A Road Mishaps Analysis using Decision Tree and Random Forest Algorithms

S. Nandhini, V. Hima Bindu, Sanjit Yadav, Rajan Singh

Abstract: AI (ML) is the investigation of calculations and factual models that PC frameworks use to play out a particular activity without utilizing guidelines and depending on designs. It is communicated as subset of man-made brainpower. In this, the sample data is split into test set and the training set. Major drawback for the deaths in world is recorded by the road accidents. Most of the deaths are occurred in the middle-income countries. These studies result in finding the major factors for road accidents using decision tree and random forests. Decision tree is a choice help device that is a like a tree model which contains just control explanations. Random forest corrects the decision tree for overfitting to their training set. In this, the decision tree and the random forest algorithms are used to find the severity and the factors for the road-accidents using driver's personal information. Results conclude that the possibilities for the road accidents using the machine learning algorithms.

Keywords: Data Pre-processing, Feature Extraction, Feature Selection, Random forest, Decision tree, Test set, Training set.

I. INTRODUCTION

Most of the deaths are recorded due to road accidents that to mostly in the middle-income countries. Every human for his day to day work needs a vehicle. With the increase in the usage of vehicle the rate of accidents is also increasing. Motor vehicles are highly efficient means of transportation. So, with the increase in the demand for the vehicles the drivers and the owners are also increasing and the traffic rules are also being violated. Rash driving increases with the ease of going fast these factors leads to the high possibilities of causing road accidents. Many works have been done to take measures for controlling the road accidents. The reasons for the road accidents can be gender, age, wearing seat belt or not, increasing speed, alcohol consumption, over taking, exciding seating capacity and many.

Many works are being recorded using the machine learning techniques for finding the main causes for the road accidents.

S. Nandhini*, Assistant Professor, Department of ME CSE, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India. E-mail: nandhiniseec@gmail.com

V. Hima Bindu, Student, Department of CSE, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India. E-mail: hbindu597@gmail.com

Sanjit Yadav, Student, Department of CSE, SRM Institute of Technology, Ramapuram, Chennai, Tamil Nadu, India. E-mail: yadavsanjit837@gmail.com

Rajan Singh, Student, Department of CSE, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India. E-mail: rajansingh199911@gmail.com

In this the data sets are collected and are processed then the feature extraction and the feature selection are done. Then the data sets are split into the training set and the test set. To these sets the decision is first applied and next then the random forest is applied to the results of the decision tree and finally the results are interpreted.

The remaining paper is organised as follows. SECTION 1 conducts about the system architecture. SECTION 2 conduct about the data sets collection and data pre-processing. SECTION 3 conducts about the attributes for the cause of the accidents. SECTION 4 conducts about the implementation of the decision tree. SECTION 5 conducts about the usage of random forest to the decision tree and finally SECTION 6 about the conclusion.

II. RELATED WORKS

There are few works which deals about the road accidents. This worked [I] on presenting a cross breed AI approach for panning a sheltered direction in complex street traffic situations. A 3D-ConvNet, a variation of convolutional neural system, is utilized to take in spatiotemporal highlights from an arrangement of anticipated inhabit matrices produced from forecasts of other street traffic members.

In [II] the achievability of acoustic estimations and AI calculations for separating between dry, wet and soggy street surfaces was researched. The outcomes show that the characterization can be performed with moderately high exactness, practically identical to results. The sound sign was prepared with a sliding window of 100ms lengths which fitted well with the examining pace of GPS information. In [III] read a calculation for forecast of street abnormalities. Portrayed the BDS (Bumps Detection System) and calculations to screen the street surface conditions. It utilizes an accelerometer sensor for assortment of information and GPS for plotting the area of recognized potholes in Google map. K implies grouping calculation was utilized on preparing information and Random timberland classifier for approval on testing information. Advanced cell-based technique evacuates the need to sending extraordinary sensor in vehicle or at different street intersections. In [IV] proposed a novel calculation for consequently recognizing urban street sewer vent covers utilizing MLS point cloud information. The proposed calculation has been effectively applied to three chose MLS point cloud datasets. Computational execution investigation demonstrated that, by utilizing a multi-string figuring procedure. MLS point cloud information are a promising information hotspot for quickly and precisely estimating street surface highlights and bothers.

