INFORMATICS INSTITUTE OF TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

SafeDrive: Location Based Road Traffic Accidents Prediction System

A dissertation by

Ms. S. R. B. N. Sanduni Bhagya

W1761916 / 2018434

Supervised by

Mr. Pumudu Fernando

Submitted in partial fulfilment of the requirements for the BSc (Hons) in Computer Science degree at the University of Westminster.

May 2023

ABSTRACT

Traffic accidents have been a serious problem since the introduction of the automobile. Despite this worry, there have been substantial societal and economic effects of road safety issues throughout the past century. Understanding these safety concerns and the necessity of mitigating them requires the identification of regions that reflect a higher vehicle crashes probability than an accepted standard.

In this project, a web application that predicts traffic accidents at specific points along a driver's route will be created. Using the date, time, origin, and destination of the trip, this program forecasts traffic accidents for upcoming dates. It helps users to plan their trips and view potential traffic hazards in real-time. Therefore, this research aims to design, develop, and test an application that will facilitate drivers to predict a road traffic accident at certain locations along the driver's route beforehand by getting details of the trip and the driver and save lives and properties by mitigating the traffic accident, with the use of ensemble learning technique.

Key Words: Traffic accidents Sri Lanka, Accident severity, Traffic hazard's locations, Machine learning, Ensemble learning

Subject Descriptions

- Computing methodologies >> Machine learning >> Learning paradigms >> Supervised learning >> Supervised learning by classification
- Computing methodologies >> Machine learning >> Machine learning approaches >> Classification and regression trees
- Computing methodologies >> Machine learning >> Machine learning algorithms >> Ensemble methods

DECLARATION

I attest that this dissertation and all related materials are my original work and have not been submitted previously, nor are currently being submitted, for any other academic program or degree. All sources used in this project have been appropriately cited and acknowledged.

Full name: S. R. B. N. Sanduni Bhagya
UOW No: W1761916
IIT Reg No: 2018434

Signature: 

Date: 10. 05. 2023

ACKNOWLEDGEMENT

I would like to express my gratitude to all the individuals who provided support during the completion of this research. Special thanks are due to my research supervisor, Mr. Pumudu Fernando, for his unwavering support, guidance, and motivation throughout the research period. Despite his busy schedule, he made time to ensure that I was putting my best efforts into the project. Without his expertise, this research would not have been feasible.

I also want to express my appreciation to the evaluators who dedicated their time to provide feedback and encouragement for the project. Their insights were invaluable in strengthening my research.

Finally, I would like to thank my family for their tremendous support throughout the last four years. Their unwavering encouragement kept me focused on my goals.

Table of Contents

List of Tables

List of Figures

Abbreviations

ML	Machine Learning
IoT	Internet of Things
GDP	Gross Domestic Product
RMSE	Root Mean Square Error
DFD	Data Flow Diagram
RTA	Road Traffic Accidents
GSM	Global System for Mobile communication
IDE	Integrated Development Environment
SDG	Sustainable Development Goals
MAE	Mean Absolute Error
GPS	Global Positioning System
UN	United Nations
MSE	Mean Squared Error
SSADM	Structured Systems Analysis and Design Method
API	Application Programming Interface
VS Code	Visual Studio Code
UI	User Interface
DBSCAN	Density Based Spatial Clustering of Applications with Noise

1. INTRODUCTION

1.1 Chapter Overview

In this chapter author has discussed about the problem domain of this project which is a major mortality and morbidity cause in all around the world. The problem statement and the aims and objectives of the project addressed in this introduction chapter to the study. The Novelty of the research and the research gap are included in this chapter along with the contribution to the body of knowledge and research challenges.

1.2 Problem Domain

1.2.1 Road Traffic Accidents Globally

Nearly 1.3 million individuals are killed in traffic-related incidents annually. Further 20-50 million individuals suffer non-fatal injuries, a large number of them becoming disabled as a consequence of the trauma (WHO, 2022). Considerable financial harm is caused by traffic accidents to individuals, their families, and entire nations. Around 3,700 individuals worldwide pass away every day as a result of accidents involving automobile, bus, motorbike, bicycle, truck, or pedestrian. Over 50% of all fatalities involve motorbikes, pedestrians, and cyclists. As per statistics, accident injuries are the ninth most common causes of death worldwide f along with all age categories and the top cause of death for kids and young people between the ages of 5 - 29. Nowadays, vehicle accidents cause more fatalities than AIDS and HIV put together. From 2015 to 2030, it is estimated that the world economy will lose $1.8 trillion due to both deadly and nonfatal collision casualties. That amounts to a 0.12% annual tax on the global GDP (gross domestic product). In low-income nations compared to high-income nations, the crash death rate is more than three times higher. From 2013 to 2016, there were no decreases in any low-income country's collision fatalities (CDC, 2020).

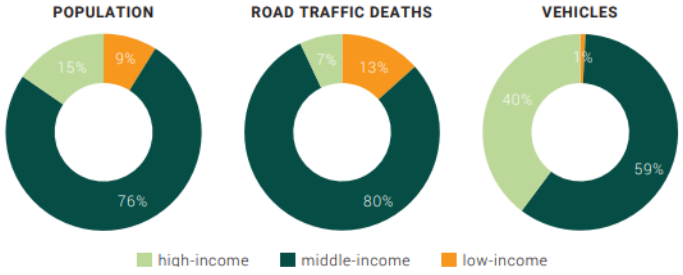


Figure 1: Proportion of population, road traffic deaths, and registered motor vehicles by country 2016

1.2.2 Road Traffic Accidents in Sri Lanka

There are 40,887 traffic accidents in Sri Lanka each year, resulting in an average of six fatalities each day. According to the records, there are approximately 740 pedestrian fatalities annually, around two on average each day (Kushan and Chandrasekara, 2020). According to estimates, road crash deaths in Sri Lanka are five times higher than those in the world's top performers and twice the rate of high-income countries. Sri Lanka has the most traffic deaths as one of its close neighbors in the region of South Asia (World Bank, 2020).

1.2.3 Taking Action for Prevention

There is still room for improvement in the prevention of road accidents and the injuries they cause. And many international leaders have begun to pay attention to the need to prevent traffic accidents. The Global Plan for the Decade of Action for Road Safety 2011-2020 was presented by the 2010 UN General Assembly. The decade's mission is to maintain lower traffic fatalities globally by 2020 by enhancing road infrastructure, vehicle design, traffic user behavior, post-crash response, and management capability for managing traffic safety (WHO, 2010). The Sustainable Development Goals (SDGs) are the upcoming significant benchmark for traffic control. To create a healthier and more sustainable future by the year 2030, UN representatives created 15 SDGs. The SDGs address The SDGs address numerous important global challenges which includes inequality, poverty, the environment, prosperity, peace, and justice. Road safety is a priority of two of the goals. SDG 3.6, the first goal, is to decrease worldwide fatalities and injuries from traffic accidents by 50% by 2020. The second goal is, by 2030 everyone is expected to have access to secure, reasonable, accessible, and sustainable transportation systems. According to the UN, enhancing the global transportation system is essential for enhancing road safety and reducing traffic accidents (United Nations, 2015). Many transportation issues are being solved using cutting-edge technologies. By utilizing cutting-edge applications of developing technology, it may be possible to address some of the issues in the domain of road safety that appear to be intractable.

1.3 Problem Definition

Since the invention of the vehicle, traffic accidents have been a major concern. In spite of this concern, there have been significant economic and societal consequences associated with traffic safety issues during the past century. The identification of areas that display higher accident incidence (severity and frequency) than an established norm is important in order to

understand these safety issues and the significance of reducing them. In order to improve road safety, government authorities must accurately identify hazardous areas with high accident rates. Currently, there is no method to identify these kinds of areas prior to starting the driver's journey, unless otherwise seeing boards on roads while driving.

Accidents don't just happen at random. A variety of circumstances, including the health of the drivers, the types of cars people are driving, their speeds, the structure of the road, and the weather, all have an impact on when occurrences happen. It would be possible to develop an accident predictor by better understanding the correlations between these parameters and traffic accidents by studying past accident records.

1.3.1 Problem Statement

It is really challenging to develop a ML model to predict traffic accidents that happens in Sri Lanka due to lack of resources and technology which is available in government organizations and also since automobiles accidents are a significant cause of death globally, there is no technological application which everyone could simply access, to predict a traffic accident based on locations with a history of higher traffic accidents.

1.4 Research Motivation

Road traffic accidents (RTAs) are a significant cause of death worldwide, resulting in 50 million injuries and 1.25 million fatalities per year (Abdulhafedh, 2017). Worldwide, transportation authorities have been working to put mechanisms in place to reduce RTAs. This is a challenging effort, nevertheless, as RTAs have not considerably decreased despite the implementation of numerous regulations and safety measures. The inability to accurately forecast where and when RTAs would occur is a contributing factor in this failure. Therefore, the author believes that a system which predicts an accident within the driver's route, for a future date and time will be important for mitigating a traffic accident. The results of this initiative will be advantageous to the general population of Sri Lanka and assist the traffic authority in developing methods to decrease RTAs.

1.5 Research Gap

After critically evaluating prior researchers' works, the author has identified that currently there is no system available to predict a traffic accident at specific places by considering driver's

details, vehicle details and weather condition before the driver starts his trip to a destination. Several works in predicting traffic accidents, especially in Sri Lanka, researchers have classified the severity of the accident. And also, most Sri Lankan researchers have used foreign datasets and not Sri Lankan data because of the lack of data available in Sri Lanka. Researchers from University of Moratuwa have done research on predicting accident severity using a USA dataset (Vijithasena and Herath, 2022). Researchers from South Eastern University of Sri Lanka found the primary contributing factors of the accident severity in Sri Lanka and determined the importance of the elements in model development (Dhananjaya and Alibuhtto, 2016). Researchers from University of Kelaniya have done research on factors that belong to accidents and classification of accident severity. This research approach is very limited because it is limited to Colombo-Katunayake expressway and also vehicles are also limited because motorcycles, three wheelers, tractors. etc. are not allowed to drive on expressways. In this project, the author believes in creating a system which predicts an accident during a driver's journey to a destination through a machine learning approach and by integrating google maps. The novel approach of this project is it will make a request to the Google Maps API with an origin and destination to get the most efficient driving route between them. The model which has been built using ensemble learning technique will then be fed this data, and a map showing the possibility that a road accident will occur along the user's route will be displayed. The author believes that a system that can predict occurrences of road accidents or predict accident prone zones can be able to save lives.

1.6 Contribution to The Body of Knowledge

1.6.1 Contribution to the problem domain

The problem domain that belongs to this project is the Road Traffic Accidents domain. Thousands of lives have been lost as a result of road accidents, which have also caused significant property damage. From this innovative approach, the author tries to present a user-friendly web application which can be used by drivers. By just entering a few details about the vehicle and the driver, users will be able to successfully register to the system. Users will be able to get to know predicted accidents during his/her driving route by entering the date, time of the trip and origin location to destination location. This way it will help to reduce the impact of a road accident and will be able to save millions of lives and properties.

Digitalization of data is very poor in Sri Lankan government organization. Dataset is a major component for a machine learning project. But unfortunately, there were no proper computerized accident dataset available in Sri Lanka. Therefore, the author had to create the dataset of her own by manually entering data with referring records books.

1.6.2 Contribution to the research domain

This research aims to develop and introduce ensemble learning approach which includes bagging, voting techniques to predict traffic collisions based on locations with a history of having a lot of traffic accidents. Algorithm benchmarking will be done to find the best performing algorithm for this prediction and finally, providing drivers a system that is accurate and well-trained. As a result, this research contributes to the discovery of patterns that accurately forecast when and where traffic accidents are likely to occur.

1.7 Research Challenges

Road traffic accident domain is a widely researched area in the world. But when it comes to Sri Lanka there are very less number of researches done in this area because of lack of resources available. According to the background study, the author has found that there are many problems and challenges to be overcome during this research process, and they are described below.

- Creating the dataset about traffic accidents is a challenging task in Sri Lanka. All the accident details are recorded in books as written records. Therefore, researchers must enter data one by one manually. And, this research is mainly focused on locations of accidents. The mechanism of keeping records of locations in Sri Lanka Police is an old map which has north coordinate and east coordinate points invented in 2002 and it is only usable for police stations. Therefore, the author must find a method to convert these points into latitude and longitude one by one.

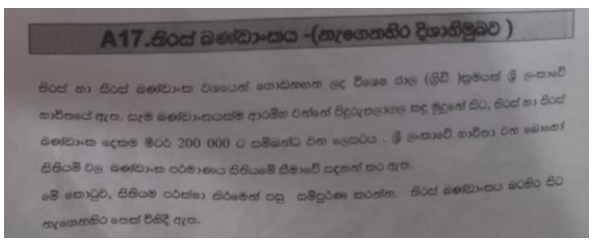


Figure 2: Police Map coordinates

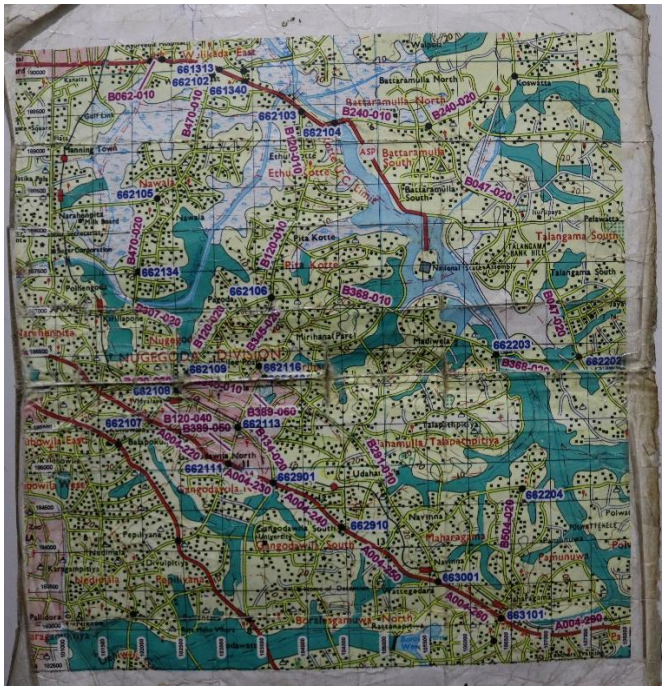


Figure 3: Police Map

- Including non-accident data is essential for the dataset. For that purpose, the author has to conduct a separate research on creating non- accident data.
- Evaluating the process and selecting the best technology to build and train these ML models.
- Finding most correlated features/hyper parameters from the dataset.
- Implementing the novel approach for the prediction model.

1.8 Research Questions

RQ1: How to identify probable accident hazard locations in real time?

RQ2: What methods can be used to identify most important features that contribute to the accident severity?

RQ3: What is the most accurate technique/algorithm that should be used to develop/train a machine learning model to find when and where accidents are likely to occur?

1.9 Aims and Objectives

1.9.1 Aims

This research aims to design, implement, and evaluate an application that will facilitate drivers to predict a road traffic accident at certain locations along the driver's route beforehand by getting details of the trip, driver and the vehicle and save lives and properties by mitigating the traffic accident, and mainly target Sri Lankan roads.

To further elaborate on the aim, this project will invent an interactive web application for drivers which predicts road traffic accidents at certain locations along drivers' route. This application predicts traffic accidents for future dates by entering date and time of the trip, origin location and destination location. To predict, details of the driver, details of the vehicle and weather will be taken into consideration. In order to develop the proof of concept, the required fields will be examined and researched, components will be implemented, and performance will be tested.

1.9.2 Research Objectives

The research objectives to complete the research successfully, are listed in the below table together with the learning outcomes and research questions for each.

Research Objectives	**Description**	**Learning Outcomes**	**Research Questions**
Literature Review (LR)	Carry out a comprehensive LR, **RO1**: To identify the existing solutions for road traffic accidents prevention and find their weaknesses. **RO2**: To research and analyze current road traffic accident prediction systems.	LO1, LO3, LO4	RQ1, RQ2

	RO3: To analyze methods used to identify dangerous areas with high accident rates. **RO4**: To discover research gaps in the road traffic accidents domain.		
Requirement Elicitation	Conduct a thorough user requirement gathering, **RO1**: To determine the essential attributes and factors in road traffic accidents. **RO2**: Gather requirements by conducting a survey focusing on the target audience. **R03**: Gather requirements related to the domain and technology by conducting formal meetings with domain experts and technical experts. **R04**: Analyze functional and non-functional requirements.	LO1, LO3, LO4, LO6	RQ1
Design	Design a system that can use the suggested methods to solve the problems that have been identified. **RO1**: To design the high level architecture of the proposed solution. **RO2**: To create the backend and frontend including UI of the application.	LO2, LO3, LO5, LO8	RQ1, RQ2, RQ3

	RO3: To design appropriate design diagrams belongs to the SSADM design paradigm, such as data flow diagrams.		
Implementation	Conduct a review on existing technologies to select most suitable ones for the system, **RO1:** To do data exploration and pre-processing with multiple techniques. **RO2**: To train multiple ML models and implement the optimized model and identify the most accurate model. **RO3**: To develop the proposed traffic accident prediction interactive application using the best approach.	LO2, LO5, LO7	RQ1, RQ2, RQ3
Evaluation	Use appropriate evaluation techniques to test and evaluate the system, **RO1**: Evaluate and test each component of implemented application to predict road traffic accidents. **RO2**: Make test plan to perform unit testing, integration, functional testing. **RO3**: Get feedback from evaluators from technical experts and domain experts and also get evaluated by end users.	LO6, LO8	RQ1

Table 1: Research Objectives

1.10 Chapter Summary

This chapter gave contextual information on the chosen matter and argued for its importance in this situation. The author has further clarified the project's aim and objectives in this part, along with the accompanying novelty of the research. To represent the contribution and underlying structure of the entire system, the suggested solution's outline is shown by a contribution to the body of knowledge. The software requirement specification will be covered in the next chapter.

2. LITERATURE REVIEW

2.2 Concept Map

Concept Map can be found in APPENDIX A.

2.3 Problem Domain

2.3.1 Risk factors of road traffic accidents

For children and young adults aged 5 to 29 years, tragic road accidents are the main cause of death and an estimated 1.3 million individuals each year pass away in automobile accidents (WHO, 2022). Identifying risk factors can help to prevent fatalities and limit serious injuries, ensuring an effective transportation system for everyone who uses the roads.

Figure 4: Risk factors

2.3.1.1 Speeding

The risk of an accident happening as well as the seriousness of the effects of an accident are strongly correlated with an increase in average speed. For instance, the chance of fatal collisions increases by 4% and the risk of a major crash increases by 3% with each 1% rise in average speed. The probability of fatality for pedestrians struck by moving cars increases quickly (by 4.5 times from 50 to 65 km/h). At 65 km/h, there is an 85% chance of death for passengers in car-to-car side accidents (WHO, 2022).

2.3.1.2 Driving while intoxicated or using other psychotropic substances

Driving while under the influence of alcohol or any other drug or psychoactive substance raises the possibility of an accident that causes fatalities or catastrophic injuries. When a driver is intoxicated while operating a vehicle, the probability of an accident rises gradually at low blood alcohol concentrations (BAC) and sharply when the BAC is less than 0.04 g/dl. Based on the psychoactive drug used, driving while under the influence of drugs increases the chance of being in an accident with another vehicle to varying degrees. For instance, persons who have taken amphetamines are nearly five times more likely to be involved in an accident that results in death than those who have not (WHO, 2022).

2.3.1.3 Lack of seat belts and motorcycle helmets

Proper use of helmets can reduce the likelihood of fatalities by 42% and the incidence of harm to the head by 69%. The probability of fatalities and severe injuries amongst back seat occupants is reduced by 25% whereas the probability of fatality and severe injuries amongst drivers and front passenger decreases by 45 to 50% with using a seatbelt (WHO, 2022).

2.3.1.4 Driving while distracted

Distractions of all kinds might affect a person's capacity to drive. Mobile phone distraction is an increasing issue for the safety of the road. Drivers who are on their phones while driving are around four times more probable to be in an accident than those who aren't. Using a phone when driving impairs reaction times, particularly those related to braking and traffic signal’s reaction times and makes it more challenging to maintain the proper lane and following distances. Texting greatly raises the danger of an accident, and hands-free phones cannot be considered less dangerous than handheld phone (WHO, 2022).

2.3.1.5 Improper road construction

The safety of roadways can be significantly impacted by the way they are built. All road safety concerns should ideally be taken into consideration while designing roads. This involves making fully aware that there are sufficient facilities for bicycles, motorcyclists, and pedestrians. To lessen the danger of injury for these road users, strategies including walking paths, bicycle paths, safety crossing places, and various traffic-calming initiatives may be essential.

