# Road Accidents Prediction Using Machine Learning

**Pallavi N[1], Mayur raj[2]**
[1]Assistant Professor, Dept of Computer Science & Engineering
[2]Dept of Computer Science & Engineering
[1, 2] Atria Institute of Technology, Bangalore

***Abstract-*** *As we can see that automobile engineers and researchers have tried their best to design & build safe automobiles, but the traffic accidents on roads are unavoidable. We cannot neglect it that day by day the number of traffic accidents in the cities are increasing in a amount. Patterns involved in various dangerous crashes could be detected if we develop a prediction model.One of the key objectives in accident data analysis is to identify the main factors associated with road accidents.*

*This paper presents a novel machine learning strategies which uses KNN and many other algorithms and techniques for the prediction of traffic in better way to decrease the accidents and there after effects. This paper presents some models to predict the severity of injury that occurred during accidents using three machine learning approaches. It includes neural networks trained using hybrid learning approaches, decision trees and a concurrent hybrid model involving decision trees and neural networks. Experiment results reveal that among the machine learning process considered the hybrid decision tree-neural network approach outperformed the individual approaches.*

***Keywords***- Road Accidents, KNN Algorithm, Machine Learning, Hybrid Learning,Neural Network.

## I. INTRODUCTION

Road Accidents is becoming a worldwide problem nowadays. In a rush of being on time and to save a few minutes people faces crucial and very dangerous situations which leads to their death and severe injuries. The Transport Research Wing of the Ministry of Road Transport and Highways publishes an annual report of accidents, related deaths and injuries,India, however ranks 1st in the number of road accident deaths across the 199 countries reported in the World Road Statistics, 2018 followed by China and US. As per the WHO Global Report on Road Safety 2018, India accounts for almost 11% of the accidents related deaths in the world.A total of 4,67,044 road accidents have been reported by States and Union Territories (UTs) in the calendar year 2018, claiming 1,51,417 lives and causing injuries to 4,69,418 persons. In percentage terms, in 2018, the number of accidents has increased by 0.46 percent, persons killed has increased by 2.4% and injuries have decreased by 0.33 percent over that of the previous year i.e 2017. Road accident severity measured by the number of persons killed per 100 accidents, has seen an increase of 0.6 percentage points in 2018 over the previous year. With this increasing death rates and injuries, my aim is to help the system to get the prediction of road accidents so that it can be controlled in a positive way

## II. LITERATURE SURVEY

Multiple research papers were observed and overviewed. To deal with the limitations of statistical methodologies, the machine learning methods, including Artificial Neural Network(ANN), Support Vector Machine(SVM) models, and deep learning models, have been applied to various traffic safety problems and used as data analytic methods because of their ability to work with massive amounts of multidimensional data. In addition. Because of the modelling flexibility, learning and generalization ability, and good predictive ability, the machine learning has been considered as generic, accurate, and convenient mathematical models in the field of traffic safety.

[1]In this study, we conducted an analysis of road vehicle collisions in the city of Montreal using open data provided by Montreal city and the Government of Canada. Using three different datasets, we built road vehicle collision prediction models using tree-based algorithms. Our best model can predict85%ofroadaccidentsintheareaofMontrealwitha false positive rate of 13%. Our models predict the occurrence of a collision at high space resolution and hourly precision. In other words, it means our models can be used to identify the most dangerous road segments every hour, in order to take actions to reduce the risk of accidents. Moreover, we believe that our work can easily be reproduced for other cities under the condition that similar datasets are available. One can freely use our source code on GitHub for reference. Finally, our study shows that open data initiatives are useful to society because they make it possible to study critical issues like road accidents.

[2] In this paper, we studied the GES automobile accident data from 1995 to 2000 and investigated the performance of neural network, decision tree, support vector

machines and a hybrid decision tree- neural network for predicting drivers' injury severity in head-on front impact point collisions. The classification accuracy on the test results reveals that, for non-incapacitating injury, incapacitating injury, and fatal injury classes, the hybrid approach performed better then neural network, decision trees and support vector machines. For no injury and possible injury classes, the hybrid approach performed better than neural network. The no injury an possible injury classes could be best modeled by decision trees.

