



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 8, August 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com



A Road Accident Prediction Model Using Data Mining & Machine Learning Techniques

Dr Aravinda T V, Dr Krishna Reddy K R, Prof. B E Ramesh, Prof. Basantha Kumari, Prof. Shruthi M K, Prof. Beena Sheril, Spoorthi M

Professor, Computer Science and Engineering, SJM Institute of Technology Chitradurga, India

Professor, Computer Science and Engineering, SJM Institute of Technology Chitradurga, India

Asso. Professor, Computer Science and Engineering, SJM Institute of Technology Chitradurga, India

Asso. Professor, Computer Science and Engineering, SJM Institute of Technology Chitradurga, India

Asso. Professor, Computer Science and Engineering, SJM Institute of Technology Chitradurga, India

Asst. Professor, Computer Science and Engineering, SJM Institute of Technology Chitradurga, India

M.Tech Student, Computer Science and Engineering, SJM Institute of Technology Chitradurga, India

ABSTRACT: Road Accidents have consistently been fatal events throughout history. According to data from India's road safety project, 17 individuals per hour lose their lives in traffic-related incidents. Rollovers, head-on collisions, and rear-end collisions are just a few of the numerous possible outcomes of a car accident. Details on accidents may be found in police reports filed with the state, sometimes as FIRs. Anyone involved or the state police may file a report. The frequency patterns of automobile collisions are estimated using the SVM and KNN Classifier. With this trend in mind, the government or non-governmental groups might implement efforts to increase road safety and reduce accidents in heavily populated areas. The safety of people using public transport is one of the world's most pressing social issues. For every fatality, predict four incapacitating injuries including brain or nervous system damage, ten major wounds, and sixty minor wounds. The geometric development in the number of vehicles out and about has prompted an alarming increase in the daily accident rate. The capacity of the transport department to forecast the accident rate over a lengthy period of time is vital in light of the increased number of traffic-related fatalities and accidents in recent years. Here, data mining techniques such as the KNN Classifier and the Support Vector Machines are used to develop an accident prediction model. Considered in analyses of traffic accidents, road conditions, and the influence of environmental factors on accident probability. The government's public works agency, private contractors, and even rival automakers might all benefit from the estimates derived from this.

KEYWORDS: Road traffic accidents, Government, Models that make accident predictions, Algorithm towards forecast, Patterns of data, Safety on the roads, Death.

I. INTRODUCTION

The rapidly rising number of accidents in India is a major reason for worry. Recent data suggests that India is responsible for almost 6% of all road accidents worldwide while having just 1% of the world's total car population. Over-speeding is a major element in the high number of reported accidents involving motorcycles and scooters. Drunk driving and other forms of traffic law violation very often lead to accidents. Many accidents still occur despite laws and rules being in place because drivers and passengers don't pay attention to the road, don't maintain their vehicles, and don't even bother to wear helmets. There may be more cars on the road, but road quality and other external variables also play a significant part in contributing to the problem.

Concerning numbers of people have lost their lives in traffic accidents in India. More than 137,000 individuals have lost their lives as a result of injuries sustained in traffic accidents, making this an extremely bleak situation. More than four times the average yearly death toll from terrorism, this number is staggering. Some of the most tragic and senseless accidents that occur are those involving large commercial vehicles like trucks or even public transit vehicles like buses. Rain, fog, and other forms of bad weather may all increase the likelihood of accidents. As a result, understanding accident hotspots and contributing variables is essential for developing mitigation strategies. To do so, it need to build better accident prediction algorithms and conduct in-depth research on past incidents.

Having an optimized accident prediction model that can analyze potential issues arising due to infrastructure fallbacks and estimate the effect of existing models in reducing the occurrence of accidents is often desired when implementing a well-designed road framework management system for looking into road security aspects. The primary difficulty in developing such a model is determining how best to include the influence of all relevant variables and assigning relative importance to the many factors that may have contributed to the disaster. Credit risk management, fraud detection, healthcare informatics, recommendation systems, and other fields have all found use for data interpretation using data mining methods and models in the past. Artificial intelligence and machine learning methods have been used to supplement these investigations. Here, to explore how several environmental and infrastructure elements interact with one another to increase the likelihood of traffic accidents. Data mining methods are utilised here since these massive datasets are worthless without proper interpretation in order to get useful information from them.

