

Accident Prediction and Analysis using Machine Learning models

U. Vivek Krishna, S. Sudhakaran, S. Sanju, E. Vignesh, R. Kaladevi

Department of Computer Science and Engineering, Saveetha Engineering College, India

vivekkrish71@gmail.com, sudhakaran140701@gmail.com, sanju.satish.raman2002@gmail.com,
vignesh2622vicky@gmail.com, kalaramar26@gmail.com

Abstract—Road safety is a major concern due to the rapidly growing number of automobiles. Every year, 1.2 million individuals die in road accidents. According to statistics, between 2017 and 2020, traffic accidents in Tamil Nadu resulted in the deaths of almost 22,000 individuals. According to the "Accidental Deaths & Suicides in India—2021" report released by the National Crime Records Bureau (NCRB) shows Tamil Nadu at second in the number of fatalities on roads caused by negligence. The State came in second to Uttar Pradesh, which recorded 18,972 fatalities in 18,228 accidents, with 15,384 fatalities in 14,747 occurrences. The proposed work is used to predict the severity of an accident at a place and time using few machine learning algorithms namely decision tree, random forest, logistic regression, and decision tree hyper parameter tuning. Various parameters are found, such as the speed limit, age, weather, vehicle type, light conditions, and day of the week had been used to train the model. Upon inferring the results, we found that random forest algorithm showed the highest accuracy of these algorithms, with a score of 88.89%, followed by logistic regression (86.23%), decision tree hyper parameter tuning (85.74%), and decision tree (66.67%), respectively. The major challenges faced in this project was collecting proper samples for training the models and modifying it as per the needs of the proposed model. This model will be crucial for traffic planning and management and will aid in the future reduction of many traffic accidents.

Keywords- Road safety, Deaths, Automobiles, Machine Learning.

I. INTRODUCTION

Recent claim by the World Health Organization (WHO), tells that number of accidental deaths worldwide annually is alarming. Every year 1.2 million people lose their lives in roads and 50 million people are injured. Every day, as far as 3,300 people lose their lives and 137,000 people are injured every day. Traffic accidents which are occurring continuous will directly impacts in life of an individual and which responds to loss of billions.

These days the principal research areas are in accident prediction. The geometric properties of the road, movement in traffic, operator characteristics, and the surrounding environment all these has an significant influence on the accidents. Brian Zirkel et al. proposed a paper in which they investigated the correspondence on sight distance parameters, accident rates, and accessing speeds.[1]

Numerous research on the frequency of accidents and the elements that contribute to them have been conducted, studies including dangerous locations/identification of major accident areas, severity of wound and in while some concentrated in the mechanisms of accidents. Other considerations include the weather and road visibility. The authors of "Recognition of Urban Traffic Accident Prone Zones Using GI's" have determined the location and causes of accidents in Dhaka.[2]

This paper about accidents will play an essential part in the overall designing and traffic management, the reason for which includes several nonlinear components, such as individuals, heap, thoroughfare, climate, and many.

This paper is organized as follows chapter 2 provides brief insight about the works of different author. Chapter 3 explains about proposed methodology. Chapter 4 has the snippet of the dataset used along with brief explanation of the algorithm used and results. Chapter 5 provides the conclusion and insights and future works.

II. LITERATURE SURVEY

As vehicle count rises, road safety becomes a huge concern. An average road accident kills around 1.2 million people per year. Crashes estimates worldwide around \$518 billion, costing individual countries from 1- 2% of their economy. In Tamil Nadu a total of 15,384 people got killed by road accidents in 2021 compared with the year 2020 which takes a count at a total of 14,527. Steps are being taken to minimize the occurrence of accidents but they are becoming inefficient.

T. Augustine and S. Shukla introduced a system of accident prediction which helps to examine the potential safety issues and predict the possibility of accidents [3].

Asghar Pasha, Vijayalakshmi, and other authors propose clustering techniques and other machine learning models to identify factors which are statically significant to predict the probabilities of crashes and injury that can be used to perform a risk factor and reduction. [4].

Koteswara Rao Ballamudi proposes some machine learning algorithms which is based on classification and clustering methods for accident analysis based on factors involved in the accident [5].

The author proposes many exploration method, such as ARIMA model, time series model, and missing values detection model the check different angles of traffic accidents. [6].

The authors have identified the risk factors that affect the severity of a vulnerable road users injured when involved in a motor vehicle crash. [7].

