International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

ROAD ACCIDENT ANALYSIS USING MACHINE LEARNING

Dr. Vilas Joshi[1]

Shivansh Gautam[2], Prashant Singh[3], Sonal Raj[4], Prince Singh[5]

Computer Engineering Department, ISBM College of Engineering Nande, Pune-412115,

India Savitribai Phule Pune University

Abstract: **Today, one of the top concerns for governments is road safety. Although there are various safety precautions in place to prevent auto accidents, they cannot be completely avoided. To lessen the harm caused by traffic accidents, the primary goal now is to determine what causes them. In this study, we use machine learning techniques to identify the causes of traffic accidents. By creating precise prediction models that can automatically separate distinct unintentional instances, patterns involved in diverse situations can be identified. The development of safety measures and the application of these classification approaches will help avoid accidents. Although there are numerous inventories in the automotive sector to create and construct safety features for cars, road accidents are inevitable. Both urban and rural regions see a high rate of accidents. By creating precise prediction models that can automatically separate distinct unintentional instances, patterns involved in diverse situations can be identified. These clusters will help create safety precautions and prevent mishaps. We think we can use some ML techniques to reduce accidents as much as possible while using limited resources.**

I. INTRODUCTION

Accident-related fatalities and injuries are predicted to be an increasingly common problem. Since the invention of the vehicle, traffic safety has been a major problem. In India, road accidents claimed the lives of 1,53,972 individuals in total in 2021, according to the Ministry of Road Transport and Highways (MoRTH). This equals an average of 422 fatalities per day. Approximately 67 percent of all unintentional fatalities occur in people between the ages of 18 and 45, which is the age group most frequently affected by traffic accidents.. Statistics have also shown that young individuals, who make up a significant portion of the workforce, have a relatively high death rate in traffic accidents. Various road safety measures are required to solve this issue.

Research interest in figuring out the important impact of the severity of accident caused by traffic accidents has increased in recent years. Accident analysis is built on accurate and thorough accident records. The correctness of the data, record retention, and data analysis are some of the aspects that affect how well accident records are used. Numerous methods have been used to analyse this issue using this scenario.

The prime goal of this research paper is to analyze the road accidents and determines the severity of an accident by applying advanced machine learning techniques. There exist so many developed methods in machine learning to examine this problem.

II. PROBLEM DEFINITION

To handle the enormous number of road accidents in India a precise analysis is required. This analysis will be done more deeply to determine the intensity of road accidents by using different machine learning techniques like supervised learning, unsupervised learning, etc.

Road accident analysis using machine learning to identify the factors that contribute to accidents and develop interventions to address these factors. This can be done by using machine learning to analyze historical accident data and identify patterns that would be difficult to see with the naked eye. Once the factors that contribute to accidents have been identified, interventions can be developed to address these factors. For example, if it is found that speeding is a major factor in accidents, interventions could be developed to educate drivers about the dangers of speeding and to enforce speed limits.

Machine learning is an effective tool that can be used to examine data on traffic accidents and spot trends that are difficult to spot with the unaided eye. Machine learning can help to make our roads safer

by identifying the factors that cause accidents and creating interventions to address these factors.

III. LITERATURE SURVEY

Author: Kundan Meshram and H.S. Goliya

The research paper "Accident Analysis on National Highway-3 Between Indore to Dhamnod" by Kundan Meshram and H.S. Goliya[1]. A study of accident data on National Highway-3 between Indore and Dhamnod, India, found that the number of accidents had increased over the past few years. The most common causes of accidents were speeding,drunk driving, and driver fatigue. The majority of accidents occurred during the night and on weekends. The study concluded by making a number of recommendations to improve safety on this highway, including increasing the number of speed bumps, installing more streetlights, and increasing the number of traffic police patrols.

Author: Md. Farhan Labib, Ahmed Sady Rifat, Md. Mosabbir Hossain, Amit Kumar Das, Faria Nawrine.

