

ROAD SAFETY PREDICTION USING MLA

A PROJECT REPORT

Submitted by

BAIRAVI G (810015205014)

LALITHAVANI K (810015205036)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHNOLOGY



UNIVERSITY COLLEGE OF ENGINEERING

BIT CAMPUS, TIRUCHIRAPPALLI 620024

ANNA UNIVERSITY::CHENNAI 600025

APRIL 2019

UNIVERSITY COLLEGE OF ENGINEERING

BIT CAMPUS

TIRUCHIRAPPALLI-620024

BONAFIDE CERTIFICATE

Certified that this project report “**ROAD SAFETY PREDICTION USING MLA**” is the bonafide work of “**BAIRAVI G (810015205014)** and **LALITHAVANI K (810015205036)**” who carried out the project work under my supervision.

SIGNATURE

Dr. D. Venkatesan

HEAD OF THE DEPARTMENT

Assistant Professor
Department of Information Technology
University College of Engineering,
Anna University-BIT Campus,
Tiruchirappalli -620 024.

SIGNATURE

Dr. D. Asir Antony Gnana Singh

SUPERVISOR

Teaching Fellow
Department of Computer Science and
Engineering,
University College of Engineering,
Anna University-BIT Campus,
Tiruchirappalli -620 024.

Submitted for the project Viva voce examination held on

Internal Examiner

External Examiner

DECLARATION

We hereby declare that the work entitled “**ROAD SAFETY PREDICTION USING MLA**” is submitted in partial fulfillment of the requirement for the award of the degree in B.TECH, in University college of Engineering, BIT Campus, Anna University, Tiruchirappalli. It is the record of our own work carried out during the academic year 2018-2019 under the supervision and guidance of **Dr. D. ASIR ANTONY GNANA SINGH**, Teaching Fellow, Department of Computer Science and Engineering, University college of Engineering, BIT Campus, Anna University, Tiruchirappalli. The extent and source of information are derived from the existing literature and have been indicated through the dissertation at the appropriate places.

BAIRAVI G (810015205014)

LALITHAVANI K (810015205036)

I certify that the declaration made above by the candidates is true

Signature of the Guide,

Dr. D. Asir Antony Gnana Singh

Teaching Fellow,

Department of Computer Science and Engineering,

University college of Engineering,

Anna University- BIT Campus,

Tiruchirappalli-620 024

ACKNOWLEDGEMENT

We would like to thank our honorable Dean **Dr. T. SENTHI KUMAR**, professor for having provided us with all required facilities to complete our project without hurdles.

We would like to express our sincere thanks to **Dr. D. VENKATESAN**, Head of the Department of Computer Science and Engineering, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of this project work.

We would like to thank our project Coordinator **Mr. M. PRASANA KUMAR** Teaching Fellow, Department of Computer Science and Engineering for his kind support.

We would like to thank and express our deep sense of gratitude to our project guide **Dr. D. ASIR ANTONY GNANA SINGH**, Teaching Fellow, Department of Computer Science and Engineering, for his valuable guidance throughout the project. We also extend our thanks to all other teaching and non-teaching staff for their encouragement and support.

We thank our beloved parents and friends for their full support in the moral development of this project.

ABSTRACT

Every year, there are lot of accidents happening in our country. This leads to increase the fatality rate and lot of injuries to a people. This road accidents may not be recovered completely, but it can be reduced. This may be occurred due to various factors such as time, driver's carelessness, weather conditions, traffic, and speed of the vehicles. Machine learning plays a significant role in pattern recognition and prediction. Moreover, this project uses machine learning approach to implement a road safety predict system by considering the various factors including road surface, number of vehicles, accident date, time, weather condition, and casualty. In order to improve the accuracy of the proposed system, the feature selection method is incorporated with the proposed system. The proposed system is tested on various benchmark datasets and it is identified that the proposed system produces better accuracy compared to other method compared.

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	v
	LIST OF ABBREVIATIONS	viii
	LIST OF TABLES	ix
	LIST OF FIGURES	xvi
1	INTRODUCTION	1
2	LITERATURE REVIEW	2
3	METHODOLOGY	
	3.1 FEATURE SELECTION	15
	3.1.1 Subset Selection	16
	3.1.2 Ranker Selection	17
	3.2 MACHINE LEARNING	18
	3.2.1 Naive Bayes Classification	19
	3.2.2 Sequential Minimal Optimization	24
	3.2.3 Decision Tree	24
	3.3 PERFORMANCE EVALUATION METRICS	26
	3.3.1 Confusion Matrix	26
	3.3.2 Accuracy	26
	3.3.3 TP Rate	27
	3.3.4 FP Rate	27
	3.3.5 Precision	27
	3.3.6 Recall	27
	3.3.7 F-Measure	27
	3.3.8 MCC	28
	3.3.9 ROC Area	28