II. LITERATURE REVIEW

The model proposed by the author uses four psychological instruments to assess the behavior of the driver and personal traits in the hopes of identifying a relationship between the considered constraints and the occurrence of road accidents. The participants were made to fill out a demographic survey and in addition to the described behaviors, correlation coefficients were calculated and four hierarchical regressions were designed to throw light on the predictive power that personality traits can have on the occurrence of road accidents. In terms of impulsivity and aggressiveness, it has been found that personality traits have the highest influence.

The advantage of this model is that it can be mainly used in fields like the design of procedures for selecting a professional driver, the building of a system for the prevention of traffic accidents and violations of law, and the rehabilitation of drivers who have been deprived of their driving license.

This research mainly focuses on the behavior of the driver to draw a conclusion as to whether an accident is going to occur or not. Hence, this model cannot be deployed in case there are other factors responsible for an accident to occur. [1]

This research determines associations that road accidents have with each other and predicts the type of accidents that might occur for current roads as well as upcoming roads by finding ways in mining for frequent patterns from relevant data sets which are responsible for road accidents. For this purpose, association and classification rules have been found to be the most suitable. The Apriori algorithm is used to regulate associations that can be found among road accidents. Also with the help of the Naïve Bayes algorithm prediction of common accidents that can occur with the cause of it being new roads was made. This system overcomes the disadvantage of the current system used by the government that maintains a ledger and the data is analyzed manually. [2]

this paper mainly focuses on the road crashes that occurred in the Anambra state of Nigeria. Their main intention was to develop a predictive model with maximum accuracy for forecasting crash frequency using ARIMAX and ARIMA modeling techniques. They have considered eleven contributing factors, some of which are over speeding weather conditions, loss of control, etc. All of these have significant effects on the number of traffic crashes. The advantage found was an increase in any of these considered parameters will lead to an increase in the number of accidents in Anambra State. Result: Finally, the time series predictive model is developed, which performs well for crash analysis. [3]

In this article, they have collected weather datasets of Visakhapatnam city for a period of 97 days, from the Indian Meteorological Department. This dataset consists of several attributes, such as wind pressure, minimum and maximum temperature, humidity, forecast, and type. The main objective of their project was evaluating weather conditions for a short period of days that is for fifteen days. This was intended towards predicting the weather forecast. The Forecasting experiment achieved this by using the ARIMA model prediction algorithm. Interventional analysis was done based on autocorrelation models and partial autocorrelation models. These two models were used to predict wind pressure and humidity for the next 15 days. The result was a comparison between predicted results and real data displayed only for a single city. The prediction of unfavorable weather conditions can be valuable information to take precautions to prevent an accident. [4]

Regularizing traffic in both developed as well as underdeveloped countries is done by traffic signals. The author has proposed a model that combines cluster analysis and Intelligent Agent-Based System that regulates the traffic signal based on the traffic volume. Traffic signal changes color at regular intervals. But an intelligent traffic light system senses the presence or absence of vehicles and reacts accordingly. it provides an intelligent transportation system to control and avoid accidents. This article gives a proposal that avoids traffic accidents from the perspective of



Data Mining. The results of clustering were presented by making use of WEKA and other dimension-reduction techniques. The use of Data Mining tools such as Waikato Environment for Knowledge Analysis makes it convenient for the Transportation Department to devise a plan to avoid future accidents. [5]

This article suggests that data mining has proven to be a very useful technique for analyzing vehicle crashes and rendering effective outcomes. Many models use these techniques by identifying the elements that influence the criticality of the crash. A factor considered in this article is the location of the occurrence of accidents. In this paper, they have first made use of the K-means technique to cluster crash areas into three divisions, the largest number of accident-prone areas, relatively fewer number of accident-prone areas, and the least number of accident-prone areas. It takes the count as a factor to group the areas. Association mining rule which is a popular technique was used to characterize these areas. While the data limits itself to only a particular type of parameter, this analysis was useful in finding hidden information. [6]

