



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 13, Issue 3, March 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijrset@gmail.com

www.ijrset.com

Data-Driven Road Safety Enhancement

K.Parisuddha Babu¹, M.Mahalakshmi², M.Raju³, P.Poojitha⁴, G.Tirumala Rao⁵, M.L.Pranuthna⁶

Assistant Professor, Dept. of IT, VVIT, Andhra Pradesh, India. ¹

Student, Dept. of AI&DS, VVIT, Andhra Pradesh, India. ²

Student, Dept. of AI&DS, VVIT, Andhra Pradesh, India. ³

Student, Dept. of AI&DS, VVIT, Andhra Pradesh, India. ⁴

Student, Dept. of AI&DS, VVIT, Andhra Pradesh, India. ⁵

Student, Dept. of AI&DS, VVIT, Andhra Pradesh, India. ⁶

ABSTRACT: The escalating number of vehicles on the road has led to a concerning rise in daily traffic accidents. Predicting these accidents is crucial for informed decision-making by the transportation department. Despite accidents being inherently uncertain, there is a degree of regularity in their occurrence over time. Exploring the relationships between road conditions, environmental factors, and accidents, employing data mining techniques. Utilizing datasets, this offers valuable insights for stakeholders such as government public work departments, contractors, and automobile industries. These insights can aid in optimizing road design and vehicle safety based on accident predictions.

KEYWORDS: Traffic accident prediction, Traffic safety, Data mining, Machine learning

I. INTRODUCTION

India faces a pressing road safety challenge characterized by a high rate of accidents attributed to factors such as negligence, over-speeding, alcohol influence, and traffic violations. Need for accident prediction models to analyze road safety issues and infrastructure problems. Challenges include evaluating the impact of variables and designing effective models using data mining techniques. Investigating the inter-relationship between accidents, road conditions, and environmental factors. Data mining is used to extract meaningful details from vast datasets. to identify risks in road accident scenarios and proposes a model to estimate accident possibilities in specific user-chosen areas. Building a prediction model to take user inputs and process the inputs by the saved model which has been tested on data to have good accuracy and give out the output of prediction

Paper is organized as follows. Section II describes the various models of work done on road safety enhancements mostly uses machine learning algorithms and comparing the performance. The flow diagram represents the step of the algorithm. The detailed process of raw data to useful insights and build model to predict that is given in Section III. Section IV presents experimental results showing results of predictions. Finally, Section V presents conclusion.

II. RELATED WORK

The steady increase in the rate of accidents in India have prompted many researchers to look into the factors affecting road accidents and study about it. Since data mining techniques do not require certain assumptions between dependent and independent variables which are required in traditional statistical techniques, various categories of data mining techniques have been made use of in creating prediction models for road accidents in the past. Researchers have focussed on different sets of attributes in developing such models. Srivastava et al. [2] and Ghazizadeh et al. [3] have mainly concentrated on studying the accidents occurring at intersection points. While the former has looked into categorizing the accidents based on their levels of seriousness using a Multi-layered perceptron (MLP) technique that was found to be more effective, the latter has made use of a feed forward MLP that utilizes back propagation learning to analyse the effect of several factors such as day or night, traffic conditions etc. on accidents. Chen et al. [4] in their studies have found highways to be the common area where a majority of accidents had been reported to occur. Williams et al. [5] have found through their studies that the age and experience of a driver also play a major role in the occurrence of accidents. Suganya, E. and S. Vijayarani [6] in their paper have analysed the road accidents in India and



compared the performance of different classification algorithms such as linear regression, logistic regression, decision tree, SVM, Naïve Bayes, KNN, Random Forest and gradient boosting algorithm using accuracy, error rate and execution time as a measure of performance. They have found the performance of KNN to be better than that of the others. Sarkar et al. [7] have done a comparative study on the type of roads that are prominent in accidents. While exploring the other components associated with accidents, they have found that the occurrence of accidents in highways is more common than in a normal road similar to [4]. Stewart et al. [8] have utilized original data in building a neural network model to predict accidents. They found that this model was able to give quicker results than those being used in the models built on Indian roads. Zheng et al. [9] have studied the range of injuries that come forth in a motor vehicle accident and have also analyzed the emotions of the drivers involved in the accidents that could have been a causal factor. Arun Prasath N and Muthusamy Punithavalli [10] have conducted an extensive survey on the different techniques used in road accident detection over the years, the approaches implemented in them and discusses their merits and de-merits.

