JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ROAD ACCIDENT SEVERITY PREDICTION USING MACHINE LEARNING

[1*]Tushar Goud, [2]Pallavi Kumari, [3]Vedant Patil, [4]Rishabh Tiwari
[1*]B. Tech (8th Sem) Student, Ajeenkya D Y Patil University, Pune, Maharashtra-411014
[2]Academic Tutor, Ajeenkya D Y Patil University, Pune, Maharashtra-411015
[3]B. Tech (8th Sem) Student, Ajeenkya D Y Patil University, Pune, Maharashtra-411015
[4]B. Tech (8th Sem) Student, Ajeenkya D Y Patil University, Pune, Maharashtra-412207

ABSTRACT:

Several studies and research show that one of the most frequent causes of death in world is road accidents. Several measures are taken by people and introducing high tech system in vehicles for Safety still deaths and accidents are unavoidable. Road traffic crashes kill approximately 1.19 million people each year. Road traffic injuries are the greatest cause of death among children and young adults aged 5 to 29 years. Reducing the number of events is crucial for saving lives and creating sustainable cities and communities. Machine learning and data analysis tools evaluate the causes of car accidents and recommend ways to reduce them. According to the Global Status Report on Road Safety 2023, 1.19 million less individuals die in traffic accidents each year than in previous years. The research demonstrates that initiatives to improve road safety are having an impact, with considerable decreases in road traffic. This paper provides an overview of traffic accident prediction techniques, including traditional statistical analysis, machine learning, and time series analysis. It also evaluates the benefits of each method. These methods can assess factors impacting traffic accidents, create prediction models, increase accuracy, and help prevention efforts for urban traffic.

Keywords: **Road Accident, Machine learning, Time series analysis.**

1. INTRODUCTION

Road accident is unpredictable incident that took hundreds of lives per day. Each year millions of people lost their lives because of the road accident. Traffic accidents are unavoidable and can happen at any time or place. According to the World Health Organization (WHO), transportation accidents kill almost 1.35 million people annually. However, government takes several safety measures to reduce road accidents by determining the adequate reasons and causes. Road accidents are mainly caused due to these external factors, such as weather conditions, road structured, road condition according to weather. Age of driver and driver skills. According to analysis, severe injuries and fatal death happened in darkness where there is no street light and dry and uneven surface condition. Wang et al. [4] did an analysis in China and found that serious fatalities were more common in winter than in autumn due to adverse weather conditions including rain and snow, Potoglou et al. [5] found that traffic accidents are more prevalent throughout the summer and autumn seasons in Italy and Europe, which is important to this study as it uses data from the UK. In India, most of the accident happens in winter and rain season because of slippery roads and fog (it leads to road blindness). Drowsiness is also a leading factor in road accident. According to the National Highway Traffic Safety Administration (NHTSA), sleepy driving is responsible for up to100,000 collisions each year, resulting in 1,550 deaths and 71,000 injuries.

Traffic accident data patterns can be explored and studied using machine leaning, big data analysis and time series analysis. As we know that traffic accident data is not static, it changes rapidly, it is unstructured and variant. So, old and traditional methods have no benefits in data processing and storage. Machine learning library named ML libraries, which includes several machine learning algorithms for classification, regression, decision trees, random forest, clustering method and so on.

This Paper gradually focus on predictive data analysis for road accident. The primary goal is to propose a predictive machine learning model that have capacity and abilities to determine people drowsiness, speed of vehicle, casualty count, accident severity, and the number of vehicles involved. Four machine learning algorithms (decision trees, random forest classification, multinomial logistic regression, and naïve Bayes classification) were assessed to determine the most effective predictive model for road accidents. Machine learning techniques were used to create a prediction model based on training data. We tested and evaluated the prediction accuracy of each machine learning technique to determine the best fit with the highest accuracy.

1.2. Motivation:

The motivation behind developing a predictive model for accident severity using machine learning (ML) techniques is driven by the need to improve road safety and mitigate the adverse impacts of road accidents. Accidents on roads can result in loss of life, injuries, and significant property damage, and can have far-reaching social and economic consequences. By accurately predicting the severity of accidents, it is possible to take proactive measures to prevent accidents, allocate resources effectively for emergency response, and implement targeted interventions to mitigate accidents.