The research paper "Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning " by Md. Farhan Labib, Ahmed Sady Rifat, Md. Mosabbir Hossain, Amit Kumar Das, Faria Nawrine, investigate the use of machine learning to analyze and predict accident severity in Bangladesh. The paper uses a dataset of road accidents from 2016 to 2018 to train and evaluate a number of machine-learning models. The results of the study show that machine learning can be used to predict accident severity with a high degree of accuracy. The paper concludes by discussing the implications of these findings for road safety. They discovered that machine learning can accurately predict the severity of accidents. The behaviour of the driver, the state of the roads, flaws in the vehicle, and environmental factors are the most significant factors that affect how serious an accident is. The results of this study can be used to create efficient road safety improvement strategies.

Author: Dr. Anitha Patila, Prithvish Kumbleb, Naresh Kc, Sriharid

The research paper "Road Accident Analysis" by Dr. Anitha Patila, Prithvish Kumbleb, Naresh Kc, Sriharid, investigates the causes of road accidents in India. The most frequent causes of accidents are identified by the paper using a dataset of traffic accidents from 2010 to 2014. The study's findings indicate that driver errors, such as speeding, drunk driving, and distracted driving, are the most frequent causes of accidents. weather, the condition of the road, traffic congestion, defects in the vehicle, such as bad tires, brakes, or steering, and environmental elements like rain, fog, and darkness. The implications of these findings for India's road safety are covered in the paper's conclusion. The study's authors urge the government to take action against the most frequent causes of collisions, including educating motorists about the risks of speeding, drunk driving, and distracted driving, enforcing speed limits and other traffic laws, enhancing road conditions, making cars safer, and improving weather forecasting and procedures for road closures. raising public awareness of issues related to road safety. The study's authors are of the opinion that the government can lessen the frequency and severity of road accidents in India by taking these actions.

IV. ASSUMPTIONS AND DEPENDENCIES

Here are some of the assumptions and dependencies that we have to consider while creating an analysis model using machine learning:

Data quality: The quality of the data used to train the model is critical. The data should be accurate, complete, and representative of the real world.

Model complexity: The complexity of the model should be appropriate for the data. A complex model may be more accurate, but it may also be more difficult to train and interpret.

Training data: The training data must be accurate and realistic. A wide range of potential accident-causing variables, such as driver behaviour, road conditions, and weather, should be included in the data.

Evaluation data: The evaluation data should be separate from the training data. This data is used to evaluate the accuracy of the model.

Deployment: The model should be deployed in a way that is accessible to users. The model should be easy to use and interpret.

Bias: The model should be free of bias. The model should not be biased against any particular group of people.

Interpretability: The model should be interpretable. The model should be able to explain its predictions. By considering these assumptions and dependencies, we can create a road accident analysis model that is accurate, reliable, and fair.

V. DATA FLOW

A data-flow diagram(DFD) is a visual representation of how data moves through a process or system (usually an information system). Additionally, the DFD provides details about the inputs and outputs of each entity as well as the process itself for Road Accident Analysis using Machine Learning. A data-flow diagram lacks loops, decision rules, and control flows. Using a flowchart, specific operations based on the data can be depicted. Dataflow diagrams can be displayed using a variety of notations. A process must contain at least one of the endpoints (source and/or destination) for each data flow. Another data-flow diagram that divides a process into sub-processes can be used to represent a process in more detail. The structured analysis modeling tools include the dataflow diagram.

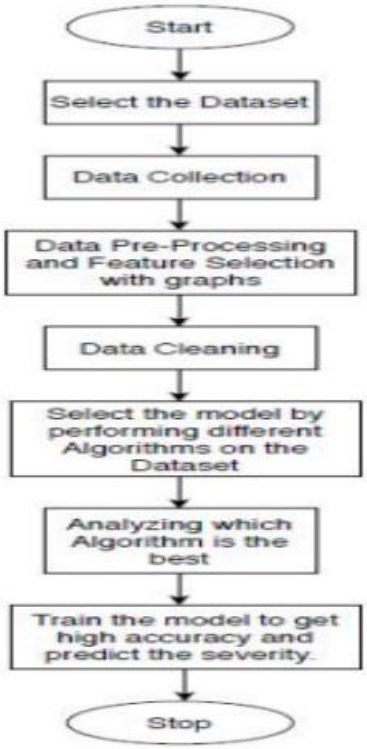

Figure 1: Data Flow Diagram

Data is collected from a variety of sources, such as police reports, government web pages, and researchers. The data is cleaned, pre-processed, and engineered to create features that are used to train a machine-learning model. The model is evaluated and deployed so that users can use it to predict the probability of an accident. The model is monitored over time to ensure that it is still accurate.

