iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IV **Month of publication:** April 2022

DOI: https://doi.org/10.22214/ijraset.2022.42010

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

An Enhanced Approach of Finding Probability of Road Accidents By using ANOVA

Sneha[1], Shubhi Gupta[2], Mariya Khurshid[3], Ranjeet Kumar[4]
[1, 2, 3, 4]*Department of Computer Science and Engineering, Meerut Institute of Engineering and Technology, Meerut 250005, U.P., India*

Abstract: ***As we know, thousands of accidents take place every day. These are very disturbing events which leads to major losses. There are various factors such as weather conditions, non-use of helmets, time period, type of accident, etc. due to which accidents take place. It is necessary to predict the areas which are more accident-prone. This model aims to predict such are considering all the above parameters using Machine Learning Concept i.e. EDA (Exploratory Data Analysis). It can further be improved and we can send the report of the accidents to the authorities like, hospitals, ambulance and insurance agencies and this will be very helpful in bringing down the accident fatality rates in the country****.***

I. INTRODUCTION

Road accidents causes a huge impact on the society. As the rate of accidents are increasing with the time, our attention has shifted towards road accidents analysis for determining the factors which significantly affects the accidents.

According to the reports given by World Bank, India accounts for 11% of death due to road accidents globally with only 1% of the world's vehicles.

[1] It is important to do analysis of Road accidents to find the causes of accidents and severity of injuries. Many Engineers and researchers have tried to design a model but road accidents are unavoidable. [2]

Here, we will study and analyse the datasets that are containing details of different states based on various factors such as weather conditions, non-use of helmets, time period, type of accidents, etc. and develop a model that will helps us in preventing road accidents by Exploratory Analysis of Data which will discover the patterns and graphical representations to understand the data better.

The collected data will be analysed and will combine the data on the various parameters using the best algorithm. Through this estimation we can analyse and identify the flaws and reasons of the accidents.

This will be very helpful in making roads so that we can avoid the problems which have been faced earlier. Through this model (Road Accidents Analysis Using Exploratory Data Analysis) we can make roads more secure and accident free**.**[7]

II. RELATED WORK

1) *S. Shanthi:* Introduced a technology based on gender classification i.e. data mining classification, to provide high precision results they used AdaBoost Meta classifier in Rnd Tree and C4.S. The Fatal Analysis Reporting System (FARS) provided The Critical Analysis Reporting Environment (CARE) used by the training data set.[13]
2) *Tessa K. Anderson:* Introduced a method of clustering technique that determines stochastic indices which exist in some clusters and can be compared in time and space to identify the high densities accident hotspots. To create the basic spatial unit of the hotspot clustering method, the kernel density estimation tool enables the visualization and manipulation of density based event.[12]
3) *Sachin Kumar:* To detect the high frequency accident locations and the root factors that effects road accidents on these locations they used data mining techniques. To identify the relationship between distinct attributes which are in accident data set and to know the characteristics of location the Association Rule mining Algorithm is
4) applied.[14]
5) *Ali Moslah Aljofey:* In this paper, they analysed the factors which are responsible for the road accidents on highway locations regularly and tries to Analyse ways to reduce the cause of accidents in such areas. The proposed framework consists of a k-means clustering technique to get the frequencies of highway location accident.[9]

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

III. PROPOSED METHODOLOGY

In our project we have used Exploratory Data analysis. Exploratory data analysis is an approach of analysing data sets to summarize their main characteristics. It is a critical process of performing initial investigations on data. Used to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

The exploratory data analysis steps that analysts have in mind when performing EDA include:

1) Asking the right questions related to the purpose of data analysis
2) Obtaining in-depth knowledge about problem domains.
3) Setting clear objectives that are aligned with the desired outcomes.

A. Exploratory Data Analysis

EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset. It generates graphical representation and summary of statics for numerical data to understand the dataset.

B. ANOVA

Analysis of Variance is a statistical technique which is used to check if the means of two or more groups are significantly different from each other. It shows the impact of one or more factor by comparing the means of different samples.[8]
The formula for ANOVA is:
F = MST/MSE
MST = SST/ p-1
MSE = SSE/N-p
SSE = $\sum (n-1)\ S^2$
Where, F = ANOVA Coefficient
MSB = Mean sum of squares between the groups
MSW = Mean sum of squares within the groups
MSE = Mean sum of squares due to
error SST = total Sum of squares p = Total number of populations n = The total number of samples in a population
SSW = Sum of squares within the groups
SSB = Sum of squares between the groups
SSE = Sum of squares due to error
S = Standard deviation of the samples
N = Total number of observations
Typical Graphical Techniques Used in EDA are: Box plot, Histogram, Multi-variate chart, Run chart, Pareto chart, Scatter plot, Parallel coordinates, Odds ratio.
ANOVA is used to calculate the impact of different parameters.
There are some steps in this method:
1.First All the required data is collected.
Datasets are:

1) Year-wise data set (1970-2017) 2. state-wise data set in 2017 based on: severity of accidents, non-use of helmet, responsibilities of driver, weather condition, time period in a day, type of accidents [4]
2) Import all the required libraries such as pandas, Matplotlib, NumPy, etc.
3) Merge all the data set.
4) Correlation is check between different feature of Data set.
5) Perform Exploratory Data Analysis to get the summary and pattern of numerical data in data set.
6) Apply ANOVA to get the mathematical Value which shows the impact of different Factors
7) At last, we get the features which are more responsible for Road accident Rate.Then enter the value for these parameters in the form of 0 and 1.
8) Calculate Percentage of no. of One's, this percentage is the probability of Accident.
9) At end we get the probability of Accidents.

