

# A Comprehensive Analysis of Indian Road Accidents

Arbaz Hussain<sup>1</sup>, Adil Saifi<sup>2</sup>, Shaalam Tyagi<sup>3</sup>, Lomash Pal<sup>4</sup>

UG Scholars, Department of Computer Science and Engineering

Sunder Deep Engineering College, Ghaziabad, India<sup>1,2,3,4</sup>

**Abstract:** *This study proposes a data-driven approach utilizing Python and machine learning techniques to investigate the trends and underlying causes of traffic accidents in India. Road safety has become a major public concern due to the quick increase in vehicle density and urbanization. In order to find important trends, the study makes use of publicly accessible accident records that include characteristics like location, time, weather, vehicle type, and human behavior indicators. To find hidden patterns and insights, data pretreatment, exploratory data analysis (EDA), and visualization approaches are used. To determine predictive accident risks and accident frequency, a random forest machine learning model is employed. The model's performance is assessed using common metrics to guarantee accuracy and dependability. The results provide insights into the temporal and geographical distributions of road events throughout various regions of India and identify important risk variables impacting accident incidents. The study demonstrates how predictive modeling can help urban planners, traffic authorities, and legislators make well-informed decisions with the goal of lowering accident rates and enhancing road safety. Through the application of analytical techniques and computational tools, this study advances the creation of proactive accident management and preventive measures*

**Keywords:** Traffic Accident Analysis, Machine Learning, Random Forest, Road Safety, Predictive Modeling.

## I. INTRODUCTION

In recent years, road transport in India has rapidly expanded due to population growth and urban development. As a result, issues such as traffic congestion and road accidents have also risen. We are Studying road-related data of major cities in India. effectively is crucial for understanding the primary causes of these issues and developing improved solutions. This research paper utilizes Python for data analysis and machine learning methods to investigate a dataset related to Indian roads.

The dataset utilized in this research includes critical information concerning roads, traffic situations, and accident specifics. By studying this data, we can pinpoint patterns and trends, including which road types are more likely to have accidents or under what circumstances accidents occur more often. Basic analysis techniques might fall short in grasping intricate relationships within the data, which is why machine learning models are utilized to achieve improved outcomes and forecasts.

This research utilizes Python due to its user-friendly nature and the abundance of valuable libraries available. Libraries such as Pandas and NumPy are utilized for cleaning and processing data, whereas Matplotlib and Seaborn are employed for visualizing data. Scikit-learn is used to construct machine learning models for data analysis and predictions.

The primary goal of this research is to demonstrate how machine learning can enhance the understanding of road data. The findings of this study can aid in enhancing road safety, decreasing accidents, and fostering improved traffic management.

In general, this study emphasizes the significance of employing data analysis and machine learning to address practical issues in the transportation industry.



Quantify the year-over-year changes in accident frequencies and severities across multiple cities in India. Evaluate the real-world efficacy of emerging vehicle safety technologies (e.g., Autonomous Emergency Braking, Lane Keep Assist) that reached critical market penetration during this timeframe. By addressing these objectives, this paper provides policymakers, urban planners, and automotive engineers with empirical evidence to guide the next generation of road safety initiatives.

## II. LITERATURE REVIEW

Many researchers have worked on analysing road accident data using machine learning techniques. These studies mainly focus on predicting accidents, identifying risk factors, and improving road safety.

Hemalatha and Dhuwaraganath (2024) developed a machine learning model to predict road accidents using factors like speed, traffic conditions, and weather. The study showed that machine learning models can improve prediction accuracy and help in reducing accidents.

Another study by Yuan Yuan et al. (2020) reviewed different machine learning methods used in road safety analysis. It was found that models like Artificial Neural Networks perform well in identifying accident patterns, especially when environmental and road-related factors are considered.

Research by Savolainen et al. (2011) focused on predicting injury severity in road accidents. The study concluded that Random Forest gives better performance in many cases.

Another paper analyzed accident data to find major causes of road accidents. The study found that weather conditions, road infrastructure, and human errors are the main reasons for accidents. It also suggested that proper data analysis can help in improving road safety policies.