Previous researches had focus mainly on no injury and injury(including fatality) classes. In this paper, the research is done to possible injury, non-incapacitating injury, and fatal injury. The experiments showed that the model for fatal and non-fatal injury performed better than other classes. The ability of predicting fatal and non-fatal injury is very important since drivers' fatality has the highest cost to society economically and socially.

One very important factor of causing different injury level is the actual speed that the vehicle was going when the accident happened. The dataset doesn't provide enough information on the actual speed since speed for 67.68% of the data records' was unknown. If the speed was available, it might help to improve the models performance. From an intelligent systems point of view it is interesting to note about the failure of SVMs to model the complexity of the different injury classes.

[3] In the road accident prediction project use the dataset is in terms of values and some data is plain English word so, the numerical values data is easily predicted and also the calculation are easily done but, the normal word are display as it is or the non predicted data are drop in the table. So, This dataset are many columns and rows and all numbers of null values will be fulfill in forward fill method and also use the classification algorithm entire dataset.

In that classification algorithm we will use Logistic Regression Algorithm The logistic algorithm will make the prediction in terms of percentage, to find accuracy level in percentage and Error percentages. This Algorithm is only for the yes and no type of result or successful and unsuccessful. The equation for combinations of all 15 input variables.

The classification algorithm of the entire dataset. In the Road Accident prediction final result is to find the percentage of accident in particular area. Having lower number of features helps the algorithm to converge faster and increases accuracy. In the Road Accident prediction final result is to find the percentage of accident in particular area. Then we apply logistic regression on these features and obtain the least error.

[4] The information we get from this paper is regarding the results shown about the feature learning module identifying relational information between input variables and output feature representations. The proposed model that includes a MVNB regression layer in the supervised fine tuning module can better account for different patterns in crash data across injury severities and provide superior traffic crash predictions. In addition, the proposed model can perform multivariate estimations simultaneously with a superior fit. The model has superior performances in terms of prediction power compared to the deep learning model without a regression layer and the SVM model. The overall performances of the proposed model for all crashes shows an 89.824% RMSD improvement over the deep learning model without a regression layer and an 121.378% RMSD improvement over the SVM model.The findings also suggest that the proposed model is a superior alternative for traffic crash predictions. The proposed model can better account for heregeneity issues in traffic crash prediction.

## III. PROPOSED SYSTEM

Requirements analysis focuses on the tasks that determine the needs or conditions to meet the new or altered product or project, taking account of the possibly conflicting requirements of the various stakeholders, analyzing, validating and managing software or system requirements. Requirements analysis mainly deals with communication with users or customer. A Software Requirement Specification(SRS) is a comprehensive description of the intended purpose and environment for software under development. SRS minimizes the time and efforts requirement by developers to achieve desired goals and also minimizes the development cost.
The proposed system includes various steps which are as mentioned below:

A. Interface Requirements:

In the initial stage of the application, the system authorization is required for the users. This login process is same for all users. The will summit valid authorized username and password. In the below sections, user interface overview is described in a general manner.

The interface requirements includes the sign in, when the user will access the system webpage the user will face with login screen where he/she needs to access the website with the username password. In case of new users they must need to access it after registering to the website. Then comes the

uploading of the data sets. The user has to upload the dataset into the system database for analysis and prediction of accidents based on parameters. Then comes the data visualization process where the system visualize suggestion on the basis of uploaded data set system process data and predict the results. The user interface will show the suggestion and also shows in the form of graphical visualization. After the data visualization the user needs to sign out from the website whenever the use click on sign out button then all process of the system will be closed and the user will be redirected to login page.

B. System Design:

User Interface Layer: It basically comprises of interaction of the external user with the system , here the basic manifesto is composed by highlighting the registration module and login module. The login module is used by specific entities those who already have a authenticate account and the registration module is for the new user who wants to sign up to a visualized view of the generated report and select the specific parameters to make future prediction that will lead to reduction in loss of life.

Database Layer: At this layer data required to carry out analysis and to generate a visualized report behind the causality or fatal death of the individual. The data would be uploaded at real time basis too by the user with the help of user interface by user side.

Road Accident Analysis System: At this layer the data manipulation is mad, Basically it comprises of classification and prediction modules.

