Ensemble Learning Algorithms based on Road Accident Data Prediction

Dr. Gondi Konda Reddy[1], Dr. Nagagopi Raju Vullam[2], Dr. G. Soma Sekhar[3], Dr. Dheeraj Sundaragiri[5], and Dr. Subhani Shaik[5]*

[1]Dept. of ME, Sreenidhi Institute of Sciences and Technology(A), Hyderabad, India.
[2]Dept. of AIML, Chalapathi Institute of Science and Technology(A), Guntur, A.P., India.
[3]Dept. of CSE, Vardhaman College of Engineering(A), Hyderabad, India.
[4]Dept. of CSE, Sreenidhi Institute of Sciences and Technology(A), Hyderabad, India.
[5]Dept. of IT, Sreenidhi Institute of Sciences and Technology(A), Hyderabad, India.

Abstract. Road accidents are common nowadays due to traffic, and road conditions. It affected life threats to propel lives. So to follow instructions of specific road conditions and safe driving. All this information about accidents, to learn lessons from these incidents and implement novel strategies to avoid accidents under any conditions. Environmental factors also play a key role in road accidents even if drivers are carefully driving vehicles. In our research paper, we construct an accident prediction model using ensemble machine learning techniques. Due to this classification system to implement of safety measures and accident prediction. Different types of elements like weather, road conditions, lighting, and vehicle conditions are used to predict road accidents. Our dataset's attributes are accidents, vehicles, and casualties. Our statistical results generate Logistic Regression, which gives the highest accuracy rate among all machine learning algorithms.

1 INTRODUCTION

Road safety is an important public health issue worldwide, with road accidents being the leading cause of injury and death [2]. As urbanization and automobile use increases, so does the frequency of road accidents, and when the need for efficient road access has emphasized an emphasis on security how To meet this challenge, data-driven techniques have become increasingly valuable, providing insights that can lead to more informed decision-making and implement policies [3].

In this paper towards Safer Roads: Using Machine Learning to Predict Accidents," we examine the use of machine learning techniques to analyze accident data and predict accident outcomes [4]. By using historical accident data, we aim to identify accidents and their causes, and ultimately help develop preventive measures. This approach not only helps reduce accidents but also increases the predictability of future events, thus

creating safer roads [5]. The program uses a variety of data, including accident statistics, detailed accidental deaths by circumstances, and data on deaths due to failure to use safety devices. Through data cleaning, visualization, and statistical analysis, we extract meaningful insights from the data. Besides, machine learning models such as Logistic Regression, Decision Trees, Random Forests, Naive Bayes, and XGBoost are used to predict the severity of accidents and identify the main factors affecting accidents [6]. The findings of this study can provide valuable information for planners, traffic managers, and urban planners, enabling interventions and targeted mitigation strategies in accidentally and to implement improved public safety [7].

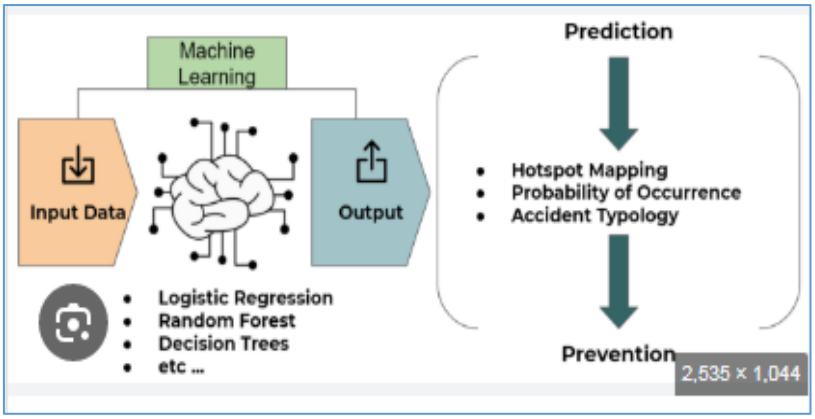

Fig. 1. Procedure for Machine Learning

The following paper continues with the next section methodology for section 2. Section 3 provides the results and analysis. The final section concludes the paper.

