International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

ROAD ACCIDENT SEVERITY PREDICTION USING MACHINE LEARNING

[1]Seela Vidya Kumari, [2]A. Gautami Latha

[1] Master of Computer Applications, Andhra University College of Engineering (A), Andhra University, Visakhapatnam, Andhra Pradesh, India, 530003.

[2] Professor, Department of IT&CA, Andhra University College of Engineering (A), Andhra University, Visakhapatnam, Andhra Pradesh, India, 530003.

[1] vidyaseela@gmail.com, [2] dr.gautamilatha@andhrauniversity.edu.in

Abstract: In our hectic society, traffic accidents pose a serious threat as they cause countless injuries, lives, and monetary losses every year. Numerous things, including as the surrounding environment, vehicle maintenance, and the actions of the driver and passengers, might lead to these occurrences. Accidents are still a problem even after many safety precautions have been put in place. In order to foresee future accidents, machine learning analyses historical accident data, providing a useful tool for tackling this issue. We may create models that forecast the severity of traffic accidents by using classification algorithms such the Decision Tree, Random Forest, KNN, Gradient Boosting, and KNN Learning algorithms. Our goal is to lower the frequency and severity of future accidents by implementing more effective techniques and precautions after learning from the causes and effects of previous ones.

Index Terms: Road Traffic Accidents, Decision Tree Algorithm, Random Forest Algorithm, KNN Learning Algorithm, Gradient Boosting Learning Algorithm, Severity prediction, Future accident reduction

1. INTROUDCTION

Unexpected events involving vehicles or other road users that cause fatalities or property damage are known as road traffic accidents, or RTAs. More than 90% of traffic fatalities take place in low- and middle-income nations, with yearly economic losses surpassing $518 billion. Traffic accidents cost developing countries over 13% of their GDP, although wealthy countries' fatality rates are falling as a result of concerted efforts. By 2030, traffic accidents may rank as the sixth most common cause of death, according to the WHO.e frequency and severity of future accidents by implementing more effective techniques and precautions after learning from the causes and effects of previous ones. wer the frequency and severity of future accidents by implementing more effective techniques and precautions after learning from the causes and effects of previous ones.

Road traffic accidents (RTAs) are a global problem that cause financial losses, psychological distress, bodily injuries, and deaths. There are over 1.3 million deaths from these incidents each year, making them the greatest cause of death for individuals between the ages of 15 and 49. RTAs have far-reaching social and economic ramifications that go beyond the acute injuries sustained, including decreased quality of life for victims and their families, lost

productivity, and legal fees. In terms of economic impact, injuries from traffic accidents account for roughly 1% of GDP in low-income nations, 1.5% in middle-income nations, and 2% in high-income nations. Road accident severity is affected by a number of elements, such as property damage, casualties, and injuries. These fall into four categories: property damage, fatalities, serious injuries, and mild injuries.

The severity of traffic accidents is divided into four categories: minor injuries, serious injuries, fatalities, and property damage. Motorization, urbanization, population increase, poor road conditions, insufficient lighting, and bad weather are some of the contributing reasons to these accidents. The goal of this research is to utilize machine learning (ML) to forecast the severity of traffic accidents by examining variables such as environmental conditions and driver behaviour. The objective is to build precise machine learning models to evaluate the likelihood of accidents, efficient preventative actions, and AI-driven applications that notify users of the danger of accidents depending on current circumstances.

2. LITERATURE SURVEY

By examining variables like weather, kind of road, and time, machine learning can predict the severity of traffic accidents, thereby improving traffic safety. For this, algorithms like Gradient Boosting, Random Forest, K-Nearest Neighbors (KNN), Decision Trees, and Random Forest are useful resources. Gradient Boosting effectively manages unbalanced data, KNN classifies based on closeness, Decision Trees and Random Forests avoid overfitting and offer interpretability. Research indicates that these models—Random Forest and Gradient Boosting in particular—perform more accurately than conventional techniques, which supports improved safety and traffic management.

The paper titled “Road accident data analysis using a data mining framework” by Sachin Kumar & Durga Toshniwal, 2015 A data mining system intended to improve the examination of data on traffic accidents is presented in this research. To enhance comprehension of accident factors, it consists of elements including trend analysis, pattern finding, and data preprocessing. The system integrates cutting-edge data mining methods and algorithms to facilitate improved policy development and decision-making for road safety.[2].