ML techniques offer the potential to analyse large-scale historical accident data and identify patterns and relationships among various factors that contribute to accident severity. By leveraging ML algorithms, it is possible to develop accurate predictive models that can aid traffic management authorities, emergency responders, and policymakers in making informed decisions. Such

decisions may include optimizing traffic flow, implementing road safety measures, and allocating resources for emergency response in a timely and efficient manner. Furthermore, the development of an ML-based accident severity prediction model can contribute to the advancement of the field of road safety research. It can provide valuable insights into the factors that contribute to severe accidents and help identify high-risk areas or conditions. The model can also serve as a foundation for further research and development of intelligent transportation systems, early warning systems, and other Gechnological solutions to enhance road safety.

Overall, the motivation behind developing an ML-based accident severity prediction model is to leverage the power of data-driven approaches to improve road safety, prevent accidents, and mitigate their impacts, ultimately saving lives and reducing the economic and social costs associated with road accidents.

1.3. Problem statement:

Today, automobile crashes are one of the leading causes of death worldwide. Measures must be taken to lessen or mitigate the impact of the accident. The determination of the most efficient measures necessitates an effective analysis of accidents that can identify and analyses the causes of an accident. Our algorithm would anticipate road accidents based on factors such as speed, alcohol consumption, and violation of traffic laws, among others.

The goal of our project is to use machine learning algorithms such as Decision Tree, KNN, and Random Forest to detect road accidents based on accidental circumstances, predict road accidents, and determine what factors influence accidents the most. The collected data will be analyzed, preprocessed, integrated, and grouped together based on various constraints utilizing the best suitable.

2. LITERATURE REVIEW

Roads and transportation are important part of our daily life. Experiment and many techniques are constantly finding ways to prevent often road accidents. Accidents are caused by many more factors, such as road condition, human status and weather. According to WHO (World Health Organization)'s report

- Road traffic crashes kill approximately 1.19 million people each year.
- Road traffic injuries are the greatest cause of death among children and young adults aged 5 to 29 years.
- Low- and middle-income countries account for 92% of all road fatalities, while having approximately 60% of the world's vehicles.
- More than half of all traffic fatalities occur among vulnerable road users, such as pedestrians, cyclists, and motorcyclists.
- The United Nations General Assembly has set an ambitious objective to halve the global number of deaths and injuries from road traffic collisions by 2030.

Accidents involving big cargo vehicles, truck and buses. Speed of vehicles, status of road, people not wearing helmet, and breaking traffic rules have all contributed to accidents inspite of having rules and regulation. Road accidents is very important and it will be improved by passing strict rules and regulation and promoting hard training session before issuing driving license. Machine learning is an artificial intelligence technique that may be used to train models to recognize patterns in data. Machine learning models may be trained to recognize face traits linked with tiredness in the instance of driver drowsiness detection. A lot of research has looked into the use of machine learning for detecting driver sleepiness. Machine learning models may be used to identify tiredness with great accuracy, according to this research. Machine learning for detecting driver sleepiness is an exciting field of research. Machine learning models have the potential to be more accurate, cost-effective, and convenient than present driver condition detection systems for drivers. This literature study will address the state of the art in detecting driver state using machine learning. The paper will go through the various machine learning approaches that have been, utilized for detecting accident factors and driver state, the many datasets that have been used for training and assessment, and the various evaluation metrics that have been employed. The review will also go through the problems and limits of applying machine learning to detect factors. The purpose of this literature review is to offer a complete overview of the state of the art in detecting accident factors and human state. In future we will be able to control road accident by the help of machine and technologies.

2.1. Gaps in existing knowledge:

There are some knowledge gaps regarding the application of machine learning for driver status detection. One shortcoming is a dearth of study on the application of machine learning for detecting driver sleepiness in real-world scenarios. The majority of machine learning research for driver sleepiness detection has been done in laboratories. More study on the application of machine learning for driver sleepiness detection in real-world situations, such as highways and motorways, is required. Another knowledge gap is the absence of research on the use of machine learning for driver sleepiness detection in various contexts. More study on the application of machine learning for driver sleepiness detection in various situations, such as varied weather conditions, at night, and in diverse traffic circumstances, is required.

3. RESEARCH DESIGN AND METHODS

Our project's primary goal is to forecast the severity of accidents. enabling the user to determine the severity using a range of numbers in a flexible and user-friendly manner. Accuracy in predicting the severity of accidents is 88%.

3.1. Research design:

Data collection: Compile accident information from pertinent sources, such as insurance providers, traffic departments, or publicly available datasets like traffic accident records. Incorporate a range of parameters, including the type of road, the weather, the features of the vehicle, and human factors (such as the driver's age and gender).