2.3.1.6 Unsafe vehicles

Vehicle safety is essential for preventing collisions and lowering the risk of serious injuries. A variety of UN regulations on automobile security may possibly save a lot of lives if they were implemented in countries' manufacturing and production regulations. These consist of requiring automobile producers to make vehicles that adhere to front and side impact standards, have electronic stability control (in order to avoid oversteering), and have airbags and seat belts installed in every vehicle. Lacking these fundamental requirements, there is a significantly higher risk of traffic injuries for people inside and outside of vehicles (WHO, 2022).

2.3.1.7 Inadequate post-crash care

The severity of injuries increases as a result of delays in identifying and treating people injured in a traffic accident. The treatment of injuries following a collision is particularly time-sensitive; delays of just a few minutes could be the difference between life and demise. In order to improve post-accident treatment, it is necessary to guarantee immediate access to prehospital assistance and to enhance the standard of both prehospital and hospitalization, for example through specialized training programs (WHO, 2022).

2.3.1.8 Insufficient traffic laws

Traffic regulations cannot result in the anticipated decrease in road traffic fatalities and injuries linked to certain behaviors if they are not implemented. This includes legislation relating to drinking and driving, seat belt use, speed limits, helmet use. As a result, it is likely that traffic laws won't be followed and won't have much of an impact on behavior if they aren't enforced or are considered to be ignored.

Establishing, maintaining, and enforcing regulations concerning the aforementioned risk factors at the federal, municipal, and regional levels are all components of effective enforcement. The definition of appropriate punishments is also included (WHO, 2022).

2.3.2 Impact of Traffic Accidents

In developing countries, the effect of traffic accidents on their economic growth is less well recognized. A lot of investigations connected to accidents are constrained in these nations due to data shortages and inconsistent database updates. However, economic, and social impacts of traffic accidents are raising serious concerns throughout the globe because they are viewed as a major public health issue that affects populations everywhere. According to the World Health

Organization's 2017 statistics, there are 1.25 million road deaths annually. However, between 20-50 million individuals also have non-fatal injuries, the majority of which result in disabilities. A significant finding of the study is that, although having only 54% of the world's automobiles, low and middle-income nations account for 90% of all fatalities worldwide, making traffic accidents the fifth biggest cause of death by 2030 (Bhavan, 2019). Asia has been one among the geographic areas where deaths from traffic accidents have risen substantially over the last few decades (Garg and Hyder, 2006).

2.3.3 The Economic Impact of Traffic Accidents – Sri Lankan Context

There is a daily trend of a substantial increase in the number of vehicle crashes recorded in Sri Lanka. The effects of traffic accidents and their aftermath are viewed as a great tragedy in Sri Lanka following the conclusion of the civil war in 2009. Three individuals suffer injuries, and one person dies in a traffic accident every three hours, based on the Sri Lankan Police Authority statistics. According to professionals, the financial and social consequences of traffic accidents are far greater than people realize because of their hidden costs. As a result, Sri Lanka's healthcare system is finding it difficult to sustain its high life expectancy as well as low rates of newborn and maternal mortality, which are key performance metrics (Bhavan, 2019).

Sri Lanka's rising rate of traffic accidents has a big impact on economic expansion and raises spending. Given that Sri Lanka is already struggling to recover from adverse economic circumstances like a protracted trade deficit, foreign debt, a negative balance of payments, a depreciating currency, an economy that is import-driven, and a shortage of product and market diversity, collisions on the roads are seen as an additional expense and a potentially detrimental consequence of the current economic situation (Bhavan, 2019).

2.4 Existing Work

2.4.1 Road Traffic Accidents Prediction Systems in Sri Lanka

There are very few numbers of publications on road traffic accidents prediction systems that are designed for Sri Lanka. Researchers from Sabaragamuwa University of Sri Lanka have developed a method to identify the driver's fault in a car crash. Drivers' lack of attention is the key factor in accidents. Drivers are often held accountable, although this is not always the case. Terrible accidents will result from a combination of weather circumstances, road/vehicle conditions, and driver/pedestrian actions. Determining the driver's fault category will therefore

be extremely useful. The decision tree approach and neural network have emerged as the most effective techniques for spotting hidden patterns in dataset. This research investigated traffic accident data collected on the Colombo-Rathnapura-Batticaloa road (A4) in Sri Lanka for the years 2012 to 2016. Research showed that the Decision Tree and the DNN classifier are higher to other AI algorithms employed in accuracy and comparability, and that many crashes occur within the first 25 km of the route (Ariyathilake and Rathnayaka, 2019).

(Kushan and Chandrasekara, 2020) have conducted research to identify the causes of accidents on the Colombo-Katunayake Expressway as well as build suitable ML models to analyze the accidents' severity. Because of the time, traffic, and the convenience of driving, many in Sri Lankans now choose to take advantage of the expressway route instead of main roads for their extended travels. As a result, there are more automobiles on expressways. As a result, there has been a significant increase in expressway traffic accidents over the past few years. The highway traffic police branch of Kaduwela, Sri Lanka have provided the dataset for this research. In this research, accident severity was predicted using the probabilistic neural network (PNN) and the naive bayes classification technique and PNN has performed little more effectively than naïve bayes. Time, driver age, automobile type, reason, number of automobiles involved, reason for the accident, and weather are identified as the key accident-causing causes on Colombo-Katunayake expressway.

2.4.2 Statistical Analysis Approach of Road Traffic Accidents in Sri Lanka

(Senaviratna and Cooray, 2020) have conducted research on statistical analysis of the severity in Sri Lankan traffic accidents. For their analysis they have used accident data between 2014 and 2016. As independent variables, alcohol test, license validity, weather and 8 other variables, totally 11 factors have been considered. Chi-Square test has been performed in order to identify the association between independent variables and the target variable. As a result, road surface and weather have been identified as factors that aren’t significantly associated with the target variable. They have performed binary logistic regression technique and revealed that high risk factors for catastrophic accidents include aggressive or careless driving, driving on straight roads, driving during the day, and operating light automobiles.

After the end of the war, the northernmost section of this country, Jaffna district, is presently seeing significant economic development. Due to a huge rise in automobiles rate and the increasing road development, this happens. This has led to a sharp rise in the number of

vehicle crashes in Jaffna in recent years. Therefore (Renuraj, Varathan and Satkunananthan, 2015) conducted a research to identify factors causing road accidents in Jaffna. For this study, they have used 692 accident cases based in Jaffna. They have determined through analysis that the variables "Vehicle type" and "Age" have a greater impact on accident severity. Additionally, the detected components were used for fitting the logistic regression model, then the reduced logistic regression model's parameters were estimated using the maximum likelihood procedure method.

Researchers from Southeastern University of Sri Lanka have also conducted a research which is similar to above discussed researches. It is about identifying Sri Lankan traffic accidents causing factors between 2010 and 2014 and with the help of logistic regression. There are several important findings from this research. 2012 was the year with the most traffic collisions happened. And the most fatal crashes happened in 2010, whereas the most non-fatal crashes happened in 2012. Additionally, the majority of collisions happened on dry roads in clear weather. It has been determined that motorcycles are more likely to result in traffic accidents. Additionally, the majority of accidents involving recently registered automobiles (age less than 10 years) have been reported. large number of traffic collisions claimed as a result of careless or aggressive driving. drivers between the ages of 18 and 30 experience the most traffic collisions. It is firmly believed that younger drivers have the greatest impact on traffic collisions. Many accidents involving private automobiles and in rural areas happened (Dhananjaya and Alibuhtto, 2016).

(Kodithuwakku and Peiris, 2022) have conducted research which is like above research and exploratory factor analysis (EFA), conformity factor analysis (CFA), and binary standard regression analysis have been performed. Results of this research revealed that there are six factors that are related RTAs in Sri Lanka. Those are overtaking, deviating from lanes, speeding, drunk driving, mechanical problems with the automobile, and pedestrian carelessness. 85% of crashes involve speeding, diverting, and overtaking. The proportion of RTAs caused by alcohol intake above the legal limit is lower than the proportion caused by overtaking, diverting, and speeding.

2.4.3 Road Traffic Accidents Prediction Systems in other countries

(Labib et al., 2019) conducted research on predicting road traffic accidents in Bangladesh. During their research they have found out that in Bangladesh, over 32 individuals are killed

each day. They also identified key factors that clearly influence traffic accidents and offer some helpful recommendations in relation to this problem. in view of their demonstrated accuracy in this area, the authors of this study work applied four of the most prominent and powerful supervised learning algorithms to analyze traffic accidents. Those are KNN, AdaBoost, Decision Tree, Naïve Bayes. AdaBoost, whose accuracy was 80%, had the highest overall performance of all these approaches. And also, results of the analysis revealed that the collision's frequency is significantly higher during rush hour (from 6 to 18) than at other times and The majority of accidents happened at those times if the surface became dry, sealed, and of excellent quality.

(Sewwandi et al., 2020) have focused on predicting the accident severity and recommending and scheduling police officers to the accident. Also, they have implemented a game-based approach to build awareness of traffic accidents. For this, they have used a UK traffic accidents dataset which is available online. For the severity prediction section, they have used random forest algorithm and they categorized severity into three classes such as slight, serious, and fatal. Secondly, they have classified the reasons for the accident and classes are speeding, breaking traffic laws, operating a vehicle while intoxicated, overtaking, and turning. For this purpose, they have used AdaBoost algorithm. Accident frequency has been classified into 4 time zones with the help of KNN algorithm. And to cluster the accident vicinity they have used K-means clustering. The game type "Hangman Game" was selected for its ability to impart and evaluate understanding, improve individual capacity for problem-solving, and increase user engagement and this has been implemented using the Flutter language.

2. 5 Technological review

2.5.1 Data exploration and preprocessing

Data exploration and preprocessing is a crucial step in a machine learning project. The primary purpose of data exploration is to gain a better understanding of the data that you will be working with, while the purpose of data preprocessing is to prepare the data for use in a machine learning algorithm. To identify and **handle missing values** in the dataset, researchers have performed several techniques. (Labib et al., 2019) have done it by replacing missing values with the mean value of the column because those may have an impact on performance. (Ariyathilake and Rathnayaka, 2019) have done this task by eliminating of missing/null values. In addition to the above mentioned steps, (Vijithasena and Herath, 2022) have eliminated columns that have a

significant amount of null values, deleted the columns that include detailed data and eliminated any columns that don't provide accident severity facts.

Outlier detection is another crucial step in machine learning. It is the process of identifying and handling data points or observations that are significantly different from the majority of the other data points in a dataset. (Ariyathilake and Rathnayaka, 2019) have performed outlier detection using Interquartile range technique.
Scaling is a **feature engineering** technique and it is about transform the numerical features in a dataset to have similar scales or ranges, which can improve the performance and convergence of many machine learning algorithms. (Sewwandi et al., 2020) have used normalization technique to scale numerical and categorical data and (Vijithasena and Herath, 2022) have also used normalization to normalize the continuous features. And also, (Malik et al., 2021) have used Z-Score approach to normalize data as it scales numerical data to have a mean of zero and a standard deviation of one.

Feature selection technique is helpful to identify the most important features that contribute to the target variable. In order to achieve this (Labib et al., 2019) have applied several techniques. Those techniques are univariate feature selection, recursive feature elimination, and feature importance. (Sewwandi et al., 2020) has also performed this task using the feature importance technique. (Ariyathilake and Rathnayaka, 2019) and (Vijithasena and Herath, 2022) have performed correlation analysis, whereas (Kushan and Chandrasekara, 2020) have done this using pearson’s chi-squared parametric test.

2.5.2 Model Creation

Previous researches have applied several algorithms to predict a road traffic accident. By reviewing previous research, researchers can identify the machine learning techniques that have been used to address similar problems or research questions. This can help to determine which techniques are most relevant and appropriate for the research project. And also, it can provide insights into the strengths and limitations of different machine learning techniques. Therefore, it is important to analyze and review various machine learning techniques from existing systems.

2.5.2.1 Naïve Bayes

Naive Bayes classifier is a probabilistic algorithm that is widely used for classification problems in machine learning. It is based on the Bayes theorem, which is a rule that describes the probability of an event, based on prior knowledge of conditions that might be related to the event. Naive Bayes classifier assumes that the features in the input data are conditionally independent of each other, given the class label. This means that the probability of a particular feature occurring does not depend on the presence or absence of any other feature, given the class label. The Naive Bayes classifier works by calculating the probability of each class label for a given input data point, based on the joint probability distribution of the input features and the class label. The class label with the highest probability is then assigned to the data point. (Labib et al., 2019) has applied this algorithm to predict the severity of an accident and have achieved 80% as the accuracy for their research project, whereas (Kushan and Chandrasekara, 2020) have applied this for the same purpose and they were able to achieve 72.14% for their overall accuracy and (Malik et al., 2021) have achieved 74.71%.

2.5.2.2 Probabilistic Neural Network (PNN)

Probabilistic Neural Network (PNN) is a type of artificial neural network that uses a probabilistic approach to classify input data. PNN is based on the Bayesian decision theory and uses a non-parametric density estimation technique to classify input data. It is computationally efficient and requires minimal training data. It is also highly accurate and robust to noise and outliers in the input data. By applying this technique, (Kushan and Chandrasekara, 2020) have achieved the highest accuracy for their research which is 74.29%.

2.5.2.3 AdaBoost

Adaptive Boosting is a popular ensemble learning algorithm in machine learning that combines the predictions of multiple weak learners to create a strong classifier. The basic idea of AdaBoost is to iteratively train a series of weak classifiers on a weighted version of the training data. In each iteration, the algorithm assigns higher weights to the misclassified data points from the previous iteration. This allows the weak classifier to focus on the difficult-to-classify data points in the subsequent iteration. (Labib et al., 2019) have applied this technique to classify accident severity and achieved 80% of accuracy. But, (Sewwandi et al., 2020) have applied this technique to classify the reasons for the accident such as such as carelessness while driving, breaking traffic laws, driving while intoxicated, speeding, passing, and turning, whereas (Malik et al., 2021) has achieved 79.74% for the accuracy.

2.5.2.4 Decision Tree

It is a tree-structured model that makes a series of decisions based on the input features of the data, with each decision leading to a particular outcome. The input data is split into subsets based on the values of the features. The algorithm then calculates the information gain of each feature, which measures how well the feature separates the classes. The feature with the highest information gain is chosen as the root node of the decision tree. (Malik et al., 2021), (Ariyathilake and Rathnayaka, 2019), (Vijithasena and Herath, 2022) and (Labib et al., 2019) have used this technique to classify accident severity and they have achieved 87.88%, 96.15%, 96.2% and 72% respectively.

2.5.2.5 Random Forest

Random forest is a popular ensemble learning algorithm in machine learning that combines the predictions of multiple decision trees to create a more accurate and robust model. works by constructing a forest of decision trees, where each tree is trained on a subset of the data and a random subset of the input features. The decision trees are constructed using a modified version of the decision tree algorithm, such as CART or ID3. To predict the class or value of a new input instance, the algorithm passes the instance through each decision tree in the forest and takes the majority vote of the predictions. In (Sewwandi et al., 2020)'s project random forest algorithm classifies three levels of accident severity: minor, major, and fatal.(Vijithasena and Herath, 2022) have used RFC for the same purpose and achieved 97.2% and (Malik et al., 2021) have got 98.8%.

2.6 Evaluation and benchmarking

2.6.1 Evaluation Metrics

The performance of a model is explained by evaluation metrics. The ability of evaluation metrics to distinguish between different model results is a significant feature. According to the literature review, most of the similar researches in predicting traffic accidents have used these metrics. Therefore, below listed techniques are currently selected as evaluation matrices.

- Accuracy: Accuracy is the most basic metric for evaluating a classification model. It measures the proportion of correct predictions over the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision: Precision measures the proportion of true positive predictions over the total number of positive predictions, which represents the ability of the model to avoid false positives.

$$Precision = \frac{TP}{TP + FP}$$

- Recall: Recall measures the proportion of true positive predictions over the total number of actual positive instances, which represents the ability of the model to detect all positive instances.

$$Recall = \frac{TP}{TP + FN}$$

- F1 Score: F1 score is the harmonic mean of precision and recall, which provides a balanced measure of the two metrics.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

- Confusion Matrix: A confusion matrix is a table that summarizes the actual and predicted classes of the model, which provides a more detailed view of the performance of the model in different classes.

2.6.2 Benchmarking

The author will do algorithm benchmarking, which means compare all the algorithms that are found from literature review with the author's proposed algorithm and find the best performing algorithm. For that, above mentioned evaluation metrics will be used.

2.7 Chapter Overview

3. METHODOLOGIES

3.1 Chapter Overview

Research is as much about the issue it addresses and the conclusions it draws as it is about carrying out the research process. Therefore, the research methodology as well as the project management methodology and the development methodology relevant to the production of the prototype provide the framework for this study. Following that, the task is divided into deliverables and arranged on a schedule. In order to prevent any consequences from the difficulties this project faces, risk mitigation techniques are also identified.

3.2 Research Methodology

The methodologies have been derived from the Saunders Research Onion Model. The table below lists the methodologies deemed suitable for the project (Saunders, 2019).

	What:	**Why**:
Philosophy	Pragmatism	This research compares both qualitative and quantitative findings based on data used to construct the hypothesis.
Approach	Deductive	The purpose of this research is to test and prove the hypothesis. Given that the research chose to apply an already-existing theory to the RTA domain, a deductive approach was chosen to be used.
Strategy	Survey, Interviews, Experiments	These strategies were picked because they would work well together and produce enough relevant data for the research.
Choice	Mixed method	This method was chosen since the traffic accident prediction model contained quantitative and qualitative results that could be compared to the model prototype that will be built in this research. These results were obtained through interviews, survey papers, and other documents such as

		journal publications and conference papers.
Time horizon	Cross-sectional	During some point in the research's Requirement Analysis phase, data collection must be completed.

Table 2: Research Methodologies

3.3 Development Methodology

The prototype model was picked for research out of the development models. This approach uses an iterative design, testing, and evaluation process that allows for changes, enhancements, or the fulfillment of new requirements.

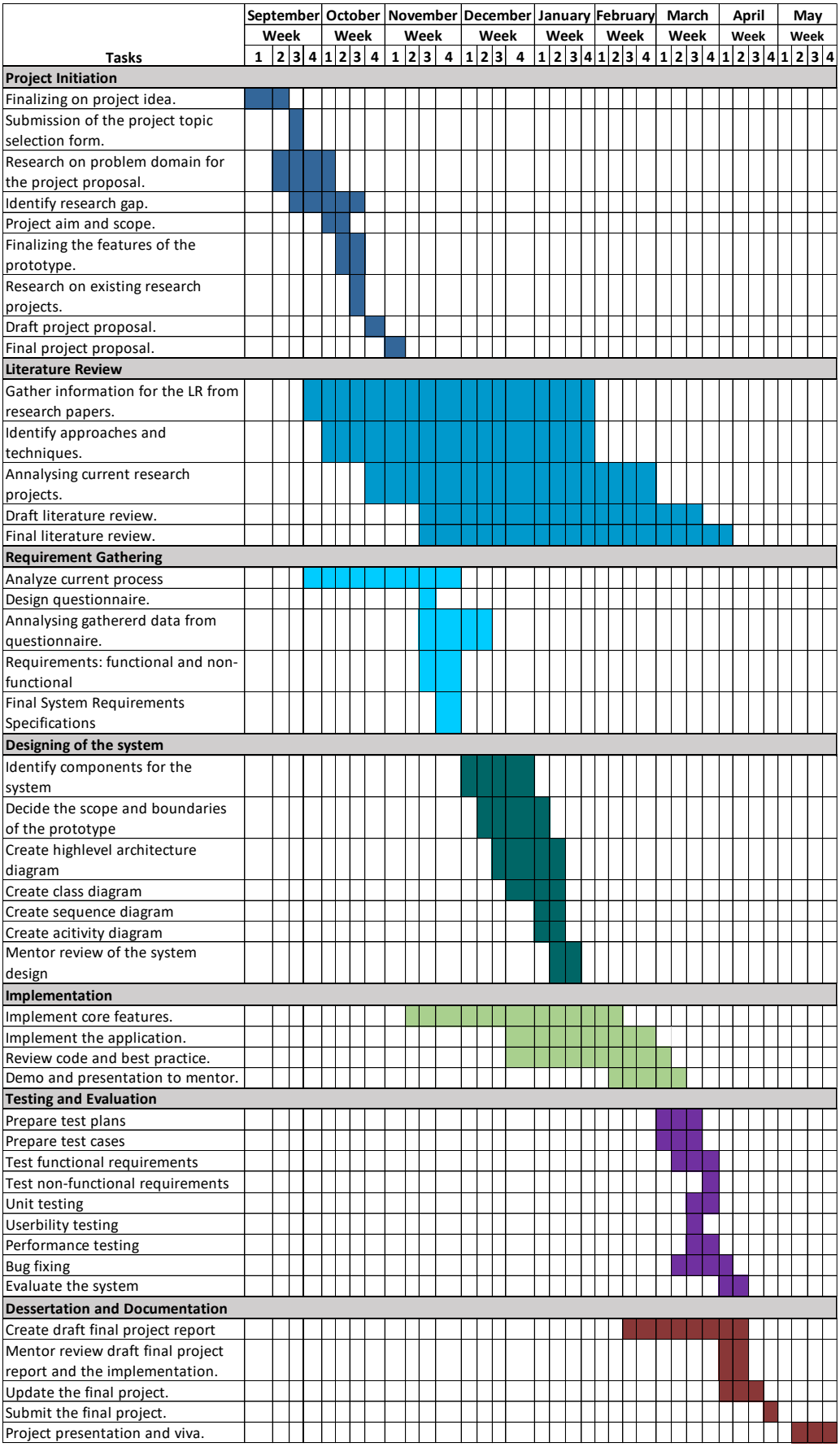
3.4 Project Management Methodology

Agile Prince2 was selected for project management because it enables management-focused planning, recursive scheduling, and flexible delivery while adjusting to risks. The agile prince2 technique was chosen by the author because it enables simultaneous management and delivery focus.

