Road traffic accident severity prediction using machine learning

[1]Sivabrahmam Talloju, [2]Suneel Kumar Duvvuri,

[1]Student, [2]Assistant Professor
Computer Science,
Government College (Autonomous), Rajahmundry, India

Abstract: **This study performed a scientometric analysis of studies on traffic accidents in India between 1977 and 2020. It sought to look at the different types of publications and how they were cited and used, the growth in publications and citations year over year, the most popular journals, the keywords that authors preferred to use, the collaboration of Indian authors, the authorship pattern and the most prolific authors, and the top contributing organisations. 1,132 research articles were published throughout 44 years of study and included in the Web of Science (WoS) bibliographic collection. A study found that scholarly literature has made considerable progress, with the number of publications rising from one (0.08%) in 1977 to 182 (16.07%) in 2018.The project's objective is to examine traffic accidents in India at the national, state, and metropolitan city levels. According to the data, the distribution of Traffic-related fatalities and injuries in India differs by age, gender, month, and Time. The population group most at risk is 30 to 59 years old, while males are more likely than females to die or sustain injuries. Additionally, severe weather and working hours are when car accidents are comparatively more common. The probability of fatalities varies greatly between states and localities, according to analysis of road accident scenarios at the state and municipal level. The risk of death exceeds the national average in 16 of the 35 states and union territories. Although the number of road accidents in India's metropolises is slightly lower, nearly 50% of the cities have a higher risk of fatalities than their metropolitan counterparts. Without greater efforts and fresh ideas, India's overall number of Traffic fatalities is predicted to surpass 250,000 by 2025.The majority of academic works (740, 65.37%) were published as articles. The number of publications climbed quickly from 11 (0.97%) in 2006 to 182 (16.07%) in 2018, The most active year for the researchers. 392.95 citations were made to an average of 25.73 published documents annually. Most of the publications—108, or 30%—were published in journal of Evaluation of Medical and Dental Sciences. The keyword "Trauma" was used the most frequently. The majority of articles (83.38%) on road traffic accidents (RTA) were authored by Indian authors, either alone or in conjunction with local authors. The Indian Institute of Technology made a significant contribution by publishing 120 documents (10.60%).**

I. INTRODUCTION

The definition of an accident is an incident during a series of events that typically results in unanticipated harm, death, or property damage. One of the leading causes of death nowadays is an accident. Worldwide, the number of fatal road traffic accidents (RTA) is rising alarmingly. The World Health Organization (WHO) was urged to act in response to the high death toll from car accidents fifty years ago. RTAs continue to be a major cause of morbidity and mortality among people, nonetheless. It is crucial to note that low- and middle-income nations are now bearing the brunt of RTAs (LMICs), Which account for 90% of the weight of the 12.5 lakh road traffic deaths and two to five crore non-fatal road traffic injuries, although having only around 50% of the world's motor cars (Lena, etal., 2019). Road traffic accidents are the sixth leading cause of death in India, accounting for 16.6% of all fatalities. They also have a significant impact on socioeconomic losses, the burden of disabilities, and hospitalisation. Poorer rural households and those without health insurance who seek treatment at private medical facilities face a disproportionately large price burden. Road accidents are a primary cause of death in India and occur often (Goli, Shruti, Siddiqui, & Gouda, 2018). In all traffic collisions, the majority of human lives are lost in traffic accidents, which are also generally regarded as the most significant global issue. Each year, almost 1.2 million people are thought to die in road traffic accidents (RTAs) worldwide, while 50 million more are thought to be wounded (Pathak, Gadhari, Chaudhari, & Devraj, 2018). By 2020, traffic-related fatalities in India will rank third in terms of causes of death. Unless there is a fresh commitment to prevention, these numbers might rise by as much as 80% in low- and middle-income nations over the next 20 years (World Health Organisation, 2004). According to predictions made by Murray and Lopez, road fatalities and injuries would rank third among all causes of disease and injury by the year 2020.