C. Classification:

Data Classification is the process of sorting and categorizing data into various types, forms or any other distinct parameters. Data classification enables the separation and classification of data according to data set requirements for various objectives. It is mainly a easy and also helps to neglect the features which does not have as much effect or response.

D. Prediction:

Prediction refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome. Prediction is been carried out using naïve Bayes, Decision Tree, K-Nearest Neighbours(KNN) and AdaBoost.

E. Objectives:

Analyzing the previously occurred accidents in the locality which will help us to determine the most accident-prone area and help us to set up the immediate required help for them. To make predictions based on constraints like weather, pollution, road structure, etc. To determine the main factors associated with road accident data analysis. Analyzing accident patterns of different types of accident on roads. To predict the root cause behind fatality.

## IV. SYSTEM ARCHITECTURE

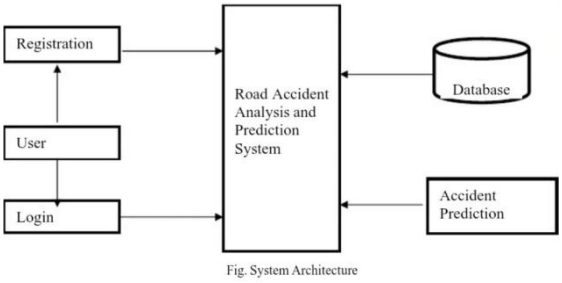


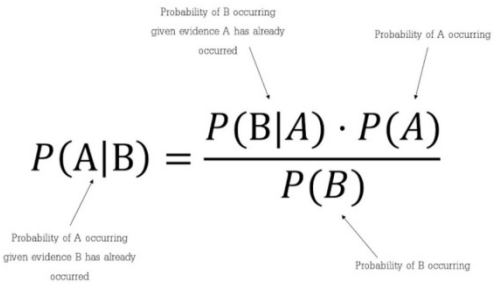

Fig. System Architecture

Basically the system architecture deals with the proper functioning of the algorithms to predict the correct and appropriate prediction. Basically the algorithms which deals with the road accidents prediction system using machine learning are KNN algorithm, Naïve Bayes Algorithm and the Decision Tree algorithms.

[A] Naive Bayes Algorithm:

Classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Bayes theorem provides a way of calculating posterior probability P(c|x) from P(c), P(x) and P(x|c). Look at the equation below:




Above,

P(A|B) is the posterior probability of class (A, target) given predictor (B, attributes).

P(A) is the prior probability of class.

P(B|A) is the likelihood which is the probability of predictor given class.

P(B) is the prior probability of predictor.

[B] KNN Algorithm:

1. Load the data.
2. Initialize K to your chosen number of neighbours.
3. For each example in the data.
   i) Calculate the distance between the query example and the current example and the current example from the data.
   ii) Add the distance and the index of the example to an ordered collection.
4. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances.
5. Pick the first K entries from the sorted collection.
6. If regression, return the mean of the K labels.
7. If classification, return the mode of the K labels.

[C] Decision Tree Algorithm:

Decision tress often mimic the human level thinking so its so simple to understand the data and make some good interpretations. Decision trees actually make you see the logic for the data to interpret(not like black box algorithms like SVM,KNN,etc..)

1. Pick the best attribute/feature. The best attribute is one which best splits or separates the data.
2. Ask the relevant question.
3. Follow the answer path.
4. Go to step 1 until you arrive to the answer.

## V. CONCLUSION

Many people lose their life due to road accident and also get injured.This project will help to reduce the possibility of accidents and increase the survival chances.This can be done by road accident analysis.The raw data contains the attributes of a particular accident spot and using the attribute we analyze the reason for accidents.Cleaning the raw data and eliminating the unwanted attributes will also help in decreasing the accidents. ANN technique lists out all the possible accidents so that the people have other options to choose the right way during their journey. The combination of artificial intelligence and the datasets from the previous accidents will help most of the drivers and the people who are using automobiles for the travel purposes. In comparison with other, Neural network is the best solution for the accident problems( such as prediction of road accidents). The application of ANN plays an important role in prediction of the road accidents based on the certain important factors and it also plays a major role in future evaluation of the predictions based on the present/past data.