Data pre-processing: To deal with missing numbers, outliers, and inconsistencies, clean up the dataset.

Use methods like one-hot encoding or label encoding to translate categorical variables into numerical representations.

Scale or normalize numerical features to guarantee consistency throughout the dataset.

Exploratory Data Analysis: Visualize accident distribution across different variables (e.g., time of day, location).

Identify correlations between accident severity and contributing factors using statistical methods and visualizations.

Model selection and Training: Try out several supervised learning techniques (including decision trees, random forests, support vector machines, and neural networks) that are appropriate for categorization tasks. Using a subset of the dataset, train several models, then use suitable methods (such cross-validation) to verify each model's performance.

Model Evaluation: In order to judge the effectiveness and dependability of machine learning models created for the purpose of predicting the severity of traffic accidents, model evaluation is essential. Using the right metrics to assess the models' efficacy, trained models are applied to test data in this step. Understanding the elements impacting accident severity and the models' forecasting abilities is made possible through the interpretation of model results.

3.2. Methods and resources:

Data Gathering: Identify sources of accident data (e.g., government databases, insurance records). Obtain necessary permissions and approvals for data access. Define the scope of data collection (e.g., geographical coverage, time period).

Pre-processing: When preparing raw data for machine learning analysis and modelling, data pre-processing is a crucial step. In order to guarantee data quality, consistency, and modelling method compatibility, it must be transformed and cleaned. In general, this procedure entails handling outliers, handling missing values, scaling numerical features, encoding categorical variables into numerical representations, and dividing data into training and testing groups. Optimizing data format and quality is the aim of data preparation, which enables machine learning algorithms to recognize patterns and generate precise predictions. Through addressing data inconsistencies and preparing the data for analysis, each preprocessing step helps to improve the performance and interpretability of the model.

Model Development: Model development in the context of road accident severity prediction involves leveraging machine learning algorithms to build predictive models from preprocessed data. This phase focuses on selecting appropriate algorithms, training models, optimizing parameters, and evaluating performance to achieve accurate predictions.

Interpretation and Reporting: In the road accident severity prediction research process, interpretation and reporting are crucial stages that concentrate on evaluating model results, drawing conclusions, and successfully conveying findings to stakeholders and decision-makers.

4. IMPLICATIONS AND CONTRIBUTIONS TO KNOWLEDGE

4.1. Introduction:

Because machine learning has a track record of accurately forecasting traffic accidents, we employ it. Based on variables such as road condition, speed, weather, light, crossing physical amenities, and accident severity, models are built using accident data. We start by gathering statistics on traffic accidents. Next, by finding the missing values, filling in the blanks, and doing an analysis of the data, pre-processing procedures assist clean up the raw data and transform it into a clean dataset. We obtain the cleaned data after these procedures. Next, the trained dataset is used to train the model. The model is trained using a training dataset in order to predict the outcomes. Eighty percent of the data were used as training data in our research. Following Training, the model is trained using a training dataset in order to predict the outcomes. Eighty percent of the data were used as training data in our research.

A testing dataset is used to test the model once it has been trained. Twenty percent of the data were used for testing. Subsequently, various machine learning techniques such as KNN, decision tree, and Random forest are employed to evaluate the trained and test data in order to achieve accuracy for the provided dataset. Ultimately, the model is fitted using the algorithm that yields the best accuracy.

4.2. Algorithm and Process Design:

We are going to use a Supervised method as we are going to provide labelled data. As our ML project is based on prediction of occurrence of road accidents, we want the output as either yes or no. Hence, we will use a classification method.

Algorithms used are:

- Decision Tree
- Random Forest
- Logistic Regression

Decision Tree: Partitioning the source set into smaller groups according to an attribute value test allows a tree to be "taught." Every derived subset undergoes a cyclical repetition of this process. The recursion ends when splitting no longer improves the predictions, or when every member of the subset at a node has the same value for the target variable.

Random Forest: One of the reliable methods for predicting a lot of datasets is this one. Before combining or aggregating subtrees, the random forest generates a large number of random, shallow subgroup trees. More accurate predictions are also obtained when used with huge datasets.

Logistic Regression: All regression analysis, also known as logistic regression, is a prediction analysis in which the dependent variable is binary integers, or "Os and 1s." The quality—Accident Severity will be utilized to forecast the accident's severity.

5. TESTING OF THE PROPOSED DESIGN

A range of test cases should be used when evaluating a machine learning model to forecast accidents in order to guarantee the model's accuracy and generalizability. The following are a few test situations that we thought about:

Normal conditions: To make sure the model doesn't incorrectly anticipate an accident, test it using data that represents typical driving conditions without any incidents.

