

**Predicting the Severity of Vehicle Accidents Based on Traffic Accident Attributes Using  
Machine Learning**

**Segun Abayomi SOFOLUWE**

**LCU/PG/002221**

**Being a MSc Thesis Submitted to the Department of Computer Science, Faculty of Natural  
and Applied Sciences, Lead City University, Ibadan, Oyo State, Nigeria**

**In Partial Fulfilment of the Requirements for the Award of Master of Science Degree  
(MSc) in computer science**

*Do Not Copy, Lead City University, Nigeria*

2023

*Do Not Copy, Lead City University, Nigeria*

**Certification**

This is to certify that Segun Abayomi SOFOLUWE with matriculation number LCU/PG/002221 carried out this research work titled “Predicting the Severity of Vehicle Accidents Based on Traffic Accident Attributes Using Machine Learning” in the Department of Computer Science, Faculty of Natural and Applied Sciences, Lead City University, Ibadan, Oyo State, for the award of Master of Science (MSc) in Computer Science and that this has not been previously submitted.

.....  
Dr. A.A Waheed  
**Supervisor**

.....  
**Date**

.....  
Dr. Wilson Sakpere  
**Head of Department**

.....  
**Date**

Do Not Copy, Lead City University, Nigeria

## **Dedication**

This research work is dedicated to God almighty, my Parents, and my Sisters.

*Do Not Copy, Lead City University, Nigeria*

## **Acknowledgement**

Foremost, I would like to express my gratitude to the leadership of the Lead City University, Ibadan for creating a medium for us to acquire knowledge for self-reliance.

I acknowledge my supervisor Dr. A.A Waheed, for the continuous support for M.Sc study and research, for his motivation, enthusiasm, to guide me through the research. The immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. Besides my supervisor, my sincere thanks go to the Head of Department Dr. Wilson Sakpere, the Postgraduate Coordinator Dr. Azeez Waheed, all other lecturers and staff members in the department of computer science for their guidance encouragement, and insightful comments.

I thank my fellow coursemates, Mr. Abiodun Akanni, Mr. Lucky Samuel, Miss. Iyanuoluwa Fatoki, Mr. Jimmy Lawal, Mr. Folahan Jiboku and others numerous to mention.

I thank my Parents Mr. Yomi Sofoluwe and Mrs. Bola Sofoluwe for their guidance, parental advice and commitment instilled in me towards learning. May you both live long in good health to enjoy and reap fruits of your labour. My sincere thanks to my Sister Miss Sarah Sofoluwe for the understanding, sacrifice, and prayers during my study.

## Table of Contents

Content	Page
Certification	i
Dedication	ii
Acknowledgement	iii
Table of Content	iv
List of Tables	vii
List of Figure	viii
List of Acronyms	x
<b>Chapter one: Introduction</b>	<b>1</b>
1.1 Background to the Study	1
1.2 Statement of the Problem 7	
1.3 Aim and Objectives of the Study 8	
1.4 Significance of the Study 8	
1.5 Scope of Study 9	
1.6 Limitation of the Study	10
1.7 Operational Definition of Terms	11
Endnotes	
<b>Chapter Two: Literature Review</b>	<b>15</b>
2.1 Conceptual Review	15

2.1.1	Road Traffic Accident	16
2.1.2	Machine Learning	21
2.1.3	Classification of RTA Using Machine Learning	28
2.1.3.1	Gradient Boosting Machine	29
2.1.3.2	Support Vector Regression	30
2.1.3.3	Extreme Learning Machines	32
2.1.3.4	K Nearest Neighbour	33
2.1.3.5	Gaussian Naive Beyes	35
2.1.3.6	Multilayer Perceptron	37
2.1.3.7	Esemble Technique	39
2.1.4	Neural Network	43
2.1.5	Application of Neural Network	45
2.1.6	Classification of (ANN)	46
2.2	Methodological Review	49
2.2.1	Random Forest Algorithm	49
2.2.2	Decision Tree	54
2.2.2.1	Type of Decision Tree Algorithm	54
2.3	Review of Related Works	55
2.3.1	Classification Based on Decision Tree Algorithm	55
2.3.2	Classification Based on Random Forest Algorithm	63
2.3.3	Classification of Vehicle Accident Using Machine Learning	66
2.4	Summary of Gaps in Literature Reviewed	78
	Endnote	
	<b>Chapter Three: Methodology</b>	<b>91</b>
3.1	Research Approach	91
3.2	Requirements Specification	91
3.3	System Design	93