3.4.1 Schedule

3.4.1.1 Gantt chart

The Gantt chart below shows the authors' project plan, including the steps to take and their corresponding due dates.


Tasks
September
October
November
December
January
February
March
April
May
Week
1 2 3 4
Project Initiation
Finalizing on project idea.
Submission of the project topic selection form.
Research on problem domain for the project proposal.
Identify research gap.
Project aim and scope.
Finalizing the features of the prototype.
Research on existing research projects.
Draft project proposal.
Final project proposal.
Literature Review
Gather information for the LR from research papers.
Identify approaches and techniques.
Annalysing current research projects.
Draft literature review.
Final literature review.
Requirement Gathering
Analyze current process
Design questionnaire.
Annalysing gathererd data from questionnaire.
Requirements: functional and non-functional
Final System Requirements Specifications
Designing of the system
Identify components for the system
Decide the scope and boundaries of the prototype
Create highlevel architecture diagram
Create class diagram
Create sequence diagram
Create acitivity diagram
Mentor review of the system design
Implementation
Implement core features.
Implement the application.
Review code and best practice.
Demo and presentation to mentor.
Testing and Evaluation
Prepare test plans
Prepare test cases
Test functional requirements
Test non-functional requirements
Unit testing
Userbility testing
Performance testing
Bug fixing
Evaluate the system
Dessertation and Documentation
Create draft final project report
Mentor review draft final project report and the implementation.
Update the final project.
Submit the final project.
Project presentation and viva.

3.4.1.2 Deliverables and dates

Deliverable Component	Tentative Delivery Date
Project Proposal	3rd November 2022
Literature Review Document	13th November 2022
Software Requirement Specification	24th November 2022
Project Specifications Design and Prototype	2nd February 2023
Test and Evaluation Report	23rd March 2023
Final Thesis	10th May 2023

Table 3: Deliverables and Dates

3.5 Resource requirements

3.5.1 Software resources

- Operating System – A powerful operating system that can run the necessary ML related tools without experiencing any compatibility problems.
- Python or R – The primary programming language used to build the ML models.
- Scikit-learn, NumPy, Pandas
- Flask or Django – Backend of the application for the prototype.
- React or Angular – Frontend of the application for the prototype.
- VS Code or IntelliJ – IDE for the frontend and backend development.
- Anaconda Jupiter Notebook or Google Colab– Development environment to build ML models.
- Zotero or Mendeley - Tool for managing references and saving backup copies of research artifacts.
- MS Office Package,Figma,Canva or Google Docs/Sheets/Slides,Adobe XD - Tools for producing reports, figures, and documentation.

3.5.2 Hardware resources

- Core i7 8th generation or above - To supply the necessary powerful processing power.
- 8GB RAM or above - To be capable of handling huge data volumes.
- Disk space of 40 GB or above – To keep the implementation code and testing files.

3.5.3 Data requirements

Traffic accidents dataset was created by the author from scratch since all the data are stored as written records. For this, traffic accidents of Mirihana police division have been considered.

3.5.4 Skill requirements

- Machine learning knowledge.
- Application design and development skills.
- Research documentation skills.

3.6 Risk and mitigation

Risk	Probability of Occurrence	Magnitude of the loss	Mitigation Plan
Due to the limited time, gaining knowledge on ML will be challenging.	5	5	Through learning and practicing, the author will be able to mitigate this risk.
The system may require advanced computational power & processing speed, which limits the hardware resources available.	4	5	Cloud services such as Google Collab will be able to be used.
Code files and documents can be deleted permanently due to a technical issue in the machine.	5	4	Backup all the code files and documents in Cloud and version control systems.
Failure to finish all required deliverables in the time allotted.	4	4	Prioritize all the features accordingly and finish according to the priority list.
Unexpected problems such as illness/natural disasters may arise.	4	5	Work according to a schedule and try to make self-deadlines and achieve targets.

Table 4: Risk and Mitigation

3.7 Chapter Summary

A research technique was developed since the research process and the research are of equal value. To handle additional parts of this project, such as the creation of the prototype, the project management methodology and the development methodology were also defined in conjunction with it. The job was then divided into deliverables, and a Gantt Chart was used to plan and schedule the work. Furthermore, potential dangers were noted along with their significance and mitigation tactics.

4. SOFTWARE REQUIREMENTS SPECIFICATION

4.1 Chapter Overview

The focus of this chapter is on choosing the partners for the suggested arrangement and getting feedback from the chosen requirement elicitation methods. In this chapter, there will be coverage of several requirement elicitation techniques, and the most practical method or methods will be chosen with appropriate rationale. This chapter will also go into detail about how the user feedback was mapped to the previously established research goals, including both the functional and non-functional requirements of this structure.

4.2 Rich Picture

Fig. 5 shows the detailed picture of the suggested system. The motorist has the option to sign up for the program and enter personal information as well as vehicle information, as seen in the image above. Drivers can enter the date, time, origin, and destination locations after logging into the app. These inputs for the web application's traffic accident prediction system will be collected by four form fields and sent over an API in order to estimate the accident location. The machine learning algorithm will return the projected value, which the web application will display.

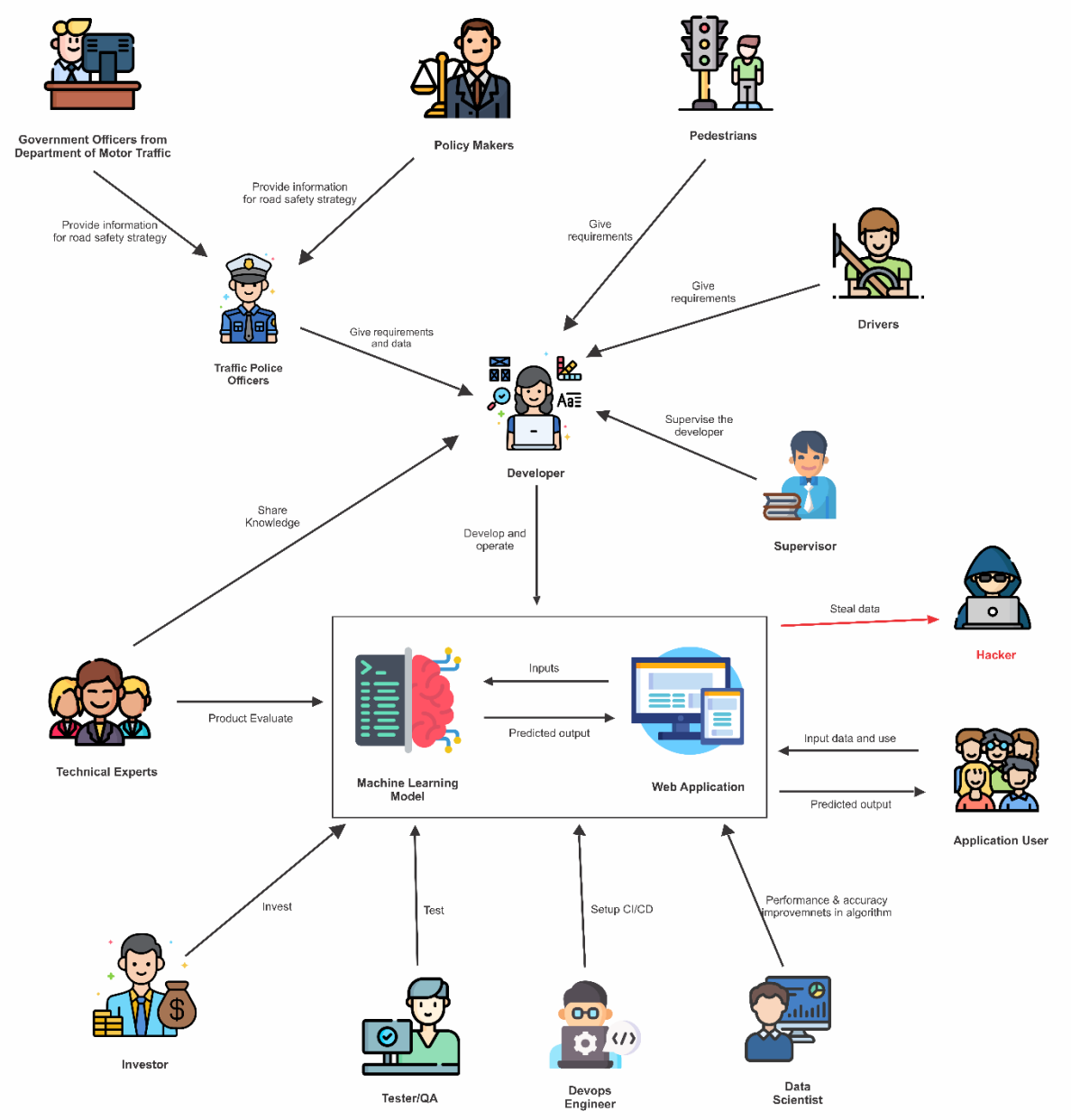


Figure 5: Rich Picture

4.3 Stakeholder Analysis

4.3.1 Stakeholder Onion Model

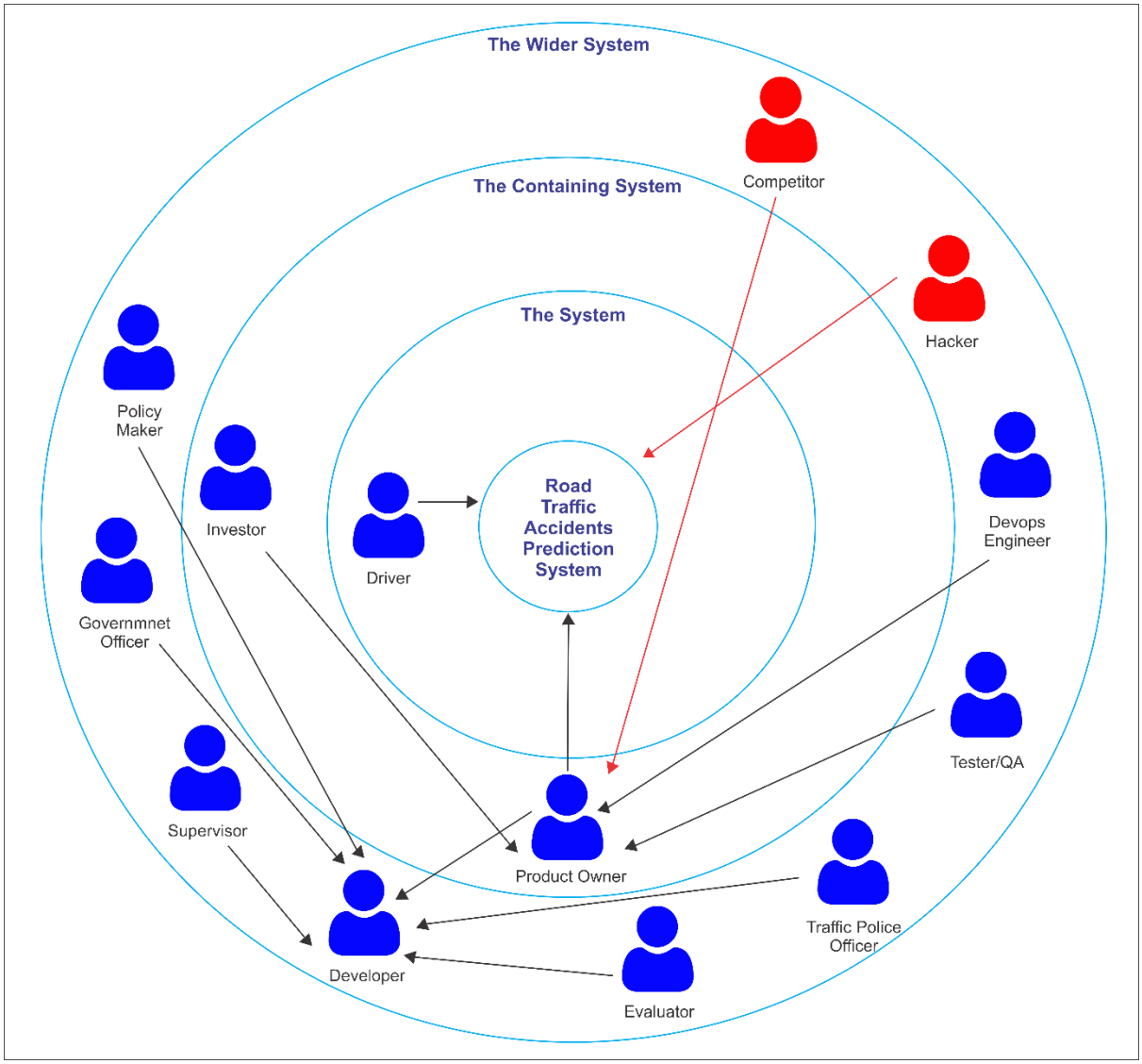


Figure 6: Stakeholder onion model

4.3.2 Stakeholder Viewpoints

According to their respective roles, each system stakeholder's point of view is presented in this section.

Stakeholder	Role	Benefits/ Role Description
Driver	Functional beneficiary	It makes drivers easier to find the areas where an accident may occur during his/her driving root and save lives

		and properties by preventing the accident.
Developer	Development staff and operational maintainer	Implement the research project concept as a product in order to prevent road traffic accidents.
Supervisor Evaluator	Quality regulator/ Advisor	Give feedback after evaluating the developer's developments. Assist the developer through the process of developing the product.
Traffic Police Officer Government Officer Policy Maker	Functional Beneficiary/ Operational benefactors	Due to the prediction, the cost of police investigation, policymaking, and hospital admission for medical treatment will be reduced. This will be considered as operational benefactors also because based on the behavior of the system, they will be getting benefit to their work operations.
Product Owner	Operational benefactors	Control the commercial operations and keep an eye on the developers. Accountable for coordinating with internal stakeholders to integrate or make the system publicly accessible.
Investor	Financial beneficiary	Funds to support prospective trials or advancements, the

		creation of new idea, and the realization of financial rewards.
Data Scientist	Expert/ Quality regulator	Enhances the performance of applied machine learning models and techniques.
DevOps Engineer	Product deployment & Maintenance	Guarantees that the prototype is functioning properly in the cloud and is not being blocked while providing services to users.
Tester/QA	Quality inspector	Examines and tests the system to verify if it is suitable for production use.
Competitor	Negative stakeholder	Creates features that compete with the established system in a comparable or more direct way.
Hacker		Infiltration attempts that aim to mess with the data and flow.

Table 5: Stakeholder viewpoints

4.4 Requirement Elicitation Methodologies

Technique 1 – Literature Review
As an initial step of this research project, the author has conducted a comprehensive evaluation of the literature. This offers a solid basic knowledge of the current systems and the methods used and also a thorough understanding of the traffic accidents domain will be gained. Nearly all research is published online, making it accessible to anybody at a reasonable cost.
Technique 2 - Survey

Using a survey, the necessary information as well as suggestions from possible users (drivers) of the planned system was gathered. The author has conducted a web-based online survey that is given to a specific population and consists of a list of questions related to a particular domain. Since it is an online survey, despite the geological obstacles, it can spread among a large population.
Technique 3 - Interview
Interviews are important to gather requirements and to obtain expert opinions on predicting accident prone locations to determine the most effective strategy to address the issue and solve the problem using the domain knowledge from domain experts such as traffic police officers and technical knowledge from technical experts. The author has interviewed one of the main parties which involves when traffic accident occurs. From this technique, able to gather in-depth facts from the subject-matter experts.

4.5 DTable 6: Requirement elicitation methodologiesiscussion of Findings

4.5.1 Findings

Literature Review

Finding	Citation
The severity of traffic accidents in the USA has been analyzed through this project. The main findings of this research show that mediate severity accidents occur more frequently than accidents with extremely low and high risks, and that these accidents are affected differently by a variety of factors including weather, the day of the week, and infrastructural factors.	(Vijithasena and Herath, 2022)
This study predicted whether the driver is at fault for the collision or not. The attributes of traffic accidents were empirically analyzed. That makes numerous trends about the time, day, and other accident-related elements at each chosen place.	(Ariyathilake and Rathnayaka, 2019)
From 2013 to 2019, 704 accident cases in Colombo Katunayake expressway were taken into consideration. Classified the accident severity and identified factors that cause accidents.	(Kushan and Chandrasekara, 2020)
Researchers predicted the accident severity. Recommending and scheduling police officers to the accident. Implemented a game-based approach to build the awareness of traffic accidents.	(Sewwandi et al., 2020)

Detects any incidents that may have occurred and automatically alerts the appropriate authority or person via wireless technologies. This system uses several sensors to feel the physical changes caused by accidents in order to identify casualties. Using a GPS sensor to identify the accident's location, this device then uses a GSM module to send the area's coordinates to a backup crew.	(Kapilan, Bandara and Dammalage, 2020)

Table 7: Findings

Interview

This section discusses the interview themes and conclusions for the requirement elicitation. The author has conducted several interviews with various stakeholders such as traffic police officers and as a technical expert senior software engineer.

Codes	**Themes**
Traffic accidents, Accident severity and locations	Research Problem and Gap
Dataset, Existing work	Requirements
Machine learning, Algorithms, Traffic Accidents Prediction	Methodology
Training, Testing, Evaluation metrics, High performance	Implementation and Evaluation
Web application, Google maps, Efficiency	Features in Prototype

Table 8: Themes and Codes

Themes	**Conclusion**
Research Problem and Gap	The technical expert and all the police officers were very happy with the research gap and the domain. According to a senior software engineer, the proposed solution to the selected research challenge is not available in Sri Lanka. All the police officers agreed that this research issue has a significant impact on the public since it helps to reduce traffic accidents.
Requirements	The technical expert mentioned that it is crucial to follow the principles and directions provided in the literature when creating the dataset and putting the solution into practice. According to police officers, there are no computerized data

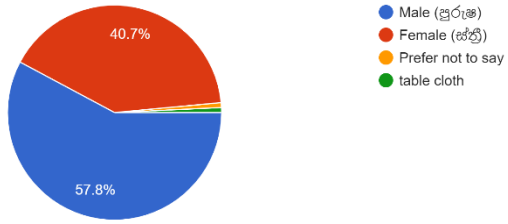
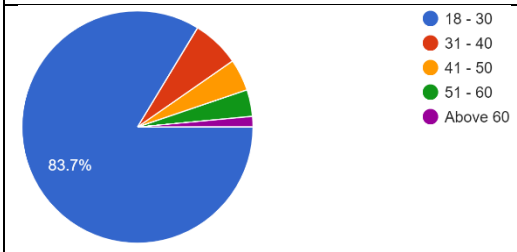
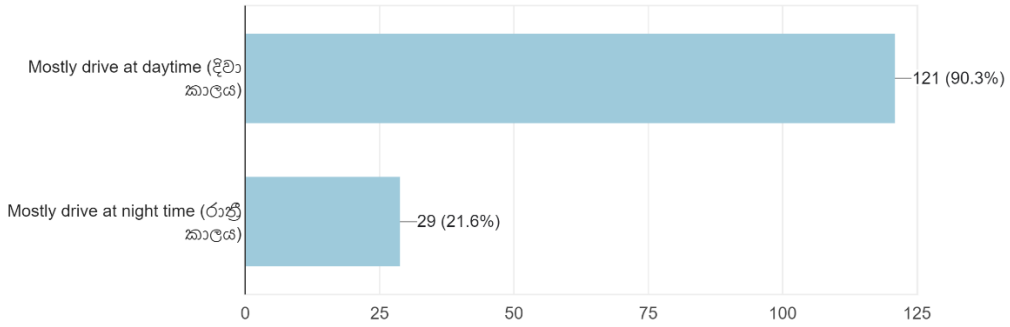
	available about traffic accidents. Therefore, author must create it from scratch with referring to written police records.
Methodology	According to the technical experts, classification approach would be more suitable if the objective is to precisely estimate a location as a accident hazard location in real time. In order to identify this, the first step would be the decision of whether an accident will occur or not according to the user data, vehicle and trip data. For that purpose, classification of the severity of each location that driving will pass during his journey would be useful. According to the state-of-the art, decision to apply ensemble learning would be beneficial to strong the contribution to the body of knowledge since it has been not applied to this domain before.
Implementation and Evaluation	According to the technical expert, with the data that has been cleaned and prepared, train the appropriate model and use an appropriate validation technique, such as cross-validation or holdout validation, to verify the training model. To enhance the model's performance, tweak its parameters.
Features in Prototype	Different ideas and suggestions were put forth for the prototype. One police officer asked to propose similar kind of solution for human-elephant conflict which protects human lives from elephant attacks while driving. However, the project scope does not cover that aspect.