Every 23 seconds, there will be a death, followed by 43,640 every month and 7,22,917 per year. South East Asia has the highest rate of road fatalities per 1,000 people worldwide. The number of fatalities from traffic accidents is gradually increasing. The rate of mortality, however, has stayed steady in relation to the expansion of the global population. Over 231,000 people are killed in traffic accidents in India each year, according to the Global Status Report on Road Safety for 2013. (World Health Organisation, 2013). 50% of all fatalities on American roads are caused by motorcyclists, pedestrians, and cyclists, who represent the majority of the most vulnerable road users. The high fatality rates observed on Indian roads are a result of a heterogeneous traffic mix, which includes high-speed cars sharing the road with vulnerable road users, unsafe road infrastructure, and vehicles that are in poor condition (World Health Organization, 2018) [1].

In India, vehicle traffic accidents are a serious and expanding public health issue that cause both deaths and injuries. Every week, traffic accidents result in 9,000 injuries and about 2,650 fatalities. In India, vehicle accidents caused 137,423 fatalities and 469,900 injuries in 2013, the most recent year for which statistics is available. With approximately 140,000 deaths per year, India has surpassed China to take the top spot in the world for road fatalities as a result of traffic accidents. India is the only country in the

world where traffic accidents result in more than 15 fatalities and 53 injuries every hour. While the situation is generally improving in many developed and developing nations, including China, India is seeing a worsening situation. If the pattern holds, there will be a 100% increase in road traffic fatalities in India between 2013 and 2027. By 2025, India's total number of traffic fatalities is anticipated to surpass 250,000 without more effort and innovative programmes. This study's primary goal is to examine the national, state, and metropolitan city levels of road traffic accidents in India.

The main focus would be on identifying the significant concerns with road safety and discussing solutions that could be able to address the particular challenges. Accidental Deaths & Suicides in India, 1970 to 2013, a publication of the National Crime Records Bureau, Ministry of Home Affairs, Government of India, New Delhi, serves as the study's main source of data. According to the data, India's road accident fatalities have climbed by 5% annually over the past 10 years, despite the country's population only growing by 1.4% annually. As a result, the probability of fatalities from traffic accidents has increased from 7.9 in 2003 to 11.2 in 2013. The chance of death is not just quadruple in India compared to some other industrialised nations, such as Sweden and the United Kingdom, but it is also rising quickly. Additionally, it has been discovered that the distribution of traffic accident fatalities and injuries differs by age, gender, month, and time. The most vulnerable age group among all age categories is the 30- to 59-year-old economically engaged group. But when we looked at accidents and fatalities by gender, we discovered that in 2013, men made up 85.2% of all fatalities and 82.1% of all injuries. Road accidents are also more frequent in May–June and December–January, demonstrating how harsh weather can affect their frequency. Between 9 AM and 9 PM, accidents are often continuous and high, while between midnight and early in the day, they are variable but low. This does not, however, indicate that driving during the day is more dangerous than driving at night. The study also looks at the distribution of traffic accidents by causes. Drivers' negligence, which accounted for 78% of all incidents in 2013, is the most significant of the several factors that cause accidents. In 2013, three states and union territories—Tamil Nadu (22.8), Haryana, and Uttar Pradesh—recorded more fatal traffic accidents than any other (17.2), and Andhra Pradesh (16.9), faced a 50% higher risk of fatalities than the average for all of India (11.2). The prevalence of traffic accidents is determined to be very low in India's major cities (cities with a population of 1 million or more), yet the risk of deaths varies from 3.0 per 100,000 in Kolkata to 25.5 per 100,000 in Jaipur. Six out of the 21 metro areas that were chosen saw a greater increase in fatality risk from 2003 to 2013 than the rest of the nation. The probability of fatalities increased the most in Ahmedabad (from 0.6 to 4.2), then in Varanasi (from 9.5 to 17.9), Patna (9.2 to 17.4), Chennai (8.8 to 14.3), Jaipur (15.9 to 25.5), and Vishakhapatnam (15.5 to 22.0) [2].