Existing System

Predictions of fatalities and injuries from traffic accidents were traditionally made using statistical models. The mixed logit model, the ordered Probit model, and the logit model are all examples of traditional statistical research methods. Some research suggests that the more conventional statistical method is superior for identifying both causal and contributing factors in accidents. Extracting hidden predictive information from massive datasets, data mining is a novel and powerful technology that may help businesses zero in on the most important data in their database systems. It's a helpful resource for extracting hidden patterns and other valuable information from databases. As the number of accidents in India continues to rise, scientists are beginning to investigate and understand what causes them. Data mining approaches have been used in the past to create prediction models for traffic accidents since they do not rely on the same assumptions between dependent and independent variables as conventional statistical methods. In creating such models, researchers have prioritized various sets of characteristics.

Disadvantages

- 1) There is not enough data for the system to train and test on.
- 2) Poor categorization models mean that the system overestimates the frequency with which traffic accidents occur.

Proposed System

Based on past road accident data, the suggested system has developed an application that can foresee the likelihood of an accident happening. This road accident data undergoes data pre-processing in order to produce a dataset. Feature selection occurs after data preparation, during which only relevant features from the original dataset are chosen to be included in the final dataset by removing the null and trash values and normalising the data. The dataset is subsequently processed using various data mining methods. This dataset undergoes clustering analysis. Next, the clusters are analysed using various methods, such as Support Vector Machines (SVM) and K-Nearest Neighbours (KNN).

The former (SVM) is used to forecast the likely probability of accidents since the data utilised for the research has an uncertain distribution, and to separate the common and rare elements in the dataset.

Accidents of variable severity and frequency on various road types and in different weather conditions have been taken into account while formulating the rules. The selected support and confidence levels for the frequently occurring item sets indicate a greater possibility that the specific combination of traits in question led to an accident.

For instance, the training dataset indicates that there is a significant probability of an accident happening even in excellent weather at a crossroads due to over-speeding. Each accident has been assigned to a high- or low-risk group using SVM classification. The accident dataset is processed using a number of data mining and exploratory visualisation methods, from which the interpreted findings are drawn.

Advantages

- 1) These optimized models can be efficiently utilized by the government to reduce road accidents and implement policies for road safety.

2) The overall model has helped to give an understanding of the combinations of factors that have proven fatal in accident scenarios.

III. SYSTEM DESIGN

System Architecture

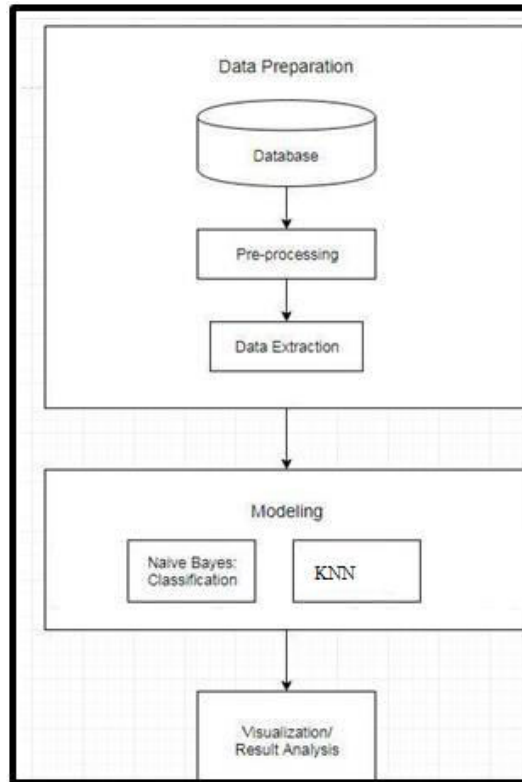


Fig: Proposed system architecture

A. Data Preparation:

Each model creation begins with data preparation. Any and all data that lacks a value for the specified characteristics is purged. The data dictionary in the user manual is used to translate all numbers into their corresponding nominal values. The rate of fatalities is determined and then sorted into High and Low bins. By doing preprocessing, may clean up the dataset and prepare it for algorithmic analysis.

Preparing Data Consists of the Following Steps: -

Step 1 - Gather Data: Finding relevant data is the first step in the data preparation process. This may either be preexisting data or newly contributed information.