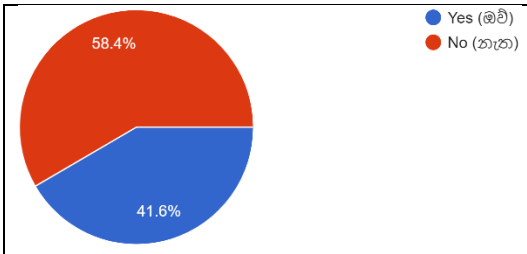
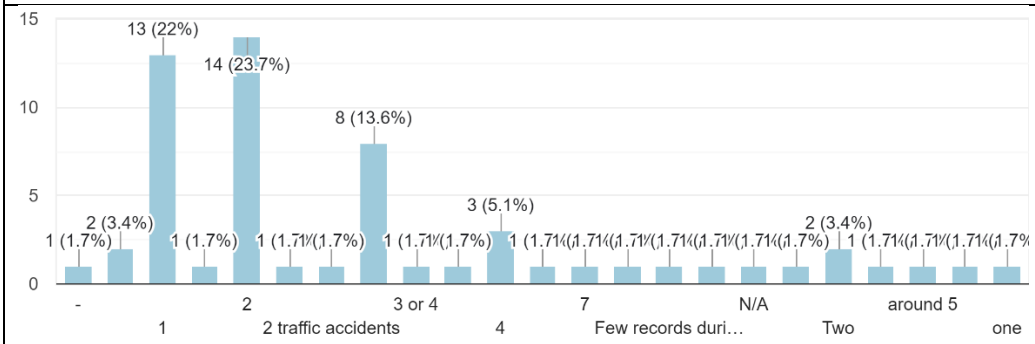
Table 9: Themes and Conclusions

Survey

This section analyses the survey results that were gained after conducting a survey with end users of the proposed system as a requirement elicitation method.

Question	**What is your gender?**
Aim of question	To determine which sex category has the most proportion of drivers and who will be benefited

	from implementing this solution in terms of gender.
Observations & Conclusion	
	In this survey, responses were mostly from men (57.8%), and most of the information was gathered from men who are driving. Therefore, it can be assumed that men will use the project's final result in such way.
Question	**Choose your age group?**
Aim of question	To determine the different age groups of drivers
Observations & Conclusion	
	The 18–30 age group received 83.7% of the votes, which indicates that this age group made up the majority of the drivers who participated in the survey.
Question	**At what time do you mostly drive?**
Aim of question	To determine the time of day when respondents often drive the most.
Observations & Conclusion	
	Most of the drivers are likely to use this product during the daytime since 121 (90.3%) of participants selected the option to mostly driving at day.
Question	**Have you ever met with a traffic accident while driving?**
Aim of question	To identify whether these participants have faced any traffic accident while driving
Observations & Conclusion	

	The majority of votes are for the No option, but there is only a slight difference between the number of votes for Yes and number of votes for No. Therefore, there is a considerable number of drivers who have faced a traffic accident while driving.
Question	**If yes, how many traffic accidents have you faced before?**
Aim of question	To get an idea about the number of traffic accidents participated drivers have faced.
Observations & Conclusion	
	
In here, particpants who have faced traffic accidents while driving has put a number indicating how many accidents they have faced. There were lots of answers for this questions, but majority of the participants which means 14 particpants have mentioned 2 and 13 particpants mentioned as 1, 8 participants have mentioned 3. Likewise there are lots of answers for this question. Therefore, lots of particpants have faced at least 1 accident while driving.	
Question	**From this application, users will be able to get to know predicted accidents during his/her driving route by entering the date, time of the trip and origin location to destination location.** **Do you think it's better to have a high accuracy web application to predict Road Traffic Accidents based on locations with a history of having a lot of traffic accidents?**

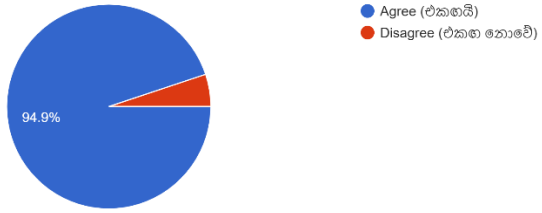
Aim of question	To learn the driver's thoughts on the concept of establishing a web application to forecast traffic accidents based on areas with a history of experiencing many accidents.
Observations & Conclusion	
	Most drivers 94.9% believe it is preferable to have a very accurate web application to forecast traffic accidents based on places that have a history of having a high number of accidents. As a result, there is a good likelihood that drivers will choose to purchase the project's final product.
Question	**If you agree, what kind of features would you like to have in that web application?**
Aim of question	To obtain feedback from drivers regarding the web application existing finalized features and allowing users to request additional features they want to see in the system.
Observations & Conclusion	
a function to see possible places that can cause accidents in my route, rate accidents by the type of it The time that most accidents occur To show areas where is high probability of accidentsv Better if you can add the types of accident/ the way which the accident is happening අනතුරු බහුල ස්ථානයන් පිලිබඳව දැනුවත් වීමට Possible reasons for accident	
Above row shows few answers from 137 responses. Most common answer for this question was to show the areas where accidents might happen with a high probability. Some participants have offered some suggestions. Those suggestions have been taken into consideration.	

Table 10: Survey Results

4.5.2 Summary of Findings

ID	Finding	LR	Interview	Survey
1	The most important and influential factors that influence the likelihood of an accident are time, the driver's age, the type of vehicle, the cause of the accident, the number of cars involved, and rainfall.	✓		
2	Relevant and important information for using advance technology in traffic accident prevention strategies, specifically in accident prevention approaches are important.	✓	✓	
3	This research issue has a significant impact on the public since it helps to reduce traffic accidents.	✓	✓	✓
4	A hybrid approach which has supervised and unsupervised can be used for the implementation.	✓	✓	
5	Predicting the exact location where the accident might happen would be more beneficial for end users.		✓	✓
6	In terms of the quantity, most number of accidents are faced by Males.		✓	✓
7	To enhance the model's performance, tweak its parameters, appropriate validation technique, such as cross-validation or holdout validation, to verify the training model can be used.		✓	
8	The proposed solution to the selected research challenge is not available in Sri Lanka and it has a great contribution to the research domain.	✓	✓	✓

Table 11: Summary of Findings

4.6 Context Diagram

The below diagram shows the level 0 data flow diagram which consists of the system and the end user of the system.

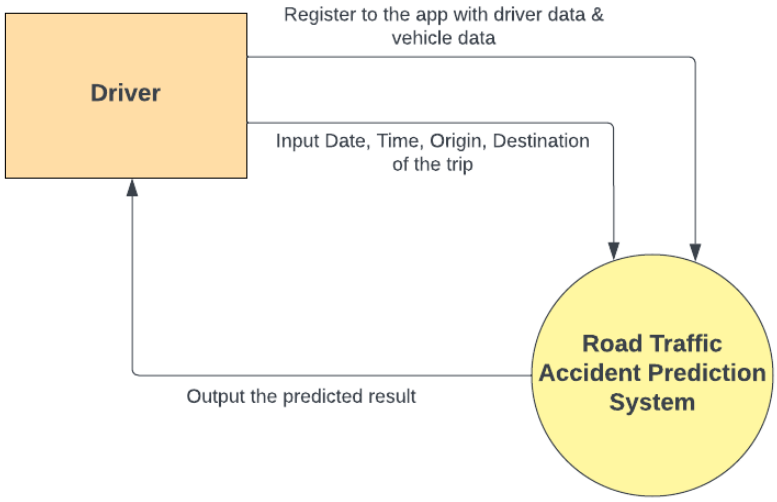


Figure 7: Context Diagram

4.7 Use Case Diagram

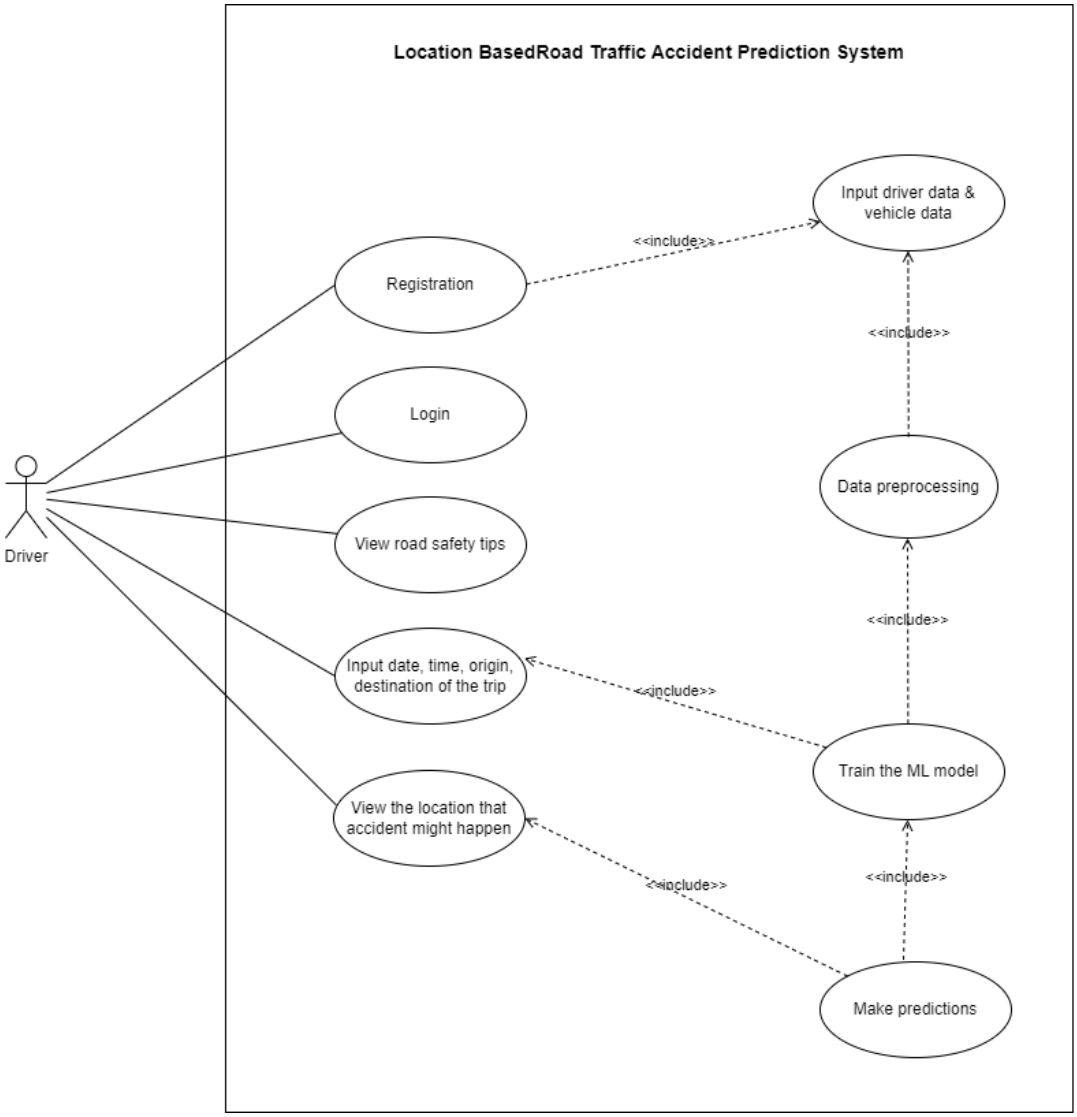


Figure 8: Use Case diagram

4.8 Use Case Descriptions

Use case name	Input date, time, origin, destination of the trip
Description	To make a prediction, it is necessary to gather information about the trip, including the date, time, origin, and destination locations.
Participating actors	Driver
Preconditions	The user needs to be signed in or registered with the app. Prior to entering the trip details, the user should have registered and provided information about themselves and about their vehicle.
Extended use cases	None
Included use cases	None
Main Flow	• The user enters the date, time, origin location and destination location of the trip. • Click on submit button.
Alternative Flows	None
Exceptional Flows	EF1- Cannot enter the information: Show an error. EF2- Internet connection is lost: Show an error.
Post Condition	Locations where accident might be happened will be displayed on the map.

Table 12: Use case description 1

Use case name	Make prediction
Description	From the pretrained model, when the relevant details are provided, the user should be able to see the potential accident-prone location.
Participating actors	Driver
Preconditions	The user needs to have provided correct details of the user, vehicle, and details about the trip.
Extended use cases	None
Included use cases	View the location that accident might happen.
Main Flow	• Provide the right information for the trained model. • Make predictions. • The use case comes to an end.

Alternative Flows	None
Exceptional Flows	EF1- Inability of the system to obtain the model's predictions: Show an error.
Post Condition	User will be able to view the predicted locations where accident might be happened will be displayed on the map.

Table 13: Use case description 2

4.9 Requirements

In this section all the requirements have been prioritized using the MoSCoW principle.

Priority Level	**Description**
M – Must have	The prototype's basic functional need is the criterion at this level, and it must be met. Implement the road traffic accident prediction model using appropriate techniques and train machine learning models and implement the optimized model that predicts high accuracy.
S – Should have	Important developments that are not absolutely required for the prototype. However, they will have a commensurate impact on the prototype. Producing an interactive application for drivers which predicts traffic accident locations.
C – Could have	Optional, non-essential desirable needs are crucial to the project's scope.
W – Will not have	Requirements that won't be part of the system and won't significantly affect it.

Table 14: Requirements

4.9.1 Functional Requirements

FR No.	**Requirement Description**	**Use case**	**Priority**
FR1	Driver registration to the app.	Registration	M

FR 2	Registration data such as date of birth, gender of the driver and vehicle data such as vehicle type, year of manufacture must be stored to cloud storage.	Input driver data & vehicle data	M
FR 3	After successfully login, driver must be able to go to the prediction page.	Input date, time, origin, destination of the trip, View the location that accident might happen.	M
FR 4	User must be able to input date, time, origin location and destination location of the trip.	Input date, time, origin	M
FR 5	Data must be sent to web API.	Input date, time, origin, Train the ML model	M
FR 6	Date, Time data must be sent to the weather API to forecast the weather.	Input date, time, origin, Train the ML model	M
FR 7	Data pre-processing must be done using raw inputs.	Train the ML model	M
FR 8	Processed data must be sent to the pretrained model.	Train the ML model, Make predictions	M
FR 9	Predicted output must be gone to the web app through web API.	Make predictions	M
FR 10	User must be able to view the location where an accident might be happened through google map in the prediction page.	View the location that accident might happen.	S
FR 11	The user needs to be able to see the road safety tips on the road safety page.	View road safety tips	C

Table 15: Functional Requirements

4.9.2 Non-Functional Requirements

NFR No.	Requirement	Requirement Description

NFR 1	Performance	The web application needs to be quick enough to calculate the location in a shorter amount of time.
NFR 2	Accuracy	The forecasts should be reasonably accurate so that the drivers can take the necessary actions to reduce their chance of being involved in a traffic accident.
NFR 3	Usability	The application must be simple, engaging, and user-friendly as appropriate.

Table 16: Non- functional requirements

4.10 Chapter Summary

Capturing the requirements from many stakeholders of this proposed structure was the primary goal of the software requirement specification. The beginning of this chapter covered requirement elicitation approaches, and appropriate elicitation methods were used to compile the project's needs. The response from traffic police officers and technical expert have then been carefully examined and demonstrates the necessity of the suggested system. An onion model was presented to demonstrate a thorough stakeholder study, that covers the various responsibilities and points of view. The functional and non-functional requirements for the product were then finalized and used to demonstrate the use case diagram.

5. SOCIAL, LEGAL, ETHICAL, AND PROFESSIONAL ISSUES

5.1 Chapter Overview

Consideration of research ramifications is just as crucial as making novel, ground-breaking discoveries. A study loses credibility if it does not meet the requirements in these areas. In light of the BCS Code of Conduct, this chapter details the considerations that were given to potential social, legal, ethical, and professional difficulties.

Social	**Legal**
The addition of questionnaire replies to the thesis did not reveal the respondents' personal opinions. There was only a recording of the responses' summary. Interviewers were informed that their comments will be included in the thesis and given the option to have their names and positions added.	The usage of any programming languages, tools, or frameworks was subject to an open-source license. The GPL3 license will be used to license all of the research's source code, including that for data gathering and preprocessing. The system was created in a way that ensures no personal information is needed to get the desired results.
Ethical	**Professional**
The project and their role in it were explained to the participants who answered the questionnaires. The thesis does not contain any fabrication, falsification, or plagiarism. The knowledge and facts that were gathered were correctly attributed and referenced, and all of the data and information provided are true.	There was no illicit or pirated software or equipment used to make the prototype. Throughout the procedure, only open-source or student licenses were applied. The project's results were accurate representations of nature that were recorded without any alterations. The entire study procedure was conducted in accordance with very high standards.

Figure 9: Social, Legal, Ethical and Professional Issues

5.3 Chapter Summary

According to the BCS Code of Conduct, this chapter looked at the social, legal, ethical, and professional aspects of this study effort. The social component included user and data privacy concerns in the next prototype, as well as consent from interview participants in requirements engineering and evaluation. The legal element touched on the appropriate usage of software

and data in accordance with their conditions of use. In this study, the ethical aspect was concentrated on moral behavior that would be seen. Last but not least, the professional element discussed some of the best practices that should be used to meet the high standards of professionalism anticipated from this activity.

6. DESIGN

6.1 Chapter Overview

The proposed system's design will be described in this chapter. Especially, this chapter will include a thorough explanation and defense of the selected design paradigm, the system's high-level representation. Additionally, the design of the suggested arrangement will be explained using necessary diagrams to support the selected design strategy and suggested system.

6.2 Design Goals

Design Goal	Description
Accuracy	In order to assist users in making decisions about their trip, the system must have the ability to forecast the most accurate location with the greatest degree of accuracy. If the system is unable to deliver precise output, it cannot be considered accurate.
Performance	In order for the system to be considered productive, it must create predictions in a less amount of time. Various datasets should be supported by the prototype as well, in order to deliver information in the anticipated time.
Interactivity	All users' needs to be capable of using the system and comprehend all of its components in order to maximize its potential. The working prototype includes a system guide from the viewpoint of the user to assure interactions.
Scalability	When responding to requests from the client side, the system should instantly complete all calculations and predictions for large datasets.

Table 17: Design Goals

6.3 High Level Design

6.3.1 Architecture Diagram

Because of its increased maintainability, reusability, and scalability, the proposed system was created in accordance with the 3-tier software architecture. The independence of each layer makes it simpler to change one layer without affecting others. A summary of each tier of each tier included in the proposed system architecture can be found in the following section.

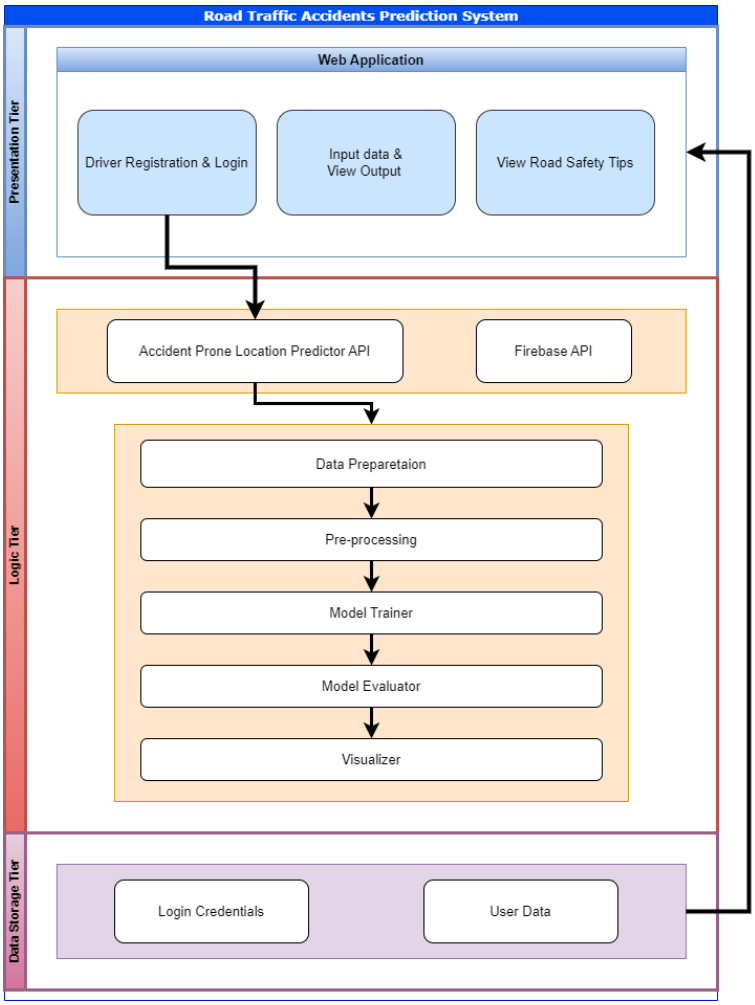


Figure 10: High Level Architecture

6.3.2 Discussion of tiers

Presentation Tier

Web application interface with a presentation tier allows users to carry out common tasks. The basic elements of this layer can broadly be divided into:

1. Driver login and registration - The authentication and registration of users are handled by this module.
2. Input data & View Output - This enables the users to enter the data, time, origin, destination locations of the trip. After submitting, the user will be able view the locations where accidents might happen on their driving route.