Numerous injuries and fatalities are thought to be caused by traffic accidents, can significantly harm the country's economy. According to the World Health Organization (WHO), there are 1.25 million yearly traffic accident fatalities and 20–50 million injuries. As a result, road accidents are responsible for 23% of direct fatalities and 2.1% of all fatalities. Typically, 3% of gross domestic product is used to make up for accident-related losses. Although developing nations account for 54% of the world's automobiles, these areas account for more than 90% of fatal accidents.[3]

Understanding and identifying the patterns of road accidents in India is the goal of our research. We have concentrated on the causes of traffic accidents in order to offer insightful information that will help decision-makers better plan their initiatives. We focus on four key factors affecting traffic accidents based on the literature study presented in section 2: Festivals, tourists, the seasons, and the age and gender of the driver are all factors.

The study responds to the following inquiries:

1. What connection exists between festivals and traffic incidents?
2. Are there any connections between yearly seasons and traffic accidents?
3. What effect does tourism have on auto accidents?
4. How do the driver's gender and age affect traffic accidents?[4]

problem statement

Furthermore, in recent years, traffic-related problems have been the primary cause of serious accidents and fatalities. The issue comes with carelessness and the incorrect methodology used to better analyse traffic incidents. Road accidents are liable for a variety of reasons, including drunk driving, weather-related causes, pedestrian crossings, braking failure, tyre blowouts, etc. If factors could be evaluated, understood, and acted upon to take the proper actions, a solution would be simple to find. It is necessary to control traffic accidents by analysing cause-and-effect laws in addition to improving road conditions in order to prevent accidents [5].

objectives

1. To investigate the various publications' RTA categories, citation patterns, and usage.
2. To determine the growth in publications and citations on RTA by year.
3. To identify the publications that are preferred for RTA research.
4. To learn which keywords RTA publications prefer to use.
5. To learn about the authors' collaboration on RTA research.
6. To be aware of the RTA literature's authorship pattern and prolific writers.
7. To research the leading institutions that contribute to RTA research [6].

project scope

This project needs a coordinated scope of work. These scopes will help to focus on this project. The scopes are:

i. Modified existing machine learning algorithm.

ii. Make use and classify of a data set including data preparation, Classification and visualization.

iii. Score of data to determine the accuracy of RTA.

machine learning

The process by which computers learn to recognise patterns, or the capacity to continuously learn from and make predictions based on data, then make adjustments without being specifically programmed to do so, is known as machine learning (ML), a subcategory of artificial intelligence.

How Machine Learning Works

The operation of machine learning is quite complicated and varies according to the task at hand and the algorithm employed to do it. However, at its foundation, a machine learning model is a computer that analyses data to spot patterns before using those realisations to better fulfil the work that has been given to it. Machine learning can be used to automate any operation that depends on a set of data points or rules, including more difficult tasks like answering customer service calls and analysing resumes. Machine learning algorithms work with more or less human intervention/encouraging depending on the circumstance. Supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning are the four main machine learning models. By giving the computer a labelled collection of data, supervised learning enables it to teach itself how to perform a human task. Given that it aims to mimic human learning, this model is the simplest. When unsupervised learning is used, the computer uses unlabelled data to discover patterns and insights that were previously undiscovered. Machine learning algorithms accomplish this in a variety of ways, including:

•Creating "clusters" is the result of the computer finding comparable data points within a data collection and grouping them accordingly.

• Density estimation, in which a computer analyses the distribution of a data set to find insights.

• Data point anomaly detection, whereby a computer finds data points in a dataset that differ noticeably from the rest of the data.

• Principal component analysis (PCA), in which a data collection is analysed and summarised by a computer so that precise predictions can be made from it.