## VI. ACKNOWLEDGMENT

I feel immense pleasure and privileged in expressing my deepest and most sincere gratitude to my guide Prof. Pallavi N, for her excellent guidance throughout the research work. Her kindness, dedication, hard work and attention to detail have been a great inspiration to me. I would particularly like to thank her for all her help in patiently and carefully correcting all of the manuscripts.

I also want to acknowledge the support received from my Computer Science & Engineering Department.

I am also very thankful to my family and friends, for their support and suggestions during the course of research paper work in the final year of my undergraduate course.

## REFERNCES

[1] Dia, H., & Rose, G., Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data. *TransportationResearch C*, Vol. 5, No. 5, 1997, pp.313-331.

[2] Yang, W.T., Chen, H. C., & Brown, D. B., Detecting Safer Driving Patterns By A Neural Network Approach. *ANNIE '99 for the Proceedings of Smart Engineering System Design Neural Network, Evolutionary Programming, Complex Systems and Data Mining*, Vol. 9, pp 839-844, Nov. 1999.

[3] Martin, P. G., Crandall, J. R., & Pilkey, W. D., Injury Trends of Passenger Car Drivers In the USA. Accident Analysis and Prevention, Vol. 32, 2000, pp.541-557.
[4] Mayhew, D. R., Ferguson, S. A., Desmond, K. J., & Simpson, G. M., Trends In Fatal Crashes Involving Female Drivers, 1975-1998. Accident Analysis and Prevention, Vol. 35, 2003, pp.407-415.
[5] M.Chang,L.Y.,&Chen,W.C.(2005)."Datamining of tree- based models to analyse freeway accident frequency". Journal of Safety Research, 36(4),365-375
[6] Tesema, T. B., Abraham, A., & Grosan, C. (2005). "Rule mining and classification of road traffic accidents using adaptive regression trees". International Journal of Simulation, 6(10),80-94.
[7] C. Dong, D. B. Clarke, S. S. Nambisan, and B. Huang, "Analyzing injury crashes using random-parameter bivariate regression models," *Transportmetrica A:Transport Science*,vol.12,no.9, pp. 794–810,2016.
[8] M.G.KarlaftisandE.I.Vlahogianni,"Statistical methods versus neural networks in transportation research: differences, similarities and some insights," *Transportation Research Part C: EmergingTechnologies*,vol.19,no.3,pp.387–399,2011.
[9] J.C.Principe,N.R.Euliano,andW.C.Lefebvre,*Neuraland Adaptive Systems: Fundamentals through Simulations*, vol. 672, Wiley, New York, NY, USA,2000.
[10] W.Huang,G.Song,H.Hong,andK.Xie,"Deeparchitecture for traffic flow prediction: deep belief networks with multitask learning," *IEEE Transactions on Intelligent TransportationSys- tems*,vol.15,no.5,pp.2191–2201,2014.
[11] Y.Lv,Y.Duan,W.Kang, Z.Li,andF.-Y.Wang,"TrafficFlow Prediction with Big Data: A Deep Learning Approach," *IEEE Transactions on Intelligent Transportation Systems*,vol.16,no.2, pp. 865–873,2015.
[12] N. G. Polson and V. O. Sokolov, "Deep learning forshort-term traffic flow prediction," *Transportation Research Part C: EmergingTechnologies*,vol.79,pp.1–17,2017.
[13] G. W. H. Organization, Ed., *Global Status Report on Road Safety 2018*. Geneva:WorldHealthOrganization,2018.[Online].Availabl e:https://www.who.int/violence injury prevention/road safety status/2018/en/
[14] SAAQ,"Roadsafetyrecord,"2018.[Online].Available:https ://saaq.gouv.qc.ca/en/saaq/documents/road-safety-record/
[15] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management(IJIM)*,vol.35,no.2,pp.137–144,2015.
[16] Government of Canada, "Open data 101," 2017. [Online]. Available:https://open.canada.ca/en/open-data-principles
[17] L.-Y. Chang, "Analysis of freeway accident frequencies: Negative bino- mial regression versus artificial neural network," *Safety Science*, vol. 43,no.8,pp.541–557,2005.
[18] https://www.wikipedia.org/