2 Methodology

The following methodology for machine learning algorithms is implemented using Python. Programming using Pandas, Numpy, matplotlib and sklearn.

2.1 Procedure for implementation [12]

- Reading the CSV Files
- Initial Data Inspection
- Handling Missing Values
- Checking for Duplicates
- Examining and Modifying df1
- Mean Calculation for Accident Data
- Pie Chart of Road Accidents
- Bar Chart of Accidents by State in 2015
- Mean Calculation for Gender-based Deaths
- Pie Chart for Gender-based Deaths
- Renaming Columns and Data Inspection

- Dropping a Row and Handling Missing Values

2.2 Models for Our Research

- Logistic Regression Model

Logistic Regression: Used to predict the number of accidents in 2016

will be higher than the average of 2003-2015. The model is evaluated using accuracy [8].

- Decision Tree Model

Decision Tree: Another classification model used to predict accident severity. It's evaluated using an accuracy and confusion matrix [9].

- Random Forest Model

Random Forest : A more complex model that generally provides better accuracy. It uses an ensemble of decision trees to make predictions [10].

- Final Machine Learning Models

Logistic Regression, Decision Tree, Random Forest, Naive Bayes, and XGBoost models are used to predict the severity of accidents. These models are trained on the cleaned data frame and evaluated using accuracy, confusion matrix, and classification reports [11].

3 Results and Analysis

This classification system implements safety measures and accident prediction. Different types of elements like weather, road conditions, lighting, and vehicle conditions are used to predict road accidents. Our dataset's attributes are accidents, vehicles, and casualties. Our statistical results generate Logistic Regression, which gives the highest accuracy rate among all machine learning algorithms.

3.1 Dataset's Description

Figure 1 shows the State-wise Total Number of Accidents during-2013,2014,2015 2016 and 2023. Based on these statistical results predict results in the future also.

Table 1. Dataset's [1]

```
RangeIndex: 12316 entries, 0 to 12315
Data columns (total 32 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   Time                         12316 non-null  object
 1   Day_of_week                  12316 non-null  object
 2   Age_band_of_driver           12316 non-null  object
 3   Sex_of_driver                12316 non-null  object
 4   Educational_level            11575 non-null  object
 5   Vehicle_driver_relation      11737 non-null  object
 6   Driving_experience           11487 non-null  object
 7   Type_of_vehicle              11366 non-null  object
 8   Owner_of_vehicle             11834 non-null  object
 9   Service_year_of_vehicle      8388 non-null   object
 10  Defect_of_vehicle            7889 non-null   object
 11  Area_accident_occured        12077 non-null  object
 12  Lanes_or_Medians             11931 non-null  object
 13  Road_allignment              12174 non-null  object
 14  Types_of_Junction            11429 non-null  object
 15  Road_surface_type            12144 non-null  object
 16  Road_surface_conditions      12316 non-null  object
 17  Light_conditions             12316 non-null  object
 18  Weather_conditions           12316 non-null  object
 19  Type_of_collision            12161 non-null  object
 20  Number_of_vehicles_involved  12316 non-null  int64
 21  Number_of_casualties         12316 non-null  int64
 22  Vehicle_movement             12008 non-null  object
 23  Casualty_class               12316 non-null  object
 24  Sex_of_casualty              12316 non-null  object
 25  Age_band_of_casualty         12316 non-null  object
 26  Casualty_severity            12316 non-null  object
 27  Work_of_casuality            9118 non-null   object
 28  Fitness_of_casuality         9681 non-null   object
 29  Pedestrian_movement          12316 non-null  object
 30  Cause_of_accident            12316 non-null  object
 31  Accident_severity            12316 non-null  object
dtypes: int64(2), object(30)
memory usage: 3.0+ MB
```

