PROJECT REPORT

ON

Road Accident Analysis

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF

BACHELOR OF ENGINEERING

In

COMPUTER ENGINEERING

By

Shivansh Gautam-35

Prashant Kumar Singh-26

Prince Kumar Singh-29

Sonal Raj-39

Under the Guidance of

Prof. Veer Bhadra Pratap Singh

Submitted to

Peoples Empowerment Group

ISB&M COLLEGE OF ENGINEERING, NANDE, PUNE

DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

This is to certify that the Project Report entitled

Road Accident Analysis

Submitted By

Shivansh Gautam-35

Prashant Kumar Singh-26

Prince Kumar Singh-29

Sonal Raj-39

Is a bonafide work carried out by them under the supervision of **Prof. Veer Bhadra Pratap Singh** and it is approved for the partial fulfillment of the requirement of **Savitribai Phule Pune University** for the Project in the Final Year of Computer Engineering.

Prof. Veer Bhadra Pratap Singh	**Prof. B. B. Gite**	**Dr. P.K. Srivastava**
Guide	**H.O.D**	**Principal**
Dept. of Computer Engg.	**Dept. of Computer Engg.**	**ISB&M-COE Nande, Pune**

Place: Pune	**External Examiner**	**Prof. Veer Bhadra Pratap Singh**
Date: 23 / 11 /2022		**Project Coordinator**

ACKNOWLEDGEMENT

We would like to take this opportunity to thank all the people who were part of this seminar in numerous ways, people who gave un-ending support right from the initial stage.

In particular we wish to thank **Prof. Veer Bhadra Pratap Singh** as internal project guide who gave their co-operation timely and precious guidance without which this project would not have been a success. We thank them for reviewing the entire project with painstaking efforts and more of his, unbanning ability to spot mistakes.

We would like to thank our **H.O.D Prof. B. B. Gite** for his continuous encouragement, support, and guidance at each and every stage of the project.

Finally, we would like to thank all my friends who were associated with me and helped me in preparing my project. The project named **"Road Accident Analysis"** would not be possible without the extensive support of people who were directly or indirectly involved in its successful execution.

Project Group Members:

Shivansh Gautam-35

Prashant Kumar Singh-26

Prince Kumar Singh-29

Sonal Raj-39

ABSTRACT

Today, one of the top concerns for governments is road safety. Although there are various safety precautions in place to prevent auto accidents, they cannot be completely avoided. To lessen the harm caused by traffic accidents, the primary goal now is to determine what causes them. In this study, we use machine learning techniques to identify the causes of traffic accidents. By creating precise prediction models that can automatically separate distinct unintentional instances, patterns involved in diverse situations can be identified. The development of safety measures and the application of these classification approaches will help avoid accidents.

Although there are numerous inventories in the automotive sector to create and construct safety features for cars, road accidents are inevitable. Both urban and rural regions see a high rate of accidents. By creating precise prediction models that can automatically separate distinct unintentional instances, patterns involved in diverse situations can be identified. These clusters will be helpful in creating safety precautions and preventing mishaps. We think we can use some scientific techniques to reduce accidents as much as possible while using limited resources.

Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction

The problem of deaths and injuries because of accidents is to be a global phenomenon. Traffic safety has been a serious concern since the start of the automobile age, almost one hundred years ago. It has been estimated that over 300,000 persons die, and 10 to 15 million persons are injured every year in road accidents throughout the world. Statistics have also shown that mortality in road accidents is very high among young adults that constitute a major part of the workforce. To overcome this problem, there is a need for various road safety strategies. [5.8][12]

In recent years, there is an increase in research attention to determine the significant effect of the severity of driver injuries which is caused due to road accidents. Accurate and comprehensive accident records are the basis of accident analysis. [5.8][7] The effective use of accident records depends on some factors, like the accuracy of the data, record retention, and data analysis. There are many approaches applied to this scenario to study this problem. [5.8][9]

A recent study illustrated that the residential and shopping sites are more hazardous than village areas as might have been predicted, the frequencies of the casualties were higher near the zones of residence possibly because of the higher exposure. [5.8][11] A study revealed that the casualty rates in residential areas are classified as relatively deprived and significantly higher than those from relatively affluent areas.

1.2 Machine Learning

Machine Learning is a branch of the broader field of artificial intelligence that makes use of statistical models to develop predictions. It is often described as a form of predictive modeling or predictive analytics and traditionally, has been defined as the ability of a computer to learn without explicitly being programmed to do so. [5.8][13]

In basic technical terms, machine learning uses algorithms that take empirical or historical data, analyze it, and generate outputs based on that analysis. In some approaches, the algorithms work with so-called “training data” first and then they learn, predict, and find ways to improve their performance over time. [5.8][13]

Types of Machine Learning

There are three main approaches to machine learning:

a) Supervised
b) Unsupervised,
c) Reinforcement learning.

There are also hybrid approaches including semi-supervised learning. Each approach has specific strengths and weaknesses, and some techniques are better suited to particular types of problems than others. [5.8][13]

a) Supervised:

In supervised learning, the computer is trained on a set of data inputs and outputs, with a goal of learning a general rule that maps the given inputs to the given outputs.

Two main types of supervised learning are:

1) classification, which entails the prediction of a class label.
2) regression, which entails the prediction of a numerical value. [5.8][13]

b) Unsupervised:

In unsupervised learning, the learning algorithm is not given this type of guidance instead, it works to discover the pattern or structure in the input on its own.

Two main types of unsupervised learning are:

I. Clustering is involved in discovering groups within the dataset that share similar characteristics.
II. Density estimation, which involves evaluating the statistical distribution of the data set.