	3.3.10 PRC Area	28
4	PROPOSED SYSTEM	
	4.1 FEATURE SELECTION USING ONE-R	29
	4.2 K-STAR	36
	4.3 ARCHITECTURE DIAGRAM	41
5	SOFTWARE AND HARDWARE SPECIFICATION	
	5.1 HARDWARE SPECIFICATION	43
	5.2 SOFTWARE SPECIFICATION	43
6	RESULTS AND DISCUSSION	
	6.1 EXPERIMENTAL SETUP	44
	6.2 EXPERIMENTAL PROCEDURE	45
7	CONCLUSION	91
8	REFERENCES	92

LIST OF ABBREVIATION

S.NO	ABBREVIATION	EXPANSION
1	J48	Decision Tree
2	One R	One Rule
3	K-Star	Korea Superconducting Tokamak Advance Research
4	NB	Naive Bayes
5	SMO	Sequential Minimal Optimization
6	TP Rate	True Positive Rate
7	FP Rate	False Positive Rate
8	MCC	Matthews Correlation Coefficient
9	ROC	Receiver Operating System
10	PRC	Precision Recall
11	WEKA	Waikaato Environment for Knowledge Analysis

LIST OF TABLES

S.NO	TITLE	P.NO
3.1	Dataset 2015 for naive Bayes algorithm	19
4.1	Frequency table for no of vehicles with the class Type of vehicle	30
4.2	Frequency table for 1 st road class with the class Type of vehicle	30
4.3	Frequency table for road surface with the class Type of vehicle	31
4.4	Frequency table for lighting condition with the class Type of vehicle	31
4.5	Frequency table for weather condition with the class Type of vehicle	32
4.6	Frequency table for casualty class with the class Type of vehicle	32
4.7	Frequency table for casualty severity with the class Type of vehicle	33
4.8	Frequency table for sex of casualty with the class Type of vehicle	33
4.9	Frequency table for sex of casualty with the class Type of vehicle	33
4.10	Confusion matrix table for lighting condition	34
4.11	Confusion matrix table for casualty severity	35
4.12	Confusion matrix table for sex of casualty	35
6.1	Details of datasets	44
6.2	Accuracy on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	47
6.3	TP Rate on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	47

6.4	FP Rate on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	48
6.5	Precision on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	48
6.6	Recall on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	49
6.7	F Measure value on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	49
6.8	MCC on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	50
6.9	ROC Area on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	50
6.10	PRC Area on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	51
6.11	Accuracy on Leeds 2009 dataset by CAE with respect to J48, Naive Bayes, K Star, One R algorithm	51
6.12	Accuracy on Leeds 2009 dataset by IGAE with respect to J48, Naive Bayes, K Star, One R algorithm	52
6.13	Accuracy on Leeds 2009 dataset by GRAE with respect to J48, Naive Bayes, K Star, One R algorithm	52
6.14	Accuracy on Leeds 2009 dataset by ORAE with respect to J48, Naive Bayes, K Star, One R algorithm	53
6.15	Accuracy on Leeds 2009 dataset by RAE with respect to J48, Naive Bayes, K Star, One R algorithm	53
6.16	Accuracy on Leeds 2009 dataset by SUAE with respect to J48, Naive Bayes, K Star, One R algorithm	54

6.17	Accuracy on Leeds 2010 dataset by CAE with respect to J48, Naive Bayes, K Star, One R algorithm	54
6.18	Accuracy on Leeds 2010 dataset by IGAE with respect to J48, Naive Bayes, K Star, One R algorithm	55
6.19	Accuracy on Leeds 2010 dataset by GRAE with respect to J48, Naive Bayes, K Star, One R algorithm	55
6.20	Accuracy on Leeds 2010 dataset by ORAE with respect to J48, Naive Bayes, K Star, One R algorithm	56
6.21	Accuracy on Leeds 2010 dataset by RAE with respect to J48, Naive Bayes, K Star, One R algorithm	56
6.22	Accuracy on Leeds 2010 dataset by SUAE with respect to J48, Naive Bayes, K Star, One R algorithm	57
6.23	Accuracy on Leeds 2012 dataset by CAE with respect to J48, Naive Bayes, K Star, One R algorithm	57
6.24	Accuracy on Leeds 2012 dataset by IGAE with respect to J48, Naive Bayes, K Star, One R algorithm	58
6.25	Accuracy on Leeds 2012 dataset by GRAE with respect to J48, Naive Bayes, K Star, One R algorithm	58
6.26	Accuracy on Leeds 2012 dataset by ORAE with respect to J48, Naive Bayes, K Star, One R algorithm	59
6.27	Accuracy on Leeds 2012 dataset by RAE with respect to J48, Naive Bayes, K Star, One R algorithm	59
6.28	Accuracy on Leeds 2012 dataset by SUAE with respect to J48, Naive Bayes, K Star, One R algorithm	60
6.29	Accuracy on 2009 dataset by CAE with respect to J48, Naive Bayes, K Star, One R algorithm	60