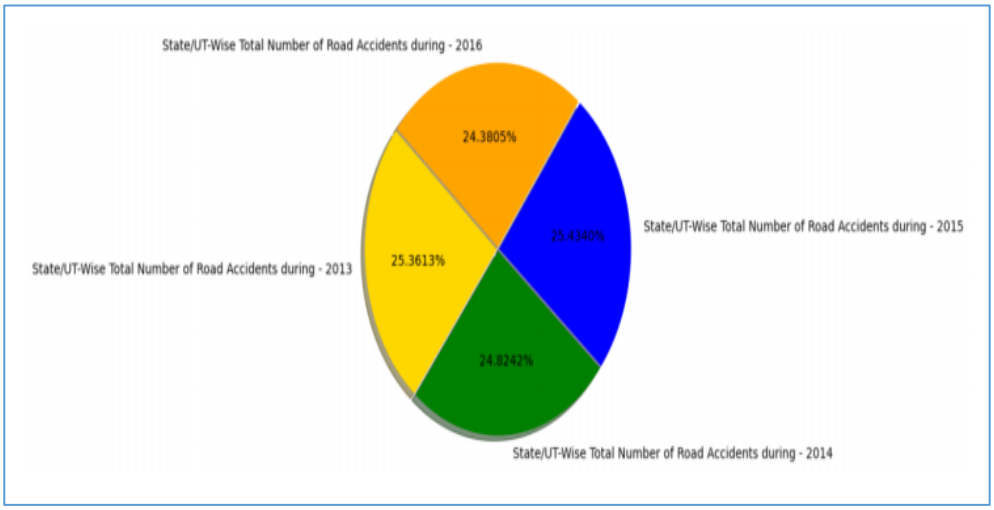

Fig 2. State-wise Total Number of Accidents

Figure 2 shows the total Number of Accidents in each State in India. Different states represent in this diagram on number of accidents.

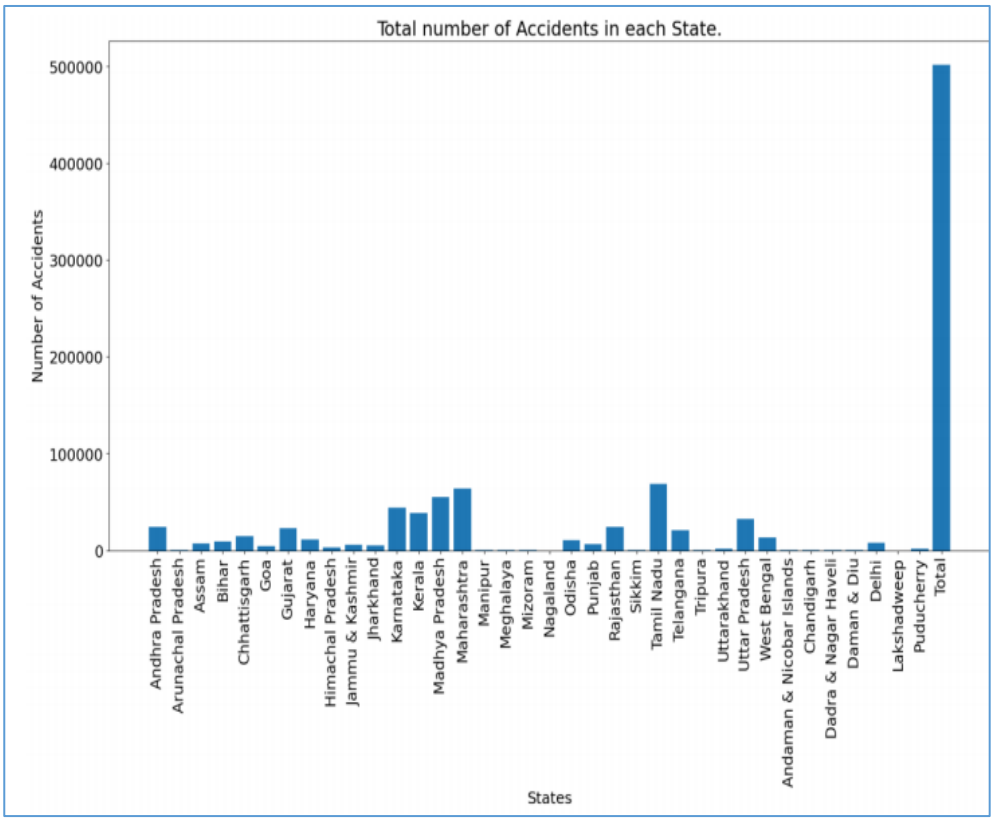

Fig. 3. Number of Accidents in each state

The following figure 3 discusses with Percentage of Deaths occurring due to not wearing helmets. 15% of deaths occur due to not wearing helmets among males and females.