Step 2 - Discover and assess data: It is crucial to collect all datasets before attempting to organize the data. This stage entails making preparations to recognize the information and learning what must be accomplished before the data may be used in a very specific setting.

Step 3 - Cleanse and validate data: Improving the data has traditionally been the most time-consuming aspect of the data preparation process, but it is essential for correcting mistakes and filling in missing information. After the information has been scrubbed, it must be validated by verifying that no mistakes were made throughout the knowledge preparation process leading up to the present use case. In most cases, a system error will become obvious at this stage, allowing for it to be fixed before proceeding.

Step 4 - Transform and enrich information: Information may be remodeled by altering its structure or value entries to achieve a desired result or to make the data more accessible to a larger audience. When data is "enriched," it is augmented with new or related information in order to provide more meaningful results.

Step 5 - Store data: After the information has been cleaned, it may be saved or sent to another programme, such as a business intelligence suite, where further processing and analysis can take place.

B. Modeling

Knowledge modelling refers to the process of creating a model of the information that will be kept in a database. Knowledge items, the links between data objects, and the governing principles might all be conceptualized in this data model. Business rules, regulatory compliance, and government standards may all be enforced on data with the use of data modeling. Data Models guarantee the integrity of data by standardizing things like names, defaults, meanings, and protections. The data model focuses on what data is needed and how it should be organized rather than what actions must be taken on the data. The data model serves a similar purpose to a construction design in that it establishes the relationships between various data elements and contributes to the development of a conceptual model. It begins by calculating a number of statistics from the dataset that is believed to highlight the most salient features of fatal incidents. Then SVM and KNN is used to discover patterns by analyzing the interrelationships between the characteristics