3.4	Research Methods	95
3.4.1	Data Collection and Pre-Processing	95
3.4.2	Dataset Details	96
3.4.3	Data Mining/Preprocessing	97
3.4.4	Feature Engineering	98
3.4.5	Model Design, Training and Validation	98
3.4.6	Evaluation Process	99
3.5	Ethical Consideration	99
	Endnote	
	<b>Chapter Four: Results and Discussion of Findings</b>	<b>103</b>
4.1	Result on Data Collection	103
4.1.1	Data Mining and Pre -processing and exploratory Data Analysis	104
4.2	Model Design, Training, and Validation	107
4.3	The Evaluation Process	111
4.3.1	Comparing the Two Models	115
4.4	Discussions of Findings	116
	Endnote	
	<b>Chapter Five : Conclusion</b>	<b>120</b>
5.1	Summary	120
5.2	Conclusion	121
5.3	Recommendations	122
5.4	Contribution to Knowledge	124
5.5	Suggestions for Further Research	125
	Bibliography	127
	Conference Proceedings	133
	Dissertation	138

Appendix	140
Bio Data	171
University Compliance Form	173

#### **List of Tables**

<b>Table</b>	<b>Title</b>	<b>Page</b>
4.1	Result of Precision, Recall, f1 Score for Random Forest Classifier	111
4.2	Result of Precision, Recall, f1 Score for Decision Tree Classifier	113

### List of Figures

<b>Figure</b>	<b>Title</b>	<b>Page</b>
2.1.	Number of road traffic injuries and deaths in Nigeria	20
2.2.	Sequential Ensemble Approach	22
2.3	Sequential Ensemble Approach	29
2.4	Support Vector Regression	32
2.5	Architecture of ELM	33
2.6	Operation of K-NN Algorithm	34
2.7.	Bagging and Boosting Methods	41

2.8	Stacking Workflow	42
2.9	Framework for Artificial Neural Networks Classification	45
2.10	Framework For Artificial Neural Networks Classification	46
2.11	Two-Layered Feedforward Neural Network	47
2.12	Feed-Backward Neural Network	48
2.13	Random Forest Flow Chart	49
2.14	Random Forest training Flow Chart	50
2.15	Decision Tree Flow Chart	53
2.16	Decision Tree	53
3.1.	The Anaconda Interface	93
3.2	Proposed Prediction Model	93
3.3.	Flowchart of the Proposed Method	95
4.1	Snapshot of the Sample Dataset	103
4.2	Description Function on the Dataset	104
4.3	Count Plot for Accident Severity	105
4.4	Distribution of Accident Severity Over the Road Surface Type	106
4.5	Confusion Matrix for the Predictions	108
4.6	Performance Metrics Table	109

*Do Not Copy, Lead City University, Nigeria*

### **List of Acronyms**

WHO-	World Health Organization
ITS-	Intelligent Transportation Systems
AI-	Artificial Intelligence
ML-	Machine Learning
ERP-	Effective Radiated Power

RTA-	Road Traffic Accidents
PICO-	Problem Identification, Interventions, Comparisons, and Outcome
SVR-	Support Vector Regression
SVM-	Support Vector Machine
ELM-	Extreme Learning Machines
SLFM-	Single Hidden Feed-Forward Artificial Neural Network
KNN -	K-Nearest Neighbour
NB-	Naive Bayes
GNB-	Gaussian Naive Bayes
MLP-	Multilayer Perceptron
XGBoost-	Extreme Gradient Boosting
Adaboost-	Adaptive Boosting
LightGBM-	Light Gradient Boosting Machine
CATboost-	Category boosting
EFB-	Exclusive Feature Bundling
GOSS-	Gradient-Based Onside Sampling
SHAP-	SHapley Additive exPlanations

Do Not Copy, Lead City University, Nigeria