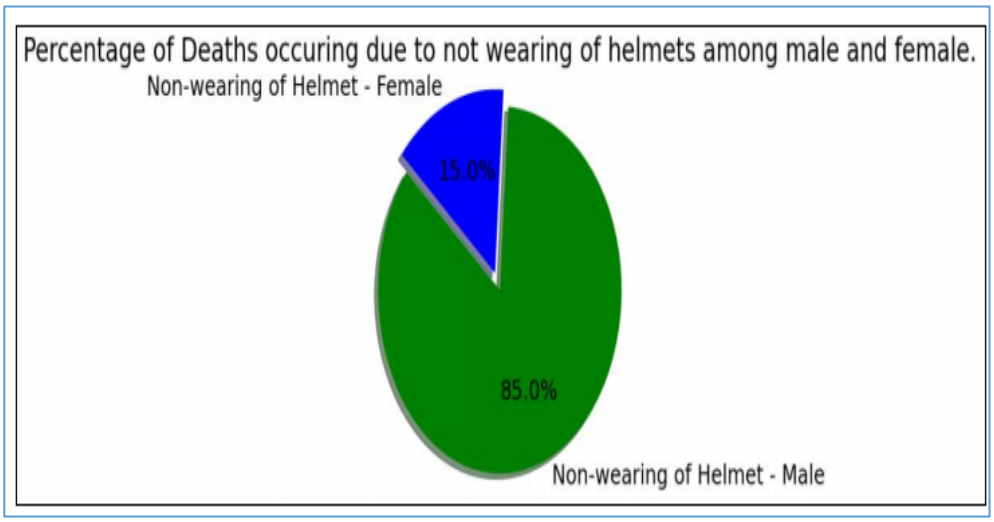

Fig. 4. Number of Accidents in each state

The following table 1 states the accuracy of two machine learning algorithms on the prediction of a number of accidents. The logistic regression algorithm predicts more accuracy (84.77%) among the two algorithms on a number of accident predictions.

Table 1. Prediction of Number of Accidents

Algorithm	**Accuracy**
Logistic Regression	84.77 %
Decision Tree Classifier	75%

Table 2 provides the information about accuracy of each algorithm on accident severity. Logistic regression predicts the highest accuracy (84.77%) among all machine learning algorithms on accident severity [13-17].

Table 2. Model accuracy on Accident severity

Model Name	**Precision**	**Recall**	**F1 Score**	**Accuracy**
Logistic Regression	0.64	0.11	0.19	84.77%
Decision Tree	0.34	0.49	0.30	71.79%
Random Forest	0.54	0.09	0.18	78.18%
Naïve_Bayes	0.23	0.07	0.15	64.32%

XGBoost	0.45	0.08	0.30	70.97%

Table 3. Model accuracy on Accident severity

Algorithm	Accuracy
LR and DT	81.17%
RF and XGBoost	72.69%
LR and Naïve_Bayes	62.67%
RF and LR	74.43%

From Table 2, Compared to ensemble learning on road accident data analysis, Logistic regression analysis gives more accuracy for any combination of accuracy. But one thing it is only applicable to our dataset. It may give different results same problem statement.

4 CONCLUSION

In this paper, we successfully developed a machine learning model to predict the likelihood of traffic-based accidents on different factors such as weather conditions, time of day, road conditions, and traffic density. The model demonstrated an acceptable level of accuracy and reliability, which suggests that machine learning can be a valuable tool for accident prediction and prevention. Our analysis highlighted several key factors significantly influencing accident risk, allowing for targeted interventions to reduce accident rates. For instance, certain weather conditions and road types were found to have a higher correlation with accident occurrences.

The results of this study can benefit traffic management authorities by helping them implement data-driven strategies for accident prevention. Moreover, the model can be further enhanced by incorporating additive data sources and more precocious machine learning methods to ameliorate its predictive powerlessness. In conclusion, the project underscores the potential of machine learning in enhancing road safety and provides a foundational model that can be built upon for more comprehensive accident prediction systems.

