



# VEHICLE COLLISION PREDICTION SYSTEM

<sup>1</sup> T.Sai Prasad Reddy, <sup>2</sup> M. Bhargavi

**Abstract:** Vehicle collision prediction system is for delivering the data of previous accidents and knowing the people about the accidents and reducing the accidents. Vehicle collision means colliding of any objects, the object may be movable or non-movable like a car collides a current pole or an object collides a person, or a vehicle will collide with another vehicle. Road accidents are a regrettable fact of life. Most days, there is at least one occurrence of a traffic collision reported in the daily news. The number of fatalities and the extent of damage might be worrying, depending on the accident's severity. Road accidents cause significant property damage and loss of life. Everyone should be more cautious and watchful on the roads to prevent accidents. Driving requires extreme caution since one mistake might result in the deaths of innocent people. The number of deaths due to road accidents in India is indeed a cause for worry. The scenario is very dismal with more than 137,000 people succumbing to injuries from road accidents. The negligence of people towards the speed of the vehicle, the vehicle condition and their own negligence. Road accidents is due to the increasing number of vehicles, the role played by the condition of the roads and other environmental factors cannot be overlooked. It will be good to analyze the occurrence of accidents so that this can be further used to help us in coming up with techniques to reduce them.

**Index Terms** -Road accident prediction, data analysis, feature engineering, machine learning.

## I.INTRODUCTION

In Vehicle collision prediction system, we have an data set. First, data set will be uploaded.This prediction system uses supervised machine learning algorithms for data pre- processing and analysing the data. Here, Random Forest gives the highest accuracy. Using Tkinter, graphs will appear. Graphs include weather conditions, location wise analysis, day wise analysis. We can also have new data entry where we can enter recently happened accidents. It can also give accuracy comparison graph.The aim of project is to deliver the data of previous accidents and know the people why the accidents are happening more and what are the main causes of road accidents.

## II.Literature review

### Prediction of the cause of accident and accident prone location on roads using data mining techniques

Pointing out the traffic collision data of roads the frequency of traffic collision of roads is analyzed using correlation analysis and exploratory visualization techniques. Government authorities will get to know the causes of accidents, the dominant factors such as weather conditions, transport infrastructure etc. that are inflicting upon the various accident prone regions and will help to provide assistance in sketching out the association between the various factors that directly or indirectly have a part in causing the accident.

### Text mining analysis of railroad accident investigation

The output from these analyses indicates that the railroad accidents can be successfully grouped into different topics. The output also suggests that recurring accident types are track defects, wheel defects, grade crossing accidents, and switching accidents. normalization of the data, followed by feature selection, where only relevant features from the original dataset are selected to be included in the final dataset. The dataset is then subjected to different data mining techniques. Clustering is performed on this dataset. The clusters are then subjected to other algorithms like Support Vector Machines (SVM) and Apriori.

## Analysis of road accidents in India using data mining classification algorithms.

This work analyzes the road accidents in India data set using classification algorithms namely linear regression, logistic regression, decision tree, SVM, Naïve Bayes, KNN, Random Forest and gradient boosting algorithm. Performance measures used are accuracy, error rate and execution time. This analysis is done in R data mining tool. This is followed by a summarized description of the methodology used in this work. Further, the different components of implementation including the system architecture, software and languages used, simulation, user interface

### A review on road accident detection using data mining techniques

Transportation has evolved greatly over time. With modern technology, the automobile industry has obtained new heights with respect to comfort, speed, efficiency and security. Despite improvement in technology, there has been increase in the rate of accidents. Accidents involving heavy goods vehicles like trucks and even those involving commercial vehicles used for public transportation like buses are some of the most fatal kind of accidents that occur, claiming the lives of innocent people.

### A Road Accident Prediction Model Using Data Mining Techniques

Due to the exponentially increasing number of vehicles on the road, the number of accidents occurring on a daily basis is also increasing at an alarming rate. With the high number of traffic incidents and deaths these days, the ability to forecast the number of traffic accidents over a given time is important for the transportation department to make scientific decisions. In this scenario, it will be good to analyze the occurrence of accidents so that this can be further used to help us in coming up with techniques to reduce them. Even though uncertainty is a characteristic trait of majority of the accidents, over a period of time, there is a level of regularity that is perceived on observing the accidents occurring in a particular area. This regularity can be made use of in making well informed predictions on accident occurrences in an area and developing accident prediction models. There are inter relationships between road accidents, condition of a road and the role of environmental factors in the occurrence of an accident. We have made use of data mining techniques in developing an accident prediction model using Apriori algorithm and Support Vector Machines.

### III. PROPOSED METHODOLOGY

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning. Which is process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

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 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.