II. LITERATURE SURVEY

related work

According to Shami (2005), about seven million Indians experience traffic-related injuries each year, ranging from minor mishaps to fatalities. Most of the victims were among the vulnerable road users, such as cyclists and pedestrians. Traffic advertising, however, emphasised the safety of drivers of four-wheel vehicles and the efficient flow of traffic. Nearly 2% of India's GDP was lost annually to traffic accidents, which was an unacceptable expense. A well-informed, multi-sectoral strategy supported by a strong commitment to minimising road accidents might make a significant difference. Shami (2005)'s findings were corroborated by Sundar and Ghate's (2013) research, which also made clear that among all nations, The majority of fatalities from traffic-related accidents occurred in India. Road accidents were the sixth greatest cause of mortality in the nation in 2012, accounting for approximately 1,40,000 fatalities. Road safety had not gotten the attention it deserved while being a significant public health concern that impacts both the most vulnerable and productive segments of society. Additionally, they talked about how the government had been sluggish to pass legislation to set up the institutional structures to improve road safety because it had failed to understand the importance of roads for mobility, health, and equity.[7]

In 2012, Zangooei Dovom et al. [8] looked at the statistics of fatal road accidents in Mashhad, another major city in Iran's east. Similar to this, senior people had the highest accident rate. However, the most common type of vehicle used in fatal road accidents was a motorbike for those who were under 30 years old and a car for those who were older. According to the current study, hemodynamic instability, chest tube placement, orotracheal intubation, and brain injuries were all factors that contributed to a patient's death while they were hospitalised after a car accident.

Hamzeh et al.'s[9] attempt to pinpoint the conditions that elevated the probability of traffic-related deaths across a nine-year period in Western Iran were consistent with our findings and pointed to head and face injuries as well as unstable vital signs as the main causes of death.

The most significant causes of death in traffic accidents, according to Sánchez-Mangas et al.'s[10] investigation of a sizable population in Spain in 2010, were head and facial injuries, unstable vital signs upon arrival, older age, and the requirement for artificial ventilation. Ratan Tata Trust (2011), Bayan et al. (2013),

Kanchan et al.[11] (2012), Bayan et al. (2013) and Ratan Tata Trust (2011) deduce that younger males and older females experience higher number of accidents in India. Shibata & Fukuda (1994) indicate that male unlicensed drivers are more prone to injury due to road accidents. Regarding age, (Mohan, 2009) children comprise only 6% of the fatalities, though their share in the Indian population is 32%; however, the middle-age groups (30-59 years) are over-represented by about 70%. These studies highlight the role of demographic variables like age and gender in mediating occurrence of traffic accidents.

III. ANALYSIS

data source

Fig.1. List of datasets by Kaggle

Figure 1 shows the list of datasets in Kaggle. Due to several restrictions, such as the volume of data and the throughput needed for accurate and timely intake, gathering data is incredibly challenging. The dataset I utilised for this project is genuine data that can be downloaded from a site that houses machine learning data.