Logic Tier

Services Layer:

This tier holds the services that interact with the interface, business logic, and data tiers. Between the web application and the prediction module, the accident severity and location predictor API manages the input and output flow of data that user has inputted and forecasts. The Firebase API manages identification procedures for users who want system access, user registration, and permits data retrieval and storage in the Firebase.

Domain Logic Layer:
The main functionality of the application is supported by this layer, which is in charge of handling the application logic. The basic operations occurring within this layer are generally broken down into the following categories:

1. Preparation - This involves getting the dataset ready for the machine learning model to be trained (80% of the dataset as test data, 20% of the test data is split sequentially).
2. Preprocessing - This refers to the standard preprocessing procedures used before training the model such as handling null values, feature scaling, outlier detection.
3. The machine learning algorithm is trained in this step and the most basic step in this tier, the model trainer. The Scikit-learn libraries are used by this module.
4. Model Evaluator: Several machine learning models are tested and evaluated using this component. Evaluation metrics and error metrics were primarily used to check the evaluation of the trained algorithms.

Data Tier:
The layer mentioned above oversees the components used for data storage of the proposed system. Driver data, user credentials are all kept on this layer. Google Firebase was chosen to preserve the solution's persistency.

6.4 System Design

6.4.1 Choice of Design Paradigm

SSADM is considered to be the ideal fit for this project because of its execution nature, according to the author's analysis of several sorts of design paradigms. The suggested system's sequential key execution involves the user entering driver data, vehicle data, receiving forecasted results, and storing data. The aforementioned procedures take place in the order described, keeping the data unchanged and preventing simultaneous access from other sources.

As a result, the program displays a sequential data flow, blocking access to data by other processes while it is being used by another. Furthermore, since most of the significant and stable requirements were identified in the initial stage, this project is less poor possibility to go through frequent requirement changes. Therefore, SSADM is used as the design paradigm.

6.5 Design Diagrams

6.5.1 Data Flow Diagram – Level 1

The data flow across the key parts of the road traffic prediction app is shown in Fig. 6.2.

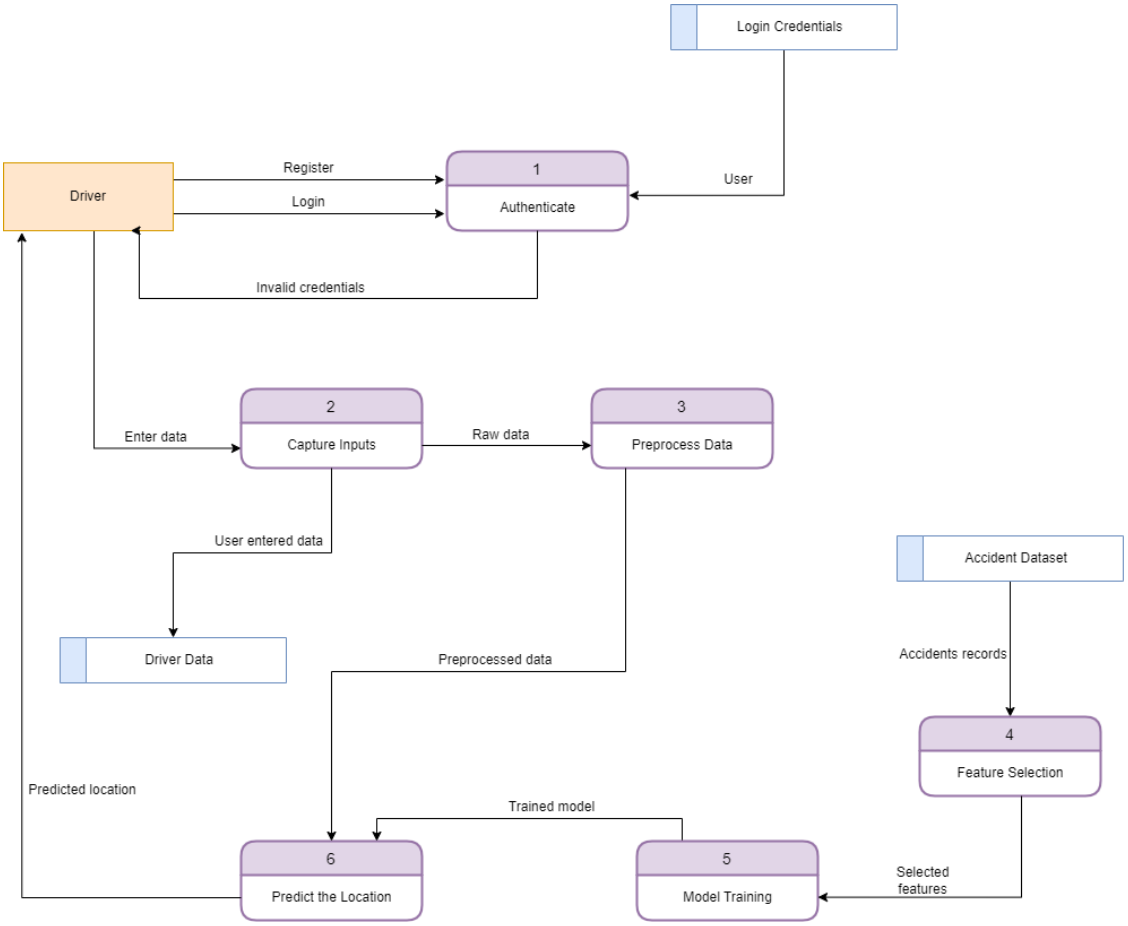


Figure 11: Data flow diagram 1

6.5.2 Data Flow Diagram – Level 2

In this diagram, several of the Level-1 DFD's identified primary components were dissected into their individual components and discussed.

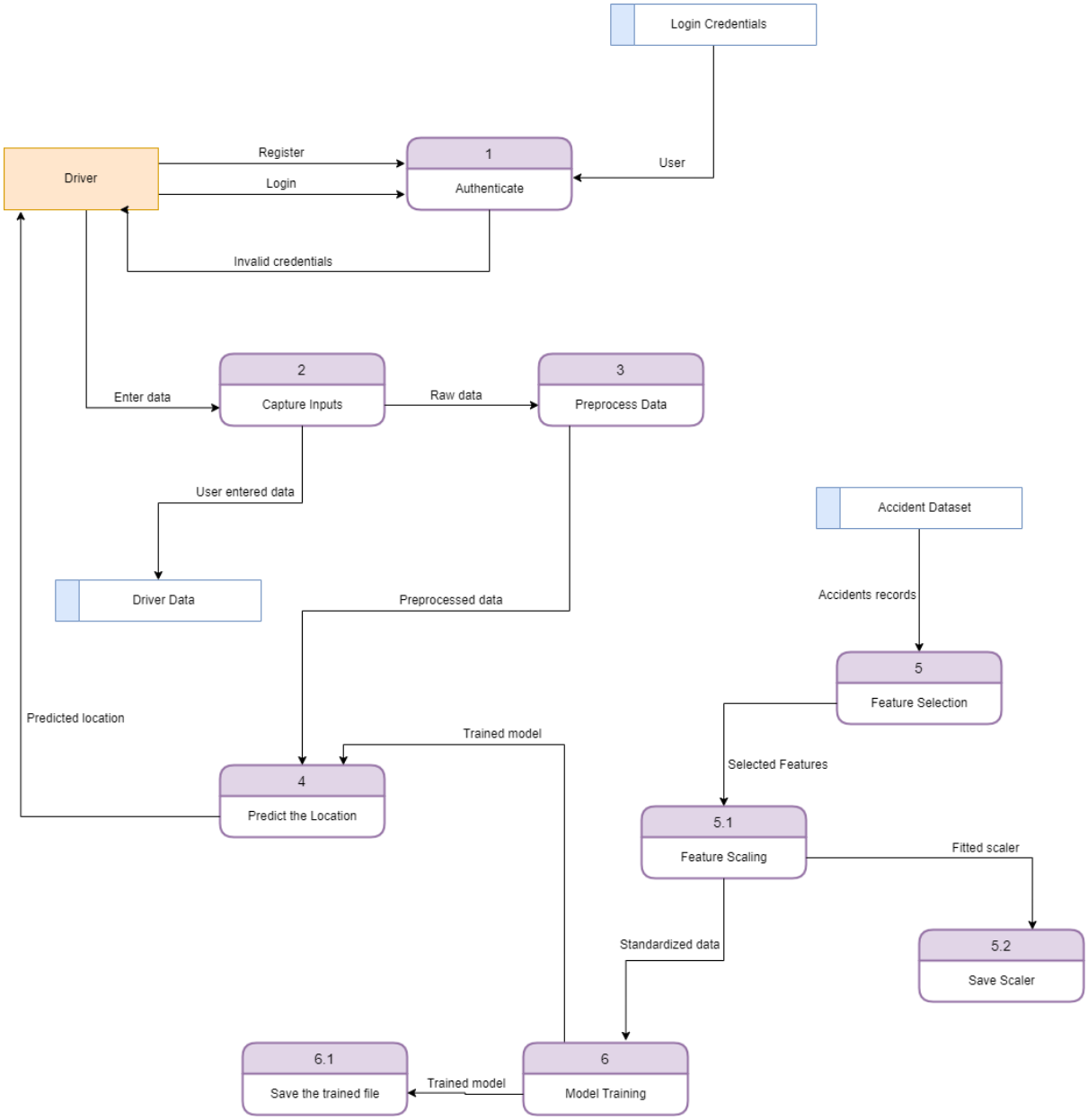


Figure 12: Data flow diagarm 2

6.5.3 System Process Flowchart

The workflow of the proposed methodology is displayed in the system process flow chart for the suggested system that is given below. It will clearly indicate the options, actions, data inputs, and results presented to normal users.

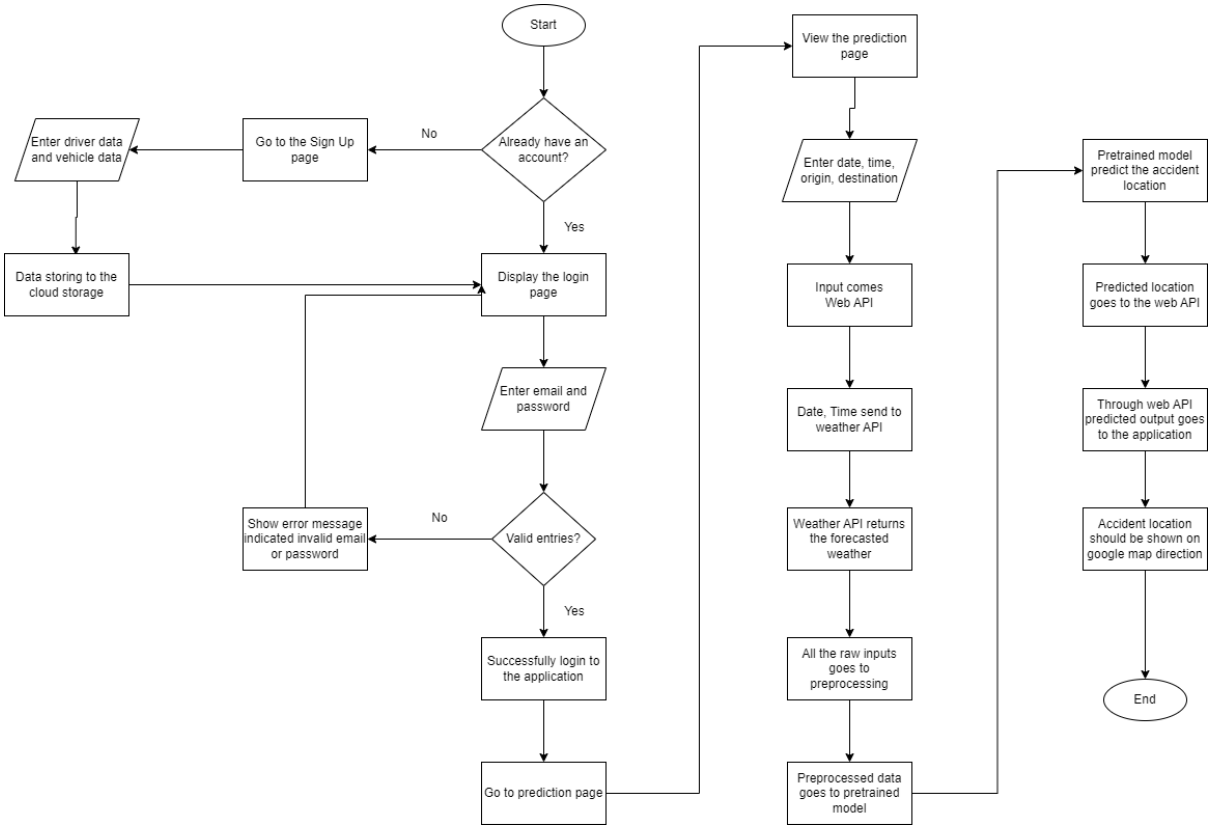


Figure 13: Flow Chart

6.5.4 User Interface

Before the front end of the proposed system, which will progress during the following period, user interface high fidelity wireframes are attached to the APPENDIX A.

6.6 Chapter Summary

The justification of the chosen design process opens this chapter. The most practical design methodology is SSADM because data flow occurs consecutively. The DFD and Context diagram was deemed the best design diagrams based on the design methodology. Following that, an example is used to expound on a general overview of the design architecture of the proposed project. In conclusion, each layer is examined in detail, outlining how it contributes to the overall result. The implementation methods used with the design approaches found in this phase will be thoroughly discussed in the following chapter.

7. IMPLEMENTATION

7.1 Chapter Overview

The proposed solution's design strategies and high-level architecture were covered in the previous chapters. The implementation chapter will deliver a general review of the libraries, tools, and technologies relevant to this project and will critically assess and rationalize the selected option(s) with the proper technical justifications. The system's key features will also be highlighted, corresponding with the supporting code sections.

7.2 Technology Selection

7.2.1 Technology Stack

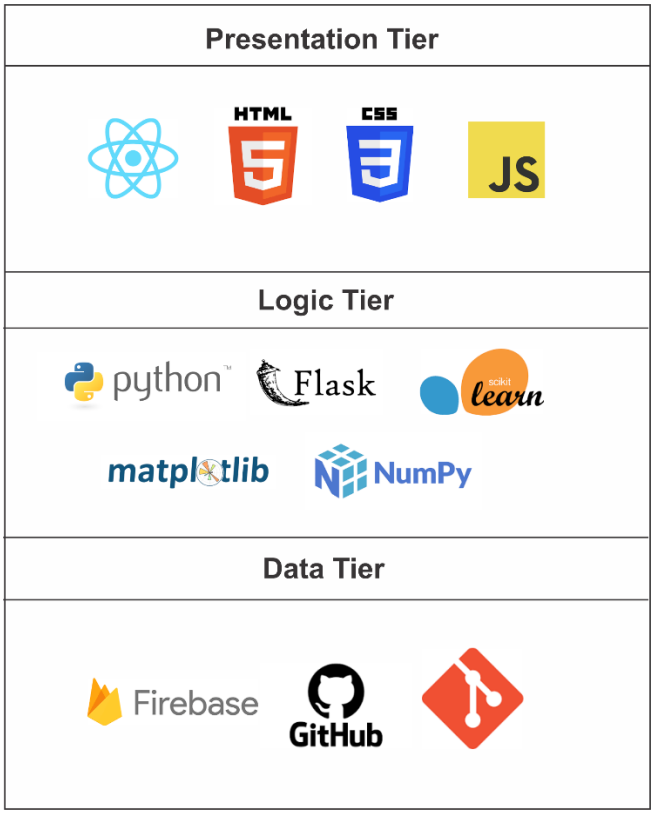


Figure 14: Technology stack

7.2.2 Dataset Selection

Considering primarily a machine learning project, it was essential to select the appropriate information sources to gather sufficient data for analysis and to produce the most effective predictions. The identified data for this research project were traffic accidents records of Mirihana police division of Sri Lanka.

In Sri Lanka, it is difficult to create a dataset on road accidents. The details of the accident are all documented in books as written documents. As a result, the researcher must manually enter each piece of data. Additionally, the primary focus of this research is the locations of accidents. An old map with north and east coordinate points that was created in 2002 is the Sri Lanka Police's system for keeping track of locations; it can only be used in police stations. As a result, the author must devise a mechanism to transform each of these points into a latitude and longitude. All the gathered data will be pre-processed before building the machine learning model.

7.2.3 Development Frameworks

Framework	Justification
React	Because of its flexible structure, react code is more customizable and easier to maintain and also it very useful when building rich user interfaces.
Flask	It provides a structure for programmers to adhere to when building the backend of a web application.

Table 18: Development Frameworks

7.2.4 Programming Languages

Programming Language	Justification
Python	Python is used for machine learning model development. Python is a general-purpose language that has been applied in a lot of machine learning initiatives. It has a huge variety of supporting libraries that make many machine learning coding easier.
JavaScript	JavaScript will be used as the react framework which is to build the user interface.

Table 19: Programming Languages

7.2.5 Libraries

Since Python was decided as the optimum programming language for creating models, training them, and performing other tasks, it was used together with its packages and modules. The following is a list of some of the chosen libraries.

Library	Justification

Scikit-learn	The most efficient and dependable machine learning library is Python's Scikit-learn (Sklearn) package. It provides a number of efficient methods for statistical modeling and machine learning, including dimensionality reduction, clustering, and classification, through a uniform Python interface.
NumPy	NumPy includes a number of beneficial characteristics for matrices in addition to being able to do mathematical and logical operations on arrays.
Pandas	It is helpful for handling real-world messy data and data cleaning, analyzing.
Matplotlib	When developing static, animated, and interactive visualizations, matplotlib is helpful.

Table 20: Libraries

7.2.6 IDEs

Google Colab	It is straightforward to design and test ML models, build and test data retrieval studies, and work across many devices due to the cloud development environment.
VS Code	It is exceptionally flexible because to its extensions and code snippets while also simplicity of use and effective for front-end programming and backend programming.

Table 21: IDEs

7.2.7 Summary of Technology Selection

The main technologies and equipment picked for the project are listed in the table below.

Component	**Tools**
Development frameworks	React, Flask
Programming languages	Python, JavaScript
Libraries	Scikit-learn, NumPy, Pandas, Matplotlib
IDEs	VS Code

Table 22: Summary of Technology Selection

7.3 Implementation of the Core Functionality

Creating the dataset

It is challenging to create a dataset on traffic accidents in Sri Lanka. All of the accident's circumstances have been written down in books. The researcher must manually enter each piece of data as a result. The locations of accidents are also the focus of this investigation. The Sri Lanka Police uses an outdated map with north and east coordinate points that was made in 2002 which can only be used in police stations. Therefore, the author has come up with a method to convert each of these locations into latitude and longitude using google earth pro.

Below figure is an example accident record which was given by Mirihana police division of Sri Lanka police.