S. Seid and Pooja implemented ANN model to analyze injury based accidents and classified into the category of no trauma-viable trauma-minor non immobilize injury ,immobilize and fatality [8].

Hanninen managed maritime traffic safety and made decisions through the Bayesian network model. [9].

Mrudula S P and Sarah Jaison George performed a method called as Accident Severity Index Method to identify blockspots by which they are prioritized according to severity of location and performed analysis on road safety in identified hotspots. [10].

Luo Dongyu & Yan Menghua, generated a probabilistic model linking major crash antecedents to variations in crash potentials [11].

The paper on number of vehicles, lighting conditions, road features focuses on the Traffic Collision also called MVC(Motor Vehicle Collision). More Focused on Road Lights, although the occurrence of accident cannot be controlled. It requires huge data, comprising accident severity categories based on difficulties such as slight or fatal . The techniques of Navie Byes and XG Boosting are used [12].

The authours have developed a GIS -based framework to have an detail analysis on the road accident and applied the geographical information system in reducing the blackspots and to have an idea of a safer urban road network [13].

The authors have proposed an model to classify road network based on the traffic accident which is integrated with the geographical information system(GIS) [14].

Authors had made a series of artificial neural networks to make a relationship between the crash related factors and injury severity on the traffic crash on this calculated the effects on crash factors on injury severity levels on different [15].

III. PROPOSED SYSTEM

As the road accidents are inevitable in today's world , We have proposed the use of following machine learning algorithms such as decision tree, random forest, logistic regression and decision tree hyper parameter tuning to predict the possibility of road accidents. These algorithms are used for classification and regression process. Our first process is the obtaining a raw dataset and pre-processing operations are performed such as cleaning, transformation and reduction. After pre-processing, the dataset is splitted into inculcation set and evaluation set. After splitting of dataset, some machine learning algorithms are implemented and the trained model is obtained. Then the model is evaluated and result is obtained. In this way, analysis is performed. From the analysis the result is obtained in the form of accuracy. The process is implemented step by step in the fig 3.1.

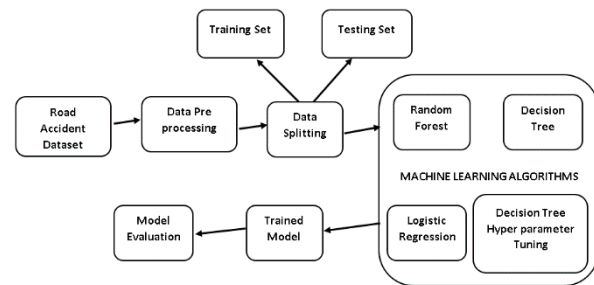


Fig 3.1: Accident Prediction Architecture.

IV. IMPLEMENTATION AND ANALYSIS

A. DATASET

The below fig.4.1. is the dataset snippet which is used for training the machine learning model.

Accident_Index	Location_Easting	OSGR	Location_Northing	OSGR	Longitude	Latitude	Police_Force	Accident_Severity
200501BS00001	525680		178240		-0.19117	51.489096	1	2
200501BS00002	524170		181650		-0.211708	51.520075	1	3
200501BS00003	524520		182240		-0.206458	51.525301	1	3
200501BS00004	526900		177530		-0.173862	51.482442	1	3
200501BS00005	528060		179040		-0.156618	51.495752	1	3
200501BS00006	524770		181160		-0.203238	51.51554	1	3
200501BS00007	524220		180830		-0.211277	51.512695	1	3
200501BS00009	525890		179710		-0.187623	51.50226	1	3
200501BS00010	527350		177650		-0.167342	51.48342	1	3
200501BS00011	524550		180810		-0.206531	51.512443	1	3
200501BS00012	526240		178900		-0.182872	51.494902	1	3
200501BS00014	526170		177690		-0.184312	51.484044	1	3
200501BS00015	525590		178520		-0.192366	51.491632	1	3
200501BS00016	527990		178690		-0.157753	51.492622	1	3
200501BS00017	526700		178970		-0.176224	51.495429	1	3
200501BS00018	526460		177460		-0.18022	51.481912	1	3
200501BS00019	524680		179450		-0.205139	51.500191	1	2
200501BS00020	527000		179020		-0.171887	51.495811	1	3
200501BS00021	527810		178010		-0.16059	51.486552	1	3
200501BS00022	526790		178980		-0.174925	51.495498	1	2
200501BS00023	525940		178680		-0.18727	51.492992	1	3
200501BS00024	524700		180880		-0.204346	51.513039	1	3
200501BS00025	526930		177490		-0.173445	51.482076	1	3
200501BS00028	527290		178710		-0.167824	51.49296	1	3
200501BS00029	527380		179280		-0.166322	51.498062	1	3
200501BS00031	523930		180330		-0.215629	51.508265	1	3
200501BS00032	524470		180980		-0.207623	51.513988	1	3
200501BS00033	525570		178190		-0.192771	51.488671	1	3

Fig 4.1: Snippet of the dataset.