III. METHODOLOGY

The proposed one that predicts the severity of road traffic accidents. It discusses data collection, feature selection, model training, and deployment. The project uses a machine learning model to predict accident severity based on 31 features. The dataset contains extensive information on 12316 accidents. It has 31 predictor variables and the target variable Accident_severity which takes three possible values: Slight, Serious and Fatal. The aim of the project is to build prediction models to classify whether the severity of an accident is slight, serious or fatal, given various relevant information in form of the 31 predictor variables, regarding that particular accident. The model was trained on a dataset of road traffic accidents from Kaggle. The dataset was preprocessed to handle missing values and class imbalance. Features were selected using correlation and chi-square methods to remove collinearity. The model was trained using XGBoost and achieved an accuracy of 74%. SHAP was used to explain the model's predictions.

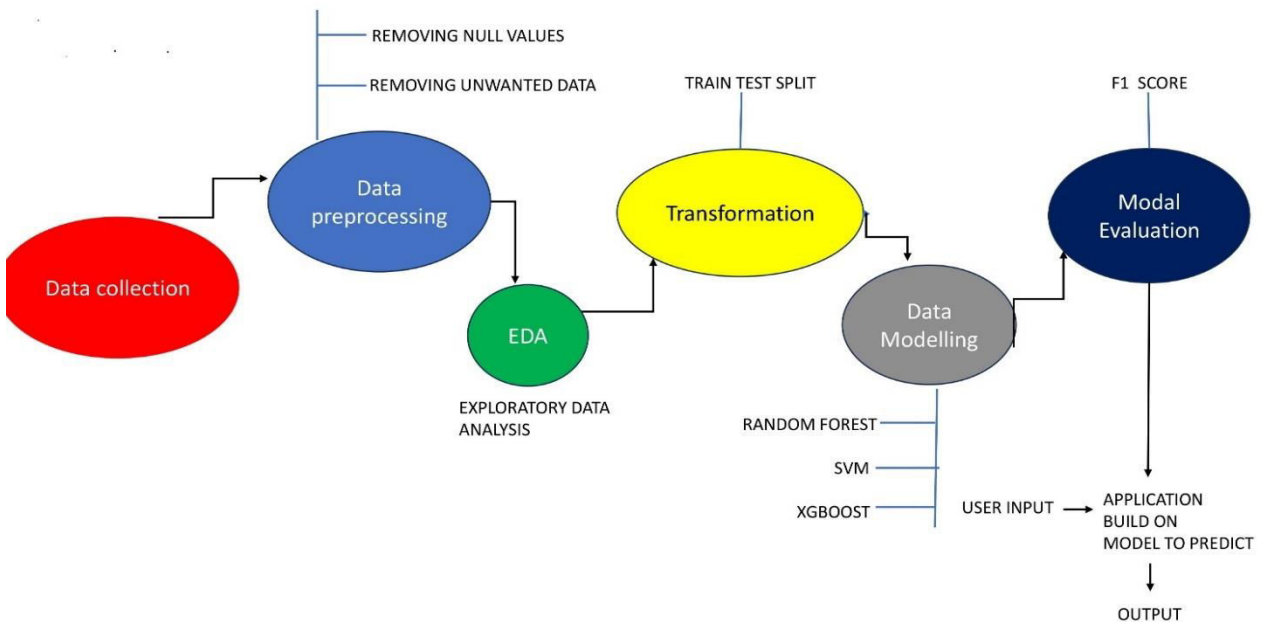


Fig- Flow chart of proposed System

III. EXPERIMENTAL RESULTS

For each severity category (slight, serious, fatal), the decision plot illustrates how different features contribute to the model's prediction of road severity. Each plot provides insights into the importance and directionality of features in



determining the severity level. By visualizing these contributions, we can better understand the factors influencing the model's classification decision for each severity category.