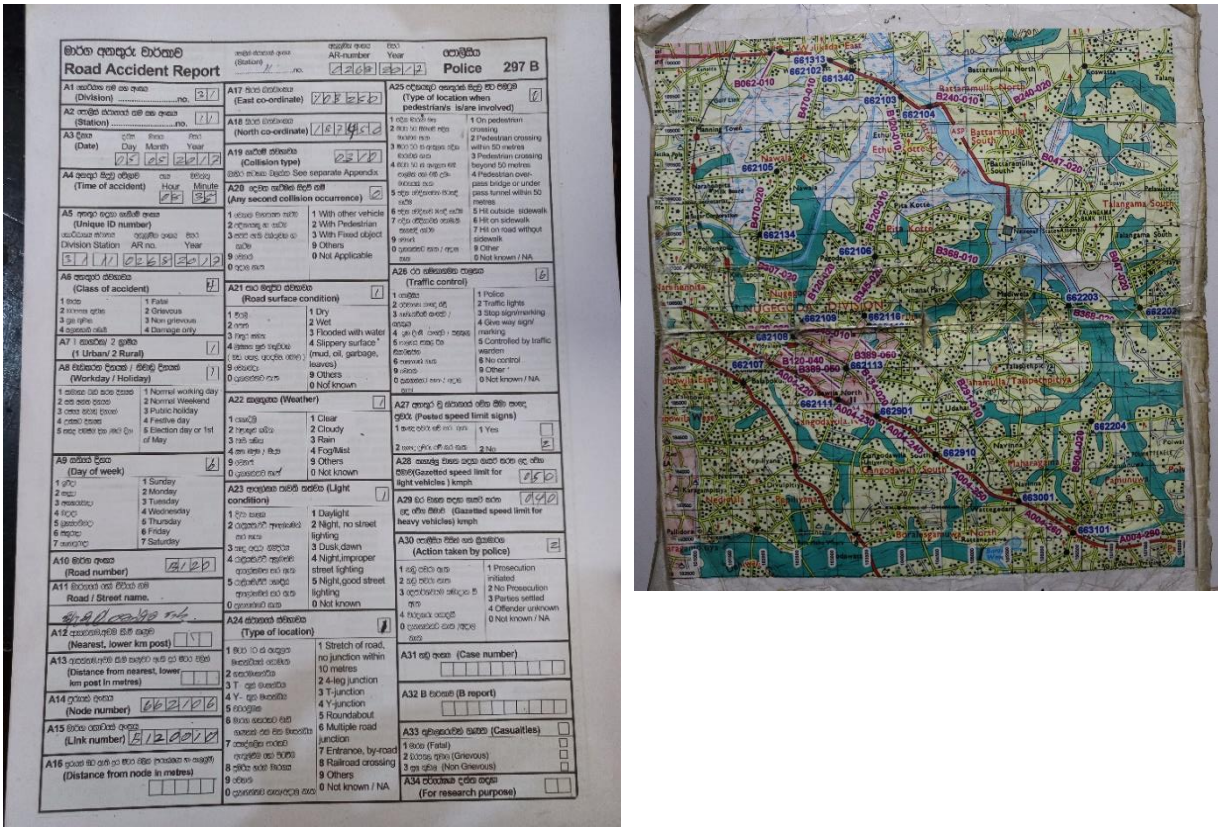


Figure 15: Police accident record

Below figure shows how the author has used google earth pro software to recognize actual latitude and longitude from east coordinate and north coordinate from police map.

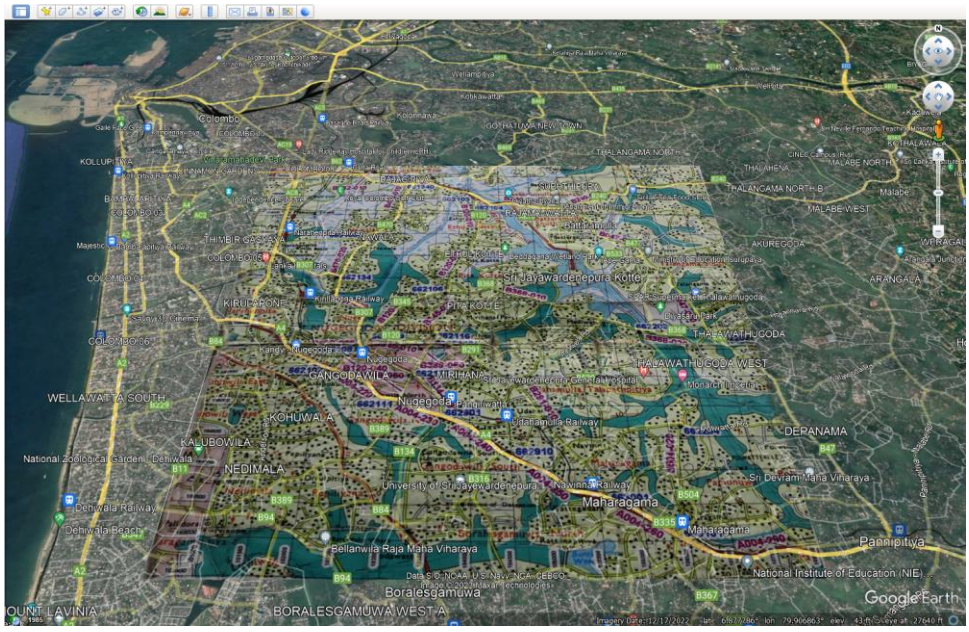


Figure 16: Google earth pro

Below figure shows the finalized dataset which was created by the author.

	A	B	C	D	E	F	G	H	I	J	K
1	Lat	Long	Day	Time	Weather	Light	Severity	Sex	Age	Vehicle_ty	Vehicle_age
2	6.871939	79.90149	1	19	1	2	3	1	26	5	2
3	6.871939	79.90149	1	19	1	2	3			1	2
4	6.884665	79.90713	7	8.5	1	1	4	1	26	1	6
5	6.884665	79.90713	7	8.5	1	1	4	1	42	1	2
6	6.877136	79.89298	5	12.55	1	1	4	1	39	2	13

Figure 17: Dataset

Implementation of the ML component

Classification algorithms are used for this purpose because they are designed to categorize data into different classes based on specific features or variables. In this case, the classification algorithm would take as input the various features such as user demographics, vehicle information, weather conditions, and the geographical location of each waypoint on the route. The algorithm would then use this information to predict the severity level of potential accidents that could occur at each waypoint location.

The classification algorithm would be trained on a dataset of historical accident data that includes information such as the severity level of each accident, the location of the accident, weather conditions, and other relevant factors. Once the algorithm is trained, it can be used to predict the severity level of potential accidents along the route based on the input features.

The author has created three machine learning models for the initial implementation. 1. Decision Tree classifier, 2. Bagging classifier and 3. Voting classifier. Ensemble learning techniques, such as voting and bagging, are powerful tools in machine learning that have been used to improve the accuracy and robustness of classification algorithms.

Several techniques have been performed under each stage from data pre-processing to evaluation of the model creation flow.

Data exploration and pre-processing:

Preprocessing is a crucial stage in the machine learning pipeline since it can significantly affect the model's performance. Several steps have been followed in the pre-processing stage such as handling missing values, feature scaling using normalization and standardization. By inputting the missing values or eliminating instances with missing values, preprocessing can assist in handling missing data. Outlier detection has been performed using the Interquartile range (IQR) technique. And also, as many machine learning algorithms are sensitive to the scale of the input features, preprocessing can help to bring features to a common scale using feature scaling and standard scaling has been used for this purpose. Feature selection technique is helpful to identify the most important features that contribute to the target variable and correlation analysis technique has been applied to achieve this.

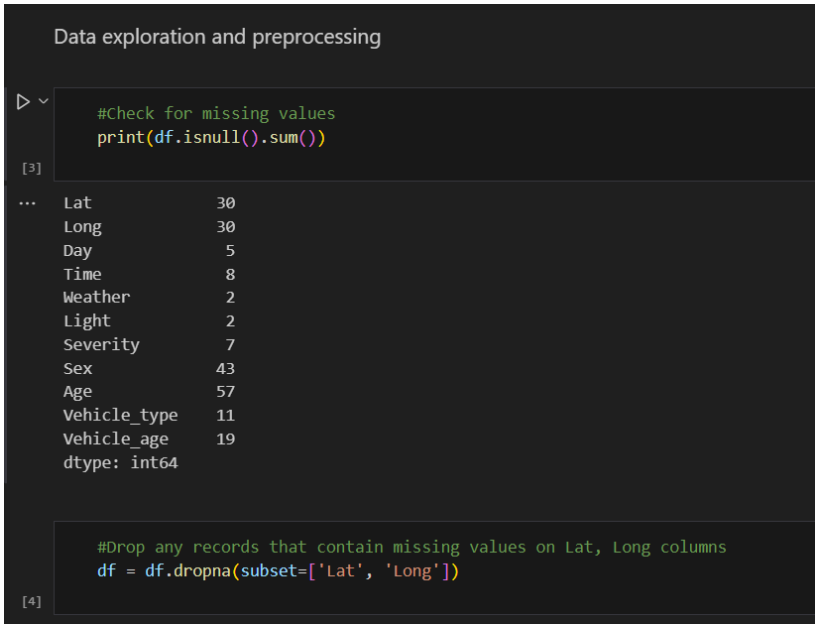


Figure 18: Pre-processing

```python
# Outlier detection using Interquartile range (IQR)
def detect_outliers(df, var):
    Q1 = df[var].quantile(0.25)
    Q3 = df[var].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers = df[(df[var] < lower_bound) | (df[var] > upper_bound)]
    return outliers

outliers_lat = detect_outliers(df, 'Lat')
outliers_lon = detect_outliers(df, 'Long')
outliers_time = detect_outliers(df, 'Time')
outliers_age = detect_outliers(df, 'Age')
outliers_vehicle_age = detect_outliers(df, 'Vehicle_age')
```

[5]

```python
# Replace missing values with mode of the column
df.fillna(df.mode().iloc[0], inplace=True)
```

[6]

Feature Engineering

```python
#Scaling
scaler = StandardScaler()

# def normalize(features):
#   scaled_data = preprocessing.normalize(features, axis=0)
#   print(scaled_data)
#   return scaled_data

def standardization(features):
  # scaler = StandardScaler()
  scaled_data = scaler.fit_transform(features);
  print(scaled_data)
  return scaled_data

# def minMaxScaler(features):
#   scaler = MinMaxScaler()
#   scaled_data = scaler.fit_transform(features);
#   print(scaled_data)
#   return scaled_data
```

[8]

Feature selection

```python
# Calculate the correlation matrix
corr_matrix = df.corr()

# Display the correlation matrix
print(corr_matrix)

# Visualize the correlation matrix using a heatmap
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

[9]

```
...                   Lat      Long       Day      Time   Weather     Light  \
Lat           1.000000 -0.080745  0.032522 -0.017865 -0.117386 -0.042416
Long         -0.080745  1.000000  0.122725  0.033529  0.108849  0.075577
Day           0.032522  0.122725  1.000000  0.003181  0.006047  0.008552
Time         -0.017865  0.033529  0.003181  1.000000 -0.027238  0.239278
Weather      -0.117386  0.108849  0.006047 -0.027238  1.000000  0.157372
Light        -0.042416  0.075577  0.008552  0.239278  0.157372  1.000000
Severity      0.247504 -0.154570 -0.112315  0.020955 -0.560406 -0.239921
Sex          -0.125424  0.064199  0.047005  0.042370  0.330607  0.146648
Age          -0.053904  0.081695  0.033445 -0.083718  0.159432 -0.060751
Vehicle_type -0.107785  0.081910  0.087140 -0.000127  0.198116  0.100597
Vehicle_age   0.012161  0.062457 -0.057282 -0.052257  0.158706  0.034315

                Severity       Sex       Age  Vehicle_type  Vehicle_age
Lat             0.247504 -0.125424 -0.053904     -0.107785     0.012161
Long           -0.154570  0.064199  0.081695      0.081910     0.062457
Day            -0.112315  0.047005  0.033445      0.087140    -0.057282
Time            0.020955  0.042370 -0.083718     -0.000127    -0.052257
Weather        -0.560406  0.330607  0.159432      0.198116     0.158706
Light          -0.239921  0.146648 -0.060751      0.100597     0.034315
Severity        1.000000 -0.482943 -0.230921     -0.354523    -0.211815
```

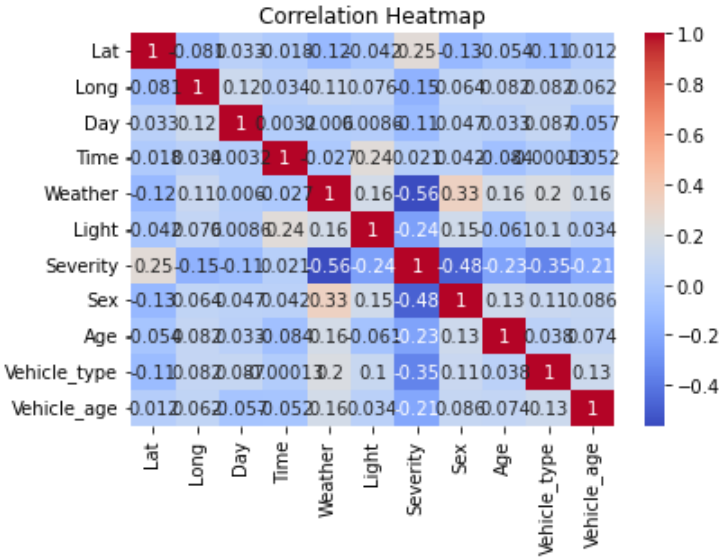


Figure 19: Correlation Analysis

Modeling and prediction:

Model 1

Below figures show how training and prediction has been done using decision tree algorithm. The model is responsible for training a decision tree classifier on the preprocessed and feature-selected dataset. The decision tree classifier is initialized using DecisionTreeClassifier() and then trained on the training data using dt.fit(). Once the model is trained, predictions are made on the test data using dt.predict().

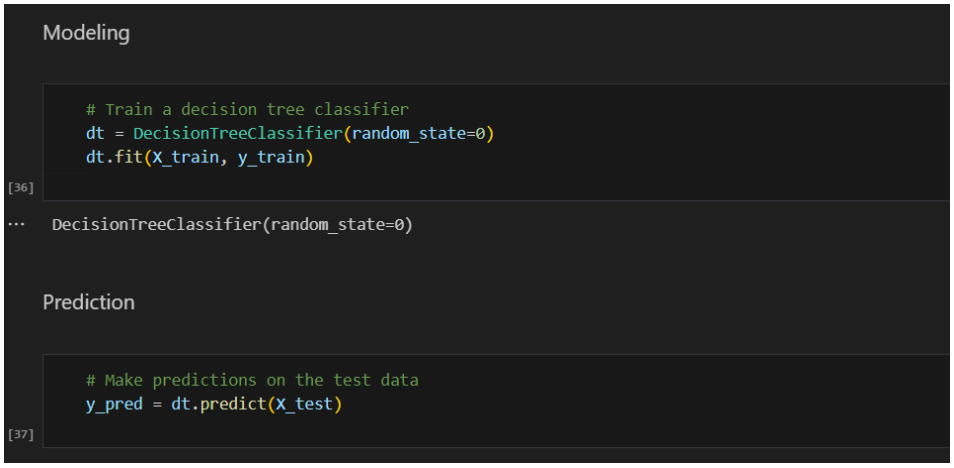


Figure 20: Decision tree classifier

Model 2

Below figure shows how training and prediction has been done using the bagging classifier. In the modeling part, two models are created using bagging technique, which is a type of ensemble

learning method. The first model is a decision tree classifier (dt) and the second model is a logistic regression classifier (lr). The bagging classifiers are created by passing the base estimator (dt or lr) and the number of estimators to the BaggingClassifier() function. The bag_dt and bag_lr models are then trained on the training set using the fit() function. Once the models are trained, predictions are made on the test set using predict() function for both models. Then, a majority vote is taken among the predictions from both models using mode() function, to make the final prediction.

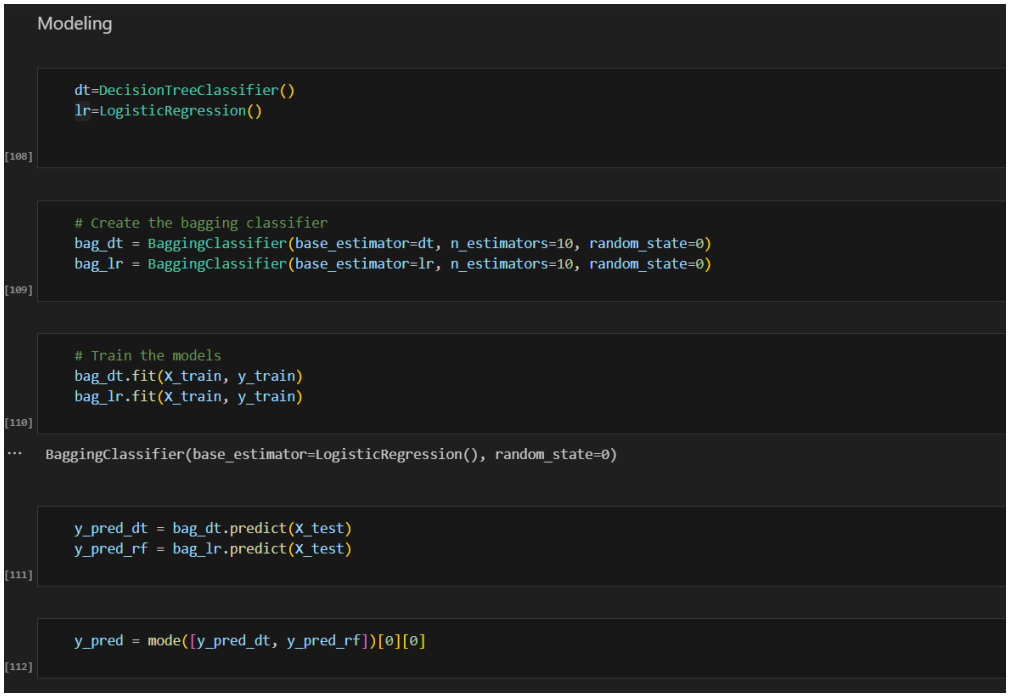


Figure 21: Bagging classifier

Model 3

Below figures show how training has been done using the voting classifier. In the modeling part, two classifiers, AdaBoostClassifier and RandomForestClassifier, are instantiated. Then, a voting classifier is created using these two classifiers as estimators and 'soft' voting method. The voting classifier is then fitted to the training data. The predict method is used to make predictions on the test data, and the results are stored in the y_pred variable.

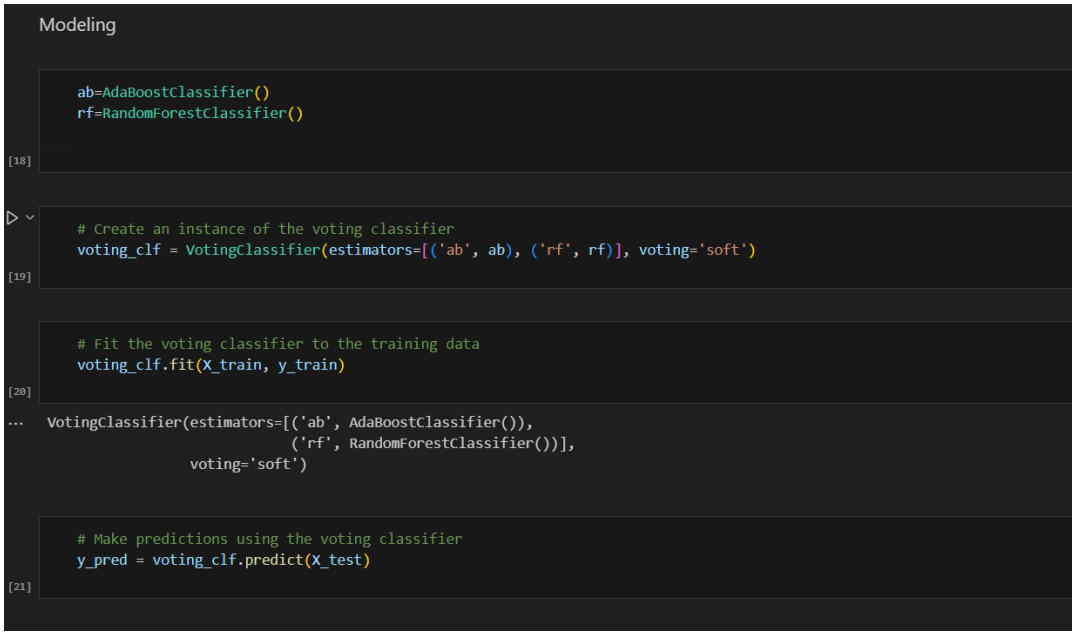


Figure 22: Voting Classifier

7.4 Web Application

The flow of the prototype is, when a new entry is made with the origin, destination and date and time, the Google Directions Service is used to find the optimal route from the start and end points. Along with the waypoints returned by the service, and the weather forecast returned by the open weather API, the other information (driver gender, age etc.) which is crucial for the accident prone location prediction is fed to the trained machine learning model through the Flask API. Subsequently, the array of waypoints were extracted and fed to the scaler, and then were iteratively used to get the final prediction of the severity status for that specific standardized waypoint input. Finally, based on the severity of the predicted output, markers were drawn and displayed on the map along the suggested route.

This code snippets shows the usage of Google Maps API to display directions and markers on a map. It also makes use of the OpenWeatherMap API to get the weather forecast for each waypoint and a Python Flask API to make predictions on severity of the route.