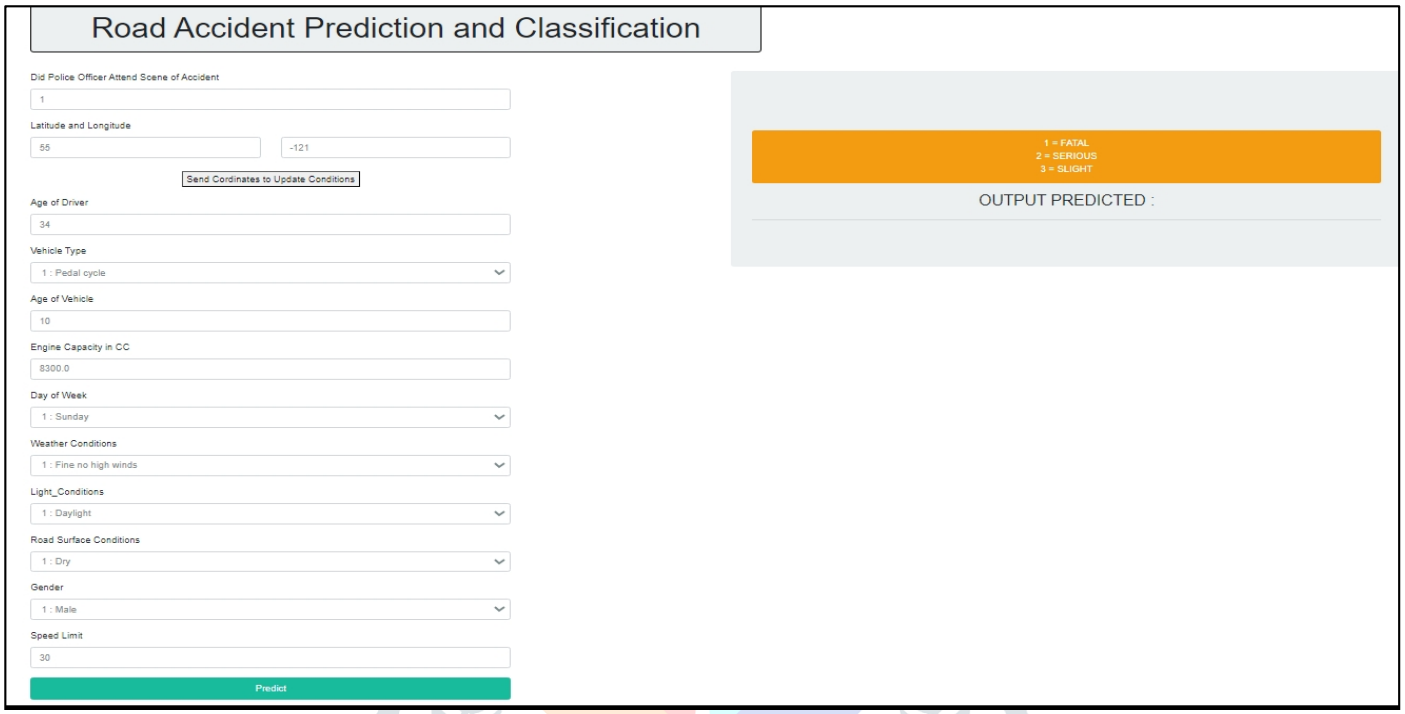
Known accidents: Using data from known accidents, test the model's predictive accuracy to determine whether accidents will occur.

Various accident types: To determine whether the model can reliably forecast various accident types, it is tested using data that represents various accident kinds (such as rear-end collisions and T-bone accidents).

Data: Lastly, evaluating the model's generalizability to novel scenarios using entirely fresh data that was not utilized during the testing or training phases. Lastly, evaluating the model's generalizability to novel scenarios using entirely fresh data that was not utilized during the testing or training phases.