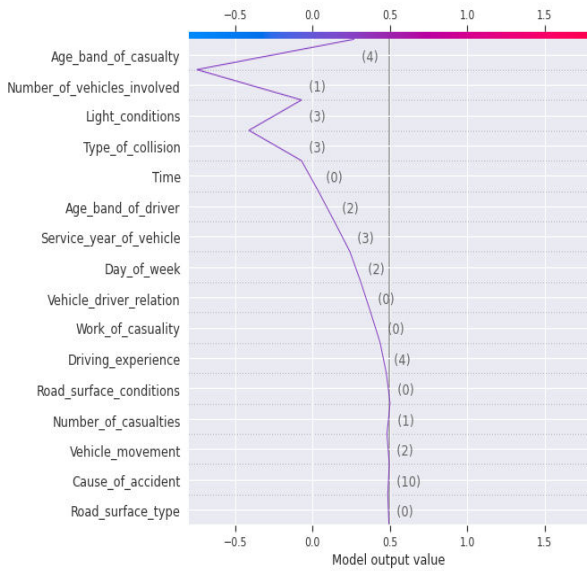


Fig: 1.1- slight Injury

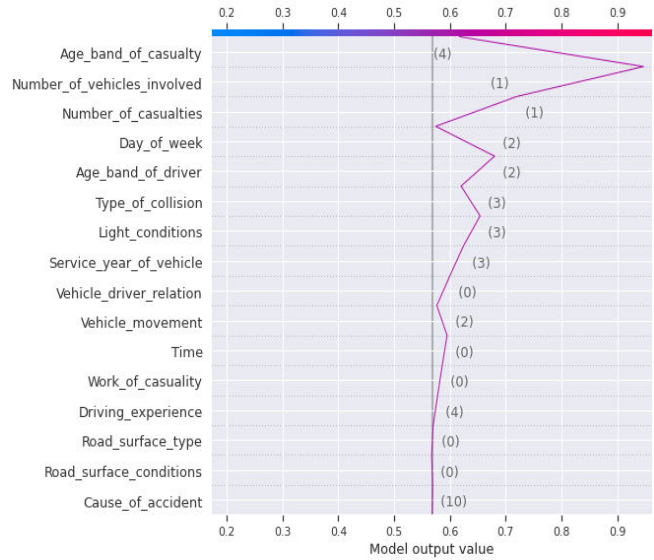


Fig: 1.2- Serious Injury

Fig-1.1: This plot illustrates the impact of various features on predicting slight injuries resulting from road incidents. Each bar represents a feature's contribution to the model's decision, with positive values indicating features that push the prediction towards a slight injury outcome and negative values indicating features that push away from it.

Fig-1.2: This plot shows the influence of different features on predicting serious injuries in road incidents. Similar to the slight injury plot, positive values indicate features favoring a serious injury prediction, while negative values suggest features that deter it.

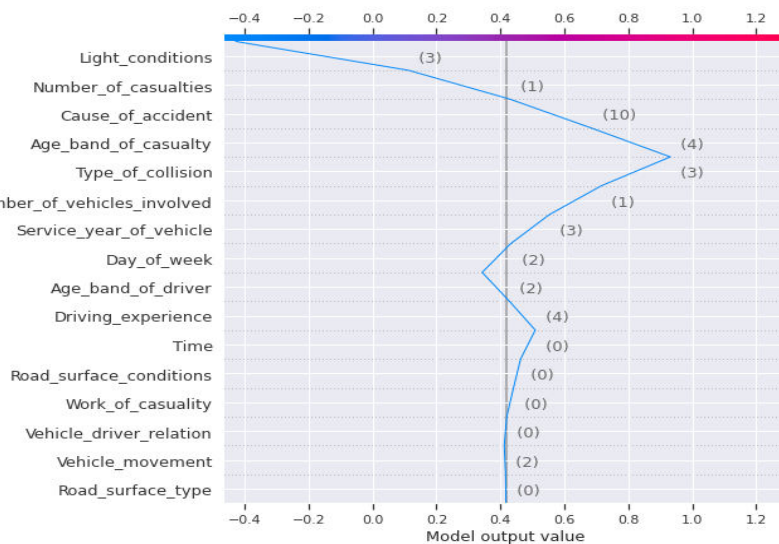


Fig-1.3- Fatal Injury

Fig:1.3- This plot displays the significance of features in predicting fatal injuries from road incidents. Again, positive values denote features supporting a fatal injury prediction, while negative values represent features opposing it.