VI. PROPOSED MODEL

PLANNING AND ANALYSIS: Our first step will be to define the problem and the desired outcome. This involves understanding the context of the problem, identifying the relevant data, and specifying the desired outcome of the model.

DATA GATHERING: To gather data for machine learning, we will use a variety of sources, including Kaggle and other trusted sources. Kaggle is a great resource for finding high-quality data sets, and it makes it easy to import data into our model.

DATA PRE-PROCESSING: In order to make raw data suitable for analysis, it must first be cleaned, formatted, and transformed. The fact that it guarantees the data's accuracy, consistency, and completeness makes it a crucial step in the data mining process. There are many reasons why data pre-processing is necessary. First, raw data is often dirty, meaning that it contains errors, inconsistencies, and missing values. These errors can make it difficult or impossible to analyze the data. Second, raw data is often in a format that is not compatible with data mining tools. Data pre-processing can help to convert the data into a format that is compatible with these tools. Third, data pre-processing can help to improve the accuracy of data mining results. By cleaning and formatting the data, data pre-processing can help to remove errors and inconsistencies that can lead to inaccurate results. A vital phase in the data mining process is data pre-processing. Data pre-processing can help to ensure that the data is accurate, consistent, and complete by cleaning, formatting, and transforming raw data. As a result, data mining results may be more accurate and data analysis may be simpler.

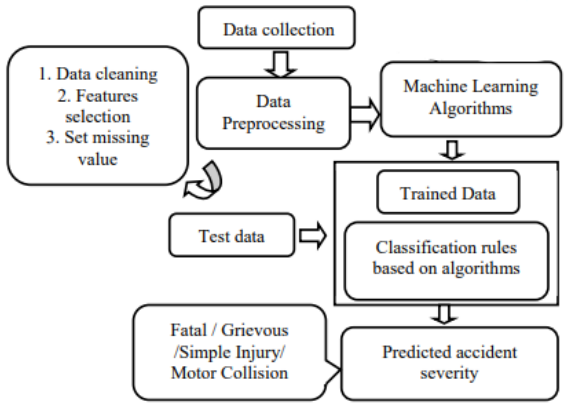

Figure 2 : Model Pipeline Process

FEATURE SELECTION: Feature selection is the process of selecting a subset of features from a dataset that is most relevant to the target variable. This is an important step in the machine learning process, as it can help to improve the accuracy and performance of the model. In our case, we are using features like speed, weather conditions, and road conditions to build our model. These features are all relevant to the target variable, which is the likelihood of an accident. By selecting the right features, we can improve the accuracy and performance of our model, which can help to improve safety on the roads.

MODEL SELECTION: Model selection is the process of choosing the best model from a set of candidate models. The best model is the one that has the highest accuracy on the training data and the testing data. We will also consider the complexity of the model and the time it takes to train the model. The model with the highest accuracy will be the final model.

MODEL TRAINING: Pre-processed data are used to train the model. Based on the engineered features, the model is trained to forecast the probability of an accident.

MODEL EVALUATION: Model evaluation is the process of assessing the quality of a machine-learning model. It involves using the evaluation data to measure the accuracy of the model. There are many different methods that can be used to evaluate a machine learning model, some of the most common methods include accuracy, precision, recall, F1 score, and area under the curve (AUC). The best method for evaluating a machine learning model will depend on the specific application.

MODEL DEPLOYMENT AND MAINTENANCE: Model deployment and maintenance is the process of making a machine learning model available for use in a production environment and ensuring that it continues to perform well over time. This involves packaging the model, making it accessible to users, monitoring its performance, updating the model with new data, retraining the model, and monitoring the model for bias. Streamlit and Flask are two popular frameworks that can be used to build user interfaces for machine learning models.