6. RESULT

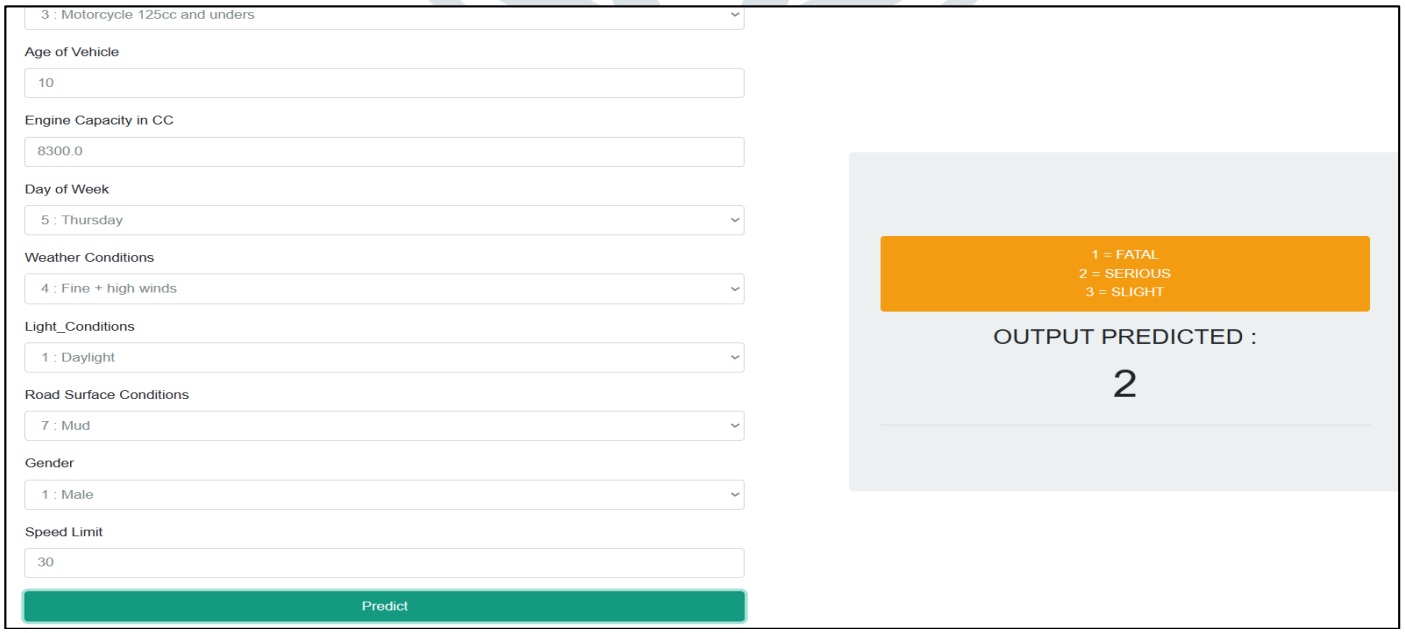
The accompanying figure shows the home page of the web application. The web domain has been secured with HTTPS, which was obtained from the certificate authority, in order to facilitate secure data transfer and make use of the Geolocation API.



7. OUTPUT

Numerous assessments were carried out to compare the three methods and determine which one is best at predicting the seriousness of traffic accidents. The Jupyter notebook was used for the panda and seaborn investigations. This section presents and discusses the experimental setup, procedures, and results for three distinct algorithms:

- Random forest
- Decision tree
- Logistic regression



8. CONCLUSION

The main purpose of proposing this machine learning model to reduce road accidents efficiently and effectively by generating a predicting model that find out the causes of accidents and alerts about it. It will save a lot of lives by detecting the major factors of accidents. The machine learning model has enabled us to analyses meaningful data to give solutions with more accuracy than humans.

The study has limitations, including the possibility of unaccounted-for factors influencing accidents. In the future, comprehensive datasets will incorporate more relevant variables. Using variables like air humidity and wind speed may increase prediction accuracy. This study's implementation can help traffic authorities reduce accidents by providing reliable and validated predictions of severity. To make this more possible, the authors suggest developing a system or mobile app that can accurately forecast and alert drivers about traffic accidents.

9. FUTURE WORK

In the future, Spatial-temporal graph neural networks have potential for expanding their use in road traffic accidents by analyzing more spatial and accurate attributes. With the help of technologies, we can generate alerts for the location where prevention can be taken by traffic police to avoid accident. Google maps that help in real time tracking can integrated in online application. If Indian government provide real time data, we will be able to Indian streets too.

10. REFERENCES

[1].https://www.researchgate.net/publication/369557935_Road_Car_Accident_Prediction_Using_a_Machine-Learning-Enabled_Data_Analysis
[2]. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3893062
[3].https://www.researchgate.net/publication/370004806_A_Systematic_Review_on_Road_Traffic_Accident_Causes_and_Control_Measures
[4].https://www.sciencedirect.com/science/article/pii/S2352146517307913