A recent study on Indian highways used machine learning models like Random Forest and Gradient Boosting to predict accident severity. It also identified important factors affecting accidents, which can help authorities in taking preventive measures.

Overall, previous research shows that machine learning is very useful for analysing road accident data. However, more dataset-specific studies are required, which this project aims to address.

## III. RESEARCH GAP

While machine learning has been extensively utilized for analysing road data, significant gaps still remain. Numerous studies primarily concentrate on creating prediction models, yet they often lack comprehensive exploratory data analysis, hindering the understanding of underlying patterns. Additionally, appropriate data preprocessing measures, such as managing missing values and selecting features, are not always explicitly discussed, which impacts the accuracy of the model. Furthermore, the results are often not clearly explained, which complicates the identification of the main factors contributing to road-related problems. In some instances, the comparison among various machine learning algorithms is also restricted. Thus, this study intends to integrate effective data analysis, preprocessing, model comparison, and straightforward interpretation to produce more significant and actionable insights.

## IV. METHODOLOGY

This research follows a structured approach to analyse the Indian roads dataset using data analysis and machine learning techniques in Python.

### Data Collection

The dataset is collected from (<https://www.kaggle.com/datasets/sehaj1104/indian-road-accident-dataset-20222025>) containing information related to roads, traffic conditions, and accident details. The data is in structured format and is used for further analysis.



**Data Preprocessing:**

In this step, the dataset is cleaned and prepared for analysis. Missing values are handled, duplicate records are removed, and unnecessary columns are filtered. Data is also converted into proper format for better processing.

**Model/Algorithm Development:**

Different machine learning algorithms such as Decision Tree, Random Forest, and Linear Regression are selected and implemented using Python libraries. These models are used to analyse the data and make predictions.

**Training & Testing:**

The dataset is divided into training and testing sets. The models are trained using training data and tested on unseen data to check their performance.

**Model Performances:**

Linear Regression -> MSE: 0.036691939187786224 R2: 0.21009374692304117

Decision Tree -> MSE: 0.03678324317639454 R2: 0.20812814921605005

Random Forest -> MSE: 0.03694262635503712 R2: 0.2046969386495895



Fig.1 A Comparison of different Machine Learning Models

**Evaluation:**

The performance of models is evaluated using metrics like accuracy and comparison is done to find the best model. Final results are analyzed to identify important factors and provide useful insights.

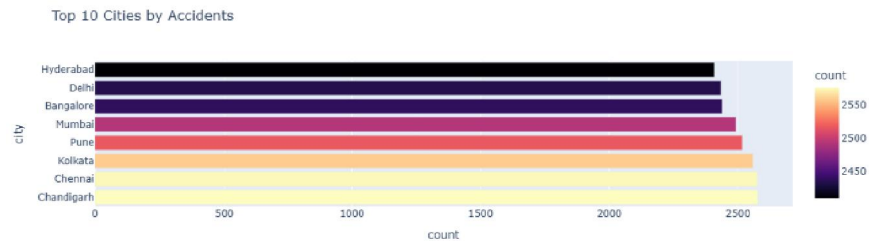


Fig. 2 Top Cities by Accidents



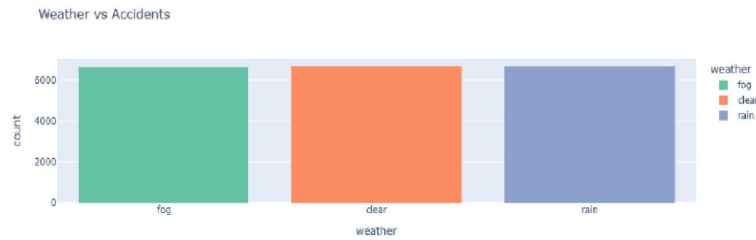


Fig.3 Weather Conditions with Count of Accidents

## V. DATA ANALYSIS AND RESULTS

Analysis of road accident data in India uncovers several key trends concerning the incidence, causes, and regional and demographic distribution of accidents.

### Overall Accident Trends

The total count of road accidents exhibits a varying yet overall upward trend throughout the study period. Even with enhancements to road infrastructure and safety policies, accident rates continue to be elevated, highlighting deficiencies in execution and adherence by the public. Fatal accidents represent a significant share of all incidents, underscoring the seriousness of collisions.