Road Accident Severity Prediction

Deploy

Accident Severity Prediction Application



Enter the input for following features:

Select time:

Day

Select day of the week:

Sunday

Fig-1.4 -screenshot showing some of the input features

Deploy

Select driver's age:

18-30

Select vehicle relation:

Employee

Select driving experience:

5-10yr

Select service year of vehicle:

Below 1yr

Select surface of road:

Asphalt roads

Select surface conditions:

Dry

Fig-1.5- screenshot of some of the input feature



Fig-1.6- screenshot of some of the input feature

Fig-1.7 - screenshot of some of the input feature and predicted value

Detailed description of FIG-(1.4,1.5,1.6,1.7) as follows

Here is the application outcome of Accident Severity build on dataset to undergo the datamining techniques and saved the model with XGboost and connect with python library to build the application which is streamlit,where it is a powerful tool for building interactive web applications with minimal effort, and integrating SHAP (SHapley Additive exPlanations) plots into a Streamlit application.

IV. CONCLUSION

In conclusion, this project aimed to predict road accident severity using machine learning techniques and explore the factors contributing to different levels of severity. By analyzing a comprehensive dataset containing various attributes such as weather conditions, road type, and vehicle characteristics, predictive models were developed to classify accidents into categories of slight, serious, and fatal injuries. Through extensive data preprocessing, feature engineering, and model evaluation, the project successfully built accurate predictive models capable of assisting in understanding and mitigating road accidents. Additionally, leveraging SHAP (SHapley Additive exPlanations) plots provided valuable insights into the decision-making processes of these models, shedding light on the importance of different features in determining accident severity. With the potential to inform policymakers, urban planners, and law enforcement agencies, this project underscores the significance of data-driven approaches in improving road safety and reducing accident-related fatalities and injuries.

REFERENCES

- [1] <https://www.statista.com/topics/5982/road-accidents-in-india/>
- [2] Srivastava AN, Zane-Ulman B. (2005). Discovering recurring anomalies in text reports regarding complex space systems. In Aerospace Conference, IEEE. IEEE 3853-3862.
- [3] Ghazizadeh M, McDonald AD, Lee JD. (2014). Text mining to decipher free-response consumer complaints: Insights from the nhtsa vehicle owner's complaint database. Human Factors 56(6): 1189-1203. <http://dx.doi.org/10.1504/IJFCM.2017.089439>.
- [4] Chen ZY, Chen CC. (2015). Identifying the stances of topic persons using a model-based expectationmaximization method. J. Inf. Sci. Eng 31(2): 573-595. <http://dx.doi.org/10.1504/IJASM.2015.068609>
- [5] Williams T, Betak J, Findley B. (2016). Text mining analysis of railroad accident investigation reports. In 2016 Joint Rail Conference. American Society of Mechanical Engineers V001T06A009- V001T06A009. <http://dx.doi.org/10.14299/ijser.2013.01>.
- [6] Suganya, E. and S. Vijayarani. "Analysis of road accidents in India using data mining classification algorithms." 2017 International Conference on Inventive Computing and Informatics (ICICI) (2017): 1122-1126.
- [7] Sarkar S, Pateshwari V, Maiti J. (2017). Predictive model for incident occurrences in steel plant in India. In ICCCNT 2017, IEEE, pp. 1-5. <http://dx.doi.org/10.14299/ijser.2013.01>.
- [8] Stewart M, Liu W, Cardell-Oliver R, Griffin M. (2017). An interactive web-based toolset for knowledge discovery from short text log data. In International Conference on Advanced Data Mining and Applications. Springer, pp. 853-858. http://dx.doi.org/10.1007/978-3-319-69179-4_61.
- [9] Zheng CT, Liu C, Wong HS. (2018). Corpus based topic diffusion for short text clustering. Neurocomputing 275: 2444-2458. <http://dx.doi.org/10.1504/IJIT.2018.090859>.
- [10] ArunPrasath, N and Muthusamy Punithavalli. "A review on road accident detection using data mining techniques." International Journal of Advanced Research in Computer Science 9 (2018): 881-885.



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