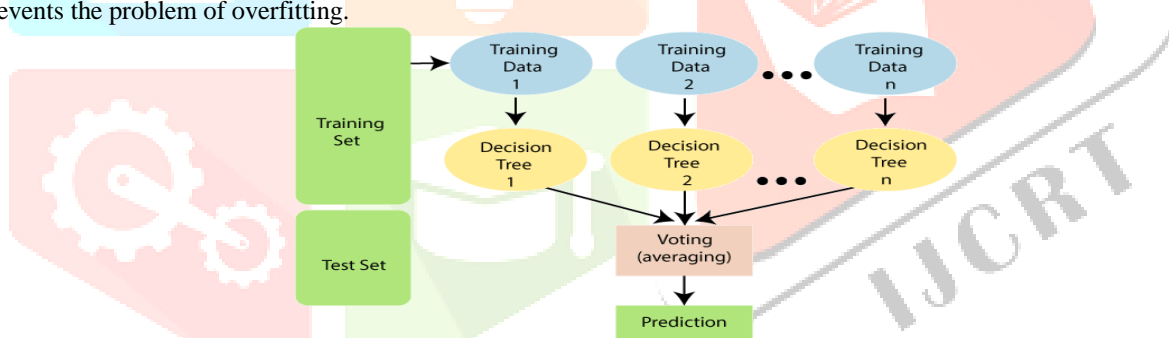


Fig 1. Random Forest

A data set is a collection of structured or unstructured data that is organized into a specific format. It can be any collection of data points, such as images, text, audio, or numerical data, that can be used to train and test machine learning models, statistical models, or other types of analytical algorithms. A data set typically includes multiple instances or examples, where each example represents a single data point with one or more attributes or features. Data sets can be gathered from a variety of sources, including surveys, experiments, social media, or other forms of data collection. They are used for various purposes, such as data analysis, research, and machine learning.

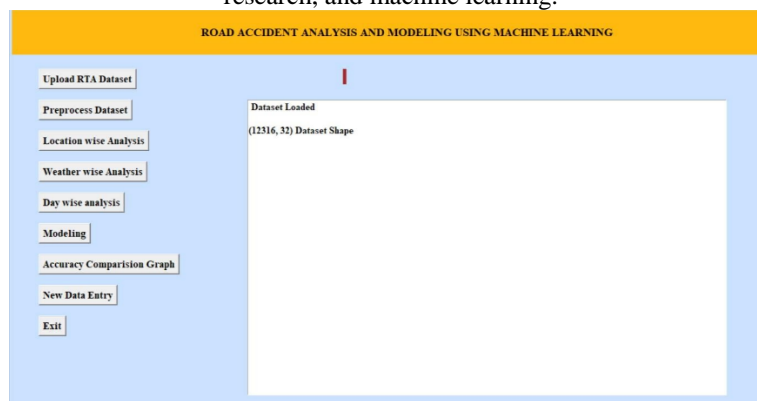


Fig 1. Dataset upload, a Tkinter Framework

Identifying major accident areas typically requires analyzing accident data from a particular region or jurisdiction. Accident-prone areas can vary depending on various factors such as traffic volume, road design, weather conditions, and driver behavior.

In general, highways and intersections tend to be high-risk areas for accidents. Some areas may have a higher frequency of certain types of accidents, such as rear-end collisions or intersection crashes. Identifying these patterns can help authorities take targeted measures to reduce the risk of accidents in those areas.

It's important to note that accident patterns can change over time due to factors such as changes in road design, increased traffic volume, or changes in driver behavior. Therefore, regular analysis of accident data is necessary to identify and address accident-prone areas.

There are several weather conditions that can potentially impact road safety and increase the likelihood of accidents. Some of the most significant weather conditions to consider when predicting road accidents include:

**Precipitation:** Rain, snow, and other forms of precipitation can reduce visibility, decrease tire traction, and make roads slippery, which can increase the likelihood of accidents.

**Fog:** Fog can greatly reduce visibility and make it difficult for drivers to see other vehicles, pedestrians, and road signs, which can increase the likelihood of accidents.

**Ice and frost:** These can make roads extremely slippery and decrease tire traction, which can increase the likelihood of accidents.  
**Wind:** Strong winds can make it difficult to control vehicles, especially for high-sided vehicles such as trucks and buses, which can increase the likelihood of accidents.

Identifying the major accident occurrence days can help authorities take appropriate measures to prevent accidents and reduce their frequency. The following are some common factors that can contribute to major accident occurrence days:

**Weekends:** Weekends, especially Friday and Saturday nights, tend to have a higher incidence of accidents. This is likely due to increased alcohol consumption and social events.

**Holidays:** Major holidays, such as Christmas, New Year's Eve, and Independence Day, can also have a higher incidence of accidents. Increased traffic volume and alcohol consumption are common factors that contribute to accidents during holidays.

**Weather conditions:** Weather conditions, such as heavy rain, snow, and fog, can increase the likelihood of accidents, especially during peak traffic hours.