The paper titled " Algorithm and Software for Identifying Accident-Prone Road Sections " by N. Zagorodnikh et al., in 2018. An algorithm and software tool for identifying high-risk road portions that are prone to accidents are presented in the study. It focuses on using traffic data analysis to identify risk indicators and identify dangerous regions. This strategy seeks to improve management of traffic safety and direct focused interventions to lower accident rates. [8].

The paper " Traffic Accidents Classification and Injury Severity Prediction " by L. G. Cuenca, E. Puertas, N. Aliane, and J. F. Andres in 2018. This research investigates different traffic accident data analysis techniques with an emphasis on accident type classification and injury severity prediction. It assesses prediction models like logistic regression, neural networks, and gradient boosting machines, which predict injury severity based on variables

including vehicle speed and road conditions. It highlights how data-driven tactics, which recognize high-risk scenarios and put preventative measures in place, can improve traffic safety. Along with outlining issues like data quality and model interpretability, the paper makes recommendations for future research paths that could lead to the development of more reliable, scalable, and real-time traffic accident analysis models [9].

The paper “Analysing the Leading Causes of Traffic Fatalities Using XGBoost and Grid-Based Analysis: A City Management Perspective” by J. Ma, Y. Ding, J. C. Cheng, Y. Tan, V. J. Gan, and J. Zhang 2019. This study investigates cutting-edge analytical methods for comprehending road deaths. It focuses on using grid-based analysis techniques along with the potent machine learning algorithm XGBoost to determine the main causes of road fatalities, The research illustrates how spatial grid analysis and XGBoost's predictive powers can be combined to improve traffic safety evaluations' accuracy and guide focused responses. This thorough investigation advances our knowledge of the factors that lead to traffic fatalities and aids in the creation of safer urban environments.[7].

In the paper titled " RFCNN: Traffic Accident Severity Prediction Based on Decision Level Fusion of Machine and Deep Learning Model" by Mubariz mansoor, Muhammad umar, Saima sadiq, Abid isaq, Saleem ullah, Hamza, and Carmen in 2021 the authors present RFCNN, a hybrid model that uses decision-level fusion to predict the severity of traffic accidents by combining Random Forests and Convolutional Neural Networks. Through the integration of various techniques, RFCNN optimizes the benefits of each to improve prediction accuracy. The fusion strategy offers better prediction accuracy and increased generalization across datasets by addressing the shortcomings of individual models. In order to improve traffic safety and management, the study emphasizes how hybrid models can improve predictive performance and offer insightful information about the severity of traffic accidents. [1].

In order to predict the severity of traffic accidents, Ahmed, Hossain, Bhuiyan, and Ray (2021) compared several machine learning techniques, such as logistic regression, decision trees, random forests, SVM, and KNN. The accuracy and resilience of each algorithm in managing variables like weather, traffic volume, and road features were assessed in their study. They discovered that while more sophisticated algorithms like random forests and SVM handle complex interactions more skillfully and offer superior prediction accuracy, simpler models like logistic regression are simpler to understand. To further improve forecast accuracy, the authors suggested that future research focus on fine-tuning these models and investigating ensemble methodologies [3].

Nour, Naseer, Alkazemi, and Muhammad (2020) used analytics to examine injury data from traffic accidents in order to find trends and variables affecting the severity of the accidents. Their study examined characteristics like driver behavior, vehicle type, and road conditions using statistical and machine learning methodologies. In order to improve traffic management and safety tactics, they advised integrating sophisticated analytics with real-time data and stressed the significance of data quality and pretreatment for accurate insights.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

In order to improve injury prediction models and road safety, future research should concentrate on improved data collecting and a variety of sources [4].

Mehdizadeh et al. (2020) examined the use of descriptive and predictive modelling in data analytics applications related to road traffic safety. Their research addressed a number of traffic data analysis approaches, such as data visualization, machine learning, and statistical techniques. They emphasized how these techniques aid in trend identification, risk assessment, and accident prediction. The study identified issues with real-time integration and data quality, and it made recommendations for more research to increase model accuracy and incorporate a variety of data sources for more thorough analysis of traffic safety [5].

