compare to other classifiers.

Table of Content

Acknowledgement	i
Abstract	ii
Table of Content.....	iii
List of Tables.....	vi
List of Figures	vii
List of Appendices	x
Chapter 1	1
Introduction.....	1
1.1 Prolegomena	1
1.2 Problem Statement	2
1.3 Aim and Objectives.....	2
1.4 Scope of Research.....	3
1.5 Proposed Solution	3
1.6 Structure of the Thesis	3
Chapter 2	4
Literature Review.....	4
2.1 Introduction.....	4
2.2 Application of Data Mining and Mobile Technologies on Accident Data.....	4
2.3 Gaps and Limitations	7
2.4 Summary	7
Chapter 3	9
Environment Setup.....	9
3.1 Introduction.....	9
3.2 Anaconda Framework	9
3.3 Packages and Tools	9
3.4 Summary	10
Chapter 4	11
A Novel Approach to Forecast Highway Accident in Sri Lanka	11

4.1 Introduction.....	11
4.2 Data Mining and Machine Learning Application Domain	11
4.2.1 Preparation of Data Set	12
4.2.2 Data Source and Description	12
4.2.3 Mining Scheme	13
4.2.4 Preprocessing	15
4.2.5 Parameter Tuning.....	22
4.2.6 Performance Metric	23
4.2.7 Boosting Techniques	24
4.2.8 Experimental Setups	24
4.3 Mobile Application Development.....	28
4.3.1 Mobile Application Development Architecture	28
4.3.2 Mobile Application Design (UI/UX)	29
4.3.3 Implement Database API (Application Program Interface).....	29
4.3.4 Functional Testing.....	29
4.4 Summary	30
Chapter 5.....	31
Experimental Result and Discussion	31
5.1 Introduction.....	31
5.2 Experimental Analysis	31
5.2.1 Experimental Analysis for LHS Dataset.....	32
5.2.2 Experimental Analysis for RHS Dataset.....	33
5.3 Experimental Summary	35
5.4 Comparison of Experiment Schemes.....	35
5.5 Comparison with Previous Work	35
5.6 Summary.....	36
Chapter 6.....	37
Conclusion and Future Work	37

6.1 Introduction.....	37
6.2 Future Developments	37
6.3 Summary	37
References.....	38
Appendix - A.....	41
Decision Tree Classifier LHS	41
Appendix - B.....	45
SVM with OVR Classifier LHS	45
Appendix - C.....	47
LR with OVR Classifier LHS	47
Appendix - D.....	49
Random Forest Classifier LHS	49
Appendix - E.....	51
Gradient Boosting Classifier LHS	51
Appendix - F	53
Decision Tree Classifier RHS	53
Appendix - G.....	57
SVM with OVR Classifier RHS	57
Appendix - H.....	59
LR with OVR Classifier RHS.....	59
Appendix - I.....	61
Random Forest Classifier RHS.....	61
Appendix - J.....	63
Gradient Boosting Classifier RHS.....	63

List of Tables

Table 4.1: Attribute Description for Both LHS and RHS Dataset	13
Table 4.2: Label Encoded Bridge_Type Column.....	18
Table 4.3: One-Hot Encoded Safety-Level Column	18
Table 4.4: Data Discretization or Binning	19
Table 4.5: Preprocessed Dataset	21
Table 4.6: Performance of Decision Tree Classifier for LHS Dataset.....	25
Table 4.7: Performance of SVM Classifier for LHS Dataset.....	25
Table 4.8: Performance of Logistic Regression Classifier for LHS Dataset	26
Table 4.9: Performance of Random Forest Classifier for LHS Dataset.....	26
Table 4.10: Performance of Gradient Boost Classifier for LHS Dataset.....	26
Table 4.11: Performance of Decision Tree Classifier for RHS Dataset.....	27
Table 4.12: Performance of SVM Classifier for RHS Dataset	27
Table 4.13: Performance of Logistic Regression Classifier for RHS Dataset	27
Table 4.14: Performance of Random Forest Classifier for RHS Dataset	28
Table 4.15: Performance of Gradient Boost Classifier for RHS Dataset.....	28

List of Figures

Figure 3.1: Anaconda Software Distribution Setup Guide	9
Figure 4.1: The Working Mechanism of Proposed Model.....	12
Figure 4.2: Correlation Heatmaps for Both Datasets.....	16
Figure 4.3: PCA for Both Datasets	20
Figure 4.4: Mechanism for Train Validation Test Splits	21
Figure 4.5: Proposed Architecture of Mobile Application.....	29

List of Abbreviations

- SLP - Sri Lanka Police
- USA - United States of America
- CART - Classification and Regression Tree
- ID3 - Iterative Dichotomiser
- UAE - United Arab Emirates
- ANN - Artificial Neural Network
- SVM - Support Vector Machine
- LR - Logistic Regression
- OVR - One-Vs-Rest
- OVO - One-Vs-One
- OVA - One-Vs-All
- GPS - Global Positioning System
- SMS - Short Message Service
- IOS - iPhone Operating System
- LHS - Left Hand Side
- RHS - Right Hand Side
- KNN - K-Nearest Neighbor
- PCA - Principal Component Analysis
- DT - Decision Tree
- ROS - Random Over Sampler
- GS - Grid Search
- RS - Randomize Search
- RF - Random Forest

- GB - Gradient Boost
- API - Application Program Interface
- AWS - Amazon Web Services
- UI - User Interface
- UX - User Experience

List of Appendices

Appendix - A.....	41
Appendix - B.....	45
Appendix - C.....	47
Appendix - D.....	49
Appendix - E.....	51
Appendix - F.....	53
Appendix - G.....	57
Appendix - H.....	59
Appendix - I.....	61
Appendix - J.....	63