Requirement Analysis

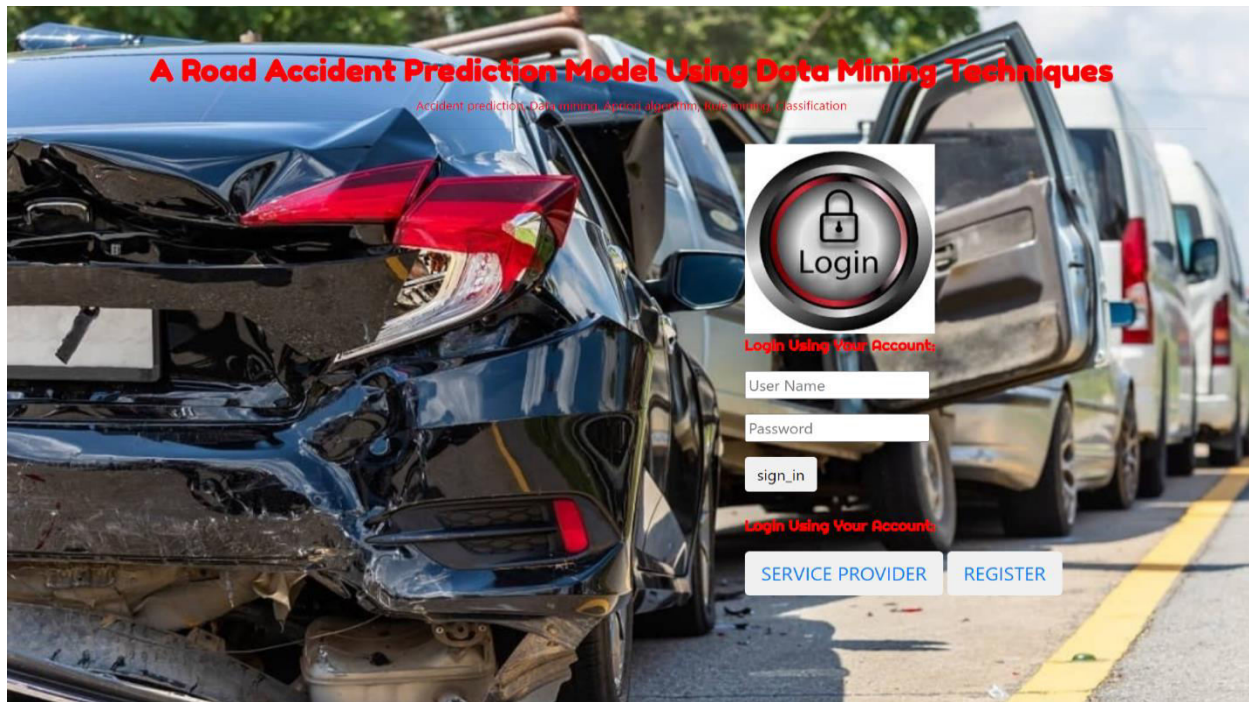
Hardware and software details

- The experiment setup is carried out on a computer system that has different hardware and software specifications. The hardware used is Processor Intel Core i7, HDD 1 TB, RAM 8 GB.
- The software used is Operating Language - Windows 11, Programming Language - Python, Front-end – HTML, CSS, Javascript, Database - MySQL, and Back-end - SQL.

IV. RESULTS AND DISCUSSION

There are three proposed system modules they are Service Provider, View and Authorize Users, Remote User. The Service Provider will need to provide their login and password to access this section. Successful login grants access to features like Browse Datasets, Train & Test, View Trained & Tested Accuracy in Bar Chart, View Trained & Tested Accuracy Results, View All Road Accident Prediction, Locate Road Accident Prediction Type Ratio, View Road Accident Ratio Results, Download Predicted Data Sets, and View All Remote Users. Admins should only use the Service Provider Module. . View and Authorize Users module allows the administrator to examine a list of all registered users. In Remote Users Module there are n participants in this section. Users need to sign up first before they can do any actions. When a user signs up, their information is added to a database. After his registration has been accepted, he will be prompted to provide a valid username and password. After successfully logging in, users may do actions such as checking the status of predicted road accidents, viewing their profiles, and logging out. A user must first create an account in the User module and provide their personal information.

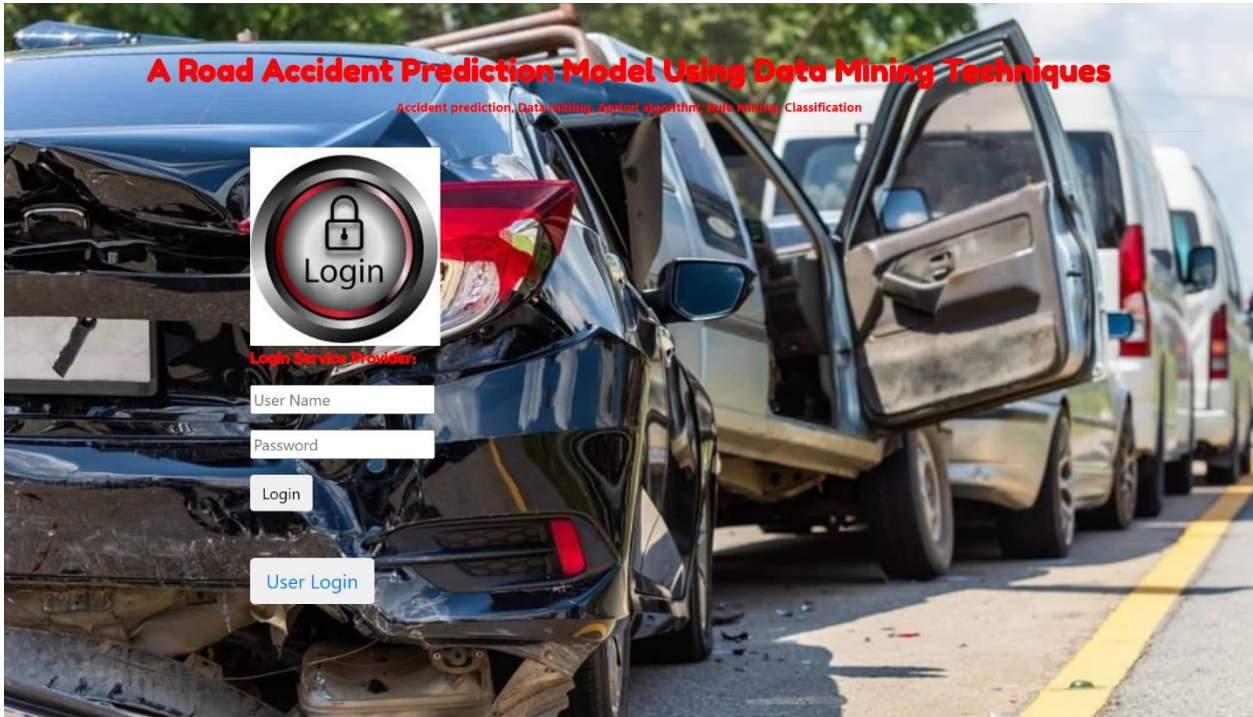
In the below fig where the User Module is seen, where a user can log in and any new user can also Register from this page.