3. METHODOLOGY

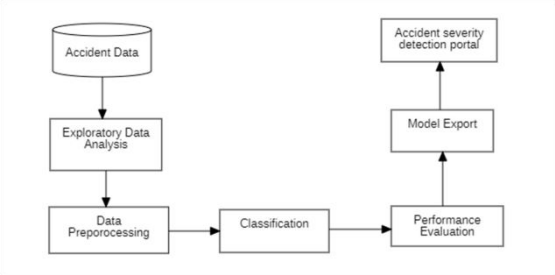


Fig 1. Block diagram of the proposed system

The project is divided into five primary stages, as shown in Figure 1. It starts with a descriptive analysis of the accident data and proceeds to pre-process the data by grouping and label encoding. To classify accident severity, a machine learning classification model is then constructed. The accuracy of the model is then verified by assessing its performance. Ultimately, a sophisticated online platform is created to categorize the intensity of accidents, offering an intuitive user interface for real-world implementation.

A. Dataset

The Kaggle Accident Severity dataset was used to implement this project. There are 3057 accident samples in the dataset. There is one goal attribute and fourteen predictive attributes in each sample. The collection includes accident data from January 1, 2009, to December 31, 2009. The dataset does not contain any null values. There are two values for the goal attribute: minor accident and major accident. There are 2736 minor accident records and 321 major accident records in the dataset.

Variables	Description	Data Type	Scale	Null Value
Reference No	Accident identity number	Integer	Serial number	No
Easting	Easting point	Integer	Map point	No
Northing	Northing point	Integer	Map point	No
Number of Vehicles	Number of vehicle in the spot	Integer	Vehicle count	No
Accident Date	Date of accident happened	Date	Date	No
Time (24hr)	Time of accident happened	Time	Time	No
1st Road Class	Road type	Varchar	Category	No
Road Surface	Surface type of the road	Varchar	Category	No
Lighting Conditions	Lighting condition in the spot	Varchar	Category	No
Weather Conditions	Weather condition in the spot	Varchar	Category	No
Casualty Class	Casuality type (Driver..ect)	Varchar	Category	No
Sex of Casualty	Sex of the casuality	Varchar	Category	No
Age of Casualty	Age of the casuality	Integer	Age in years	No
Type of Vehicle	Type of the vehicle	Varchar	Category	No
Severity	Accident severity	Varchar	Category	No

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

B. Data Preprocessing

In order to transform raw data into a format that can be analysed in data mining and machine learning, data preparation is necessary. Usually, this procedure entails multiple stages to guarantee precise outcomes. Preprocessing for the accident data comprised:

1. Data Reduction: Consolidating related qualities into a single variable, such as combining several motorcycle kinds into a single "Motorcycle" category, in order to simplify the dataset.
2. Data Encoding: Using methods such as Label Encoding, which assigns a unique integer to each category, transform categorical data into numerical representation.
3. Dropping Unwanted Columns: To increase model accuracy, eliminate characteristics that don't add to predictions.

Road Surace	Weather Conditions	Casualty Class
0-Dry 1-Frost / Ice- 2-Flood 3-Snow 4-Wet / Damp	0-Fine without high winds 1-Fog or mist – if hazard 2-Fine with high winds 3-Other 4-Raining without high winds 5-Raining with high winds 6-Snowing without high winds 7-Snowing with high winds 8-Unknown	0-Driver 1-Passenger 2-Pedestrian

Fig 2. Label Encoding

C. Algorithms for classification

The machine learning methods are used for estimating the severity of traffic accidents those are Decision Tree Learning, K-NN, Random Forest Classifier, and Gradient Boosting Classifier. By building a tree structure that divides data according to decision rules and optimizing splits using metrics like entropy and information gain, decision tree learning models both classification and regression tasks. K-NN uses distance measures to determine similarity while classifying data by comparing new examples to the closest neighbors in the training set. Using a majority voting mechanism for final classification, Random Forest combines numerous decision trees trained on random subsets of data to improve prediction accuracy. By creating models successively to fix past mistakes and combining predictions from all models to produce accurate results, gradient boosting minimizes bias. These techniques offer special advantages for managing complicated data and enhancing forecasts for road safety since they examine a variety of characteristics, including vehicle kinds, accident scenes, and meteorological conditions.