Unsupervised learning methods also include visualization with the data and projection, which reduces the dimensions of the data, a form of simplification. [5.8][13]

c) Reinforcement:

In reinforcement learning, the computer and algorithms will confront a problem in a dynamic environment and as it works to perform a given goal, it will receive feedback /rewards, which will reinforce its learning and goal-seeking effort. [5.8][13]

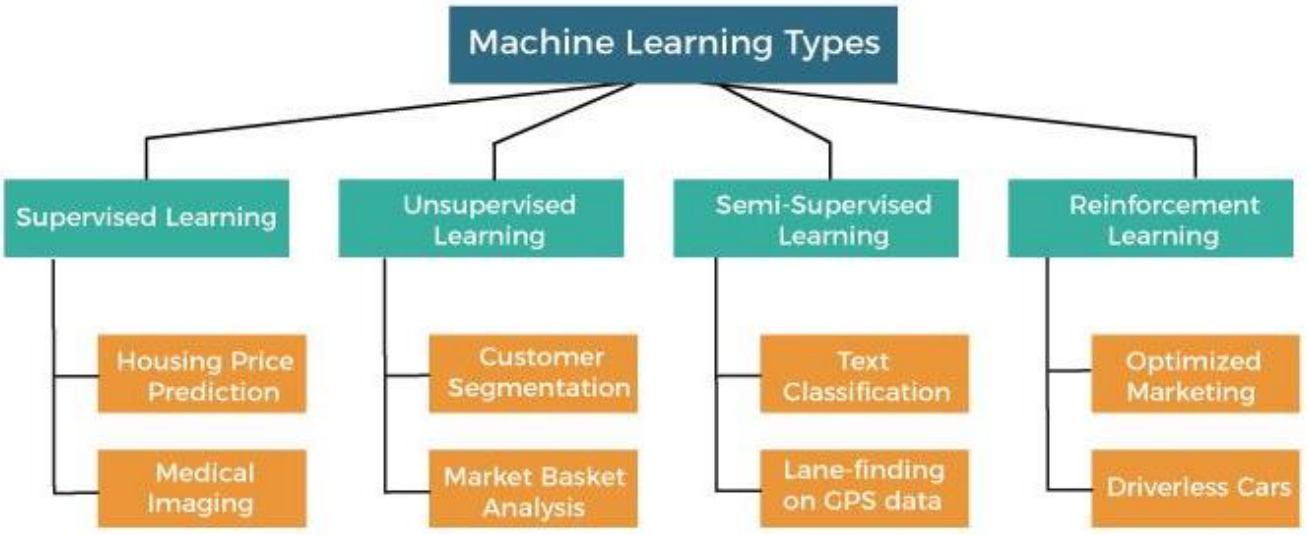

Fig-1.1 Types of machine Learning [5.8][13]

1.3 Problem Definition

To handle the enormous number of road accidents in a locality a precise analysis is required. This analysis will be done more deeply to determine the intensity of the road accidents by using supervised learning techniques like Deep Learning Neural Networks

and AdaBoost. This will classify the severity of the accidents as fatal, grievous, simple injury and motor collision. Many of agencies especially government agencies are identifying the factors that contribute to the accident roads or highways. The measurements to prevent accident speed reduction, widen divider, or other else. These different types of the accident on the critical roads or highways of agencies such as Royal Mal (JKR), Road Transport process, planning process or in remedy process for a serious part when all the road user's measurement of how the accident can occur. Theres model that has been developed to analyze the accident and can analyze all the variables.

1.4 Scope

Losses in road accidents are unbearable, to the society as well as a developing country like us. So, it has become an essential requirement to control and arrange traffic with an advanced system to decrease the number of road accidents in our country. By taking simple precautions, based on prediction of a sophisticated system may prevent traffic accidents. Moreover, to tackle this situation where every day so many people were killed in a traffic accident and to control the increasing rate of traffic accidents, implementation of machine learning is functional and great approach to take an accurate decision, with the experience to manage the situations and analysis The ML can analyze which part is responsible for accidents which can be suggested to managing authorities for precautionary action.

1.5 Project Objectives:

1. To study the causes of the accident by feature extraction.
2. To understand the severity of the accident based on these features.
3. To classify it as fatal, grievous, simple injury, or motor collision.
4. To carry out algorithms to predict the performance and accuracy of each.
5. To conclude the fastest algorithm.
6. To understand the effect of each feature on the accident and conclude how much is it responsible for the accident.

1.6 Organization of Report

This section provides information about the organization of the report distributed in different chapters, along with the sections discussed in each chapter. Chapter 1 is the introduction. It discusses about the background information, problem definition, scope, and objectives of the project. Chapter 2 is the system analysis. It discusses about the literature survey of triage Road accident analysis, proposed system, different types of the feasibility study, risk analysis, scheduling, and effort allocation. Chapter 3 is the system requirement specification. It discusses about the different types of requirements, such as functional, non-functional, hardware, software, performance, and design requirements. Chapter 4 is about system design. It focuses on the architecture of the system, data flow, and UML diagrams.

Chapter 2

Risk Analysis

2.1 Risk analysis Introduction

Risk analysis is the process of identifying and analyzing potential issues that could negatively impact algorithm or project. This process is done in order to help organizations avoid or mitigate those risks. Performing a risk analysis includes considering the possibility of adverse events caused by malicious or inadvertent human activities. An important part of risk analysis is identifying the potential for harm from these events, as well as the likelihood that they will occur.

Organizations must understand the risks associated with the use of their information systems to protect their information assets effectively and efficiently. Risk analysis can help an organization improve its security in a number of ways. Depending on the type and extent of the risk analysis, organizations can use the results to help

2.2 Steps of Risk Analysis

1 Conduct a risk assessment survey: This first step, getting input from management and department heads, is critical to the risk assessment process. The risk assessment survey is a way to begin documenting specific risks or threats within each department. [5.8][10]

2 Identify the risks: The reason for performing a risk assessment is to evaluate an IT system or other aspect of the organization and then ask: What are the risks to the software, hardware, data, and IT employees? What are the possible adverse events that could occur, such as human error, fire, flooding, or earthquakes? What is the potential that the integrity of the system will be compromised or that it won't be available? [5.8][10]

3 Analyze the risks: Once the risks are identified, the risk analysis process should determine the likelihood that each risk will occur, as well as the consequences linked to each risk and how they might affect the objectives of a project.

4 Develop a risk management plan: Based on an analysis of which assets are valuable and which threats will probably affect those assets negatively, the risk analysis should produce control recommendations that can be used to mitigate, transfer, accept or avoid the risk.

5 Implement the risk management plan: The ultimate goal of risk assessment is to implement measures to remove or reduce the risks. Starting with the highest-priority risk, resolve or at least mitigate each risk so it's no longer a threat.

6 Monitor the risks: The ongoing process of identifying, treating and managing risks should be an important part of any risk analysis process. The focus of the analysis, as well as the format of the results, will vary depending on the type of risk analysis being carried out.

2.3 Technical Feasibility

Technical feasibility of a project is performing a check on whether the development of project is possible with the available technological resources. The technical feasibility is a very important aspect to be considered before the official commencement of

the project by the organization. The technical feasibility is checked by pondering over the functional requirements of the user. To determine whether the proposed system is technically feasible, the technical issues involved behind the system should take into consideration. Proposed system uses python technology. Python is an open-source technology, it is available for free of cost and conveniently. As far as the platform for the project is concerned, it is decided to perform the project on the window OS.

Therefore, the project has to be done on any Windows OS. Thus, it becomes quite sure the project is technically feasible.

2.4 Real-World Risks

Real-world Risks are those risks that can affect property or human life. In many fields, Artificial intelligence and machine learning algorithms are making a decision for many organizations for different purposes. When this decision is made wrong the organization suffers financial/physical losses and this can happen due to faulty data or human mistake or coding error or glitches. Because of the inaccurate data, machine learning algorithm could generate inaccurate results which could lead to faulty decision-making in the program.

In road accident analysis if our program gives a faulty output or inaccurate output it may cause property or human life damage. Therefore, the risk analysis and risk management is one of the most important factor to consider.

Chapter 3

System Analysis

3.1 Literature

Many researchers have carried out research work in the area of road accidents. some of them have analyzed accident data in different ways. some of them identification of blackspot zone. some of them have developed accident models for forecasting future accident trends. they have also proposed strategies for road safety. in the present chapter literature review is carried out covering the different issues related to road accidents and road safety.

A review of the effect of traffic and weather characteristics on road safety. despite the existence of generally mixed evidence on the effect of traffic parameters, a few patterns can be observed. for instance, traffic flow seems to have a non-linear relationship with accident rates, even though some studies suggest a linear relationship with accidents. regarding weather effects, the effect of precipitation is quite consistent and leads generally to increased accident frequency but does not seem to have a consistent effect on severity. the impact of other weather parameters on safety, such as visibility, wind speed, and temperature is not found straightforwardly so far. the increasing use of real-time data not only makes it easier to identify the safety impact of traffic and weather characteristics but most importantly makes possible the identification of their combined effect. the more systematic use of these real-time data may address several of the research gaps identified in this research.

1st Published paper: Development calibration factor for crash prediction model for rural two-lane roadways in Illinois.

Authors:

- Michael Williamson
- Huaguo Zhou

Publisher: Procedia-social and behavioral sciences

Issued on 2012

Abstract:

Michael Williamson and Huaguo Zhou (2012) were the development of calibration factors for crash prediction models in the new Highway Safety Manual (HSM) for rural two-lane roadways in Illinois. The crash prediction modes (so called Safety Performance Functions (SPF)) in the HSM were developed using data from multiple states, therefore the models must be calibrated to account for local factors, such as weather, roadway conditions, and drivers' characteristics. In this study, two calibration factors were developed for two different SPFs to give a better prediction of crash frequencies on rural two-lane roadways in Illinois. This study determined the SPF that best predicts the crashes was developed specifically for rural two-lane Two-way roadways in Illinois. It is recommended that local SPFs be developed and compared to the HSM SPF when evaluating the safety of a roadway.

2nd Published paper: Accident analysis on national highway 3 between Indore to Dhamnod

Authors:

- K Meshram

- S.H Goliya

Publisher: International Journal of application or innovation in Engineering and Management (IJAIEM) Volume 2

Issued on 7-July-2013.

Conclusion:

K. Meshram and H.S. Goliya (2013) presented an analysis of accidents on a small portion of NH-3 Indore to Dhamnod. The data for analysis is collected for the period of 2009 to September 2011. More accidents occurred in Manpur region by faulty road geometry. The trend of accidents occurring in the urban portion (Indore) is more than 35 % of to rate of total accidents each year. This may be due to high speeds and more vehicular traffic. In the present study area, the frequency of fatal accidents are 2 in a week and 6 for minor accidents in a week. More number of accidents observed in 6 p.m. to 8 p.m. duration because in that time more buses are travels between villages and city. One fatal and five casualties are occurring per km per year in the study area. The volume of the trucks passing through study corridor is increasing by year. At Rajendra Nagar from 2000 onwards the traffic is reduced due to the construction of by passes in that area.

3rd Published paper: Systematic approach for formulation of a road safety improvement program in India.

Authors:

- Rakesh Mehar
- Pradip Kumar Agrawal

Publisher: Procedia-Social and Behavioral sciences

Issued on: October 2013

Conclusion:

Rakesh Mehar and Pradeep Kumar Agarwal (2013) were highlighted the deficiencies in the present state of the art and also presents some basic concepts so that a systematic approach for formulation of a road safety improvement program in India can be developed. The study presents basic concepts to develop an accident record system, for ranking of Safety hazardous locations, for identification of safety improvement measures and to determine priorities of safety measures. It is expected that this study will provide a systematic approach for development of road safety improvement program in India and thus pave the way for improving safety on Indian roads

4th Published paper: Random parameter model for accident prediction.

Authors:

- R.R. Dinu
- A. Veeraragavan

Publisher: Journal of safety research

Issued on: September-2011

Conclusion:

R.R. Dinu, A. Veeraragavan (2011) was presented Random Parameter Models for Accident Prediction on Two-Lane Undivided Highways in India. Based on three years of accident history, from nearly 200 km of highway segments, is used to calibrate and validate the

models. The results of the analysis suggest that the model coefficients for traffic volume, proportion of cars, motorized two-wheelers and trucks in traffic, and driveway density and horizontal and vertical curvatures are randomly distributed across locations. They have concluded with a discussion on modeling results and the limitations of the present study.

Book: Traffic engineering and transport planning

Publisher: Khanna publisher (seventh edition-2012)

Author:

- Dr. L.R. Kadiyali

Abstract

India's transport system has several deficiencies such as inadequate capacity, poor safety record, emission of pollutants, and outmoded technology. But as the economy is poised for big growth in the coming years, transportation engineers will have to come up with innovative ideas.

The book addresses these issues, and it is hoped that the engineering students studying transportation engineering will have a clear idea of the problems involved and how they can be overcome in their professional careers. This book has been designed keeping in mind the latest syllabi of all major Indian Universities.

Efforts have been made to include the latest developments in different areas of transportation engineering and minimize errors. The authors sincerely welcome constructive criticisms and suggestions for further improvement.

3.2 Proposed System

Classification techniques will be used for identifying the accident-prone areas. The accident data records can help to understand the characteristics of many features like driver's behavior, roadway conditions, light condition, weather conditions, and so on. This can help the users to compute the safety measures which is useful to avoid accidents. The data set can be analyzed based on Decision Trees, by comparing Gaussian naïve Bayes, Logistic regression, and Random Forest algorithms we are analyzing which algorithm will give an accurate dataset. The models are performed to identify statistically significant factors which can be able to predict the probabilities of crashes and injuries that can be used to perform a risk factor and necessary analysis on how to reduce the damage or totally prevent the accidents from happening.

3.3 Project Scheduling

A comprehensive process that outlines the project phases, tasks under each stage, and dependencies is known as project scheduling. It also considers skills and the number of resources required for each task, their order of occurrence, milestones, interdependencies, and timeline. Furthermore, it involves analyzing the resource availability and implementing the scheduling technique to ascertain timely delivery while maintaining the resource health index. Many project managers successfully generate the right schedule, yet most of them find it challenging to manage the resources intelligently. It can cause delays and

discrepancies in the deliverables as their talent pool is responsible for executing these tasks. Thus, they must master each aspect of project planning and scheduling.

The internal team conflicts are minimized when the entire team are on the same page. Resources are aware of the task dependencies and work diligently to ensure that the overall delivery is not affected. When team opt for a sophisticated project scheduling software, they get real-time updates on every project metric, which promotes proactive planning, monitoring, and coherent risk management.

Chapter 4

System Requirement Specification

4.1 System Specification

Software Requirements Specifications is the official statement of what is required of the system developers. It includes both user requirements and a detailed specification of the system requirements. Requirement analysis is done in order to understand the problem the software system is to solve.

4.1.1 Hardware Requirements

Hardware requirements give the physical component required for the proposed system. The hardware requirement includes a system with the following configurations:

1. Processor: i3 7thgen or above
2. Display Type: VGA and higher.
3. RAM: 4 GB or above
4. Storage Memory: 5 GB or above

4.1.2 Software Requirement

The Software Requirements Specification is produced at the culmination of the analysis task. The function and performance allocated to software as part of system engineering are refined by establishing a complete information description, a detailed functional description, a representation of system behavior, an indication of performance requirements and design constraints, appropriate validation criteria, and other information pertinent to requirements. The various software requirements of the system are summarized here:

1. Operating system: Windows 7/8/10/11, Linux.
2. System Type: 64-bit/32-bit operating system.
3. Web browser: chrome/Microsoft edge.
4. software: XAMPP, Anaconda, Visual Studio Code.
5. Language: Python, HTML/CSS, SQL.

4.2 Functional Requirements

- Divided the dataset into a training dataset and a testing dataset.
- Find the relation between the Data points.
- Using a classification model system divides the data into one similar class or cluster.
- Using a decision model system takes the decision to which class a particular entity belongs.
- System must perform data pre-processing to remove redundant and unwanted data.

4.3 Non- Functional Requirements

- **Privacy**: Privacy should maintain throughout the system.
- **User Friendliness**: GUI should be user-friendly.
- **Responsive:** The system should give response to the request quickly and accordingly.

Chapter 5

System Architecture

5.1 System architecture

A system architecture or systems architecture is the conceptual model that defines the structure, behavior, and more views of a system. The figure 5.1 shows the working of Road accident analysis architecture it consists of a database Machine learning model and web application for user interaction.

Registration and Login: In this module, we have designed registration and login pages for the users and admin.

Data Collection and Preprocessing: In the data collection, we gather the data from a dataset. In the data preprocessing, we perform the following operations on the dataset: Data cleaning, Data Integration, Data transformation, and Data reduction.

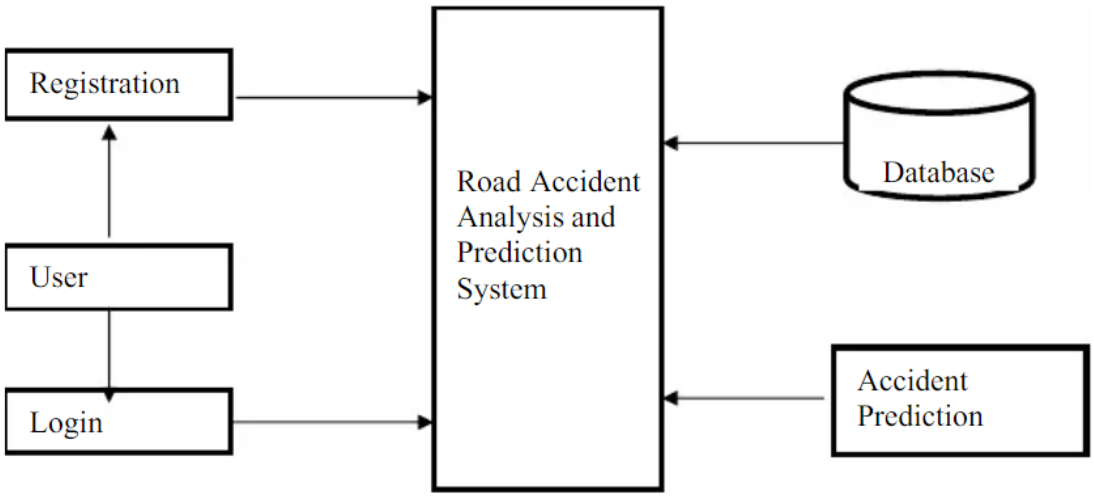

Fig.5.1
(System architecture) [5.8][10]

Feature Selection and Feature Extraction: Feature selection is the process of choosing a subset of the original dataset.
Feature extraction is the process of getting useful features from the existing dataset.

Visualization: In the module, we perform visualization on the dataset in most appropriate way.

Database: Inside the Database, all the files and previous road accident data/files are stored. A database is an organized collection of data that it can be easily accessed and managed. The main purpose of the database is to operate a large amount of information by storing, retrieving, and managing data.

5.2 Data Flow Diagram

A data-flow diagram (DFD) is a way of representing the flow of data of a process or a system (usually an information system). The DFD also provides information about Road Accident Analysis using Machine Learning outputs and inputs of each entity and the process itself. A data-flow diagram has no control flow, there are no decision rules and no loops. Specific operations based on the data can be represented by a flowchart. There are several notations for displaying dataflow diagrams. For each data flow, at least one of the endpoints (source and/or destination) must exist in a process. The refined representation of a process can be done in another data-flow diagram, which subdivides this process into sub-processes. The dataflow diagram is part of the structured analysis modeling tools.

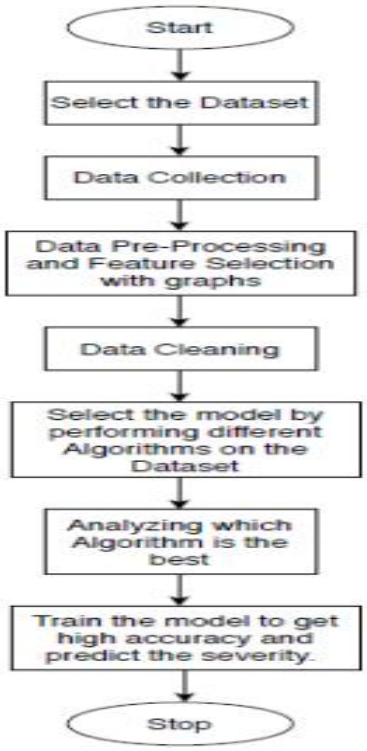

Fig-5.2
Data flow diagram [5.8][10]

5.3 Methodology

Input Dataset: First we must collect the previous accident data set from different trusted sources and analyze the dataset. The dataset we have gathered may have a different format but in the data preprocessing step we convert all data sets into a common format so we can use them to build our model.

Data preprocessing:

It is a process of converting raw data into a use full data so that data operations can be performed. The data preprocessing includes converting all datasets gather into a single common format that can be used for performing operations and the main purpose of the data preprocessing is to handle missing values to increase the accuracy and the performance of the model. There are several ways to handle missing values:

- Drop Rows and columns with Missing Values.
- Replacing With Arbitrary Value.
- Replace missing values with the mean values.
- Manually fill in the missing values.
- Filling the missing values with the most probabilistic value

Data preprocessing is one of the most important steps in model building and also the first step of model building.

Feature selection:

Feature selection plays the most important role in the ML model building while developing the machine learning model, only a few variables in the dataset are useful for building the model, and the rest features are either redundant or irrelevant. If we input the dataset with all these redundant and irrelevant features, it may negatively impact and reduce the overall performance and accuracy of the model. Hence it is very important to identify and select the most appropriate features from the data and remove the irrelevant or less important features, which is done with the help of feature selection in machine learning. "Feature selection is a way of selecting the subset of the most relevant features from the original features set by removing the redundant, irrelevant, or noisy features.

Feature Selection is the method of reducing the input variable to your model by using only relevant data and getting rid of noise in data. It is the process of automatically/manually choosing relevant features for your machine-learning model based on the type of problem you are trying to solve. We do this by including or excluding important features without changing them. It helps in cutting down the noise in our data and reducing the size of our input data. Benefits of Feature Selection:

(1) Using unnecessary feature variables for the prediction can deteriorate the performance of a predictive model. Thus, feature selection helps in improving the model performance.

(2) Algorithms like linear regression and logistic regression must avoid using correlated features. Using feature selection methods thus leads to a better fit of these models.

(3) It is an excellent practice to work with a minimum set of predictive modeling features as they significantly reduce the algorithm's complexity and computational costs.

Target attribute and Range selection:

The target of a supervised model is a special kind of attribute. The target column in the data contains the historical values used to train the model. The target column in the test data contains the historical values to which the predictions are compared. The act of scoring produces a prediction for the target.

Feature scaling is the final step in machine learning data preprocessing. It is a method for standardizing the independent variables of a dataset within a given range. In feature scaling, we place our variables in the same range and scale so that no one variable dominates the other.

Dimension Reduction:

The number of input variables or features for a dataset is referred to as its dimensionality. Dimensionality reduction refers to techniques that reduce the number of input variables in a dataset. More input features often make a predictive modeling task more challenging to model, more generally referred to as the curse of dimensionality. High-dimensionality statistics and dimensionality reduction techniques are often used for data visualization.

Test and training Dataset:

After the Data preprocessing, Dimension reduction, and scaling, we divide the data into two different parts first part we call it test dataset, and second part we call it test data set. Usually, the data set is divided in an 80:20 ratio but anyone can choose a different ratio the only condition for building an accurate model is that is that all the different data scenarios are present in the train and test datasets. The train dataset is used to build the model and the test dataset is used to test the accuracy of the model we have built.

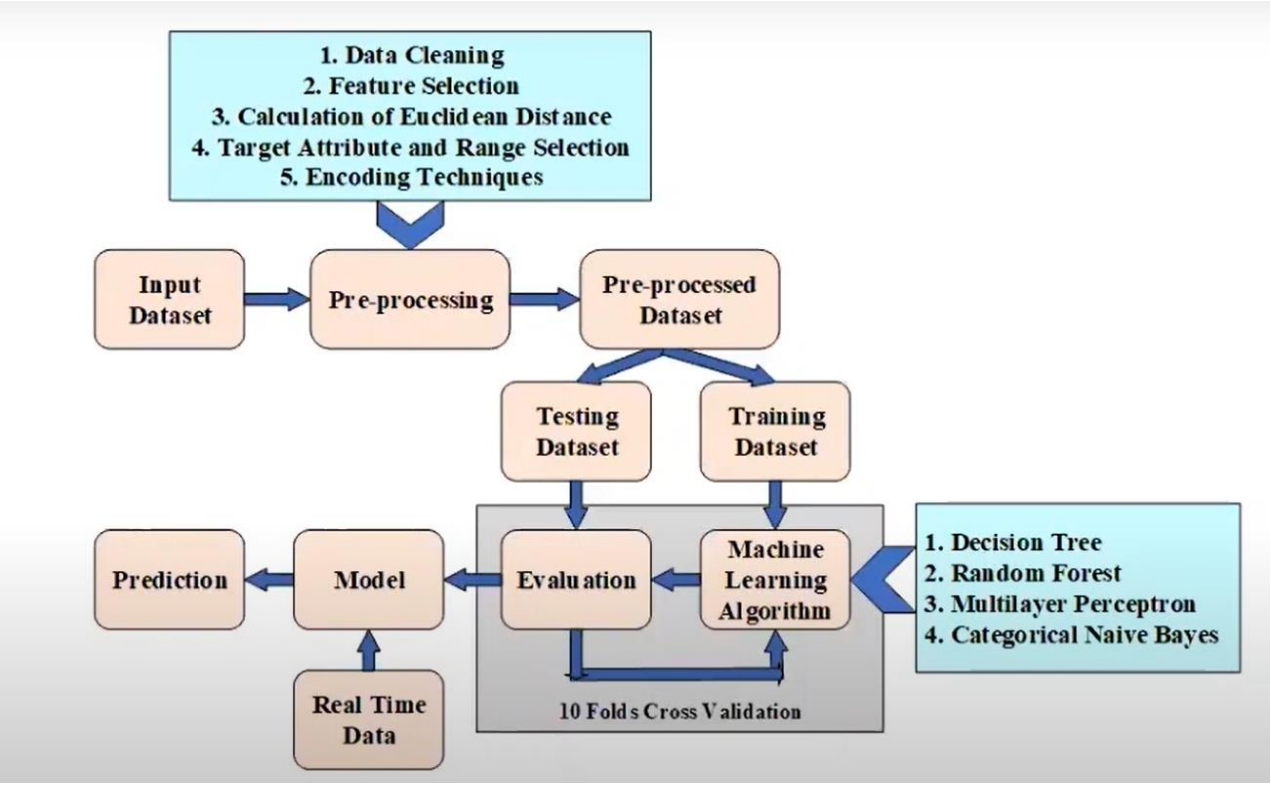

Fig-5.3 Methodology [5.8][15]

5.4 Algorithms

We are going to implement ML algorithm with the train data set to create the road accident ML model with high accuracy. To obtain the best accuracy and the performance we are going to compare the different popularly use ML algorithm and choice the one with higher accuracy to obtain the best Feasible model. Below are the some of the ML algorithm

Algorithm	Type	Accuracy	Precision
Decision Tree	Accident	99.77%	98.68%
	Severity	99.80%	99.83%
Random Forest	Accident	99.55%	99.74%
	Severity	99.56%	99.64%
Multilayer Perceptron	Accident	99.77%	99.19%
	Severity	99.82%	99.77%
Categorical Naive Bayes	Accident	93.85%	91.69%
	Severity	97.84%	98.34%

Fig-5.4 Algorithm Accuracy Comperision[5.8][14]

Gaussian Naive Bayes algorithm

It is a classification technique based on Bayes' theorem with an assumption of independence between predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

In the context of machine learning, naive Bayes classifiers are known to be highly expressive, scalable, and reasonably accurate, but their performance deteriorates rapidly with the growth of the training set. A number of features contribute to the success of naive Bayes classifiers. Most notably, they do not require any tuning of the parameters of the classification model, they scale well with the size of the training data set, and they can easily handle continuous features.

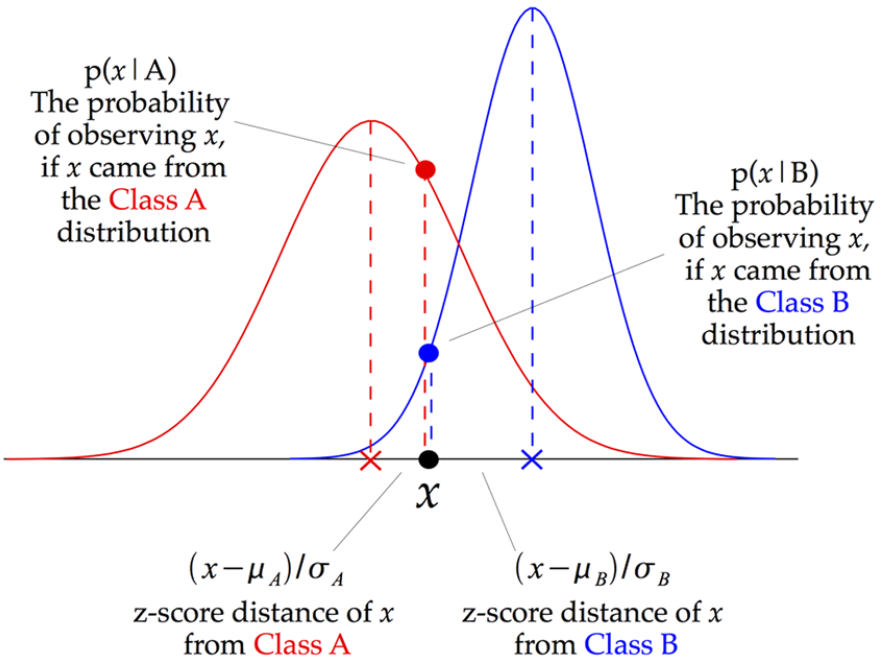

Fig-5.5 [5.8][14]

Logistic regression algorithm

Logistic regression is a classification technique borrowed by machine learning from the field of statistics. Logistic Regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The intention behind using logistic regression is to find the best fitting model to describe the relationship between the dependent and the independent variable.

It uses a logistic function to model the dependent variable. The dependent variable is dichotomous in nature, i.e. There could only be two possible classes (e.g.: either the cancer is malignant or not). As a result, this technique is used while dealing with binary data.

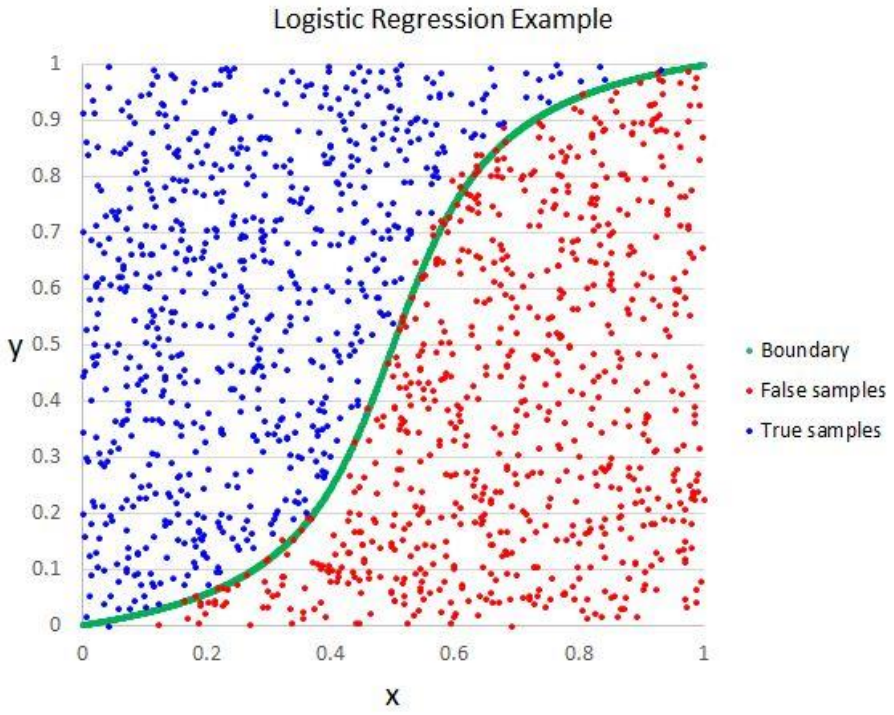

Fig-5.6 Logistic regression algorithm [5.8][14]

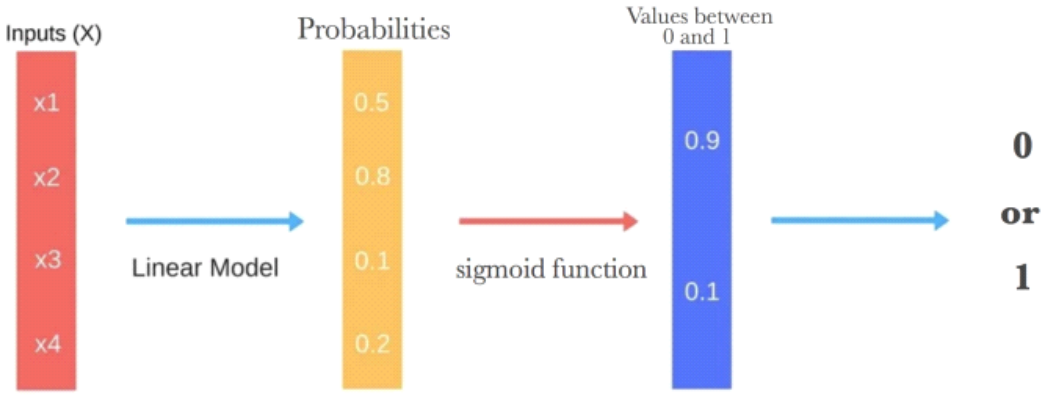

Fig-5.7 Logistic regression algorithm [5.8][14]

Random Forest generator

Random Forest is a supervised learning algorithm which is used for both classification and as well as regression. This algorithm creates Decision Trees on data samples and then gets prediction from each of them and finally selects the best solution by means of voting. It is a better algorithm because it reduces the over-fitting by averaging the result.

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

The below diagram explains the working of the Random Forest algorithm:

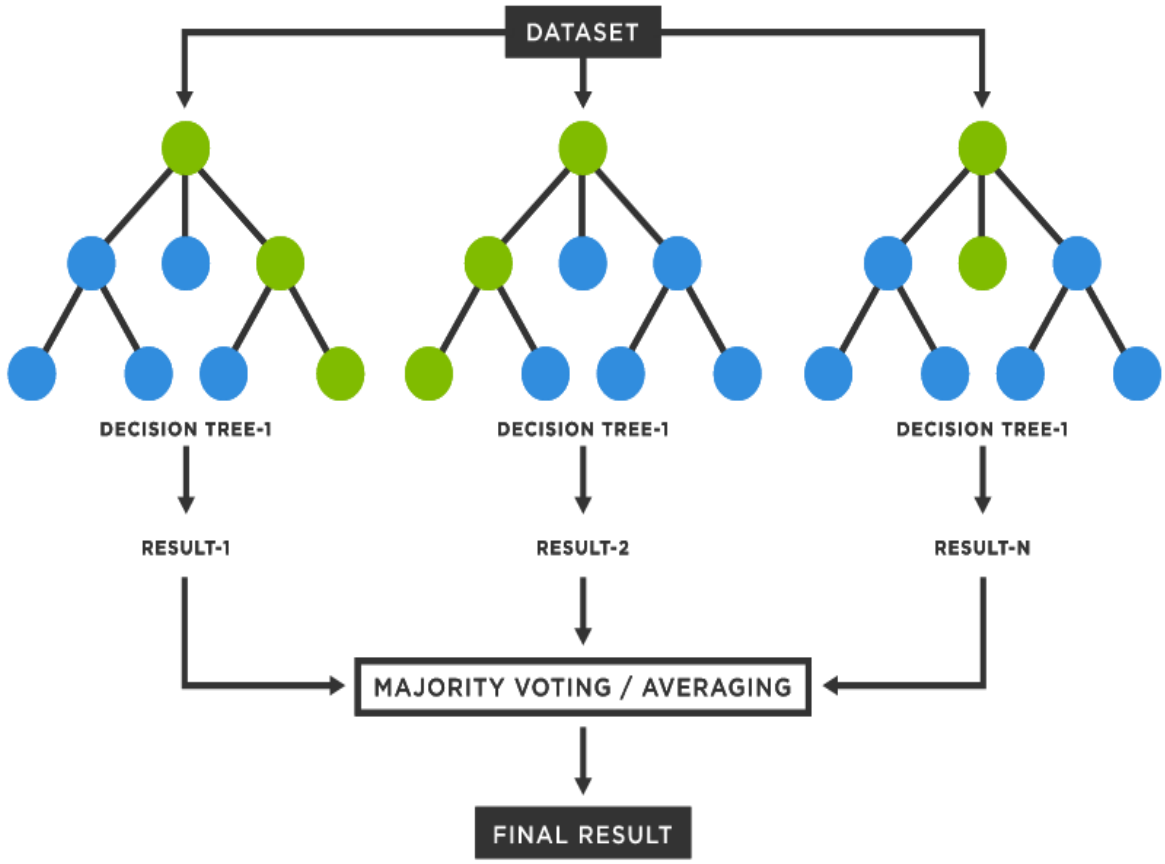

Fig 5.8 [5.8][10]

Model: Now we apply the selected ML algorithm with the train data set to create the road accident ML Model then we evaluate the model by using the test dataset.

5.5 Deployment Model

A deployment diagram shows the configuration of run-time processing nodes and the components that live on them. Deployment diagrams address the static deployment view of an architecture. They are related to component diagrams in that a node

typically encloses one or more components.

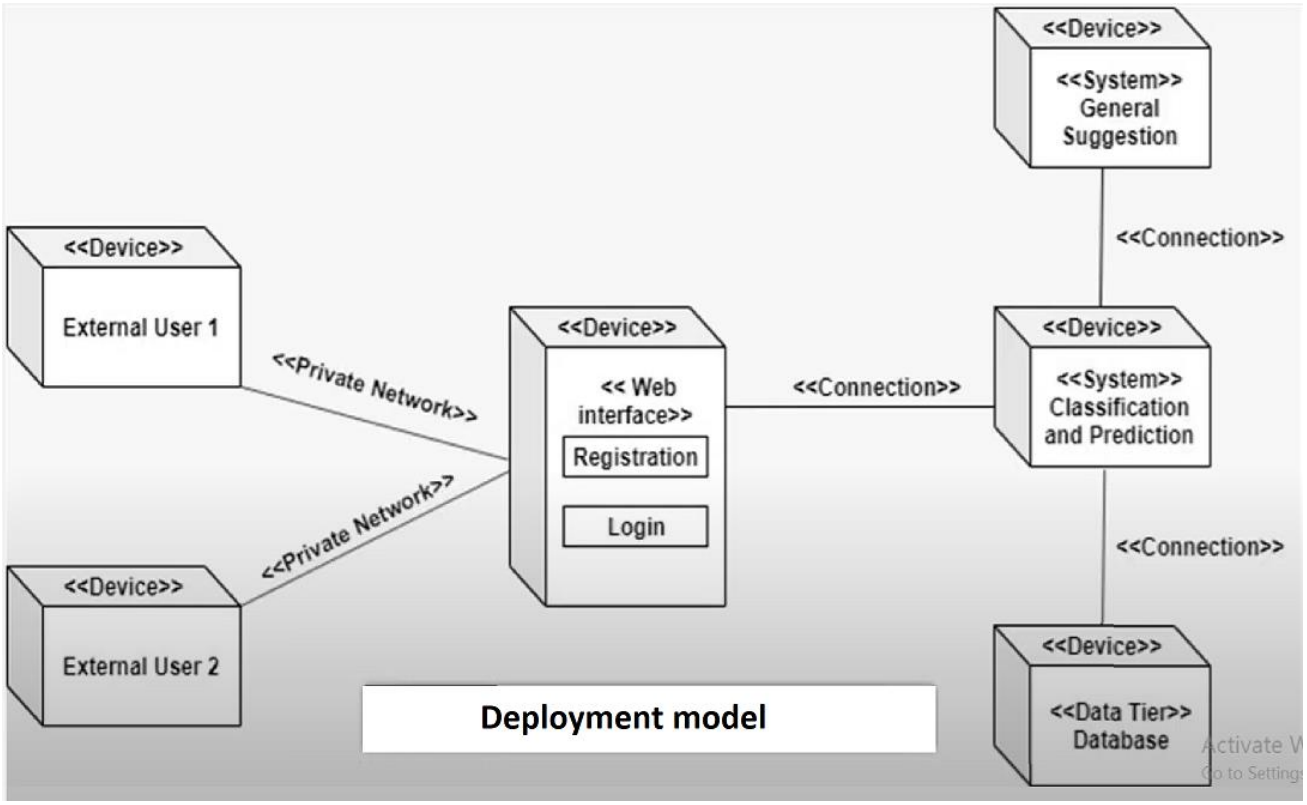

Fig-5.9
Deployment model [5.8][15]

5.6 Use case diagram:

A use case diagram is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of users the system has and will often be accompanied by other types of diagrams as well

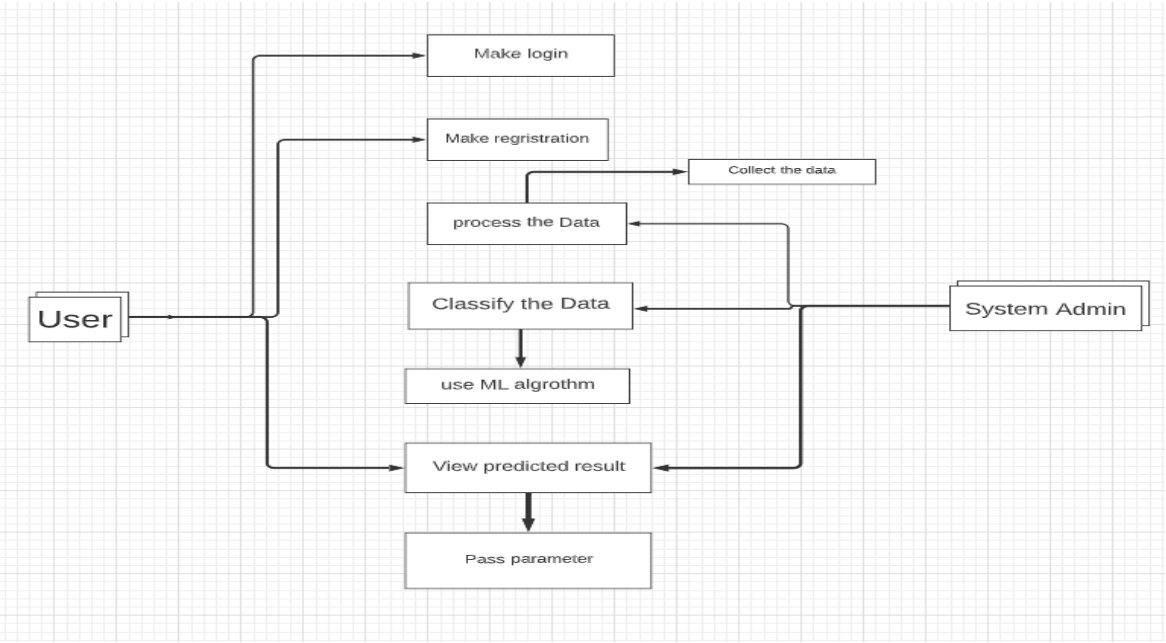

Fig 5.10 Use case diagram [5.8][15]

5.7 Conclusion

Road Accidents are caused by various factors. By going through all the research papers, it can be concluded that Road Accident cases are hugely affected by the factors such as types of vehicles, age of the driver, age of the vehicle, weather conditions, road structure, and so on. Thus, we have proposed an application that gives an efficient prediction of road accidents based on the above-mentioned factors. Due to analyzing and severity prediction, we can reduce road accidents by taking some precautions before the occurrence of accidents

5.8 References

1. Michael Williamson, and Huaguo Zhou "Development calibration factor for crash prediction model for rural two-lane roadways in Illinois".

2. K Meshram, and S.H Goliya "Accident analysis on national highway 3 between Indore to Dhamnod".

3. Rakesh Mehar, and Pradip Kumar Agrawal "Systematic approach for formulation of a road safety improvement program in India".

4. R.R. Dinu, and A. Veeraragavan "Random parameter model for accident prediction".

5. Ashwin Raj, solving classification **problems** using logistic regression. Published on https://towardsdatascience.com/

6. Road accident analysis in Tamil Nādu from January 2020 to July 2020. Published on https://www.tnsta.gov.in/

7. No. of traffic accidents in India. published on https://data.gov.in/

8. Ramesh kasimani, Road traffic accident and related factors. Published on https://www.researchgate.net/

9. Thu huyen, Road risk analysis tools. Published on https://www.researchgate.net/

10. www.google.com/image/

11. https://en.wikipedia.org/wiki/Accident_analysis

12. Ijraset.com was first indexed by google https://www.irjet.net/archives/V7/i12/IRJET-V7I12129.pdf

13. Java point **https://www.javatpoint.com/types-of-machine-learning**

14. https://www.youtube.com/watch?v=XI0619WGa_o&ab_channel=Zerinjahan

15. https://www.lucidchart.com/pages/examples/flowchart-maker