Retrieval Number: F4161049620/2020©BEIESP
DOI: 10.35940/ijitee.F4161.049620
Journal Website: www.ijitee.org

2067

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

In [V] gave a diagram of the condition of sign discovery. The location procedure has been part into division, include extraction and discovery. Ostensibly, the greatest issue with sign identification is presently the absence of utilization of open picture databases to prepare and test frameworks.

In [VI] presents a strategy for DD discovery and assessment while playing out an auxiliary errand. Discovery is executed by the AI calculation dependent on the ED figuring equation. Tackled a relapse issue in DD recognition and execution-based information combination into a solitary variable presented for the DD evaluation. A street portion is depicted by speed breaking point and street ebb and flow.

III. SYSTEM ARCHITECTURE

In this system architecture it describes about the outline of the whole
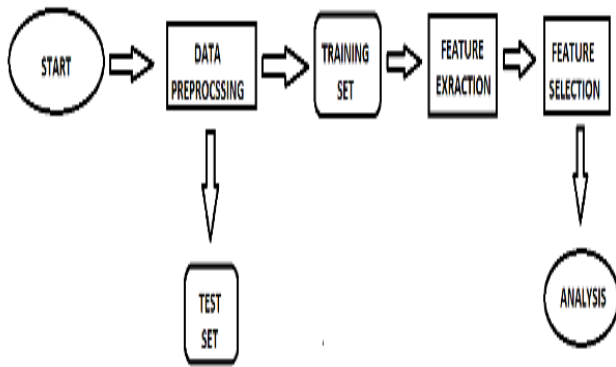
Fig 1 Block Diagram

In this architecture system, the data sets are collected from UK police and then the data pre-processing is done to these data sets collected. Then its spit into training set and the test set. After the sets are split to these sets the algorithms are implemented to find the factors for road accidents prediction. Total of 32 attributes and out of them 15 are considered for the further process. To the data sets of these 15 attributes the decision tree and the random forest algorithms are applied to find the analysis results.

IV. DATA PREPROCESSING

The data sets are collected from the UK police forces on a form names Stats19 from the year 2015-16. There are about thirty-two attributes given as the reasons for road accidents. Those are given in the table I. In this the results are declared as, if there is a possibility then it is given as negative and if there is no possibility it is given as positive.

The data sets are split randomly into training set for training the datasets and the test set for model validation. The ratio for split is taken as 75% and 25% respectively. Result interpretation and model validation are done on the testing dataset. After the data pre-processing is done the these are split into the test set and the training set, Then the feature selection is done to these sets for extracting the exact data since there might be some missing data.

V. ATTRIBUTES FOR THE CAUSE OF ACCIDENTS

Fig 2 Table

Variable	Description
A	Accident index
B	Location easting OSGR
C	Location Northing OSGR
D	Longitude
E	Latitude
F	Police force
G	Accident severity
H	Number of vehicles
I	Number of casualties
J	Date
K	Day of week
L	Time
M	Location authority (locale)
N	Location authority (thruway)
O	First class street
P	First number street
Q	Street type
R	Speed limit
S	Junction detail
T	Junction control
U	Second class street
V	Second number street
W	Pedestrian crossing-human control

Out of these thirty-two traits, 15 properties in particular number of vehicles, day of the week, time, area authority (locale), area authority (thruway), first class street, street type, speed limit, intersection detail, second class street, light conditions, climate conditions, street surface conditions, speed conditions at site and urban or country region. These qualities break down the variables for the street mishaps and anticipate the outcomes.

VI. DECISION TREE

To the training set and the test set the choice tree calculations are applied. The decision tree is a tree bolster model that resembles a tree model which contains just control explanations. In the wake of applying the choice tree calculation to the above sets as far as possible, street type, number of vehicles, climate conditions and urban or rustic zone are assimilated as the elements for the investigation of the street mishaps.

Fig 3 Improper roads leads to accidents

Fig 4 Cracks on the roads

VII . RANDOM FOREST TO DECISION TREE

After the outcomes acquired from the decision tree, we infer that coming up next are the foundations for street mishaps. To these above consumed factors, the irregular backwoods calculation is applied to locate the fundamental driver for the examination of the street mishaps since, arbitrary timberland revises the decision tree for overfitting the preparation set. After the outcomes acquired from the arbitrary backwoods it is examined that as far as possible, street type, climate conditions and urban or provincial zone assume more job for causing street mishaps than different elements referenced previously.