VII. SOFTWARE AND HARDWARE REQUIREMENTS

Language used*:* Python and HTML.

Software Requirements*:* Pycharm and VS code.

Hardware Requirements *:*

- RAM: 8MB
- Hard Disk: 100 GB working space (minimum requirement)
- Processor: Above i5-9 Gen @2.50 GHz
- Operating System: Any operating system

VIII. OTHER SPECIFICATION

A. *Advantages*

1. Identify the causes of accidents: Road accident analysis can help to identify the factors that contribute to accidents, such as driver error, road conditions, and weather. This information can be used to develop safety measures that can reduce the number of accidents and injuries.
2. Develop safety measures. Road accident analysis can be used to develop safety measures that can reduce the number of accidents and injuries. For example, if an analysis finds that a particular type of intersection is dangerous, safety measures such as traffic lights or speed bumps can be installed to make the intersection safer.
3. Governments can use road accident analysis to improve road design and construction. This can be done by identifying and addressing dangerous road features, such as poorly-marked intersections or blind spots. By making roads safer, governments can help to reduce the number of accidents and injuries.

4. Road accident analysis can be used to train automatic vehicles to be more cautious and avoid mistakes. This can help to reduce the number of accidents involving automatic vehicles and make roads safer for everyone.

B. Limitations

1. Data on road accidents is often incomplete or inaccurate. This can make it difficult to identify the causes of accidents and develop effective safety measures.
2. Road accidents are often complex events with multiple contributing factors. This can make it difficult to identify all of the factors that contributed to an accident and to develop effective safety measures.
3. Road accident analysis projects often rely on data that is collected after an accident has occurred. This can make it difficult to identify the causes of accidents and to develop effective safety measures in a timely manner. Collecting real-time data about road accidents can help to address this limitation and improve road safety.
4. Risk analysis models are only as good as the data they are trained on. If a model is not accurate, it can lead to inaccurate predictions of risk, which can put people in danger. It is important to make sure that risk analysis models are as accurate as possible before using them to make decisions about risk.

IX. CONCLUSION

Road Accidents are caused by various factors. By going through all the research papers, it can be concluded that Road Accident cases are hugely affected by these factors. Thus, we have proposed an application that gives an efficient prediction of road accidents based on various factors. Due to analyzing and severity prediction, we can reduce road accidents by taking some precautions before the occurrence of accidents.

X. ACKNOWLEDGMENT

We would like to take this opportunity to thank all the people who were part of this seminar in numerous ways, people who gave unending support right from the initial stage.

In particular, we wish to thank Dr. Vilas Joshi as an internal project guide who gave his co-operation timely and precious guidance without which this project would not have been a success. We thank him for reviewing the entire project with painstaking efforts and more of his, unbanning ability to spot the mistakes.

We would like to thank our H.O.D Prof. B. B. Gite for his continuous encouragement, support and guidance at each and every stage of the project.

And last but not least we would like to thank all our colleague who were associated with us and helped us in preparing our project. The project named “Road Accident Analysis” would not have been possible without the extensive support of people who were directly or indirectly involved in its successful execution.

XI. REFERENCES

- K Meshram, and S.H Goliya “Accident analysis on national highway 3 between Indore to Dhamnod”. International Journal of Application or Innovation in Engineering & Management (IJAIEM) ISSN 2319 – 4847.
- Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning .by- Md. Farhan Labib, Ahmed Sady Rifat, Md. Mosabbir Hossain, Amit Kumar Das, and Faria Nawrine. Published by International Conference on Smart Computing & Communications (ICSCC). 978-1-7281-1557-3/19/$31.00 ©2019 IEEE.
- Road Accident Analysis by Dr. Anitha Patila, Prithvish Kumbleb, Naresh Kc, and Srihari M. Dadkhah. Published by Turkish Journal of Computer and Mathematics Education. Vol.12 No.10(2021), 392-396.
- Road accident in India 2020 published by the Ministry of Road Transport and Highways (MoRTH).
- Road accident in India 2021 published by the Ministry of Road Transport and Highways (MoRTH).