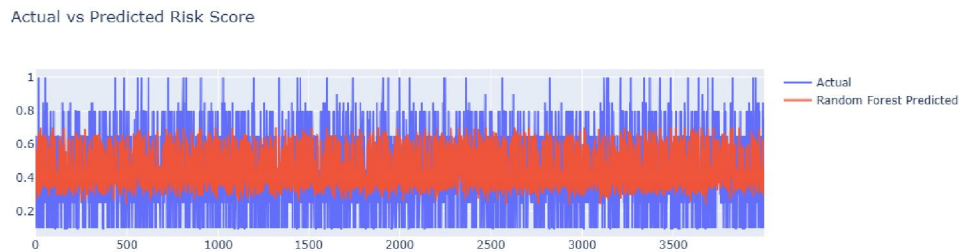


Fig.4 Predicted Risk Score by Random Forest Algorithm

### Factors Leading to Accidents

The main causes identified are Exceeding speed limits (responsible for the biggest portion of accidents) Driving while intoxicated by alcohol or drugs Distracted driving (e.g., Usage of mobile phones

Breaking traffic regulations Exceeding speed limits by itself accounts for over half of all accidents, marking it as the primary risk factor.

## VI. CONCLUSION

Road accidents are showing a decreasing trend in several major Indian cities including Kolkata, Chennai, Pune, Delhi, and Chandigarh—largely due to a multi-pronged approach involving strict, technology-driven enforcement, better traffic engineering, and targeted safety campaigns. While vehicle volume continues to rise, improvements in infrastructure and emergency care have successfully reduced the overall number of crashes and fatalities in these areas. In the current study, data analysis and machine learning techniques are used on the dataset of Indian Roads using Python. The current study aids in analyzing various patterns, trends, and relationships that exist among the variables associated with road accidents.



Various Machine Learning techniques were employed to identify the best performing model for predicting the occurrence of accidents. It is evident from the results that machine learning techniques can prove very useful for analysing the dataset related to roads and finding factors responsible for road accidents.

From the above discussion, it is clear that use of data analysis techniques plays a significant role in addressing the issues of the real world through data. The outcome of the current study can prove useful in improving road safety.

#### REFERENCES

- [1] N. Sharma, M. Bairwa, B. Gowthamghosh and S. D. Gupta, "A bibliometric analysis of the published road traffic injuries research in India, post-1990," *Health Research Policy and Systems*, vol. 16, no. 18, 2018.
- [2] M. M. Ruikar, "National statistics of road traffic accidents in India," *Journal of Orthopaedics, Traumatology and Rehabilitation*, vol. 6, no. 1, pp. 1–6, 2013.
- [3] P. Y. Patil, "Statistical study of public health disaster due to road traffic accidents in India," *International Journal of Research – GRANTHAALAYAH*, vol. 11, no. 11, 2023.
- [4] S. Abhinandithe K., S. Balasubramanian and S. K. Sahana, "Analysis of road accidents in Kerala, India, using data mining techniques," *IJBPAS*, vol. 11, no. 9, 2022.
- [5] K. V. Mhetre and A. D. Thube, "Count data modeling for predicting crash severity on Indian highways," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11816–11820, 2023.
- [6] D. Arya *et al.*, "RDD2022: A multi-national image dataset for automatic road damage detection," *arXiv preprint arXiv:2209.08538*, 2022.
- [7] D. Arya *et al.*, "Crowdsensing-based Road Damage Detection Challenge (CRDDC-2022)," in *Proc. IEEE Int. Conf. Big Data*, 2022.
- [8] K. Doshi and Y. Yilmaz, "Road damage detection using deep ensemble learning," *arXiv preprint arXiv:2011.00728*, 2020.
- [9] D. Arya *et al.*, "Global road damage detection: State-of-the-art solutions," *arXiv preprint arXiv:2011.08740*, 2020.
- [10] Ministry of Road Transport and Highways (MoRTH), *Road Accidents in India 2021*, Government of India, New Delhi, 2022.