B. ALGORITHM

We have introduced some of the following machine learning algorithms in this paper.

1. DECISION TREE

Decision Tree is a supervised learning algorithm where it can be used to solve regression problems and their classification. It is like a tree structure, which comprises of internal nodes standing in for a dataset's features, decision-making process, and each leaf node for the classification result.

The decision tree is constructed using the two nodes namely Decision nodes and leaf nodes. The Decision nodes are the vital nodes used to create decisions and have numerous branches.

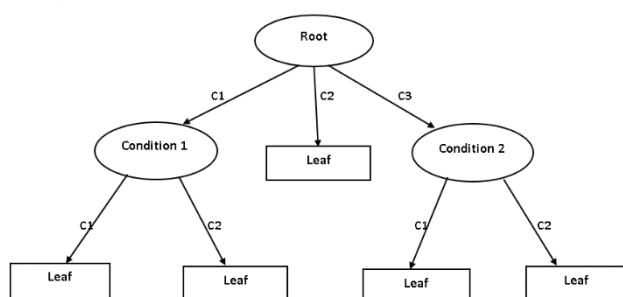


Fig 4.2 : Decision Tree Algorithm

2. RANDOM FOREST

Random Forest algorithm which used for classification and regression problem is implemented to predict accidents. The idea is to integrate various classifiers to address difficult issues and enhance model performance. It takes the average of the total number of decision trees obtained for each subset of the dataset.

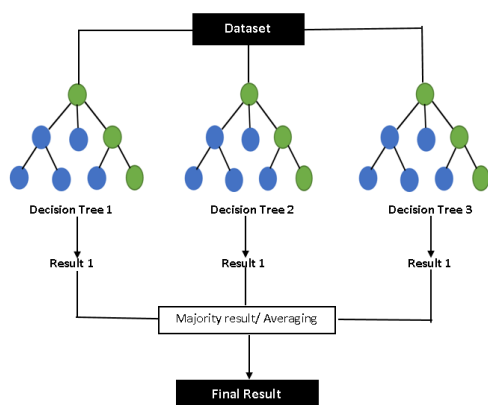


Fig 4.3: Random forest algorithm

3. LOGISTIC REGRESSION

Logistic Regression is one of the machine learning algorithm That predicts an outcome which is in the form of either 0 or 1, True or False and Yes or No. It uses a specific function called as a Sigmoid function as the value of the predicted output lies between 0 and 1. It is used for classification purpose as the outcome is based on the range whether it lies near 0 or 1.

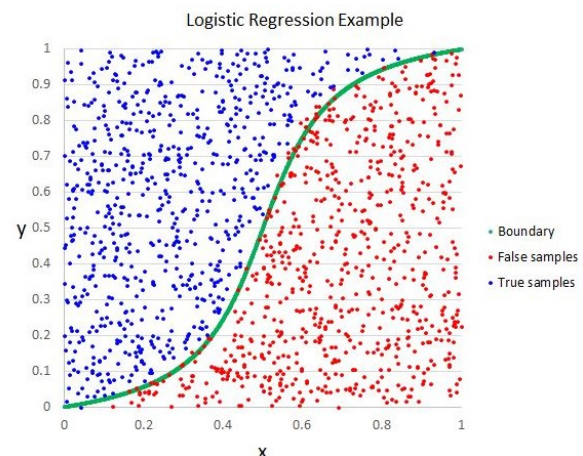


Fig 4.4: Logistic Regression Algorithm

4. DECISION TREE USING HYPERPARAMETER TUNING

Hyper parameter tuning is the process of calibrating our model by identifying the appropriate hyper parameters to make our model more widely applicable. We'll examine a handful of these hyper parameters, including :

4.1 Maximum Depth

The maximum depth of a tree is represented by this parameter. In the absence of a specification, the tree is grown until the last leaf nodes have only one value. Therefore, by lowering this meter, we may prevent over-fitting by stopping the tree from learning all training samples.