D. Build machine learning model

Using the Scikit-learn toolkit, four machine learning classifiers—Decision Tree, K-Nearest Neighbors, Random Forest, and Gradient Boosting—were used to create prediction models. The pre-processed data was split into two subsets for the training and testing process: 70% for training the models and 30% for evaluating their functionality. With this method, the models are trained on most of the data and tested on a different subset to see how well they can generalize.

E. Performance Evolution

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

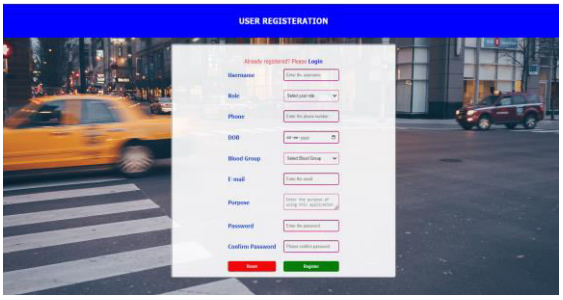
It is used to describe the methodical process of evaluating, enhancing, and optimizing a model, system, or process's performance over time. This includes a number of crucial tasks in the context of data science and machine learning, including: Confusion matrix, Accuracy, ROC curve, precession, Recall

F. User login and Authentication

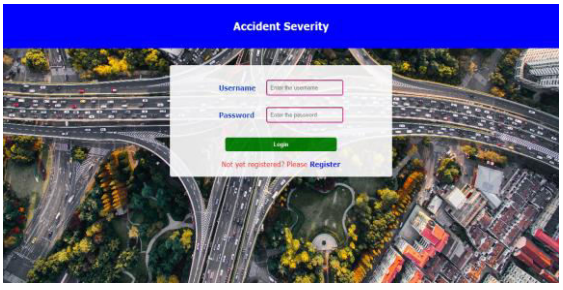
User credentials, including hashed passwords and usernames, as well as user profile data must be handled securely throughout user login and authentication. Features like session management, two-factor authentication (2FA), and login history tracking improve security. Credentials are checked and session tokens are distributed for access during authentication. In order to prevent unwanted access to user accounts, the system employs security features like recording login attempts, session timeouts, and activity monitoring.

4. EXPERIMENTAL RESULTS

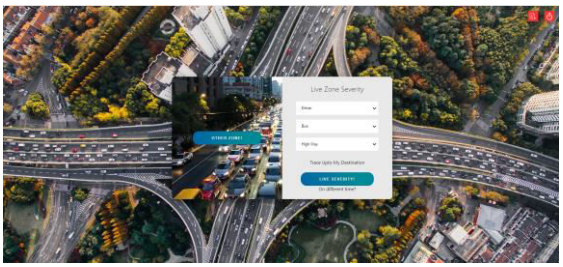
User registration: The process of creating an account on a website using personal information about a user, such as their phone number, DOB, blood type, email address, and password, is known as user registration.



User Authentication: In which it takes username and password for user authentication

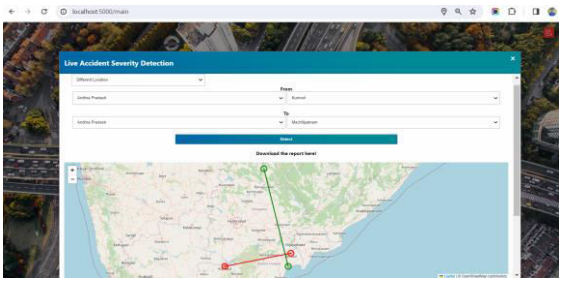


Classification page: classify the data based on type of person, which type of vehicle, and which type of road

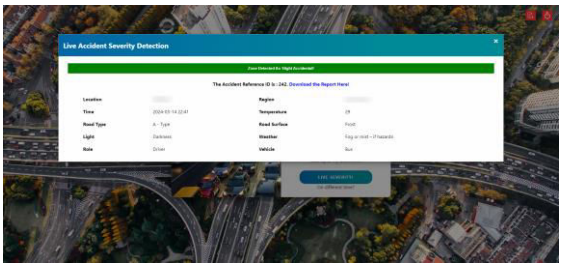


International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

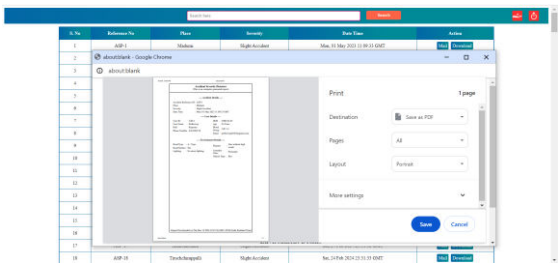
Live prediction: Here it will make the live prediction based on the data it was given and find the severity of accident to happen in that location