Références

1. Ministry of Road Transport and Highways, Government of India, "Road Accident in India," [Online]. Available: https://morth.nic.in/road-accident-in-india. [Accessed: Jan. 28, 2025].
2. R. Agarwal, A. Yadav, and D. P. Agarwal, "Road Accident Prediction Using Machine Learning," [Online]. Available:

https://www.researchgate.net/publication/379628962_Road_Accident_Prediction_Using_Machine_Learning. [Accessed: Jan. 28, 2025].

3. J. Smith and M. Brown, "Road Traffic Analysis Using Machine Learning Techniques," in Proc. IEEE Conf., 2022, pp. 1-6. [Online]. Available: https://ieeexplore.ieee.org/document/9823499. [Accessed: Jan. 28, 2025].
4. Wikipedia, "Traffic Collisions in India," [Online]. Available: https://en.wikipedia.org/wiki/Traffic_collisions_in_India. [Accessed: Jan. 28, 2025].
5. M. Balaji, "Road-Traffic Accident Severity Prediction," GitHub Repository, [Online]. Available: https://github.com/mp-balaji/Road-Traffic-Accident-Severity-Prediction. [Accessed: Jan. 28, 2025].
6. S. Shaik, "DM Algorithms Based Clustering for Road Accident Data Analysis," Int. J. Comput. Sci. Eng., vol. 6, no. 9, pp. 321–328, Sep. 2018.
7. C. S. K. Varma, S. F. Ali, S. Jishnu, and S. Subhani, "Traffic Infringement System Using Machine Learning," Dickensian J., vol. 22, no. 6, pp. 122–128, Jun. 2022.
8. World Health Organization, "World Report on Road Traffic Injury Prevention," pp. 12, 2015.
9. Y. Zou, B. Lin, X. Yang, L. Wu, M. M. Abid, and J. Tang, "Application of the Bayesian Model Averaging in Analyzing Freeway Traffic Incident Clearance Time for Emergency Management," J. Adv. Transp., vol. 2021, pp. 1–9, 2021.
10. S. A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B. W. Kim, "Aggression Detection Through Deep Neural Model on Twitter," Future Gener. Comput. Syst., vol. 114, pp. 120–129, 2021.
11. E. Azimirad, N. Pariz, and M. B. N. Sistani, "A Novel Fuzzy Model and Control of Single Intersection at Urban Traffic Network," IEEE Syst. J., vol. 4, no. 1, pp. 107–111, 2010.
12. Y. Hou, Z. Deng, and H. Cui, "Short-Term Traffic Flow Prediction with Weather Conditions: Based on Deep Learning Algorithms and Data Fusion," Complexity, vol. 2021, pp. 1–14, 2021.
13. R. V. K. Reddy, S. Shaik, G. R. Chandra, and B. S. Rao, "Breast Cancer Prediction Using Classification Techniques," Int. J. Emerg. Trends Eng. Res., vol. 8, no. 9, pp. 3218–3225, 2020.
14. R. V. K. Reddy, S. Shaik, and B. S. Rao, "Machine Learning Based Outlier Detection for Medical Data," Indones. J. Electr. Eng. Comput. Sci., vol. 24, no. 1, pp. 25–33, Oct. 2021.
15. V. K. Vijayalakshmi and S. Shaik, "Predictive Analytics of Employee Attrition Using K-Fold Methodologies," I. J. Math. Sci. Comput., vol. 2023, no. 1, pp. 23–36, Mar. 2023.
16. G. S. Chaitanya, S. Shaik, P. Visalakshi, and G. Rakshitha, "Comparative Study of Ensemble Machine Learning Techniques for Airline Delay Prediction," Int. J. Eng. Comput. Sci., vol. 6, no. 1, pp. 12–19, Jan.–Jun. 2024.
17. J. Lavanya, M. Ramesh, J. Sravan Kumar, G. Rajaramesh, and S. Shaik, "Hate Speech Detection Using Decision Tree Algorithm," J. Adv. Math. Comput. Sci., vol. 38, no. 8, pp. 66–75, Jun. 2023.