6.30	Accuracy on 2009 dataset by IGAE with respect to J48, Naive Bayes, K Star, One R algorithm	61
6.31	Accuracy on 2009 dataset by GRAE with respect to J48, Naive Bayes, K Star, One R algorithm	61
6.32	Accuracy on 2009 dataset by ORAE with respect to J48, Naive Bayes, K Star, One R algorithm	62
6.33	Accuracy on 2009 dataset by RAE with respect to J48, Naive Bayes, K Star, One R algorithm	62
6.34	Accuracy on 2009 dataset by SUAE with respect to J48, Naive Bayes, K Star, One R algorithm	63
6.35	Accuracy on 2010 dataset by CAE with respect to J48, Naive Bayes, K Star, One R algorithm	63
6.36	Accuracy on 2010 dataset by GRAE with respect to J48, Naive Bayes, K Star, One R algorithm	64
6.37	Accuracy on 2010 dataset by ORAE with respect to J48, Naive Bayes, K Star, One R algorithm	64
6.38	Accuracy on 2010 dataset by RAE with respect to J48, Naive Bayes, K Star, One R algorithm	65
6.39	Accuracy on 2010 dataset by SUAE with respect to J48, Naive Bayes, K Star, One R algorithm	65
6.40	Accuracy on 2015 dataset by CAE with respect to J48, Naive Bayes, K Star, One R algorithm	66
6.41	Accuracy on 2015 dataset by IGAE with respect to J48, Naive Bayes, K Star, One R algorithm	66
6.42	Accuracy of 2015 dataset by GRAE with respect to J48, Naive Bayes, K Star, One R algorithm	67

6.43	Accuracy on 2015 dataset by ORAE with respect to J48, Naive Bayes, K Star, One R algorithm	67
6.44	Accuracy on 2015 dataset by RAE with respect to J48, Naive Bayes, K Star, One R algorithm	68
6.45	Accuracy on 2015 dataset by SUAE with respect to J48, Naive Bayes, K Star, One R algorithm	68
6.46	Accuracy on 2011 dataset by CAE with respect to J48, Naive Bayes, K Star, One R algorithm	69
6.47	Accuracy on 2011 dataset by IGAE with respect to J48, Naive Bayes, K Star, One R algorithm	69
6.48	Accuracy on 2011 dataset by GRAE with respect to J48, Naive Bayes, K Star, One R algorithm	70
6.49	Accuracy on 2011 dataset by ORAE with respect to J48, Naive Bayes, K Star, One R algorithm	70
6.50	Accuracy on 2011 dataset by RAE with respect to J48, Naive Bayes, K Star, One R algorithm	71
6.51	Accuracy on 2011 dataset by SUAE with respect to J48, Naive Bayes, K Star, One R algorithm	71
6.52	Accuracy on 2012 dataset by CAE with respect to J48, Naive Bayes, K Star, One R algorithm	72
6.53	Accuracy on 2012 dataset by IGAE with respect to J48, Naive Bayes, K Star, One R algorithm	72
6.54	Accuracy on 2012 dataset by GRAE with respect to J48, Naive Bayes, K Star, One R algorithm	73
6.55	Accuracy on 2012 dataset by ORAE with respect to J48, Naive Bayes, K Star, One R algorithm	73

6.56	Accuracy on 2012 dataset by RAE with respect to J48, Naive Bayes, K Star, One R algorithm	74
6.57	Accuracy on 2012 dataset by SUAE with respect to J48, Naive Bayes, K Star, One R algorithm	74
6.58	Accuracy on Leeds 2011 dataset by GRAE with respect to J48, Naive Bayes, K Star, One R algorithm	75
6.59	Accuracy on Leeds 2011 dataset by IGAE with respect to J48, Naive Bayes, K Star, One R algorithm	75
6.60	Accuracy on Leeds 2011 dataset by ORAE with respect to J48, Naive Bayes, K Star, One R algorithm	76
6.61	Accuracy on Leeds 2011 dataset by RAE with respect to J48, Naive Bayes, K Star, One R algorithm	76
6.62	Accuracy on Leeds 2011 dataset by SUAE with respect to J48, Naive Bayes, K Star, One R algorithm	77
6.63	Accuracy on Leeds 2011 dataset by CAE with respect to J48, Naive Bayes, K Star, One R algorithm	77
6.64	Average Accuracy on various datasets by Ranker Selection with respect to Naive Bayes, J48, K Star, One R	78
6.65	Accuracy on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	78
6.66	TP Rate on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	79
6.67	FP Rate on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	79
6.68	Precision on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	80

6.69	Recall on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	80
6.70	F Measure value on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	81
6.71	MCC on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	81
6.72	ROC Area on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	82
6.73	PRC Area on various datasets with respect to J48, Naive Bayes, K Star, SMO, One R algorithm	82

LIST OF FIGURES

S.NO	TITLE	P.NO
4.1	Proposed system for road Safety prediction	29
4.2	Architecture diagram for Road safety prediction	42
6.1	Experimental procedure for road Safety prediction	46