Prediction report: After making live prediction, it will generate the report



Reporting: The report is generated to download



5. CONCLUSION

One of the most unfortunate hazards in our busy world is traffic accidents. Every year, there are a great deal of casualties, injuries, and fatalities from traffic accidents, in addition to large financial losses. One of the main jobs is estimating the severity of the accident. When taking into account the accident data, the gradient boosting model performs better and achieves 93% accuracy. The severity of the collision can be predicted using the following factors: number of vehicles, road class, road surface, lighting conditions, spot weather conditions, casualty class, victim's sex, age, and kind of vehicle. Police agencies, media, and highway authorities all stand to gain much from this.

6. FUTURE SCOPE

The project's technique is helpful in anticipating if an accident will happen in the future. Many road users, including drivers, passengers, and travellers, will find this data useful. For drivers, it will navigate to the desired location directly and indicate the likely severity of an accident there. Using the desired data, the driver can then modify the route or take preventative measures, such as limiting speed or changing the route from the specified

location. This information model can also increase its own accuracy by utilizing historical data to produce accurate results in the future. By taking a proactive stance, drivers can reduce hazards and better prepare for crises

7. REFERENCES

[1] Mubariz Mansoor, Muhammad Umar, Saima Sadiq, Abid isaq, Saleem ullah, Hamza, and Carmen, "RFCNN: Traffic Accident Severity Prediction Based on Decision Level Fusion of Machine and Deep Learning Model", Digital Object Identifier 10.1109/ACCESS.2021.3112546.

[2] Sachin Kumar and Durga Toshniwal, "A data mining framework to analyze road accident data", DOI 10.1186/s40537-015-0035-y

[3] Shakil Ahmed, Md Akbar Hossain, Md Mafijul Islam Bhuiyan, Sayan Kumar Ray, "A Comparative Study of Machine Learning Algorithms to Predict Road Accident Severity", 978-1-6654-6667-7/21/$31.00 ©2021 IEEE DOI 10.1109/IUCC-CIT- DSCI-SmartCNS55181.2021.00069

[4] Mohamed K Nour, Atif Naseer, Basem Alkazemi, Muhammad, "Road Traffic Accidents Injury Data Analytics", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 12, 2020

[5] A. Mehdizadeh, M. Cai, Q. Hu, M. A. A.Yazdi, N.Mohabbati Kalejahi, A.Vinel, S. E. Rigdon, K. C. Davis, and F. M. Megahed, “A review of data analytic based applications in road traffic safety”. Part 1: Descriptive and predictive modelling Sensors (Switzerland), vol. 20, no. 4, pp. 1–24, 2020.

[6] Q. Hu, M. Cai, N.Mohabbati Kalejahi, A. Mehdizadeh, M. A. A.Yazdi, A.Vinel, S. E. Rigdon, K. C. Davis, and F. M. Megahed, “A review of data analytic applications in road traffic safety. Part 2: Prescriptive modelling,” Sensors (Switzerland), vol. 20, no. 4, pp. 1–19, 2020.

[7] J. Ma, Y. Ding, J. C. Cheng, Y. Tan, V. J. Gan, and J. Zhang, “Analyzing the Leading Causes of Traffic Fatalities Using XGBoost and Grid-Based Analysis: A City Management Perspective,” IEEE Access, vol. 7, pp. 148 059–148 072, 2019.

[8] N. Zagorodnikh, A. Novikov, and A.Yastrebkov, “Algorithm and software for identifying accident-prone road sections,” Transp. Res. Procedia, vol. 36, pp. 817– 825, 2018. [Online]. Available: https://doi.org/10.1016/j.trpro.2018.12.074.

[9] L. G. Cuenca, E. Puertas, N. Aliane, and J. F. Andres, “Traffic Accidents Classification and Injury Severity Prediction,” in 2018 3rd IEEE Int. Conf. Intell. Transp. Eng. ICITE 2018, pp. 52–57

[10] Accident Data Analysis to Develop Target Groups for Countermeasures, Max Cameron

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362