**Sporting events:** Sporting events, especially those that involve large crowds, can increase traffic volume and contribute to accidents.  
**Road construction:** Road construction can cause traffic disruptions and increase the likelihood of accidents due to narrowed lanes and altered traffic patterns. Analysing historical data on accident occurrence can help identify the major accident occurrence days and the factors that contribute to them. Authorities can then take appropriate measures, such as increasing law enforcement presence, improving road infrastructure, and educating the public on safe driving practices, to prevent accidents and improve road safety.

Accurately predicting road accidents can be challenging as they are influenced by a wide range of factors such as weather conditions, driver behavior, road infrastructure, and vehicle safety features. However, machine learning algorithms can be trained using historical accident data to identify patterns and predict the likelihood of an accident occurring based on various variables.

The accuracy of accident prediction models depends on the quality and quantity of data used to train the algorithm, the relevance and significance of the features selected, and the complexity of the model. Typically, more data and more complex models tend to lead to more accurate predictions. Additionally, it's important to note that predicting accidents accurately doesn't necessarily prevent them from happening. However, accurate predictions can help authorities take proactive measures to reduce the likelihood of accidents occurring by identifying areas or factors that are more prone to accidents and taking steps to mitigate them.

#### IV. CONCLUSION

An accident can change the lives of many people. It is up to each of us to bring down this increasing number. This can be made possible by adopting safe driving measures to an extent. Since all instances of accidents cannot be attributed to the same cause, proper precautionary measures will also need to be exercised by the road development authorities in designing the structure of roads as well as by the automobile industries in creating better fatality reducing vehicle models. One thing within our capability is to predict the possibility of an accident based on previous data and observations that can aid such authorities and industries. This project was successful in creating such an application that can help in efficient prediction of road accidents based on factors such as weather condition, location wise analysis and day of the week.

#### REFERENCES

- 1) H. Song, M. J. Lynch, and J. K. Cochran, "A macro-social exploratory analysis of the rate of interstate cyber-victimization," *American Journal of Criminal Justice*, vol. 41, no. 3, pp. 583–601, 2016.
- 2) P. Alaei and F. Noorbahani, "Incremental anomaly-based intrusion detection system using limited labeled data," in *Web Research (ICWR), 2017 3th International Conference on*, 2017, pp.178–184.
- 3) M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the intrusion detection system," in *International Conference on Networked Systems*, 2015, pp. 513–517.
- 4) M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 5, pp. 516–524, 2010
- 5) S. Ashoor and S. Gore, "Importance of intrusion detection system (IDS)," *International Journal of Scientific and Engineering Research*, vol. 2, no. 1, pp. 1–4, 2011.
- 6) M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," *arXiv preprint arXiv:1312.2177*, 2013.
- 7) N. Chakraborty, "Intrusion detection system and intrusion prevention system: A comparative study," *International Journal of Computing and Business Research (IJCBR) ISSN (Online)*, pp.2229–6166, 2013.
- 8) P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly- based network intrusion detection: Techniques, systems and challenges," *computers & security*, vol.28, no. 1–2, pp. 18–28, 2009.
- 9) M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," *Procedia Computer Science*, vol. 89, pp. 117–123, 2016.
- 10) J. Zheng, F. Shen, H. Fan, and J. Zhao, "An online incremental learning support vector machine for large-scale data," *Neural Computing and Applications*, vol. 22, no. 5, pp. 1023– 1035, 2013.

- 11) F. Gharibian and A. A. Ghorbani, "Comparative study of supervised machine learning techniques for intrusion detection," in *Communication Networks and Services Research*, 2007. CNSR'07. Fifth Annual Conference on, 2007, pp. 350–358.
- 12) J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.
- 13) N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *Military Communications and Information Systems Conference (MilCIS)*, 2015, 2015, pp. 1–6.
- 14) T. Janarthanan and S. Zargari, "Feature selection in UNSW-NB15 and KDDCUP'99 datasets," in *Industrial Electronics (ISIE)*, 2017 IEEE 26th International Symposium on, 2017, pp. 1881–1886.
- 15) L. Dhanabal and S. P. Shantharajah, "A study on NSL-KDD dataset for intrusion detection system based on classification algorithms," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 6, pp. 446–452, 2015.
- 16) Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in *Proceedings of the 9th EAI International Conference on Bio- inspired Information and Communications Technologies (formerly BIONETICS)*, 2016, pp. 21–26.
- 17) M. Panda, A. Abraham, and M. R. Patra, "Discriminative multinomial naive bayes for network intrusion detection," in *Information Assurance and Security (IAS)*, 2010 Sixth International Conference on, 2010, pp. 5–10.
- 18) Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN," in *Signal Processing And Communication Engineering Systems (SPACES)*, 2015 International Conference on, 2015, pp. 92–96.
- 19) L. M. Ibrahim, D. T. Basheer, and M. S. Mahmod, "A comparison study for intrusion database (Kdd99, Nsl-Kdd) based on self organization map (SOM) artificial neural network," *Journal of Engineering Science and Technology*, vol. 8, no. 1, pp. 107–119, 2013.