```
const directionsCallback = (result, status) => {
    console.log(result);
    wayPoints = [];
    if (result !== null && count.current < 1) {
      count.current += 1;
      console.log("****" + status + " " + Object.keys(result).includes("routes"))
      if (status === "OK" && Object.keys(result).includes("routes")) {
        console.log("****")
        if (
          result.routes.length > 0 &&
          result.routes[0]["overview_path"] &&
          result.routes[0]["overview_path"].length > 0
        ) {

          result.routes[0]["overview_path"].forEach((element) => {
            const json = {
              Lat: element.lat(),
              Long: element.lng(),
              Day: queryDt().day,
              Time: queryDt().time,
              Weather: 1,
              Light: 1,
              Sex: global.userData.sex,
              Age: (new Date().getFullYear() - new Date(global.userData.dateOfBirth.toDate()).getFullYear()),
              Vehicle_type: global.userData.vehicleType,
              Vehicle_age: (new Date().getFullYear() - global.userData.yearOfManufacture),
            };
            console.log(json)
            wayPoints.push(json);
          });

        }
        setDirections(result);
      }
    } else {
      count.current = 0;
    }

    if(wayPoints.length > 0){
      predWeather(wayPoints)
    }
  };
```

```
const predWeather = (wayPoints) => {
    const url =`http://api.openweathermap.org/data/2.5/forecast?lat=${wayPoints[0].Lat}&lon=${wayPoints[0].Long}&appid=b331c99907035539ce3b056166ac8f72`
    console.log(url)
    fetch(url)
    .then(res => res.json())
    .then((result) => {
      console.log(result)
      if(result != null && Object.keys(result).includes("list") && result.list.length > 0){
        var ms = queryDt().ms
        var currentWeather = null
        var weather = null
        console.log(ms)
        for(var i =1; i < result.list.length; i++){
          console.log(result.list[i-1].dt)
          if(ms >= parseInt(result.list[i-1].dt+"000") && ms <= parseInt(result.list[i].dt+"000")){
            if(ms - parseInt(result.list[i-1].dt+"000") < parseInt(result.list[i].dt - ms+"000")){
              currentWeather = result.list[i-1].weather[0].main
              weather = result.list[i-1]
              break
            } else if(ms - parseInt(result.list[i-1].dt+"000") > parseInt(result.list[i].dt - ms+"000")){
              currentWeather = result.list[i].weather[0].main
              weather = result.list[i-1]
              break
            } else {
              currentWeather = result.list[i-1].weather[0].main
              weather = result.list[i-1]
              break
            }
          }
        }

        console.log(currentWeather)
        console.log(weather)