data sets

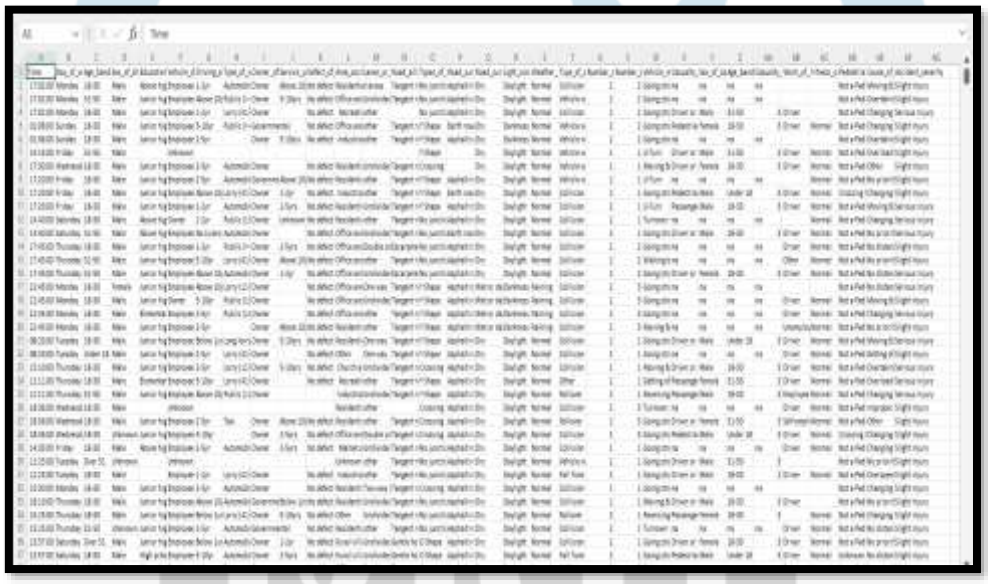

Fig.2. Dataset Details

The list of data sets offered by each website is displayed in figure 2.
For training and testing, this dataset may have more than 1000 tagged messages. The data first need to be divided into training and reformatted into.CSV. test and csv. In order to make it simpler to utilise for further processes, csv files and headers will be included.

process model

A process model is a list of actions, each with a clear description and associated choices that must be made in order to complete the project execution. The project flows must be followed in order to complete the project in the allotted time.

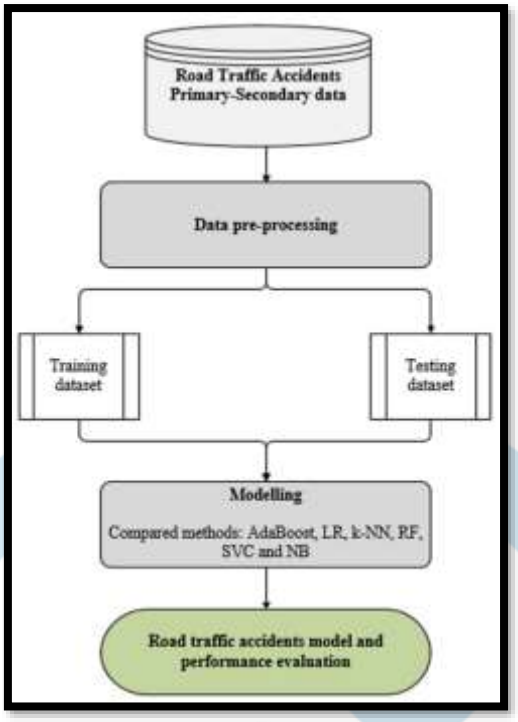

Fig.3. Process model

description

The data model flow is crucial to this project since it demonstrates how the project should be structured and how the various steps relate to one another. It aids in organising the project process in a transparent and efficient manner. Jupiter Note Book serves as the project's development platform and is based on the framework. First, examine every feature of Jupiter Note Book to ensure the project can accomplish its goals. As a training and testing dataset, make sure to download the actual, existent dataset from the machine learning data repository. The first step in preprocessing data is reformatting the dataset into two distinct files in training.csv and testing.csv formats. The prepared Code should then be uploaded into Jupiter Note Book.

Go to New in Jupyter and choose Python3 to create a notebook. We can now see that a new tab has opened with a new notebook. Without the code, the.csv Datafile in Jupiter Note Book, which is in Excel format, will not run. KNN, Logistic Regression, Random Forest, Support Vector Machine, Multilayer Perceptron, GradientBoost, and AdaBoost have all been utilised. The likelihood of a road traffic collision being detected is predicted using neural networks. After the model has been trained using the data, it must be evaluated to determine its accuracy and overstrain it to ensure that it retains the data. The confusion matrix will display the results, and you may choose the method that will produce the best results depending on its accuracy level.

confusion matrix

A matrix known as the confusion matrix is used to assess the effectiveness of the classification models for a particular set of test data. It cannot be determined until the true values of the test data are known. The matrix itself is straightforward to comprehend, but some of the language used in relation to it might not be. Since it depicts the errors in the model performance as a matrix, it is also known as an error matrix.

The Confusion matrix features listed below include:

The projected values and actual values, along with the total number of forecasts, are separated into two dimensions in the matrix. The matrix is divided into two dimensions, which are total number of predictions, anticipated values, and actual values.

• Projected values are those that the model predicts, whereas actual values are the actual values for the data that has been provided.