Fig 5 Increase in the speed limit has more possibilities of accidents.

Fig 6 Fog covering the views

Fig 7 Accident during rainy season

VIII . CONCLUSION

This work overviews the investigation of driver's hazard by the variables utilizing AI. The hazard factors are recognized by the choice tree and irregular woods calculations. Master experience is utilized in the proposed model for include extraction. Choice tree and the arbitrary timberland models are found out in which the investigation for the street mishaps is anticipated.

IX . FUTURE ANALYSIS

Future work may incorporate the accompanying insights (I) the informational indexes can be more and clear and building procedures ought to be built. (iii) it very well may be investigated all the more regularly utilizing the neural systems model. (iv) for a genuine model advancement cautiously structure the procedure.

REFRENCES

1. Amit Chaulwar, Micheal Botsch, Wolfgang, "A Hybrid Machine Learning Approach for Planning Safe Trajectories in Complex Traffic-Scenarios", "Mishap Analysis and Prediction" ,2016.
2. M. Kalliris, Olivier Haas, Stratis Kanarachos, Mike Blundell, R. Kotsakis, "AI Algorithms for Wet Road Surface Detection utilizing Acoustic Measurements", "Mishap Analysis and Prediction", 2019.
3. Manjusha Ghadge, Dheeraj Pandey, Dhananjay Kalbande, "Machine Learning Approach for Predicting Bumps on Roads", "Mishap Analysis and Prediction", 2015.
4. Yongtao Yu, Haiyan Guan, Member, IEEE, and Zheng Ji, "Automated Detection of Urban Road Manholes Covers using Mobile Laser Scanning Data", "Mishap Analysis and Prediction".
5. Andreas Møgelmose, Mohan Manubhai Trivedi, and Thomas B. Moeslund, "Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems", "Mishap Analysis and Prediction", 2012.
6. Andrei Aksjonov, Pavel Nedoma Vodovozov, Eduard Petlenkov, Martin Herrmann, "Detection and Evaluation of Driver Distraction Using Machine Learning and Fuzzy Logic", "Mishap Analysis and Prediction".
7. Jian Zhang, Zhibin Lil, Ziyuan Pu, Chengcheng Xu, "Comparing Prediction for Crash Injury Severity among Various machine Learning and Statistical Methods", "Mishap Analysis and Prediction", 2018.
8. Piotr Szczurek, Bo Xu, Ouri Wolfson, Jie Lin, "Estimating Relevence for the Emergency Electronic Brake Light Application", "Mishap Analysis and Prediction", 2012.
9. Ludovic Simon, Jean-Philippe Tarel and Roland Bremond, "Alerting the Drivers about Road Signs with Poor Visual Saliency", "Mishap Analysis and Prediction", 2009.
10. Md. Farhan Labib, Ahmed Sady Rifat, Md. Mosabbir Hossain, Amit Kumar Das, Faria Nawrine, "Road Accidents Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh", "Mishap Analysis and Prediction", 2019

AUTHORS PROFILE

S. Nandhini,Me Cse,Assistant Professor,Srmist Ramapuram-60009,Chennai,Tamil Nadu.

V. Hima Bindu Address: 54-14-41/1,Isukathota, Visakhapatnam,Andhra Pradesh530022 SRM institute of science and technology, ramapuram, Chennai, Tamil NADUBranch: Btech/ Cse

Sanjit Yadav Address: SA 14/67 a22 sarnath baraeipur sarnath varnasi pin code 221007..SRM institute of science and technology, ramapuram, Chennai, Tamil NADUBranch: Btech/ Cse .

Rajan Singh Address: LIG-2/144 Housing board colony near btm byepass jharsuguda,odisha-768202 SRM institute of science and technology, ramapuram, Chennai, Tamil NADU Branch: Btech/ Cse

Retrieval Number: F4161049620/2020©BEIESP
DOI: 10.35940/ijitee.F4161.049620
Journal Website: www.ijitee.org

2069

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication