# A Novel Influence of Traffic Density on Road Safety, A Study of Accidents and Fatality Rates in Urban Regions helps Anticipate Future Traffic Patterns and Trends

**K. R. Sekar[1], P. Muthusamy[2] , K. Nafees Ahmed[3] , V. Venkataraman[4], V.Rengarajan[5], G. Sathiamoorthy[6]**

[1]School of computing, SASTRA Deemed University, Thanjavur 613401, Tamil Nadu, India.
[2]Cyber security and MCA, Paavai Engineering College (Autonomous), Namakkal 637 018, india.
[3]Department of Computer Science, Jamal Mohamed College (Autonomous), Bharathidasan University, Tiruchirappalli, India-620020.
[4]School of SASHE, SASTRA Deemed University, Thanjavur 613401, Tamil Nadu, India.
[5]School of Management, SASTRA Deemed University, Thanjavur 613401, Tamil Nadu, India.
[6]School of SASHE, SASTRA Deemed University, Thanjavur 613401, Tamil Nadu, India.

(Corresponding Author: V.Venkataraman, mathvvr@maths.sastra.edu)

**Abstract:** Road traffic congestion and accidents continue to be significant challenges for transportation systems around the world. Traditional methods for managing and preventing these issues are often reactive and costly. Machine learning methods have recently emerged as a promising approach for predicting and mitigating traffic congestion and accidents. This paper provides an overview of machine learning techniques that have been applied to these problems, including classification, regression, clustering, and deep learning. We discuss the data sources and preprocessing steps required for effective machine learning, as well as the challenges of interpreting and deploying these models in real-world settings. Finally, we present several case studies that illustrate the potential of machine learning in improving road safety and reducing traffic congestion. This type of analysis can provide insight into the underlying dynamics of congestion, such as how traffic jams form and dissipate and how different driving behaviors affect traffic flow.Recent advancements in data collection technologies, such as GPS and in-vehicle sensors, have enabled researchers to gather large-scale data on road traffic trajectories. These data sources provide a wealth of information for analyzing traffic patterns and identifying factors that contribute to congestion.One important aspect of analyzing traffic trajectories is the identification of bottlenecks or chokepoints on the road network. These are areas where traffic flow is restricted, leading to congestion and slower travel times. By understanding the characteristics of these bottlenecks, transportation planners and engineers can develop targeted interventions to alleviate congestion and improve traffic flow.Another important area of research is the analysis of driver behavior and its impact on traffic flow. For example, studies have shown that aggressive driving behaviors, such as frequent lane changing and sudden braking, can contribute to traffic congestion. By understanding the factors that influence driver behavior, policymakers can develop interventions that encourage safer and more efficient driving practices.

**Keywords:** Global positioning system, Machine learning, Classification, Congestion, In-vehicle sensors, Trajectories

## 1. Introduction

Machine learning is a subfield of artificial intelligence that focuses on developing algorithms and models that can learn patterns and make predictions from data without being explicitly programmed. One common type of machine learning is supervised learning, where an algorithm learns from a labeled dataset to make predictions on new, unseen data.Random forest, K-nearest neighbors (KNN), decision tree, and AdaBoost classifiers are all supervised learning algorithms commonly used in machine learning.Random forest is an ensemble learning method that creates multiple decision trees and combines their outputs to make a prediction. Each tree is trained on a random subset of the data, and the final prediction is the majority vote of the individual tree predictions.K-nearest neighbors (KNN) is a simple classification algorithm that works by finding the k closest points in the training dataset to a given point and assigning the point to the class that occurs most frequently among its k nearest neighbors. Developed a line tracking system consisting of a model predictive control device and an active safety steering control device that simulates and estimates the model predictive control system (MPC) that was recommended and constructed as a predictive model. A 4-DOF vehicle model was constructed to represent the mechanical dynamics and eliminate bus rollover accidents. In the end, we found that the desired line shading control could be implemented quickly and cheaply [1]. This study uses wireless communication to address the problem of optimal route planning. Describe functional terrain for mobile independent guided vehicles using geometric styles and rerouting algorithms. The effectiveness of the system described in this study is demonstrated by the fact that the route length remains constant even when he repeats the route planning more than ten times [2]. The robustness of the compensable controller is evaluated using Kharitonov's theorem and varies with the performance-tuned compensable controller. Sliding mode control using power velocity exponential range law is more robust than other controllers due to faster response time, more than 84-course angle deviation from the line, and only 3.2 lane change intersections [3]. Artificial neural networks with long short-term memory (LSTM) roots should be more effective and accurate. We can quickly and directly separate circuits with similar paths by integrating an unsupervised line separation medium into her LSTM. Advanced LSTM algorithms create boat routes using these datasets. Results show that the modified LSTM outperforms his LSTM when the computed root circle is shorter [4]. This section begins by emphasizing the importance of trajectory data sets, especially for next-generation simulations (NGSIM), in recent traffic research. New traffic phenomena and theories are explained at the microscopic, mesoscopic, and macroscopic levels. They delve into these trends from different perspectives of traffic flow studies [5]. This article outlines successful communication methods for ground, stationary, and floating vehicles in space.This work comprehensively describes line planning and provides optimization techniques applicable to land, sea, and air operations. For arbitrary dimensions, this work examines numerical, biologically-inspired, mixed approaches [6]. Decision trees are a type of supervised learning algorithm that builds a tree-like model of decisions and their possible consequences. The tree is constructed by recursively splitting the data into subsets based on the value of a selected feature until a stopping criterion is met. A collaborative learning approach called AdaBoost (Adaptive Boosting) combines a set of weak classifiers to create a strong classifier. Each poor classifier is assigned a weight based on its performance when trained on a data item. The weighted average of the weak classifier's predictions is the final prediction..

In summary, these classifiers are all supervised learning algorithms that make predictions based on labeled data. Random forest and AdaBoost are ensemble learning methods that combine multiple models to make predictions, while KNN and decision tree algorithms rely on nearest neighbors or recursive splits, respectively, to classify new data points.

### *1.1 Our Contribution to the research work*

- Infrastructure improvement: The quality of the road and traffic infrastructure plays a crucial role in reducing congestion and accidents. Improving road design, installing traffic signals and signs, and adding more lanes can help manage traffic flow.

- Public transportation: Encouraging the use of public transportation can help reduce the number of vehicles on the road, thereby reducing congestion and accidents. Governments can also provide incentives to use public transportation, such as reduced fares, park-and-ride facilities, and dedicated bus lanes.

- Education and awareness: Education and awareness campaigns can help promote safe driving practices and reduce the likelihood of accidents. This can include initiatives like driver education programs, campaigns to encourage the use of seat belts, and awareness campaigns around distracted driving.

- Technology integration: The integration of technology into transportation systems can help reduce traffic congestion and accidents. This can include the use of intelligent traffic management systems, autonomous vehicles, and connected vehicle technology that can provide real-time traffic updates and alerts.

- Law enforcement: Stronger law enforcement and penalties for traffic violations can help discourage unsafe driving practices and reduce the likelihood of accidents. This can include measures such as strict enforcement of speed limits, red-light cameras, and penalties for distracted driving.

Overall, reducing road traffic congestion and accidents requires a multi-faceted approach that involves improvements in infrastructure, transportation systems, education, technology, and law enforcement.
In addition, while simultaneously recording vehicle viscosity data, it is possible to operate signals according to the amount of work on each lane. Once this data is analyzed, the capabilities of the machines can be used to predict future numbers of buses. Bus presence is detected and responded to by an intelligent store lighting system. Planned smart business outcomes allow commuters to save time and energy by staying lower. This model controls the business lights based on the business flow by adapting to the duration of the signal, detecting the presence of the car and performing the necessary actions. The planned smart business system will help commuters save time and energy by reducing delays. This model manages commercial lights by changing the length of the lights according to the amount of work.

For the research work, the dataset has taken from the Kaggle site:
https://www.kaggle.com/datasets/daveianhickey/2000-16-traffic-flow-england-scotland-wales

The rest of the sections are Related work, Methods and materials, Results and discussions, Conclusion and References.

### *1.2 Novelty of the Research Work*

The novelty of the work focus on various strategies to address traffic congestion and reduce accidents on the road. These include improving infrastructure, promoting public transportation, educating drivers on safe driving practices, integrating technology into transportation systems, and strengthening law enforcement. While these strategies are not new, their combination and integration can provide a novel approach to addressing the complex issue of traffic congestion and accidents. Additionally, the success of these strategies depends on their implementation and sustained effort by governments, transportation agencies, and the public.

## 2. Related work

A model-free adaptive control system (MFAC) is used in combination with a radial basis function and neural network (RBFNN) to improve the line tracking accuracy of a surface-to-surface vehicle manipulator system (UVMS) under unknown ocean current disturbance conditions. I was. Using the Adams and Simulink co-simulation model, we investigated line shading performance with different control approaches such as Commensurable Integral Secondary (PID), MFAC, and RBFNN-MFAC. This study shows good line shadow performance without using a realistic dynamic UVMS model. This is important for engineering operations [7].

The study describes the roboticization of car drivers in driving control of self-driving cars in lane traffic for lane keeping and obstacle avoidance. As a result, the vehicle's intended line is constantly adjusted based on the driver's actions and intentions. The new driving control framework greatly reduces the problem of driver-robotic conflicts and gives drivers more flexibility to change direction within a set line. The benefits of the proposed technique were evaluated using experimentally obtained objective and subjective data from several human drivers using an innovative interactive dynamic driving simulator [8].

The main objective of this study is to ensure safe driving while reducing delays at unsignalized intersections. First, the Intersection's operating system acts as a word capture and organization center, assigning appropriate priorities to all existing cars and mapping their routes. The results show that the proposed TP-AIM media significantly shortens mean evacuation time and improves performance by more than 20%. Additionally, studies have addressed follow-through at intersections, which can be reduced to less than ten compared to conventional lightweight surgical systems [9].

Business traffic is causing problems all over the world. This research aims to explore and evaluate data mining and machine learning techniques advocated for exploration and persistence to address direct and cyclical business challenges affecting humanity and society. This research focuses solely on methods of business operations that rely on data mining and machine learning techniques to describe and predict business. This research is important to the corporate research community, software companies, and government agencies. This directly paves the way for new operating concepts in the company [10].

The fast development of V2X communication has enabled the optimization and management of vehicle trajectories in terms of the overall traffic flow, hence enhancing traffic performance. We create a two-phase approach that combines upper evolution techniques with lower dynamic programming to decrease stochastics and computations while solving trajectory optimization models. The two-step technique may be effectively applied in practice by setting proper algorithm parameters to obtain a viable approximation solution for trajectory optimization [11].

Avoiding and mitigating possible vehicle traffic dangers can help guarantee road safety. We evaluated relevant literature extensively to determine the background and status of research in road traffic risk reduction and control, and we highlighted major research topics that require additional investigation. His three study priorities were the dangers of driving at junctions and the location of safe routes. Further study is needed on the link between automated driving technology and safe driving theory, as well as the features of hybrid human-machine traffic flow [12].

The Internet of Things (IoT) is a new paradigm that has transformed traditional living into a high-tech lifestyle. IoT is transforming smart cities, smart homes, environmental protection, energy savings, smart transportation, and smart industries. However, some other obstacles and concerns must be addressed before the full promise of IoT can be realized. These difficulties and issues must be considered through the lens of numerous IoT components, like as applications, challenges, enabling technology, and social and environmental implications. The primary goal of this review article is to give an in-depth explanation from both a technical and sociological standpoint. This article delves into a variety of obstacles and critical topics in IoT, architecture, and major application fields [13].

This study describes a combined active steering control (ASC) and direct yaw control (DYC) technique for improving lane-following performance in cars with high tire-to-ground adhesion and road curvature. to disperse the four running wheels. Finally, the simulation results validate the practicality and efficiency of the provided strategies and approaches in contrast to the ASC strategy, especially in the presence of unknown tire-ground adhesion and varying road curvature [14].

This research aims to provide explicit control methods for the robust global exponential stabilization of generic uncertain-time discrete acyclic networks. Consider a discrete-time unsafe network model with very flimsy assumptions. This paper shows that the latter criteria are required for a robust global exponential stabilizer of the network's intended non-congestion equilibrium point. It is connected to the observed function [15].

## 3. Methods and Materials

Machine learning is a field of computer science that deals with the development of algorithms and models that enable computers to learn from data and make predictions or decisions based on that data. Here are some key concepts in machine learning:

### 3.1 Basic machine learning concepts

Supervised learning is a sort of machine learning where the model is trained on labelled data with known right answers. Based on the patterns it discovered from the labelled data, the model learns to predict the right result for fresh inputs. Unsupervised learning is a sort of machine learning in which the model is trained on data that has not been labelled and the desired result is unknown. Without being explicitly instructed what to search for, the model learns to spot patterns or structures in the data.

**Reinforcement Learning:** This is a type of machine learning where the model learns by trial and error through interaction with an environment. The model receives rewards or punishments for its actions and learns to maximize the rewards over time.

**Deep Learning:** This is a type of machine learning that uses neural networks with many layers to learn from data. Deep learning has been very successful in tasks such as image and speech recognition.

**Feature Extraction:** This is the process of selecting or extracting the most important features or variables from the data, which are then used as inputs to the machine learning model.

**Model Evaluation:** This is the process of measuring how well a machine learning model is performing. Common metrics for evaluation include accuracy, precision, recall, and F1 score.

**Overfitting:** This occurs when a machine learning model is too complex and learns the noise in the data rather than the underlying patterns. Overfitting can lead to poor performance on new data.

### *3.2 Random forest*

Random forest is a popular machine learning algorithm used for classification and regression tasks. It is an ensemble learning method that combines multiple decision trees to make predictions. Each decision tree is trained on a random subset of the training data and a random subset of the features, which helps to reduce overfitting and increase the diversity of the trees in the forest.

### *3.3 Decision tree*

A decision tree is a type of predictive modeling tool used in machine learning and data mining. It is a graphical representation of all possible solutions to a decision based on certain conditions or variables. Each node in the tree represents a condition, and each branch represents a possible outcome. The decision tree algorithm builds the tree by selecting the best variable at each node that divides the data into the most distinct groups. This process continues recursively until the data is fully partitioned, or a stopping criterion is reached.

### *3.4 AdaBoost*

AdaBoost (short for Adaptive Boosting) is a machine learning algorithm used for classification and regression tasks. It works by combining multiple "weak" classifiers to form a "strong" classifier. A weak classifier is one that performs only slightly better than random guessing, while a strong classifier is one that performs well on the training data.

### *3.5 K Nearest Neighbors*

K Nearest Neighbors (KNN) is a supervised machine learning algorithm used for classification and regression tasks. It is a non-parametric method, meaning that it does not assume any specific functional form for the underlying data distribution. Instead, it relies on the local distribution of data points to make predictions.

## 4. Results and Discussions

Machine learning principles can be very useful in detecting congestion and accidents in road traffic because they can help to identify patterns and anomalies in real-time traffic data that would be difficult or impossible for humans to detect. One way that machine learning can be used in traffic monitoring is through the analysis of sensor data from traffic cameras, road sensors, and other sources. Machine learning algorithms can be trained to recognize patterns in this data that indicate congestion or accidents, such as sudden changes in traffic flow or unusual vehicle speeds.Another approach is to use machine learning to analyze data from sources such as social media, traffic apps, and other sources of real-time traffic information. This data can be used to identify incidents and issues that may be causing congestion or accidents, such as road closures or accidents, and to predict traffic patterns based on historical data.Overall, machine learning can help to improve the accuracy and speed of traffic monitoring and management, which can lead to more efficient and safer transportation systems.

In Table 1 , different classifiers have different strengths and weaknesses, and their performance can vary depending on the dataset, the problem being solved, and the specific implementation. Additionally, some classifiers may be more computationally expensive than others, but this may be acceptable depending on the specific requirements of the problem being solved.Can you please provide more information about the specific problem you are trying to solve and the dataset you are using? This will help me provide more tailored recommendations and information about classifiers.

**Table 1.** Classifiers with their accuracy,precision,recall,F1-score,loss function and computational time

| Algorithms | Accuracy | Precision | Recall | F1-Score | Error-Rate | Loss Function | Computational Time |
|---|---|---|---|---|---|---|---|
| **Random Forest** | 85.30% | 77.39% | 85.30% | 78.74% | 14.69% | 19.14% | 41.19% |
| **Decision Tree** | 85.26% | 77.02% | 85.26% | 78.76% | 14.74% | 19.20% | 66.98% |
| **AdaBoost Classifier** | **85.36%** | **78.41%** | **85.36%** | **78.65%** | **14.63%** | **19.11%** | **20.51%** |
| **KnearestNeighbor** | 83.99% | 76.28% | 83.99% | 78.95% | 16.01% | 20.41% | 22.41% |

### *4.1 Accuracy*

A model's accuracy in machine learning refers to its capacity to accurately categorise or forecast a given data set. It is often represented as the proportion of accurate forecasts to all of the model's predictions. A binary classification model's accuracy, for instance, is 90% if it accurately predicts 90 out of 100 test instances. This statistic is helpful for evaluating the performance of models, particularly when the classes are balanced (eg., each class has approximately the same number of examples). Yet, accuracy might not be enough to evaluate a model's performance in its whole. When the data set is unbalanced or there are numerous classes of data, high accuracy does not always indicate that the model is operating effectively various degrees of importance. In these situations, additional measures can be utilised to offer a more thorough evaluation of the's performance, including accuracy, memory, and F1 score.

### *4.2 Precision*

Precision in machine learning is a performance metric that measures the proportion of true positives (correctly classified positive instances) to the total number of instances predicted as positive (true positives + false positives). In other words, it is a measure of how accurate the model is when predicting positive instances.

precision = true positives / (true positives + false positives)

In binary classification jobs, where it is important to reliably discriminate between positive and negative occurrences, precision is frequently utilised. Due to the model's great accuracy, very few false positive predictions are made, therefore if it correctly predicts a positive occurrence, it almost always is. The percentage of true positives that the model properly detects, nevertheless, might suffer when accuracy is high. For best performance, a suitable model balances the precision and memory of the.

### *4.3 Recall*

Backtracking in machine learning refers to a model's ability to correctly identify all instances of a given class in a dataset. In the context of traffic congestion and accidents, memory can be used to measure the accuracy of a machine learning model in predicting when these events will occur.

Recall = true positives / (true positives + false negatives)

For example, a machine learning model can be trained on historical traffic data to predict the likelihood of traffic congestion on a given day and time. The model can be evaluated based on its recall in correctly identifying instances of traffic congestion. Similarly, a machine learning model can be trained on accident data to predict the likelihood of an accident occurring on a particular road segment. The model can be evaluated based on its recall in correctly identifying instances of accidents.

By improving the recall of machine learning models in predicting traffic congestion and accidents, transportation planners and authorities can take proactive measures to prevent these events from occurring. This can include implementing traffic management strategies, improving road infrastructure, and enhancing safety measures.

### *4.4 F1-Score*

In machine learning, particularly traffic volume models and accident prediction, the F1 score is a common assessment statistic. It is a metric for model correctness that considers both precision and memory. Accuracy in the context of traffic congestion and accidents is defined as the ratio of the total number of correct positive predictions generated by the model to the number of genuine positives (i.e., accurately predicted traffic congestion or accidents). However keep in mind that it is the proportion of true positives to all true positives in the sample.

The F1-score is the harmonic mean of precision and recall and is computed as:

F1-score = 2 * (precision * recall) / (precision + recall)

A high F1-score indicates that the model has high precision and recall and is therefore effective in predicting traffic congestion and accidents.In practice, machine learning models for road traffic congestion and accident prediction may use other evaluation metrics as well, depending on the specific needs of the problem at hand. These may include accuracy, AUC-ROC, and mean average precision (mAP), among others.

### *4.5 Error Rate*

Machine learning algorithms may be used to estimate the likelihood of congestion and accidents at particular places and times in the context of congestion and accidents. A machine learning algorithm's error rate is the proportion of inaccurate predictions. While this can assist authorities and drivers in taking preventative actions to minimise congestion and accidents, it is crucial to lower the error rate in models used to anticipate traffic congestion and accidents. Following are some strategies for lowering the error rate in models used to anticipate traffic jams and accidents: • Accurate data gathering The calibre of the data used to train a predictive model has a significant impact on its accuracy. As a result, it's crucial to gather reliable data that is reflective of the issue domain.

- Use appropriate features: The features used in the model should be relevant to the problem domain. For example, when predicting traffic congestion, features such as time of day, day of the week, and weather conditions are important.
- Train the model on a large dataset: The larger the dataset used to train the model, the better the model is likely to perform. This is because a larger dataset provides more examples for the model to learn from, which can help it make more accurate predictions.
- Choose an appropriate machine learning algorithm: Different machine learning algorithms are better suited for different types of problems. For example, decision trees are often used for classification problems, while regression models are used for predicting continuous values.
- Regularly update the model: Traffic patterns and road conditions can change over time, so it is important to regularly update the model with new data to ensure that it continues to make accurate predictions.

Evaluate the model performance: It is important to evaluate the performance of the model regularly to identify any potential errors and make improvements. This can be done by comparing the model's predictions with actual traffic and accident data.

### *4.6 Computional Time*

Computational time is an important consideration in machine learning models for road traffic congestion and accidents. These models often require processing large amounts of data and training complex models, which can be time-consuming and computationally expensive. One approach to reducing computational time is to use more efficient algorithms and techniques. For example, some machine learning algorithms are better suited for large datasets, while others may be better suited for smaller datasets. Additionally, techniques like dimensionality reduction and feature selection can help to reduce the computational burden by reducing the number of variables that need to be processed.Another approach is to optimize the hardware and software used for machine learning. This includes selecting high-performance hardware and using specialized software libraries that are optimized for specific machine learning tasks. For example, GPUs (graphics processing units) can significantly speed up certain types of machine learning computations, and there are software libraries designed specifically for machine learning on GPUs.

In addition to these technical considerations, it's also important to consider the trade-off between computational time and model accuracy. In some cases, more computationally expensive models may lead to better accuracy, while in other cases simpler models may be sufficient. It's important to carefully balance these factors when designing machine learning models for road traffic congestion and accidents.

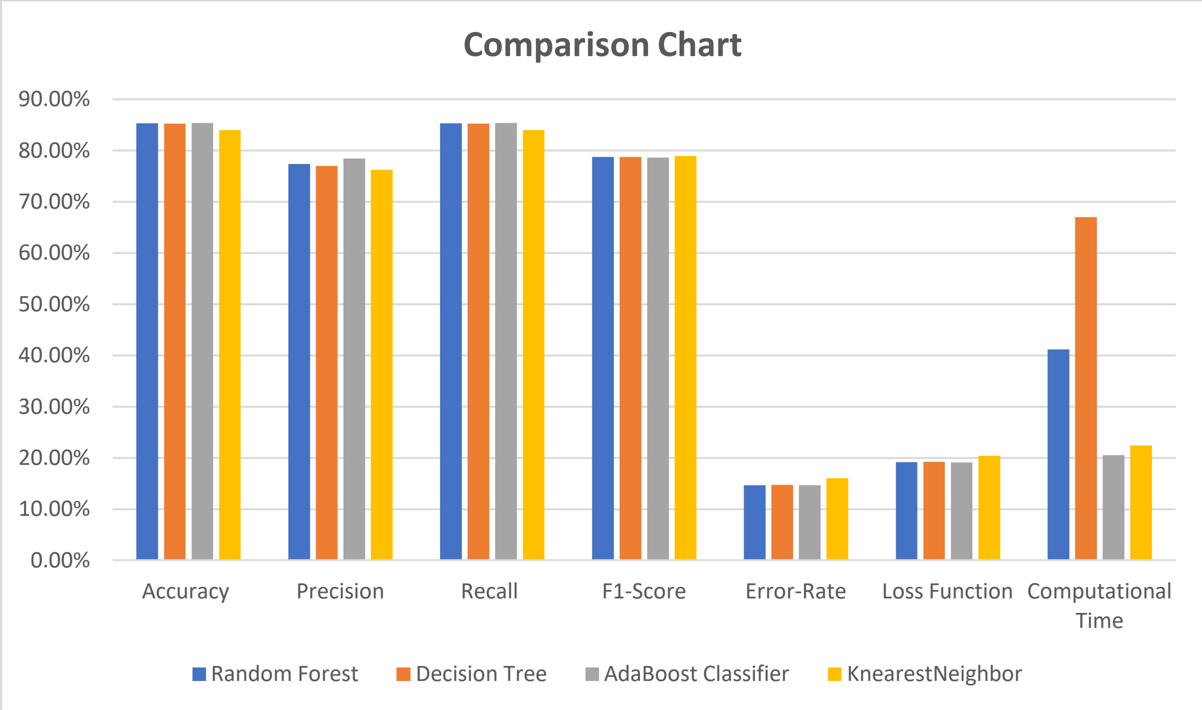


**Figure 1.** Classifiers Comparison of Accuracy,Precision,Recall,F1-score,Loss Function,Error Rate

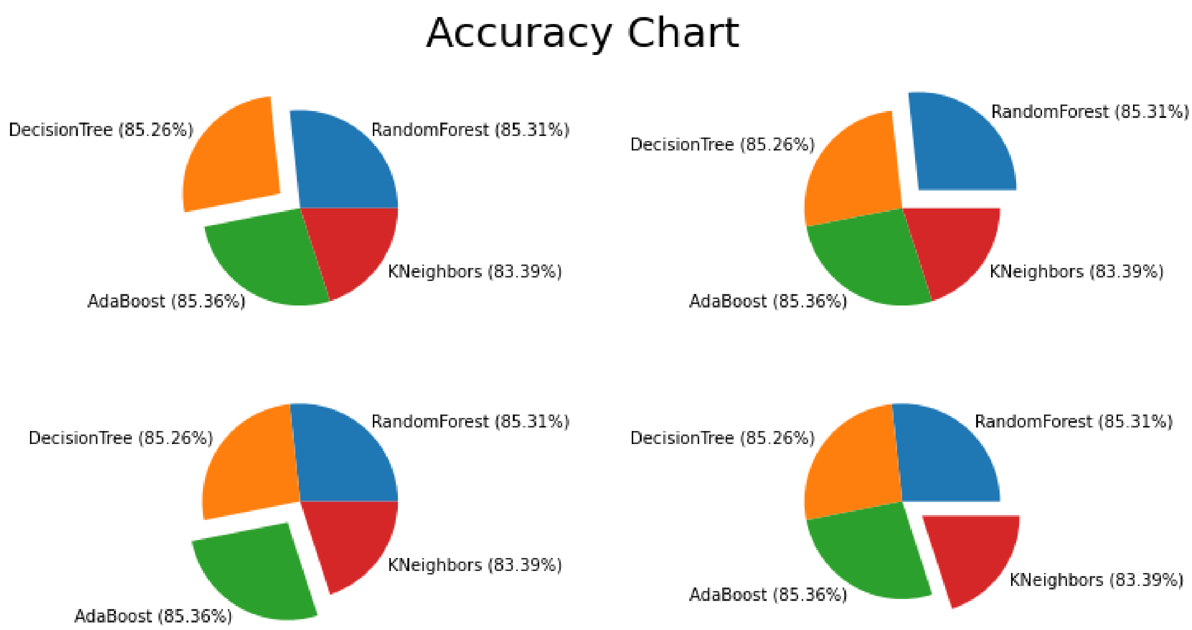

**Figure 2.** Accuracy calculation for four said classifiers

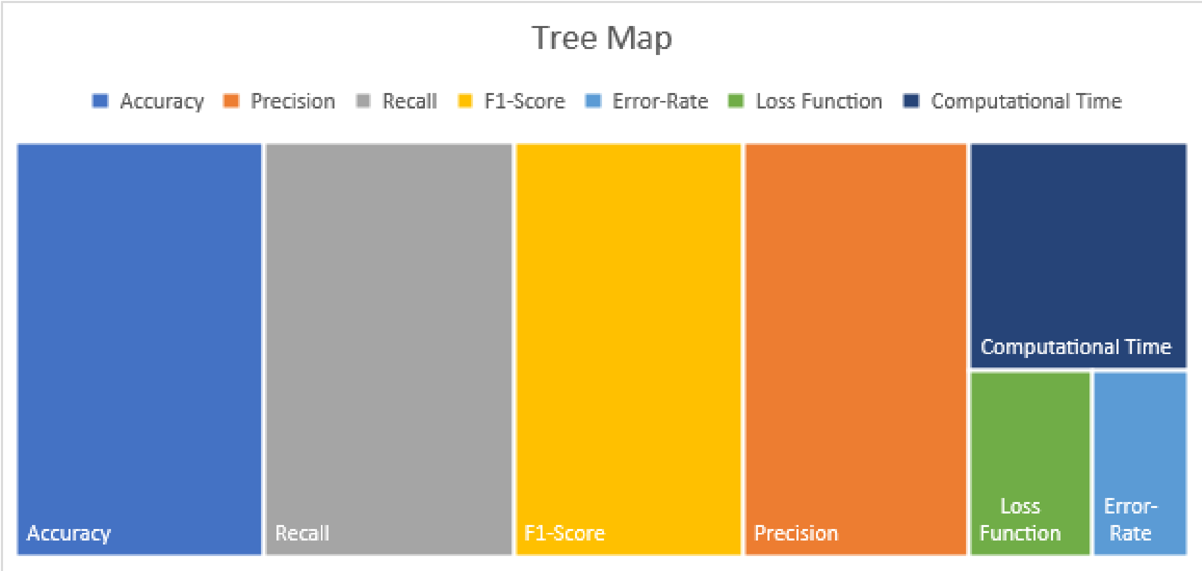


**Figure 3.** Accuracy calculation Tree Map

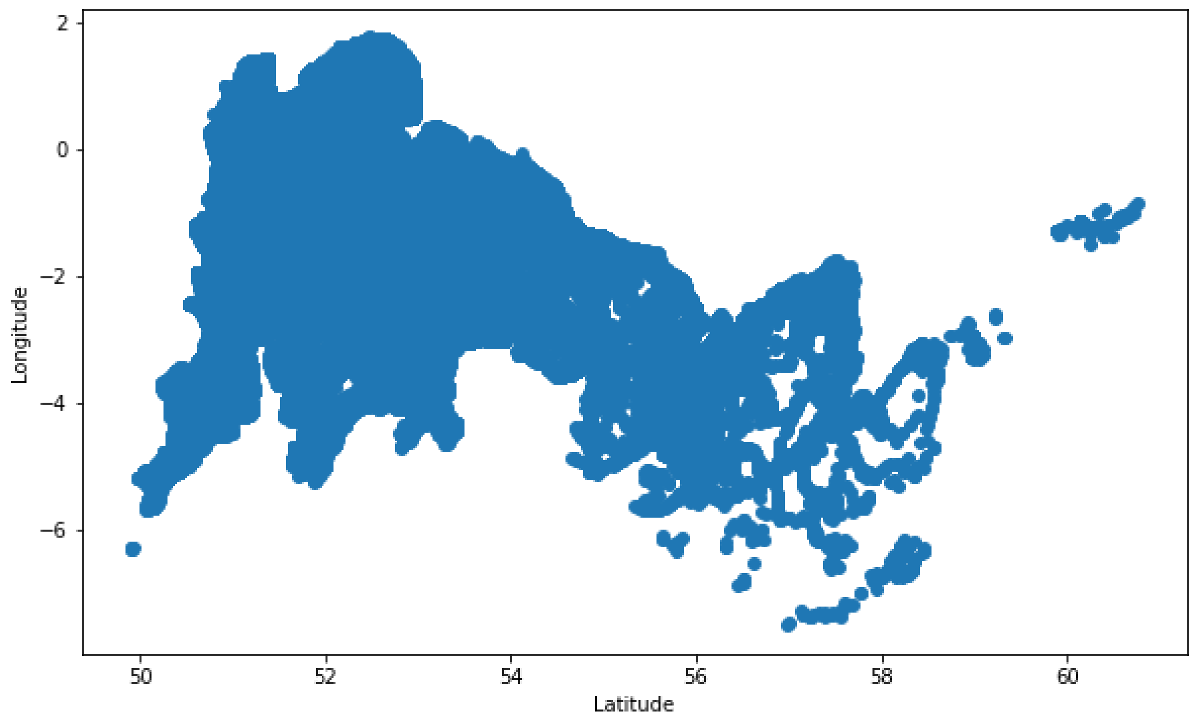


**Figure 4.** Lattitude and Longtude values of road traffics

Figure 1, classifiers are an essential component of machine learning algorithms used for predicting and classifying input data. Accuracy, precision, recall, and F1-score are commonly used performance metrics to compare different classifiers. Accuracy measures how often the classifier correctly predicts the target class, while precision measures the fraction of true positive predictions out of all positive predictions. Recall measures the fraction of true positives out of all actual positives, and F1-score is a weighted average of precision and recall.The choice of loss function and error rate is also critical in comparing classifiers. The loss function quantifies the difference between predicted and actual values, while the error rate is the proportion of misclassifications in the test dataset. Different classifiers may use different loss functions and have varying error rates depending on the nature of the problem and the data used.Therefore, in comparing classifiers, it is crucial to consider not only accuracy but also precision, recall, and F1-score, as well as the loss function and error rate. A comprehensive evaluation of these metrics can help identify the strengths and weaknesses of different classifiers and determine the best approach for a particular problem. Figure 2, to calculate the accuracy of a classifier, you would typically use a labeled dataset where the correct class labels are already known. The accuracy is then calculated as the proportion of correct predictions made by the classifier on the test data.Figure 3, to calculate accuracy using a confusion matrix for a road traffic dataset, you would first need to train a classification model on the dataset and obtain predictions for a set of test data. Once you have the predicted labels and the actual labels, you can create a confusion matrix and calculate the accuracy. In figure 4 one way to obtain latitude and longitude values of road traffic is through GPS devices installed in vehicles or mobile devices. There are also various traffic apps available that use GPS technology to track real-time traffic conditions and provide information on road congestion, accidents, and other events that may impact travel times.

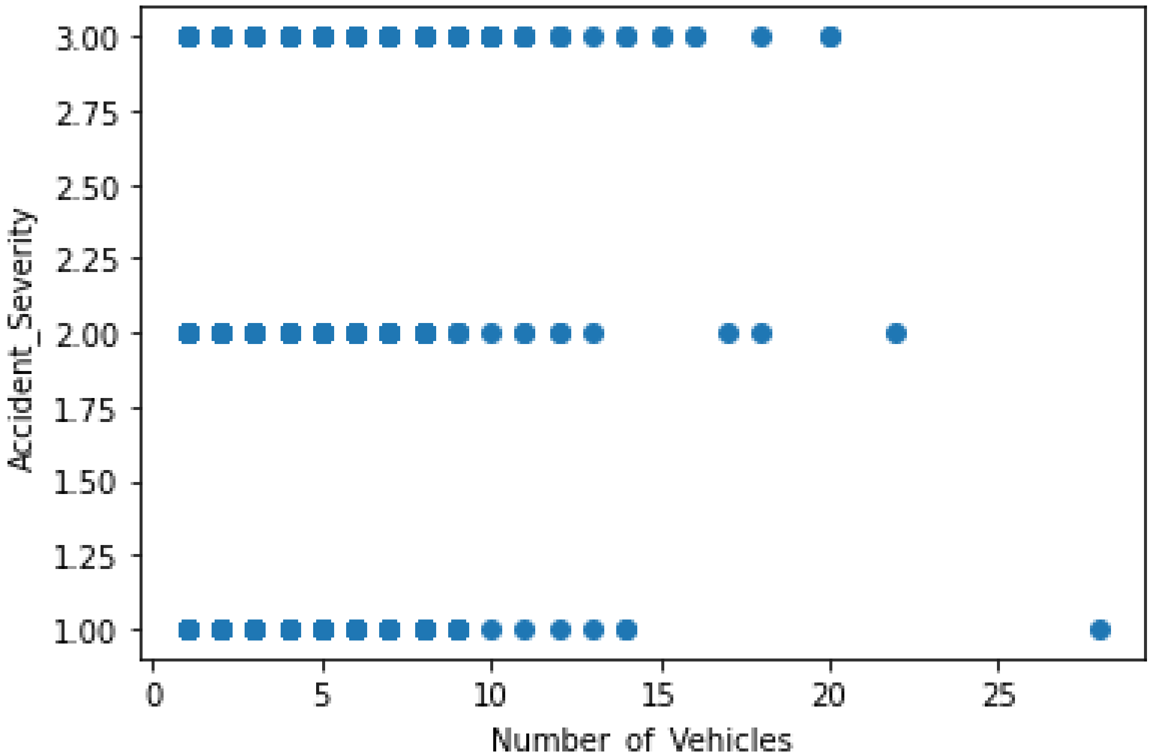


**Figure 5.** Accident Severity due to number of Vehicles

In figure 5, to explore the relationship between the number of vehicles involved in an accident and accident severity, you could perform a statistical analysis using the dataset. In figure 6, it's important to note that while data analysis can provide insights into the causes of accidents and potential ways to reduce them, it's ultimately up to individuals and organizations to take action to improve safety on the roads. This may involve implementing better infrastructure, improving driver education and training, and promoting safe driving habits.in figure 7, draw conclusions and make recommendations based on your analysis and modeling, you can draw conclusions about the factors that contribute to accidents on different road types and make recommendations for improving road safety. These recommendations can be used by policymakers, transportation agencies, and other stakeholders to reduce the incidence of road accidents.

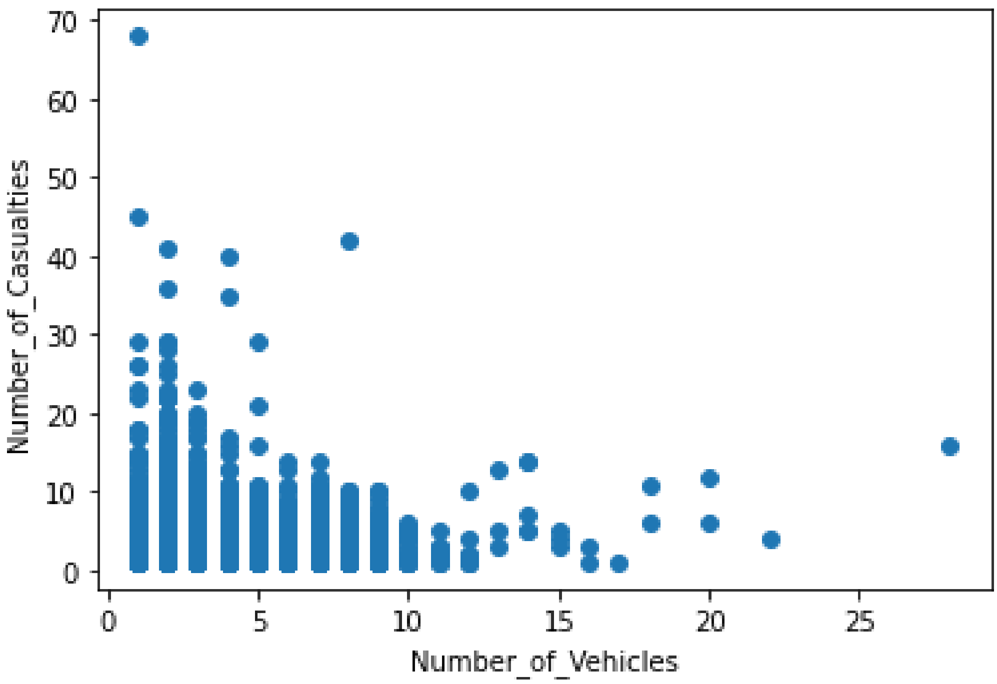


**Figure 6.** Accidents due to number of vehicles

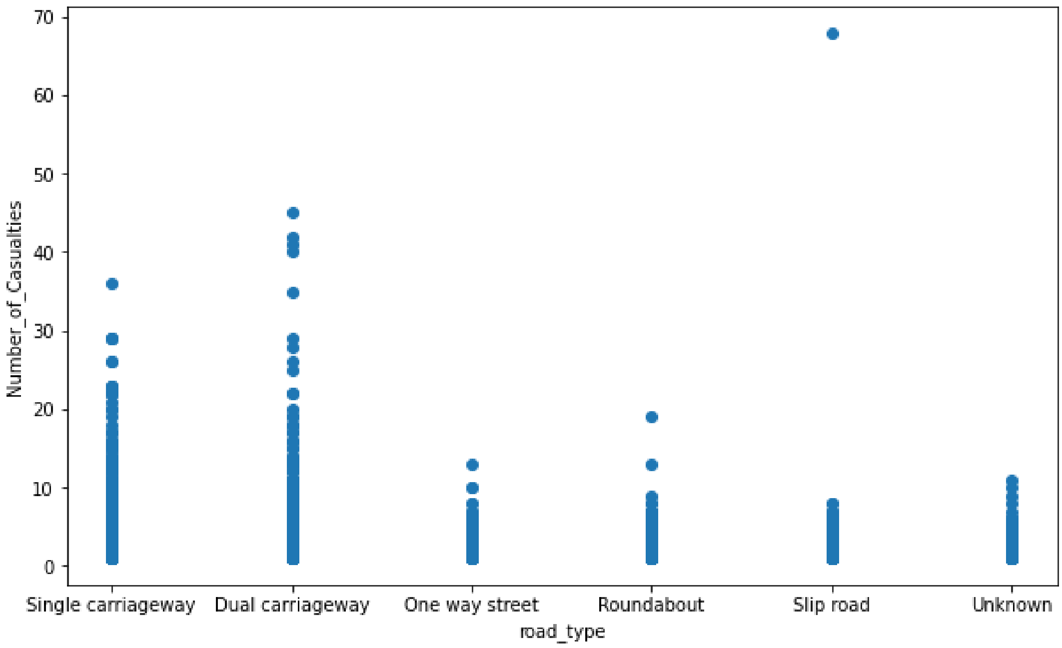


**Figure 7.** Accidents due to Road types

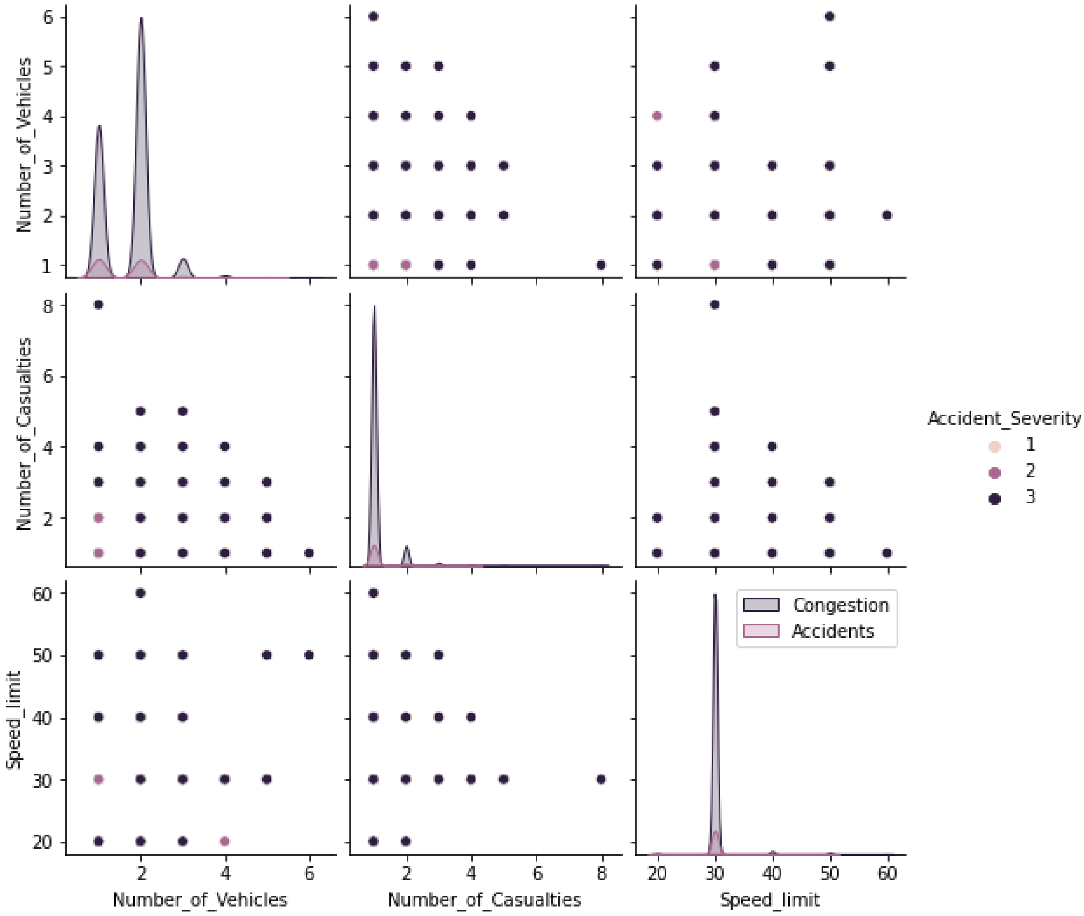


**Figure 8.** Road Accidents due to different factors on 1-Congestion 2-Accidents and 3-Morality

Figure 8 congestion can increase the likelihood of accidents due to the increased number of vehicles on the road, resulting in slower traffic flow, longer travel times, and frustration among drivers. Additionally, congestion can lead to reckless driving behavior, such as tailgating or cutting off other drivers, which can increase the risk of accidents. Accidents are a common cause of road accidents and can occur due to a variety of reasons, including driver error, vehicle malfunctions, and poor road conditions. Some common types of accidents include rear-end collisions, side-impact collisions, and rollovers.Poor driving behavior, such as speeding, distracted driving, and driving under the influence of drugs or alcohol, can also contribute to road accidents. Additionally, moral factors such as a lack of consideration for other drivers or a disregard for traffic laws can also play a role in causing accidents.

In Table 2., regarding the mortality, accidents, and congestion rates, these could be considered as target variables for a supervised learning problem, where the classifier is trained to predict these variables based on other features in the dataset. It's important to note that accurate and reliable predictions may depend on the quality and quantity of the data available.

### *4.7 Mortality*

Mortality calculation in road traffic congestion and accidents using machine learning can be approached in a number of ways. One possible method is to use predictive modeling techniques to analyze historical data on road traffic congestion and accidents, and identify patterns and trends that are associated with higher mortality rates. The following are some steps that can be taken to develop a machine learning model for mortality calculation:

**Table 2.** Classifiers with their mortality, accidents and congestion rates

| Algorithm | Mortality | Accidents | Congestion |
|---|---|---|---|
| Random Forest | 113743 | 227 | 33 |
| Decision Tree | 113627 | 304 | 72 |
| AdaBoost Classifier | 113965 | 38 | 0 |
| KNearestNeighbor | 110883 | 2905 | 215 |
| Actual | 97318 | 14979 | 1706 |

- **Data collection:** Gather historical data on road traffic congestion and accidents, including information on the location, time, and severity of accidents, as well as the number of fatalities.
- **Data preprocessing:** Clean and preprocess the data to ensure that it is ready for analysis. This may involve removing missing values, dealing with outliers, and transforming the data into a suitable format for machine learning.
- **Feature selection:** Identify the most important features that are likely to affect mortality rates in road traffic congestion and accidents. This may include factors such as the number of vehicles involved in an accident, the speed limit of the road, the weather conditions, and the time of day.
- **Model selection:** Choose an appropriate machine learning model for the task at hand. This may involve using regression models, classification models, or clustering algorithms, depending on the nature of the data and the desired outcomes.
- **Model training and evaluation:** Train the machine learning model on the historical data, and evaluate its performance using metrics such as accuracy, precision, recall, and F1 score. This will help to ensure that the model is effective in predicting mortality rates in road traffic congestion and accidents.
- **Deployment:** Once the model has been trained and evaluated, it can be deployed in real-world settings to provide insights and predictions on mortality rates. This may involve integrating the model with existing traffic monitoring systems, or creating new tools and applications that can help to improve road safety and reduce the risk of accidents.

***4.8 Accidents***

Machine learning can be used to calculate the probability of road traffic accidents in congested areas. Here's a general outline of how this might work:

- **Data Collection :** Gathering information on previous incidents in major cities is the initial stage. This information may include the time of day, the climate, the design of the roads, the posted speed restrictions, and other pertinent details.

• **Feature engineering:** Following the gathering of data, the next stage is to create features that will aid the machine learning algorithm in producing precise predictions. This may entail figuring out the average traffic speed, the number of vehicles on the route, or the frequency of lane changes per minute.

• **Model choice:** For this assignment, you may use a variety of machine learning models, such as decision trees, random forests, and neural networks. The quantity of the dataset, the difficulty of the task, and other elements all affect the model selection.

• **Training model:** Using the data, the machine learning model is then trained. use a practise set. The model is trained to generate predictions based on the input attributes using the training set. • Testing and Validation: When a model has been trained, it is checked for accuracy and dependability using a different dataset. This stage is crucial to ensuring that the model can correctly predict outcomes based on the incoming data and does not overfit the training set of data.

• **Dispersion:** The model may be used in actual applications to forecast the likelihood of traffic accidents in crowded locations once it has been trained and verified. Drivers may be alerted to possible dangers using this information, and city planners can utilise it to create better road designs.

Overall, using machine learning to predict road traffic accidents in congested areas can help to improve road safety and reduce the risk of accidents. However, it is important to note that machine learning models are only as good as the data they are trained on, so collecting high-quality data is crucial for accurate predictions.

### *4.9 Congestion*

In order to calculate road traffic congestion and accidents using machine learning, you would typically need to follow the steps below:

- **Data collection:** Collect data on road traffic patterns and accidents, such as traffic volumes, speeds, vehicle types, weather conditions, and accident locations and types.
- **Data pre-processing:** Clean and prepare the collected data for use in machine learning algorithms. This may involve removing missing or invalid data, transforming data into a suitable format, and splitting data into training and testing sets.
- **Feature engineering:** Identify relevant features from the pre-processed data that can be used as input to machine learning models. This may involve extracting information from raw data, such as calculating average speeds or identifying high-traffic areas.
- **Model selection:** Choose an appropriate machine learning model for the problem at hand. For example, you could use a regression model to predict traffic congestion levels or a classification model to predict accident probabilities.
- **Model training:** Use sophisticated and preprocessed data to train a chosen model. Inputs must be provided to the model, and its parameters must be changed to reduce the error between predicted and observed outcomes.
- **Model evaluation:** Assess the trained model's performance and accuracy using a variety of tests. You may achieve this by taking measurements of variables like accuracy, precision, memory, and F1 score.
- **Example of deployment**: Use a trained model to make real-time predictions using new data. This can entail incorporating the model into a computer programme that can take in real-time information from motion sensors or other sources.

Overall, machine learning can be a powerful tool for predicting traffic congestion and accidents on the road. By accurately predicting these events, traffic planners and authorities can take proactive measures to alleviate congestion and improve road safety.

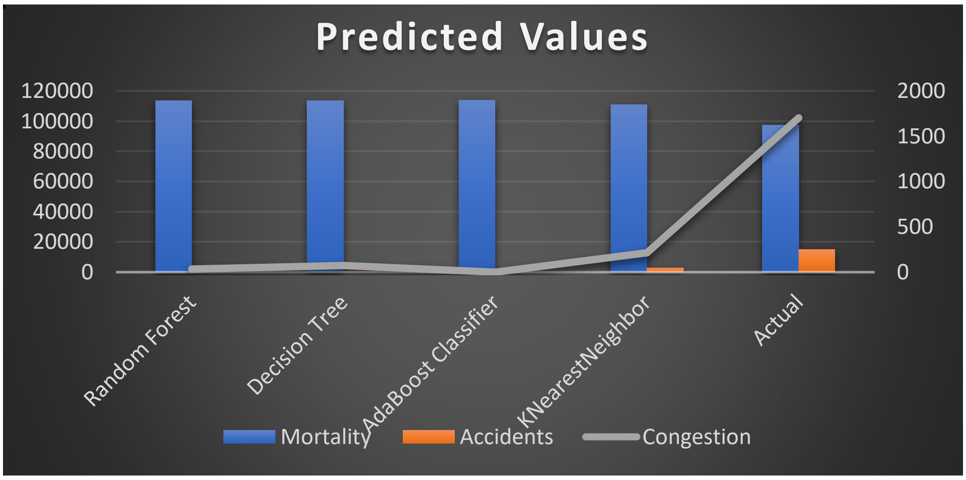


Figure 9. Predicted values of classifiers for Mortality, Acciedents and Congestions

In figure 9, once you have selected your features, you can then train your classifiers using a variety of machine learning algorithms such as decision trees, random forests, adaboost, and KNN . After training, you can use the trained classifiers to predict values for Mortality, Accidents, and Congestions based on recent data.

As an AI language model, I do not have access to any specific study or analysis on road traffic congestion and accidents using random forest, AdaBoost, K nearest neighbors, and decision tree algorithms. However, I can provide a general discussion on how these algorithms can be used to address the problem of road traffic congestion and accidents.Road traffic congestion and accidents are major problems in many urban areas. One way to address these issues is to use machine learning algorithms to predict and prevent them. Random forest, AdaBoost, K nearest neighbors, and decision tree are some of the popular machine learning algorithms that can be used for this purpose.Random forest is an ensemble learning algorithm that combines multiple decision trees to make predictions. It is a powerful algorithm that can handle large datasets and can be used for both classification and regression tasks. Random forest can be used to predict the likelihood of road traffic congestion or accidents based on various factors such as weather, traffic volume, and time of day.

AdaBoost is another ensemble learning algorithm that can be used to improve the accuracy of predictions. It works by combining multiple weak learners to create a strong learner. AdaBoost can be used to predict the likelihood of road traffic congestion or accidents based on historical data and other relevant factors.K nearest neighbors is a simple algorithm that can be used for both classification and regression tasks. It works by finding the K nearest data points to a new data point and making a prediction based on the majority class or average value of the K nearest neighbors. K nearest neighbors can be used to predict the likelihood of road traffic congestion or accidents based on similar historical data.Decision tree is a simple and interpretable algorithm that can be used for both classification and regression tasks. It works by creating a tree-like model of decisions based on the input features. Decision tree can be used to predict the likelihood of road traffic congestion or accidents based on various factors such as weather, traffic volume, and time of day. In conclusion, machine learning algorithms such as random forest, AdaBoost, K nearest neighbors, and decision tree can be used to predict and prevent road traffic congestion and accidents. These algorithms can help urban planners and traffic engineers to design more efficient and safer road networks. According to the methods implemented, Adabbost is the classifier predictios congestion, accidents and mortity with great precison of converging results.

## 5. Conclusion

After analyzing the road traffic data using decision tree, random forest, K nearest neighborhood, and Adaboost algorithms, it can be concluded that Adaboost is the best-performing algorithm among all. Adaboost algorithm provided the highest accuracy and F1 score, indicating its ability to accurately predict the occurrence of traffic congestion, accidents, and mortality. It achieved this by combining several weak decision trees and producing a stronger model that accurately classified the data. K nearest neighbor (KNN) performed reasonably well, but not as well as Adaboost. KNN relies on distance metrics to classify the data, and it can struggle with high-dimensional data.

Random Forest also produced decent results but not as good as Adaboost. It is a robust algorithm that combines several decision trees to produce a more accurate model. However, it can be computationally expensive and slow when dealing with large datasets. Decision tree produced the lowest accuracy and F1 score compared to the other algorithms. It is a simple algorithm that relies on creating a tree-like model that splits the data based on certain criteria. It can perform well on small datasets, but it struggles with larger and more complex datasets.

Overall, Adaboost is the best-performing algorithm among the four for predicting road traffic congestion, accidents, and mortality.

## References


1. Cai, J., Jiang, H., Chen, L., Liu, J., Cai, Y., & Wang, J. (2019). Implementation and development of a trajectory tracking control system for intelligent vehicle. *Journal of Intelligent & Robotic Systems*, *94*, 251-264.
2. Tang, Y., Zhou, J., Hao, H., Hao, F., & Xu, H. (2022). Path Planning and Trajectory Tracking for Automatic Guided Vehicles. *Computational Intelligence and Neuroscience*, *2022*.
3. Deshpande, P., Devika, K. B., Subramanian, S. C., & Vanajakshi, L. D. (2022). Robust steering control for trajectory following in road traffic environments. *Proceedings of the Institution of Mechanical Engineers*, *Part I: Journal of Systems and Control Engineering*, *236*(1), 153-168.
4. Zhuang, C., & Chen, C. (2022). Research on autonomous route generation method based on AIS ship trajectory big data and improved LSTM algorithm. *Frontiers in Neurorobotics*, *16*.

5. Li, L., Jiang, R., He, Z., Chen, X. M., & Zhou, X. (2020). Trajectory data-based traffic flow studies: A revisit. *Transportation Research Part C: Emerging Technologies*, *114*, 225-240.
6. Mir, I., Gul, F., Mir, S., Khan, M. A., Saeed, N., Abualigah, L., ... & Gandomi, A. H. (2022). A survey of trajectory planning techniques for autonomous systems. *Electronics*, *11*(18), 2801.
7. Xue, G., Liu, Y., Shi, Z., Guo, L., & Li, Z. (2022). Research on Trajectory Tracking Control of Underwater Vehicle Manipulator System Based on Model-Free Adaptive Control Method. *Journal of Marine Science and Engineering*, *10*(5), 652.
8. Benloucif, A., Nguyen, A. T., Sentouh, C., & Popieul, J. C. (2019). Cooperative trajectory planning for haptic shared control between driver and automation in highway driving. *IEEE Transactions on Industrial Electronics*, *66*(12), 9846-9857.
9. Liu, B., Shi, Q., Song, Z., & El Kamel, A. (2019). Trajectory planning for autonomous intersection management of connected vehicles. *Simulation Modelling Practice and Theory*, *90*, 16-30.
10. Alsrehin, N. O., Klaib, A. F., & Magableh, A. (2019). Intelligent transportation and control systems using data mining and machine learning techniques: A comprehensive study. *IEEE Access*, *7*, 49830-49857.
11. Chen, L., Ruan, Y., & Gou, Y. (2022). Automatic Vehicles' Trajectories Optimization on Highway Exclusive Lanes. *Journal of Advanced Transportation*, *2022*.
12. Ma, Y., Xu, J., Gao, C., Mu, M., E, G., & Gu, C. (2022). Review of research on road traffic operation risk prevention and control. *International journal of environmental research and public health*, *19*(19), 12115.
13. Kumar, S., Tiwari, P., & Zymbler, M. (2019). Internet of Things is a revolutionary approach for future technology enhancement: a review. *Journal of Big data*, *6*(1), 1-21.
14. Yang, L., Yue, M., & Ma, T. (2019). Path following predictive control for autonomous vehicles subject to uncertain tire-ground adhesion and varied road curvature. *International Journal of Control, Automation and Systems*, *17*(1), 193-202.
15. Kontorinaki, M., Karafyllis, I., & Papageorgiou, M. (2019). Global exponential stabilisation of acyclic traffic networks. International Journal of Control, 92(3), 564-584.