It appears as in the following table:

$$\text{Accuracy}=TP+TN/TP+TN+FP+FN . \quad (1)$$

$$\text{Precision}=TP/TP+FP. \quad (2)$$

$$\text{Recall}=TP/TP+FN . \quad (3)$$

$$F1=2TP/2TP+FP+FN . \quad (4)$$

The following cases are included in the table above:

True Negative: The real or true value was also No, as predicted by the model.

True Positive: The model's forecast and the actual value agreed.

False Negative: This error, which occurred when the model predicted no but the actual value was yes, is sometimes referred to as a Type-II error.

False Positive: Although the model expected Yes, the actual result was No. Another name for it is a Type-I mistake.

LIBRARIES NEEDED: NumPy, Pandas, Matplotlib, Seaborn, Sklearn.

IV. ALGORITHM IMPLEMENTATION

Adaboost Classifier

AdaBoost, also known as Adaptive Boosting, is a machine learning method used in an ensemble setting. Decision trees with one level, or Decision trees with only one split, are the most popular algorithm used with AdaBoost. Another name for these trees is Decision Stumps. This algorithm creates a model while assigning each data piece an equal weight. Then, it gives points that were incorrectly categorised larger weights. The next model now gives more weight to all the points with higher weights. If no lowe error is received, it will continue to train the models. [12]

Gradient Boost Classifier

Gradient Boosting Machine (GBM) is one of the most popular forward learning ensemble methods in machine learning. It is a powerful technique for building predictive models for regression and classification tasks.

We can create a predictive model using GBM by combining a number of weak prediction models, like decision trees. Gradient-boosted trees is the name of the ensuing algorithm whenever a decision tree operates as a poor learner.

By combining the predictions from different learner models, we may create a final predictive model that makes the right prediction.[13].

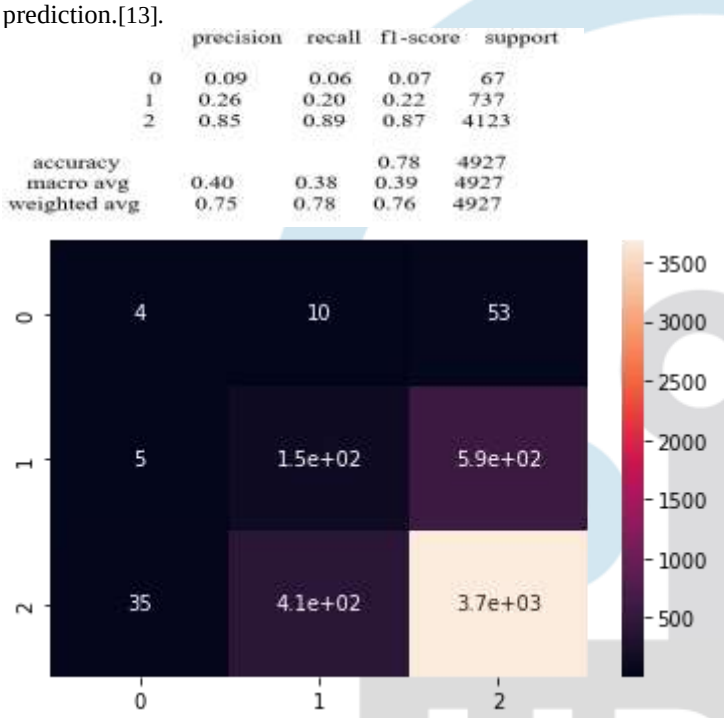

Fig.4. Confusion Matrix

V. RESULT AND DISCUSSION

Table 1 Comparision of Accurancies

ALGORITHM	ACCURACY (%)
RANDOM FOREST	84.16%
NAIVE BAYES	79.72%
DECISION TREE	72.66%
KNN	82.65%
LOGISTIC REGRESSION	84.08%
SVM	84.08%
ADABOOSTCLASSIFIER	83.87%
GRADIENTBOOSTINGCLASSIFIER	84.35%

All the eight machine learning methods accuracies are compared based on which one prediction model is generated. Hence, the aim is to use various evaluation metrics like confusion matrix, accuracy, precision, recall, and f1-score which predicts the problem efficiently. Comparing all eight the Gradient Boosting Classifier gives the highest accuracy of 84.35%.