4.2 Maximum Leaf Node

This hyper parameter, as its name implies, limits the number of leaf nodes in a decision tree. Another technique to manage the complexity of the model is to let the branches of a tree to have different depths.

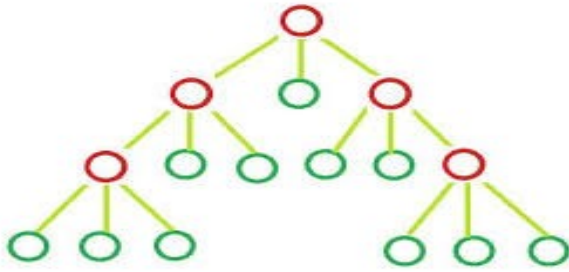


Fig 4.5: Decision Tree Hyper parameter Tuning

RESULT

Among the algorithms used to predict the road accidents random forest algorithm showed the highest accuracy .

ALGORITHM	RESULT
Decision Tree	66.67%
Random Forest	88.89%
Logistic Regression	86.23%
Decision Tree Hyperparameter Tuning	85.74%

Table 5.1 – Accuracy of algorithms.

V. CONCLUSION

The reports shows that the number of people die in road accidents increases every year. This enormously affects the economy of a country. This paper proposes the use of machine learning algorithms to predict the possibility of a accidents based on the previously occurred accidents. With the help of this model and proper dataset we hope that the number of road accidents can be reduced in future.

REFERENCES

- [1] Brian Zirkel et al, "Analysis of Sight Distance, Crash Rate, and Operating Speed Relationships for Low-Volume Single-Lane Roundabouts" in the United states, 2013.
- [2] Arif U. Ahmad, Khandaker T. Hossain, Mallik Hossain, Identification of Traffic Accident Hotspot Zones Using GIS: A Case Study of Dhaka Metropolitan Area, 2020.
- [3] T. Augustine and S. Shukla, "Road Accident Prediction using Machine Learning Approaches," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 808-811, DOI: 10.1109/ICACITE53722.2022.9823499.
- [4] Asghar Pasha, Vijayalakshmi, MD Atique, MD Hussain, Harsh namot and Bipin, "Road Accident Prediction using Machine Learning", 2021..
- [5] Koteswara Rao Ballamudi, "Road Accident Analysis and Prediction using Machine Learning Algorithmic Approaches", 2021.
- [6] Adnan Majeed "Road Accident of UK Traffic (1979-2019) DataSets Analysis Using Python Machine Learning Prediction" , International Conference of Computer Science 2021.
- [7] Vilaça, M.; Macedo, E.; Coelho, M.C. A Rare Event Modelling Approach to Assess Injury Severity Risk of Vulnerable Road Users. *Safety* 2019, 5, 29.
- [8] S. Seid and Pooja, "Road accident data analysis: Data preprocessing for better model building," J. Comput. Theor. Nanosci., vol. 16, no. 9, pp. 4019–4027, Sep. 2019.
- [9] Hanninen, M. Bayesian networks for maritime traffic accident prevention: Benefits and challenges. *Accid. Anal. Prev.* 2014, 73, 305–312.
- [10] Mrudula S P Sarah Jaison George, - "Black Spot Identification Using Accident Severity Index Method" International Journal of Current Engineering and Scientific Research (IJCESR)2018.
- [11] G. Battineni, N. Chintalapudi, and F. Amenta, "Machine learning in medicine: performance calculation of dementia prediction by support vector machines (SVM)," *Informatics in Medicine Unlocked*, vol. 16, Article ID 100200, 2019.
- [12] Rafael G .Mantovan,, Ricardo Cerri, Joaquin Vanschoren, "Hyper-parameter Tuning of a Decision Tree Induction Algorithm" INSPEC Accession Number: 16651860 DOI: 10.1109/bracis.2016.018. 2016.
- [13] Singh, N.; Katiyar, S.K. Application of geographical information system (GIS) in reducing accident blackspots and in planning of a safer urban road network: A review. *Ecol. Inform.* 2021, 66, 101436.
- [14] Rodrigues, D.S.; Ribeiro, P.J.; Nogueira, I.C. Safety classification using GIS in decision-making process to define priority road interventions. *J. Transp. Geogr.* 2015, 43, 101–110.
- [15] Topuz, K.; Delen, D. A probabilistic bayesian inference model to investigate injury severity in automobile crashes. *Decis. Support Syst.* 2021, 150, 113557.