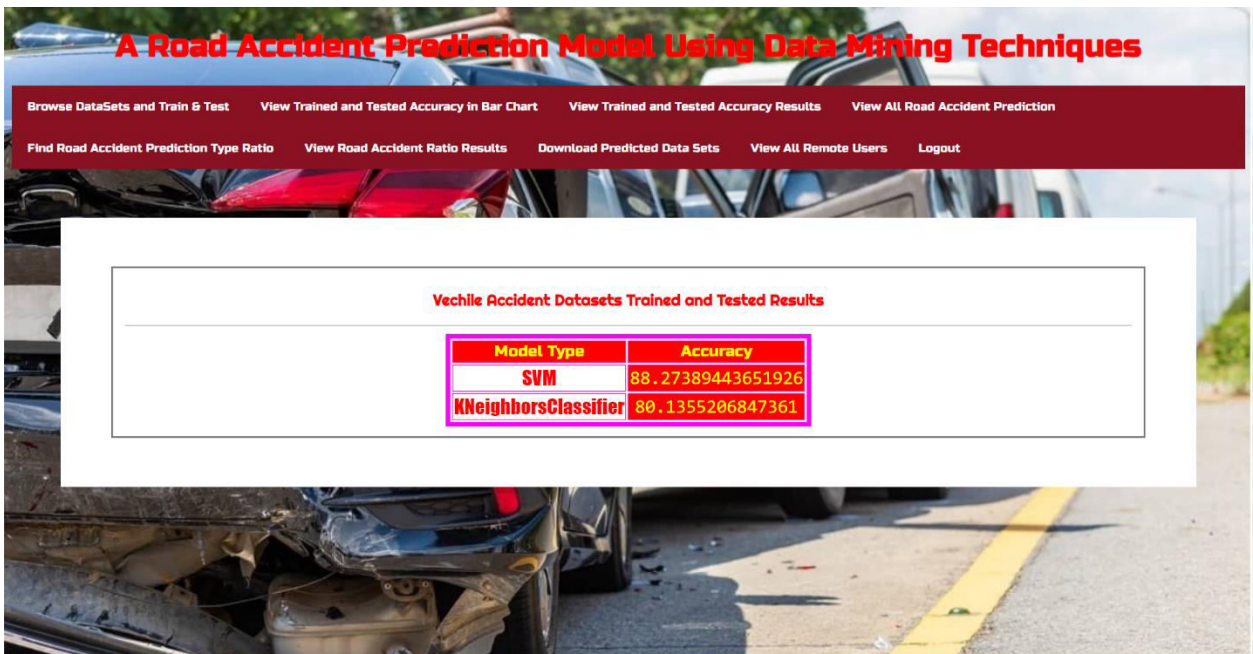
After the successful login, the User can predict the Road accident Status by inputting the details.



The service provider i.e an Admin can login in this page .



Service Provider can Browse Datasets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Road Accident Prediction, Find Road Accident Prediction Type Ratio, View Road Accident Ratio Results, Download Predicted Data Sets, View All Remote Users.



Conclusion

Many people's futures may be altered by a single accident. Each of us has some of the responsibility for reversing this rising trend. Taking precautions when driving might help achieve this goal. Since accidents can't always be traced back to one root cause, authorities in charge of road construction and automotive manufacturers both need to take extra precautions to reduce the likelihood of serious injury and death. We can help authorities and businesses prepare for the probability of an accident by making predictions based on historical data and observations. This project

was successful in developing a tool that can predict the risk probability of accidents across different areas based on factors like vehicle type, driver age, vehicle age, weather, and road structure. This model was implemented by making use of several data mining and machine learning algorithms applied over a dataset for Bangalore.