VI. CONCLUSION AND FUTURE WORK

Despite the rising number of traffic-related deaths and injuries, local, state, and federal governments have not given road safety enough priority. The primary cause of this is that no specific agency, either at the national, state, or municipal government levels, is in charge of solving the issue of traffic accidents. Numerous organisations, industries, and groups are responsible for addressing the various facets of issues, including vehicle roadworthiness tests, the design of road networks and roadways, urban planning, the development and implementation of road safety regulations, and post-crash medical care. Typically, there hasn't been a leader to make sure that they coordinate their efforts and deal with the issue.holistically. This situation needs to change so that responsibility is clearly assigned, specific roles are allocated to specific agencies, and duplication is avoided.

Since adopting a systems approach to road safety that emphasises environment, vehicle, and road user interventions rather than only concentrating on direct approaches aimed at changing the behaviour of road users, many countries, particularly from the developed world, have seen a sharp reduction in road traffic accidents and fatalities. Although India may not have the same road safety issues as other nations with high rates of motorization, some fundamental ideas would still apply. They include, for instance, enhanced vehicle standards, traffic management, safe road design, the usage of seat belts and helmets, and the enforcement of alcohol consumption limits (Margie et al., 2004). Compared to what has to be done, current attempts to solve the issues of road safety are insufficient. Even though there are numerous interventions that can save lives, political commitment and will at the federal, state, and local levels of government are crucial, and without them, not much can be done. India's drivers deserve to travel on roads that are better and safer.

REFERENCES

[1] https://www.researchgate.net/publication/344808431_Road_Traffic_Accident_Research_in_India_A_Scientometric_Study_from_1977_to_2020
[2] https://www.researchgate.net/publication/317421409_Road_Traffic_Accidents_in_India_Issues_and_Challenges
[3] https://www.researchgate.net/profile/Masumeh Safaee 2/publication/358977661_Road_Traffic_Accident_Fatality_Predictors_A_CaseControl_Study_in_Isfahan/links/6220770badd1b367ae0f6c65/Road-Traffic-Accident-Fatality-Predictors-A-Case-Control-Study-in-Isfahan.pdf
[4] https://www.researchgate.net/publication/327221931_An_exploratory_study_on_the_pattern_of_traffic_accidents_in_India
[5] https://www.slideshare.net/shraddhamane4/road-accident-alert-system
[6] https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=8106&context=libphilprac
[7] Dhaigudeetal_2018_AnexploratorystudyonthepatternoftrafficaccidentsinIndia_WTPP_24_2_.pdf
[8] Zangooei Dovom H, Shafahi Y, Zangooei Dovom M. Fatal accident distribution by age, gender and head injury, and death probability at accident scene in Mashhad, Iran, 2006-2009. Int J Inj Contr Saf Promot 2013;20:121-33.
[9] Hamzeh B, Najafi F, Karamimatin B, Ahmadijouybari T, Salari A,Moradinazar M.Epidemiology of traffic crash mortality in west of Iran in a 9 year period. Chin J Traumatol 2016;19:70-4.
[10] Sánchez-Mangas R, García-Ferrrer A, de Juan A, Arroyo AM. The probability of death in road traffic accidents. How important is a quick medical response? Accid Anal Prev 2010;42:1048-56.
[11] Kanchan, T., Kulkarni, V., Bakkannavar, S.M., Kumar, N. & Unnikrishnan, B. (2012). Analysis of fatal road traffic accidents in a coastal township of South India. Journal of Forensic and Legal Medicine, 19. 448-451.
[12] https://www.javatpoint.com/gbm-in-machine learning#:~:text=AdaBoost%20(Adaptive%20boosting)%20was%20the,tasks%20such%20as%20binary%20classification.
[13] https://www.analyticsvidhya.com/blog/2021/09/adaboost-algorithm-a-complete-guide-for-beginners/