        if(currentWeather != null){
          wayPoints.map((e) => {e.Weather = (currentWeather.toLowerCase() === "rain" ? 2 : 1)})
          console.log(wayPoints[0])
          getPreditions(wayPoints)
        }
      }
    })
    .catch(err => {
      console.log(err)
    })
  }
```

Figure 23: Frontend Development

Frontend UI

Frontend of the web application can be found in APPENDIX B.

7.5 Chapter Summary

A general summary of the technologies essential to the project's success was given at the beginning of this chapter. A thorough reason was given for choosing the desired methodology after reviewing every technical element and programming language. From that point forward, key experiments and features from the implementation phase were expanded upon with corresponding code parts. Finally, by illustrating functions and providing a basic explanation of the prototype's operational flow, this chapter highlighted the execution process.

8. TESTING

8.1 Chapter overview

After implementation, testing is performed. As a result, the test plan, test execution, and test results are all documented in this chapter. To evaluate the functionality of the Active Machine Learning, a distinct experimental setup is defined. Discussion and analysis of the experiment's findings are presented. Then, to test and confirm that the entire system is operating as planned, functional and non-functional testing is conducted along with module and integration testing.

8.2 Objectives and Goals of Testing

The primary purpose of testing is to identify and detect defects or issues in software, systems, or products, before they are released to end-users or customers. Testing helps ensure reliability, quality, user experience, etc.

According to this project, the machine learning with the full stack web application could include:

- To verify that the functional requirements of the project are met, including testing the accuracy of the ML models and ensuring that data entered through the web application is stored correctly in the Firestore database.
- To measure the performance of the ML models and the web application, including testing response times and checking that the application can handle a large number of requests.
- To ensure that the web application meets the needs of its intended users, by testing the ease of use, user interface, and overall user experience.
- To test the integration between the ML models and the web application, ensuring that data is passed correctly between the two components and that the overall system works as expected.

8.3 Testing Criteria

The testing criteria for this research project will depend on the functional and non-functional requirements of the system, as well as the goals and objectives of the testing phase.

- Functional criteria focus on testing the functional requirements of the system, which are related to the specific tasks and functions that the system is intended to perform. In the case of a web application, functional testing criteria might include testing the registration process, user login functionality, and prediction accuracy.

- Non-functional criteria, on the other hand, are related to the performance, usability, and reliability of the system. This might include testing factors such as response time, system availability, and scalability.

8.4 Model Testing

Model testing is an essential component of evaluating the performance of the trained models. It involves feeding the model with test data that it has not seen during the training phase. The performance of the model is then evaluated based on its ability to make accurate predictions on this new data. There are various metrics used for measuring model performance, including precision, recall, F1 score, accuracy.

Among the three algorithms that have been applied for this research, the voting classifier has achieved the highest level of accuracy.

Voting	
Accuracy	0.9206
Precision	0.9247
Recall	0.9206
F1-Score	0.9207

Table 23: Voting classifier evalution metrics

8.5 Benchmarking

Benchmarking involves evaluating the performance of a model or algorithm against a set of established or commonly used models or algorithms. It can help to identify the strengths and weaknesses of the new model or algorithm and compare its performance with the state-of-the-art approaches.

	Decision Tree	**Bagging**	**Voting**

Accuracy	0.9048	0.9048	0.9206
Precision	0.9153	0.9093	0.9247
Recall	0.9048	0.9048	0.9206
F1-Score	0.9067	0.9051	0.9207

Table 24: Evaluation Metrics

8.6 Functional Testing

8.6.1 Unit Testing

Table 25: Functional Testing

Test Case ID	Test Scenario	Test Steps	Expected Result	Actual Result	Pass/Fail
TC001	Sign Up	Click on Sign Up button.	Should be navigated to the Sign-Up page	Navigated to the Sign-Up page	Pass
TC002		Enter valid details (correct format) in all form items and click Sign Up.	Should be navigated to the Sign-In page	Navigated to the Sign-In page	Pass
TC002		Enter invalid information in any form item or keep empty.	Should show error messages under the relevant for item.	Showed error messages under the relevant for item.	Pass
TC003	Sign In	Enter correct email and password.	Should be navigated to the home page which consist of few form items and google map component.	Navigated to the Home page which consist of few form items and google map component.	Pass

TC004		Keep any form item empty	Should show error messages under the relevant for item	Showed error messages under the relevant for item.	Pass
TC005	Prediction	Select current date & time or future date & time and click ok.	Should be able to select current date & time or future date & time.	Were able to current date & time or future date & time.	Pass
TC006		Select past date.	Past dates should be disabled.	Past dates were disabled.	Pass
TC007		Type origin location and click on any suggestion.	While typing, suggestions of the place should come and after clicking on any place, that place should be selected and filled in the form item.	While typing, places were suggested. After clicking on one place, that place was displayed in the form item.	Pass
TC008		Type destination location and click on any suggestion.	While typing, suggestions of the place should come and after clicking on any place, that place should be selected and filled in the form item.	While typing, places were suggested. After clicking on one place, that place was displayed in the form item.	Pass
TC009		Fill all the form items and click submit.	Direction of the origin location to destination location should	Direction of the origin location to destination location was	Pass

			displayed on the map with or without markers along the route.	displayed on the map with or without markers along the route.	
TC0010		Click on submit button with empty form items.	Should show error messages under the relevant for item.	Showed error messages under the relevant for item.	Pass

8.6.2 Module and Integration Testing

Test Case ID	Test Scenario	Test Steps	Expected Result	Actual Result	Pass/Fail
Modules					
M1	Google Directions Service Test	Enter origin, destination, and date/time in the web application	Optimal route should be displayed on the map	Optimal route displayed on the map	Pass
M2	Open Weather API Test	Enter origin, destination and date/time in the web application	Weather should be forecasted	Weather was forecasted.	Pass
M3	Machine Learning Model Test	Enter origin, destination, date/time, user details and vehicle details in the web app	Severity should be predicted based on the input	Severity was predicted	Pass
Integration					
I1	Integration of Google	Enter origin, destination and	Optimal route displayed on the	Optimal route displayed on	Pass

	Directions Service and Open Weather API	date/time in the web app	map & weather forecasted.	the map & weather forecasted.	
I2	Integration of Machine Learning Model with Flask API and Scaler	Enter origin, destination, date/time, user details and vehicle details in the web app	Severity should get predicted for each way point	Severity got predicted and each way point	Pass
I3	Integration of all Modules	Enter origin, destination, date/time, user details and vehicle details in the web app	Severity should get predicted and for each way point and markers should be displayed on the route for identified accident way point	Severity got predicted and for each way point and markers displayed on the route for identified accident way point	Pass

Table 26: Non-Functional Testing

8.7 Non-Functional Testing

8.7.1 Performance Testing

Performance testing is an important aspect of software testing as it helps to ensure that the application or system can meet the performance requirements in terms of speed, stability, scalability, and responsiveness. The author has conducted a performance testing to the web application in order to identify performance bottlenecks, such as slow response times, high CPU or memory usage, network latency, and other issues that can cause the application to slow down or fail under heavy loads.

Figure 24: Performance testing

8.7.2 Usability Testing

Usability testing is important for this project as it allows to assess how easily users can interact with the web application and identify any potential areas of improvement to enhance the user experience. The author has conducted an unmoderated usability test with potential end users of the prototype. Analysis of this can be found in Chapter 9 – 9.6.2 Focus Group Testing.

8.8 Limitations of the testing process

- Limited coverage of accidents - The dataset used to train the classification algorithm may not include all possible accident scenarios, which could limit the accuracy and reliability of the predictions.
- Inaccurate location data - The accuracy of the predictions may be affected if the location data used to generate the waypoints is inaccurate or imprecise. Especially, in Sri Lankan accidents records, since police officers are using more than 20 years old map for the location identification, this could lead to false positives or false negatives in the predictions.

8.9 Chapter Summary

The test plan, the test execution, and the test outcomes from the testing that followed the implementation were all documented in this chapter. To evaluate the functionality of Active Machine Learning, a different experimental setup was created utilizing the binary Sentiment Classification task and the SST-2 dataset. Discussion and analysis of the experiment's findings were conducted. Then, to test and confirm that the entire system is operating as planned, functional and nonfunctional testing is conducted along with module and integration testing.

9. Evaluation

9.2 Evaluation Methodology and Approach

The evaluation is an essential part of the development process for any prototype or software application. It helps to ensure that the final product meets the requirements and goals of the project and is user-friendly and effective. For this prototype, the evaluation approach is particularly important due to the complex nature of the application. The integration of machine learning and other external APIs creates an intricate system that requires careful testing and evaluation.

The author has used both quantitative and qualitative evaluation approaches for this purpose. Qualitative evaluation approaches, such as usability testing and expert evaluation, can help to uncover user preferences, pain points, and areas for improvement of the prototype and quantitative approach has been applied to measure the performance with test results of the prototype.

9.3 Evaluation Criteria

The evaluation criteria identified for this research project are crucial in assessing the quality and effectiveness of the prototype. The criteria include research novelty and gap, research scope, methodology, implementation, and features in the prototype. By evaluating the prototype against these criteria, the author can identify strengths and weaknesses and make improvements where necessary to ensure a successful project outcome.

Criteria	**Reason**
Research Novelty and Gap	To evaluate the uniqueness of the proposed approach and to determine whether the approach fills a gap in the existing literature on accident prediction
Research Scope	To assess how well-defined and well-structured the research project is, in terms of its boundaries, objectives, and methods.
Methodology	To evaluate the effectiveness of the chosen techniques for the prediction of road accidents.

Development and Evaluation	To assess the effectiveness and efficiency of the implemented system in accurately predicting accident-prone zones and to evaluate the user experience of the system from the perspective of end-users.
Features in Prototype	To assess whether the developed prototype includes all the necessary features and functionalities.

Table 27: Evaluation Criteria

9.4 Self-Evaluation

Criteria	**Author's Evaluation**
Research Novelty and Gap	There are very smaller number of researches are available for Road Traffic Accidents prediction in Sri Lanka. After conducting thorough literature review, limitations in existing researches have been identified. After critically analyzing those limitations research novelty and gap have been finalized.
Research Scope	The scope was finalized by considering the schedule of the deliverables in the final year project.
Methodology	The appropriate methodologies for research, development, project management have been identified in order to achieve desired level of quality in the final prototype.
Development and Evaluation	Development was completed according to the high-level architecture which was designed in the design stage of the project. After satisfying all the components in each tier of the architecture, the author was able to implement the proposed solution with the use of appropriate tools and technologies.
Features in Prototype	All the features are functioning well and can be improved. But UI could have styled a little more to improve the engagement of the user

Table 28: Self evaluation

9.5 Selection of Evaluators

Selection of evaluators was done according to the stakeholder identification for this product. Finally, three types of evaluators were categorized as follows:

1. Technical experts who have knowledge and researched in data science and software engineering.
2. Domain experts who have knowledge in traffic accidents.
3. Potential end users of the product.

9.6 Evaluation Result

9.6.1 Expert Opinion

Evaluator	Criteria	Feedback
Technical Experts		
Mr. Dulshan Ratnayake Data Engineer Circles.Life Sri Lanka	Scope	Sanduni has carried out the research process very well. She has made a big contribution to the software engineering domain itself by collecting data and digitizing it, allowing other researchers to use her dataset for similar use cases. This contribution is highly remarkable since in the field of machine learning, good data is the equivalent of gold.
	Methodology and Architecture	She has chosen the features with justification. She has obtained good accuracy as well compared to her baseline models whereas these baseline models were self-engineered which can be considered as another research contribution. As for improvements it would be great if components such as over-fitting and under-fitting have been focused more.
	Development of the prototype	For future work real-time monitoring alerts could be provided with real time machine learning with the help of tools such as spark streaming when used with a vast number of concurrent users.

Ms. Sadunika Kularathna Senior Software Engineer Axiata Digital Labs	Scope	Drivers can be extra careful while driving through these places or they can avoid such locations.
	Methodology and Architecture	The author has used a proper methodology and technologies to implement the solution. I think this research contributes highly to the society and to the related domains.
	Development of the prototype	The proposed solution is very helpful for the drivers as they will be notified beforehand about certain locations where traffic accidents can occur. This research has proposed a solution for a very common and important issue as we loss considerable number of precious lives every year due to road traffic accidents.
Mr. Nisal Samaranayake Full stack software engineer/ Data engineer Sysco Labs	Scope	Accidents can happen anytime, anywhere. Considering factors like times when high traffic exists on the road, accident times, weather, root causes of the accidents and drivers' status, we can see there are patterns for these accidents.
	Methodology and Architecture	Analyzing such patterns and predicting accident probability for certain areas will be really useful for drivers. Many lives are lost due to accidents each day. Ability to know certain time frames and scenarios when an accident might happen when traveling might save lives.
	Development of the prototype	Prototype to prove the ability to predict is admirable. For future enhancements prefer the output csv data to be processed and output to the user while they are under navigation to a certain location and warn where they need to worry about. Good domain idea and use of tools.

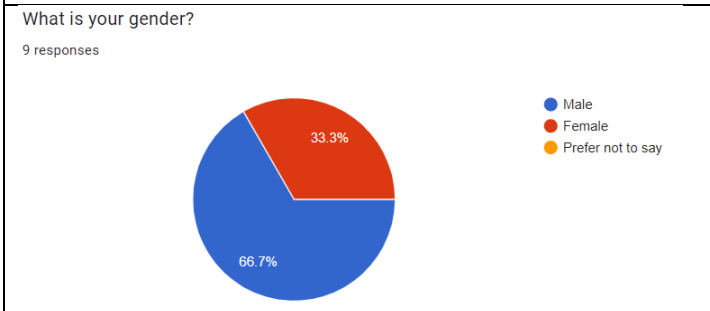
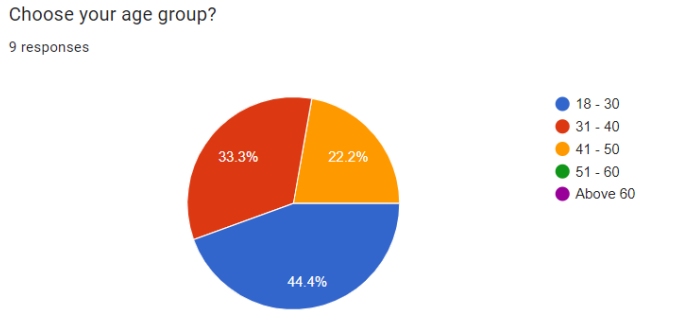
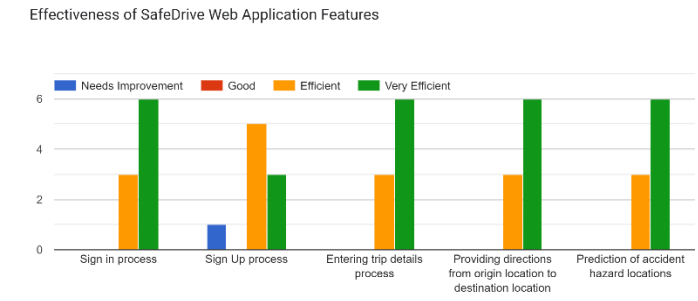
Mr. L. K. H. Chirath Didulanga Software Engineer Axiata Digital Labs	Scope	Also I think this if this project further improved, it can help to improve road safety and reduce potential accidents.
	Methodology and Architecture	Using technologies like No SQL data bases is very good.
	Development of the prototype	The presentation is very detailed and well-planned presentation. UI certainly need improvements.
Mr. Shehan Guruge Senior Software Engineer/ Machine Learning Researcher Ascentic	Scope	The idea, architecture and technologies used sound promising and the scope is well defined to match with undergraduate level.
	Methodology and Architecture	However in perspective of an end product this could be further enhanced with capturing the inputs from other third party sources like e-newspapers and related articles to capture the locations of accidents have occurred and retrain the model to gain more accuracy.
	Development of the prototype	Conceptually this is really a good solution for daily drivers. Further it could be implemented by capturing other sources which could be potential reasons for occurring accidents, such as driver emotions, drowsiness etc. to make alerts in real time.
Mr. Nuwan Sandeepa Software Engineer Axiata Digital Labs	Scope	This would be a great project idea in Sri Lanka and hope to see this application in public.
	Methodology and Architecture	There are small suggestions for this app, 1. It would be easier if you can provide mobile app as well, so that drivers can easily access this app. 2. If you can make the sign-up process easy it will be a great advantage for the users.

	Development of the prototype	Verry interested app!
Mr. Rajitha Warusavitarana Senior Software Engineer Sysco Labs	Scope	Sanduni has approached to solve a problem that has both research and business value. Road accidents are very common in the present day as she has presented in the problem, I'm really happy about the solution she has brought up with.
	Methodology and Architecture	Gathering data from the relevant police stations and creating a dataset for this problem is a very accurate and reliable source of information. The created machine learning model and applied techniques for this scenario looks very appropriate for me.
	Development of the prototype	The web application looks very user friendly but however in most cases people will be ushering their mobile phones when referring to routes when they are driving. Therefore, I think a mobile application is a more appropriate addition that she can add as a future implementation.
Domain Experts		
Mr. T.M.P Mendis, Police Inspector, Head of Division traffic OIC, Nugegoda	Concept	Concept is really good!
	Solution	I haven't seen any similar products like this before.
Mr. K. M. R. W. Karunarathne Police Inspector, PS35641	Concept	The concept is really good! It will be useful to mitigate future traffic accidents.
	Solution	App is good. It is better to get real time predictions.

Mr. D. M. Kamal Chaminda Dissanayake Police Inspector, PS	Concept	The idea is very good.
	Solution	Application is also good. It would be really great, if you can expand this to other areas in Sri Lanka.

Table 29: Expert Opinion

9.6.2 Focus Group Testing

Questionnaire Result	Analysis
	66.7% of participants who participated for the usability test are male participants.
	The majority of participants who have participated for the usability test are in 18 – 30 age range. And 2nd highest participants are from 31 – 40 age range.
	Most of participants have stated that effectiveness of the features in prototype are efficient. But some participants are not satisfied with some features and author should focus on improve them.

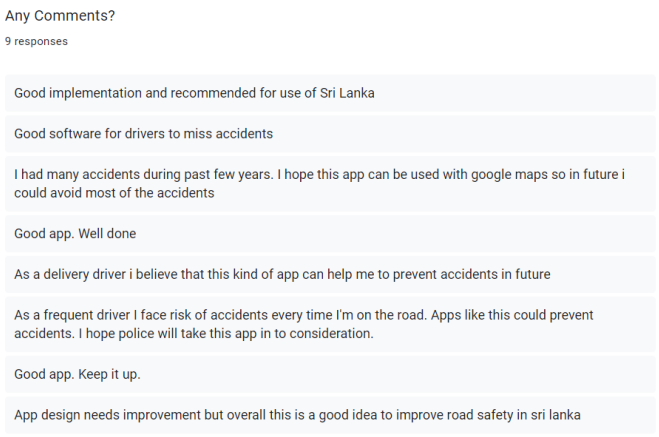
	Participants have expressed their overall feeling about the prototype. While some participants are satisfied with the current prototype, some participants have mentioned the areas that needs to be improved. Common are needs to be improved are frontend of the application.

Table 30: Usability Testing

9.7 Limitations of Evaluation

There are several limitations in expert evaluations. The experts may not be representative of the actual end-users of the product, which could impact the usability and usefulness of the product. Furthermore, the evaluation may not have captured all the relevant aspects of the product, or the evaluation criteria may not have been appropriate for the context of the product. Finally, the sample size of experts and end-users were not very large enough to capture the full range of opinions and feedback.

9.8 Evaluation on Functional Requirements

The author has performed system evaluation with respect to functional requirements.

FR No.	Priority	Pass/Fail	FR No.	Priority	Pass/Fail
FR1	Must have	Pass	FR 6	Must have	Pass
FR 2	Must have	Pass	FR 7	Must have	Pass
FR 3	Must have	Pass	FR 8	Must have	Pass
FR 4	Must have	Pass	FR 9	Must have	Pass
FR 5	Must have	Pass	FR 10	Should have	Pass
			FR 11	Could have	Fail

Table 31: Functional Requirement evaluation

$$Pass\ Rate = \frac{Number\ of\ passed\ test\ cases}{Number\ of\ test\ cases\ executed} \times 100\%$$

$$= \frac{10}{11} \times 100\% = \mathbf{91}\%$$

9.9 Evaluation on Non - Functional Requirements

The author has performed system evaluation with respect to non-functional requirements.

NFR No.	Priority	Pass/Fail
NFR1	Performance	Pass
NFR 2	Accuracy	Pass
NFR 3	Usability	Pass

Table 32: Non-functional requirement evaluation

$$Pass\ Rate = \frac{Number\ of\ passed\ test\ cases}{Number\ of\ test\ cases\ executed} \times 100\%$$

$$= \frac{3}{3} \times 100\% = \mathbf{100}\%$$

10. CONCLUSION

10.2 Achievements of Research Aims & Objectives

10.2.1 Achievements of Research Aims

Research aim of this project was to design, implement, and evaluate an application that will facilitate drivers to predict a road traffic accident at certain locations along the driver's route beforehand by getting details of the trip, driver and the vehicle and save lives and properties by mitigating the traffic accident, and mainly target Sri Lankan roads.

The author was able to successfully to complete the each component of the initially mentioned research aim. According to chapter 6, the author has conducted the design stage effectively by designing proper diagrams which are crucial for the implementation such as high-level diagram, system diagrams, data flow diagrams. As stated in each diagram, the author was able to finish each component in each diagram successfully. In chapter 7, the author has discussed how the prototype was implemented according to the proposed designed. Each step in the implementation stage has been taken properly which resulted in a full stack application. Evaluation of the developed prototype happened at the evaluation stage. The author has discussed how evaluation happened according to each criteria. As a result, the author was able to complete the prototype with desired level of quality successfully.

10.2.2 Achievements of Research Objectives

Research Objectives	**Description**	**Learning Outcomes**	**Status**
Literature Review (LR)	Carry out a comprehensive LR, **RO1**: To identify the existing solutions for road traffic accidents prevention and find their weaknesses. **RO2**: To research and analyze current road traffic accident prediction systems.	LO1, LO3, LO4	Achieved

	RO3: To analyze methods used to identify dangerous areas with high accident rates. **RO4**: To discover research gaps in the road traffic accidents domain.		
Requirement Elicitation	Conduct a thorough user requirement gathering, **RO1**: To determine the essential attributes and factors in road traffic accidents. **RO2**: Gather requirements by conducting a survey focusing on the target audience. **R03**: Gather requirements related to the domain and technology by conducting formal meetings with domain experts and technical experts. **R04**: Analyze functional and non-functional requirements.	LO1, LO3, LO4, LO6	Achieved
Design	Design a system that can use the suggested methods to solve the problems that have been identified. **RO1**: To design the high level architecture of the proposed solution. **RO2**: To create the backend and frontend including UI of the application.	LO2, LO3, LO5, LO8	Achieved

	RO3: To design appropriate design diagrams belongs to the SSADM design paradigm, such as data flow diagrams.		
Implementation	Conduct a review on existing technologies to select most suitable ones for the system, **RO1:** To do data exploration and pre-processing with multiple techniques. **RO2**: To train multiple ML models and implement the optimized model and identify the most accurate model. **RO3**: To develop the proposed traffic accident prediction interactive application using the best approach.	LO2, LO5, LO7	Achieved
Evaluation	Use appropriate evaluation techniques to test and evaluate the system, **RO1**: Evaluate and test each component of implemented application to predict road traffic accidents. **RO2**: Make test plan to perform unit testing, integration, functional testing. **RO3**: Get feedback from evaluators from technical experts and domain experts and also get evaluated by end users.	LO6, LO8	Achieved

Table 33: Achievment of Research Objectives

10.3 Utilization of Knowledge from the Course

Module	Description
Programming Principles 1	This module gave the foundation for the programming and was able to learn python language which is very important language for the data science area.
Web Development and Design	This module gave the foundation for the web development area and was able to learn the most basic language (JavaScript) in most of frontend frameworks currently in high demand.
Object Oriented Programming	This module gave knowledge on how programming modules are organized which is very crucial for the any level of programmer. This module helps to do development successfully in every aspect.
Software Development Group Project	This module gave the foundation knowledge on how a research project should be organized, conducted and maintained. This module was very helpful from idea stage to project conclusion.

Table 34: Utilization of Knowledge from the course

10.4 Use of Existing Skills

A lot of existing skills were used during the project, from idea initialization to project conclusion. Some of the existing skills that were used are listed below.

- Machine learning knowledge – Machine learning knowledge was used. Since this project is machine learning a project, existing knowledge in ML was very useful to quickly gather new techniques and aspects in the domain of machine learning.
- Programming knowledge – Since this project finally concludes with a fully functioning proof of concept, programming was crucial knowledge to have before the project initiation. The author had knowledge of programming languages like python and JavaScript. The existing knowledge was very helpful to gather knowledge about programming concepts that should be applied to this prototype.

10.5 Use of New Skills

- Data analysis with several techniques – In the machine learning component of the prototype, the author has applied several techniques which are related to dataset analyzing, data preprocessing with several statistical approaches like correlation analysis.
- Ensemble learning – Since the machine learning component is heavily implemented using different types of Ensembles learning techniques, the author had to learn and gather knowledge on ensemble learning in order to apply for the development.

10.6 Achievement of Learning Outcomes

According to the table which has been created **10.2.2 Achievements of research objectives**, author has mentioned the learning outcomes achievements.

10.7 Problems and Challenges Faced

Problem	Solution
Accident details are recorded in books as written records.	The author had to manually enter each record one by one to a excel sheet in order to create the dataset.
Accident locations are recorded according to a mechanism which was introduced in 2002 and recorded as east and north coordinates.	After taking a clear photograph of the police map, layered it on google earth pro software to recognize actual latitude and longitude from east coordinate and north coordinate from police map.
Recording non-accidents data.	To create non-accident data with the help of accident records, the author had to conduct interviews with domain experts and technical experts, and literature review.

Table 35: Problems and challenges faced

10.8 Deviations

The initial idea was to build a machine learning model to predict accident-prone location by the ML model with the use of regression techniques. But due to the nature of the model, it was predicting a one location but there may have several accident-prone locations according to the user data. And also, there was no method to identify whether an accident will happen or not according to user data. Therefore, the author decided to change the technique to classification to identify whether there will be an accident or not and the severity of the accident. After completing that, the author was able to show whether there will be an accident or not but sending each and every waypoint the direction to the machine learning model.

10.9 Limitations of the Research

- This research is only limited to the city of Mirihana in Colombo district, Sri Lanka.
- Data collection method was very poor and limited in Sri Lanka due to nature of recording data in Sri Lankan police stations.
- The software testing approaches were limited with limited sources such as with higher number of participants.

10.10 Future Enhancements

- Improve the dataset in order to get better performance and to draw conclusions about the accidents in Mirihana police division.
- The research project can be expanded to the whole country to improve road safety which is an important aspect to the economic growth in Sri Lanka.
- According to the feedback received from experts and end users, this project can be enhanced by developing a mobile application which is easy to access for every user.

10.11 Achievement of the contribution to body of knowledge

From this innovative approach, the author was able to present a user-friendly web application which can be used by drivers. By just entering a few details about the vehicle and the driver, users will be able to successfully register to the system. Users will be able to get to know predicted accidents during his/her driving route by entering the date, time of the trip and origin

location to destination location. This way it will help to reduce the impact of a road accident and will be able to save millions of lives and properties.

Digitalization of data is very poor in Sri Lankan government organization. Dataset is a major component for a machine learning project. But unfortunately, there were no proper computerized accident dataset available in Sri Lanka. The author made a big contribution to the software engineering domain itself by collecting data and digitizing it, allowing other researchers to use her dataset for similar use cases.

Currently, in the domain of road traffic accidents, there are existing works which has done using ensemble learning techniques. The author was able to introduce ensemble learning approaches to predict traffic collisions based on locations with a history of having a lot of traffic accidents.

10.12 Concluding Remarks

This research project on predicting road accidents through the use of machine learning and data analysis techniques is a valuable contribution to the field of road safety. Through the approach of integrating various data sources and APIs, the author has demonstrated the potential for technology to enhance road safety and prevent accidents.

This study has also highlighted the importance of interdisciplinary collaboration between technical experts, domain experts, and end-users to ensure that the final product meets the needs of all stakeholders. Use of both qualitative and quantitative evaluation methodologies has provided a comprehensive understanding of the strengths and limitations of the prototype, as well as valuable insights for future research in this area.

APPENDIX – A

Concept Map

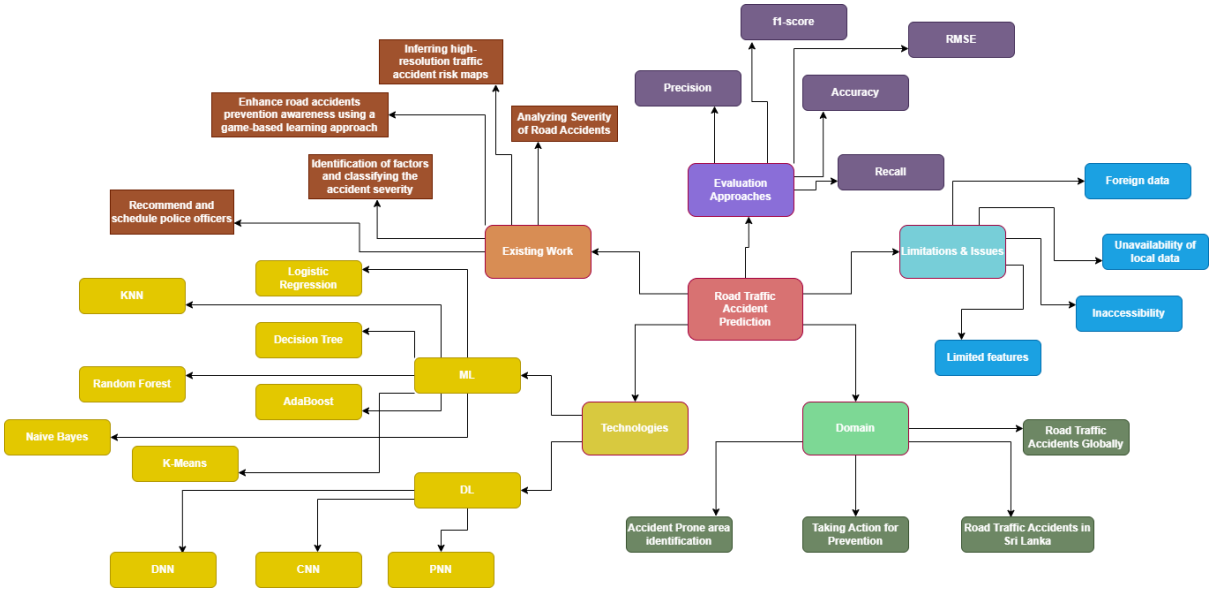


Figure 25: Concept Map

APPENDIX – B

Frontend of the Web Application

Sign-In Page

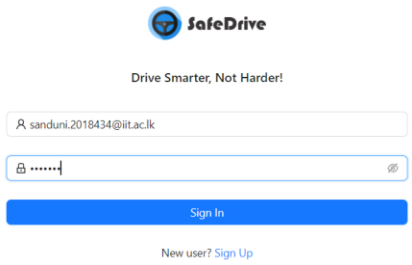


Figure 26: Sign In page

Sign-Up Page

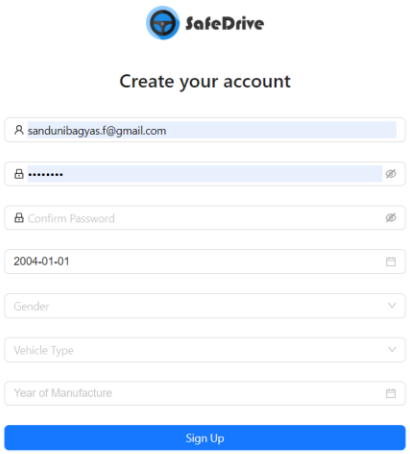


Figure 27: Sign Up page

Home Page

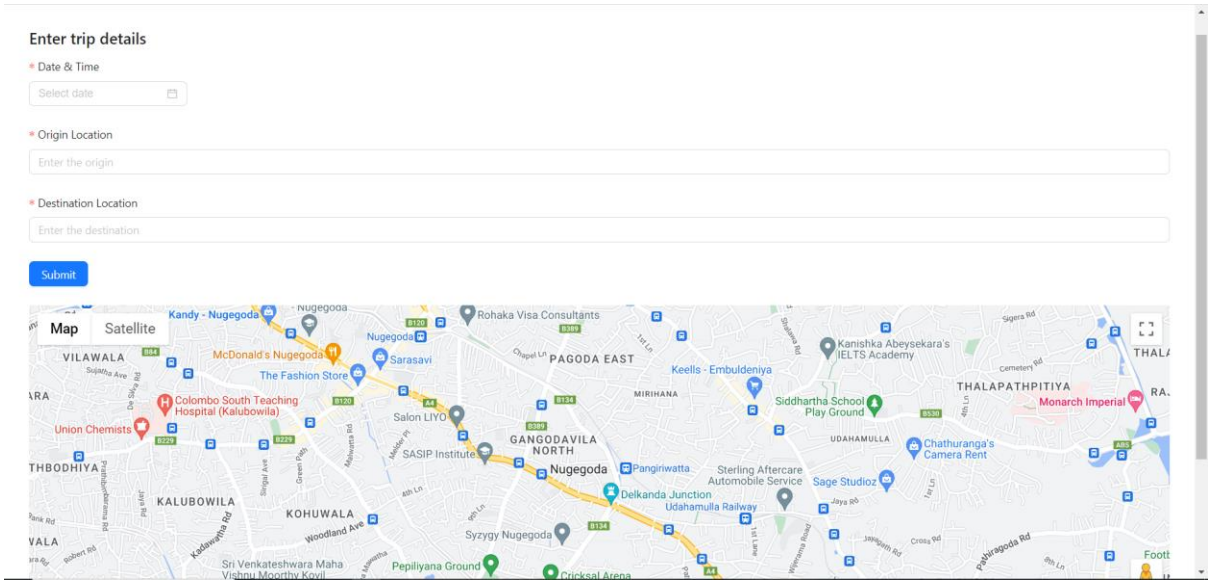


Figure 28: Home Page

References

Abdulhafedh, A. (2017). Road Crash Prediction Models: Different Statistical Modeling Approaches. *Journal of Transportation Technologies*, 07, 190–205. Available from https://doi.org/10.4236/jtts.2017.72014.

Ariyathilake, S.N. and Rathnayaka, R.M.K.T. (2019). DNN Classifier and Decision Tree-Based Novel Methodology for Analyzing Road Accidents. *2019 19th International Conference on Advances in ICT for Emerging Regions (ICTer)*. September 2019. 1–8. Available from https://doi.org/10.1109/ICTer48817.2019.9023658.

Bhavan, T. (2019). The Economic Impact of Road Accidents: The Case of Sri Lanka. *South Asia Economic Journal*, 20 (1), 124–137. Available from https://doi.org/10.1177/1391561418822210.

CDC. (2020). Road Traffic Injuries and Deaths—A Global Problem. *Centers for Disease Control and Prevention*. Available from https://www.cdc.gov/injury/features/global-road-safety/index.html [Accessed 12 October 2022].

Dhananjaya, S. a. T. and Alibuhtto, M.C. (2016). Factors influencing road accidents in Sri Lanka: a logistic regression approach. Available from http://ir.lib.seu.ac.lk/handle/123456789/2087 [Accessed 19 October 2022].

Garg, N. and Hyder, A.A. (2006). Exploring the relationship between development and road traffic injuries: a case study from India. *European Journal of Public Health*, 16 (5), 487–491. Available from https://doi.org/10.1093/eurpub/ckl031.

Kapilan, K., Bandara, S. and Dammalage, T. (2020). Vehicle Accident Detection and Warning System for Sri Lanka Using GNSS Technology. *2020 International Conference on Image Processing and Robotics (ICIP)*. March 2020. 1–5. Available from https://doi.org/10.1109/ICIP48927.2020.9367334.

Kodithuwakku, D.S. and Peiris, T.S.G. (2022). Factors Contributing to the Road Traffic Accidents in Sri Lanka. Available from https://sjhs.sljol.info/articles/abstract/10.4038/sjhs.v2i1.44/.

Kushan, M.A.K. and Chandrasekara, N.V. (2020). Identification of factors and classifying the accident severity in Colombo - Katunayake expressway, Sri Lanka. September 2020. Colombo, Sri Lanka: IEEE. Available from https://doi.org/10.1109/SCSE49731.2020.9313036.

Labib, Md.F. et al. (2019). Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh | IEEE Conference Publication | IEEE Xplore. Available from https://ieeexplore.ieee.org/document/8843640.

Malik, S. et al. (2021). Road Accident Severity Prediction — A Comparative Analysis of Machine Learning Algorithms. *2021 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*. December 2021. 69–74. Available from https://doi.org/10.1109/GCAIoT53516.2021.9693055.

Renuraj, S., Varathan, N. and Satkunananthan, N. (2015). Factors Influencing Traffic Accidents in Jaffna. Available from https://sljastats.sljol.info/articles/abstract/10.4038/sljastats.v16i2.7827/.

Saunders, M.N.K. (2019). Saunders, Research Methods for Business Students, 8/E. Available from https://www.pearson.com/nl/en_NL/higher-education/subject-catalogue/business-and-management/Research-methods-for-business-students-8e-saunders.html [Accessed 2 November 2022].

Senaviratna, N.A.M.R. and Cooray, T.M.J.A. (2020). Statistical Analysis of Severity of Motor Vehicle Accidents in Sri Lanka. Available from https://www.jetir.org/view?paper=JETIR2011198.

Sewwandi, A.K.T. et al. (2020). SmartCop: An Automated Platform to Mitigate the Impact of Road Accidents. *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*. December 2020. 1–6. Available from https://doi.org/10.1109/R10-HTC49770.2020.9357042.

United Nations. (2015). Transforming our World: The 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs. Available from https://sdgs.un.org/publications/transforming-our-world-2030-agenda-sustainable-development-17981 [Accessed 12 October 2022].

Vijithasena, R. and Herath, W. (2022). Data Visualization and Machine Learning Approach for Analyzing Severity of Road Accidents. *2022 International Conference for*

Advancement in Technology (ICONAT). January 2022. 1–6. Available from https://doi.org/10.1109/ICONAT53423.2022.9726042.

WHO. (2010). Global Plan for the Decade of Action for Road Safety 2011-2020. Available from https://www.who.int/publications/m/item/global-plan-for-the-decade-of-action-for-road-safety-2011-2020 [Accessed 12 October 2022].

WHO. (2022). Road traffic injuries. Available from https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries [Accessed 7 October 2022].

World Bank. (2020). *Delivering Road Safety in Sri Lanka: Leadership Priorities and Initiatives to 2030*. Washington, DC: World Bank. Available from https://doi.org/10.1596/33341 [Accessed 1 November 2022].