Additional optimisation of the model is possible in the future by including many restrictions not included in the present investigation. The government may make better use of these optimised models to execute road safety measures and cut down on traffic accidents. The creation of a smartphone app to aid motorists in route selection is also within the purview of this endeavour. It is possible to construct a call out to the driver through the mapping service that would declare the danger likelihood in a specified route in addition to the directions. Then, in the future, businesses like Uber, Ola, etc., may incorporate this. Better road safety instructions may be put along highways in light of the hazards revealed by this model, and better accident monitoring and emergency services will be possible as a resultant.

V. FUTURE SCOPE

Additional optimisation of the model is possible in the future by including many restrictions not included in the present investigation. The government may make better use of these optimised models to execute road safety measures and cut down on traffic accidents. The creation of a smartphone app to aid motorists in route selection is also within the purview of this endeavour. It is possible to construct a call out to the driver through the mapping service that would declare the danger likelihood in a specified route in addition to the directions. Then, in the future, businesses like Uber, Ola, etc., may incorporate this. This is helpful not just for preventing accidents but also for responding to them quickly and efficiently. Taking into consideration the dangers predicted by this model, improved road safety signs may be posted.

REFERENCES

- [1]. "A Model for Traffic Accidents Prediction Based on Driver Personality Traits Assessment" MARJANA ČUBRANIĆ-DOBRODOLAC, KRSTO LIPOVAC, SVETLANA ČIČEVIĆ, BORIS ANTIĆ, University of Belgrade, Faculty of Transport and Traffic Engineering Vojvode Stepe 305, 11000 Belgrade, Serbia.
- [2]. "Analysis of road accidents using data mining techniques" Poojitha Shetty, Sachin P C, Supreeth V Kashyap, Venkatesh Madi Dept. of Computer science and engineering, National Institute of Engineering, Karnataka, India.
- [3]. "Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria" Chukwutoo C. Ihueze, Uchendu O. Onwurah, Department of Industrial and Production Engineering, Nnamdi Azikiwe University, Awka, Nigeria. [
- 4]. "An Integrated Approach for Weather Forecasting based on Data Mining and Forecasting Analysis", G.Vamsi Krishna Research scholar, Department of CSE, GITAM University.
- [5]. "Application of Data Mining Techniques for Analyzing Road Traffic Incidents A Case Study of Libya's Road Traffic", Sayed Mujahed Hussain, Department of Computer Science, Faculty of Arts and Science Ajdabiya, University of Benghazi, Ajdabiya, Libya, Faraj A. El-Mouadib, Department of Computer Science, Faculty of Information Technology, University of Benghazi, Benghazi, Libya.
- [6]. "A data mining approach to characterize road accident locations", Sachin Kumar, Durga Toshniwal.
- [7]. "Accident prediction models for urban roads", Poul Greibe, Danish Transport Research Institute, Knuth Winterfeldts Allé, DK-2800 Kgs. Lyngby, Denmark.
- [8]. "Survey on Analysis and Prediction of Road Traffic Accident Severity Levels using Data Mining Techniques in Maharashtra, India", Baye Atnafu and Gagandeep Kaur Dept of CS/IT, Symbiosis Institute of Technology, Pune, India.
- [9]. "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques", Liling Li, Sharad Shrestha, Gongzhu Hu, Department of Computer Science, Central Michigan University, USA.
- [10]. "Development of a transnational accident prediction model", Francesca La Torre, Lorenzo Domenichini, Monic Meocci, Daniel Graham, Niovi Karathodorou, Thomas Richter, Stephan Ruhl, George Yannis, Anastasios Dragomanovits, Alexandra Laiou, University of Florence, Via S. Marta, 3, 50139, Firenze, Italy, Imperial College London, SW7 2AZ, London, United Kingdom, Technische Universität Berlin, Gustav-Meyer-Allee 25, 13355, Berlin, Germany, National Technical University of Athens, 5, Heroon Polytechniou st., 15773, Athens, Greece.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details