©IJRASET: All Rights are Reserved

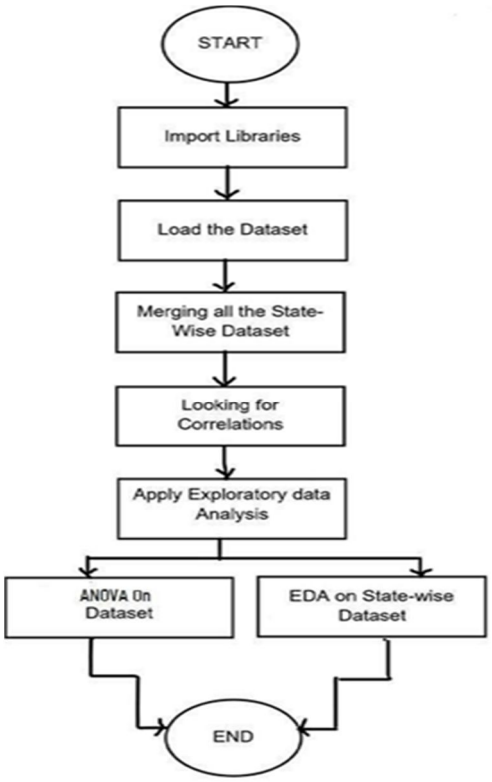


Figure 1: Architecture of the proposed methodology

IV. RESULT

1) Average Fatal accidents in all the states of India is around 3744.

	Fatal Accidents	Total Accidents	Persons killed for not using helmet	Intake of Alcohal	Exceeding lawful speed	Jumping Red Light	D w
count	36.000000	36.000000	36.000000	36.000000	36.000000	36.00000	
mean	3744.333333	12914.166667	585.972222	170.305556	2052.666667	35.00000	1
std	4565.825233	16999.059397	979.881844	460.154683	2893.060959	74.06908	2
min	0.000000	1.000000	0.000000	0.000000	0.000000	0.00000	
25%	106.000000	524.000000	2.750000	6.500000	39.500000	0.00000	
50%	2019.500000	5948.500000	141.500000	26.500000	438.500000	3.00000	
75%	5426.750000	19838.750000	722.250000	154.500000	3360.250000	24.25000	2
max	17706.000000	65562.000000	4620.000000	2716.000000	11444.000000	342.00000	12

Figure 2: Information of different parameters

2) Average Total accidents in all states are around 13000.

3) Average 585 people dies for not wearing helmets which is accompanied by average 2052 deaths due to exceeding lawful speed.

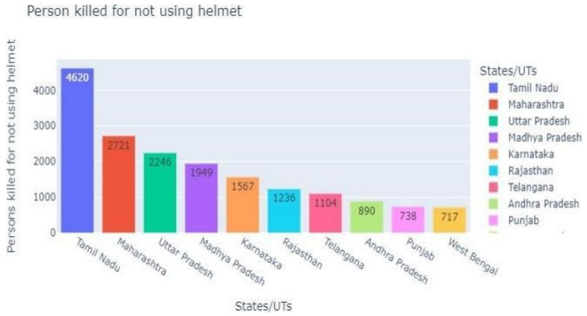


Figure 3: Person killed on the basis of Not using Helmet

a) Uttar Pradesh, Tamil Nadu, Madhya Pradesh, Maharashtra and Karnataka are the top 5 states where most fatal accidents occurred in year 2017.

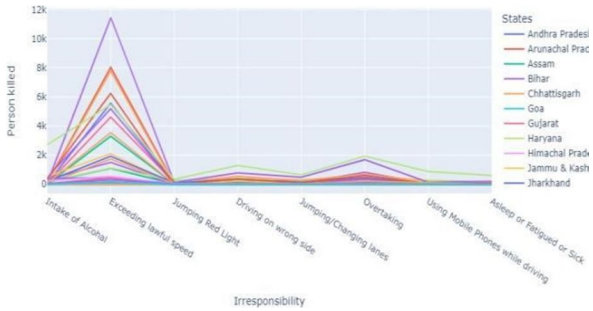


Figure 4: Person killed on the basis of Irresponsibility

b) Most of the accidents Occurred in Sunny/Clear weather that too in afternoon and evening time period.

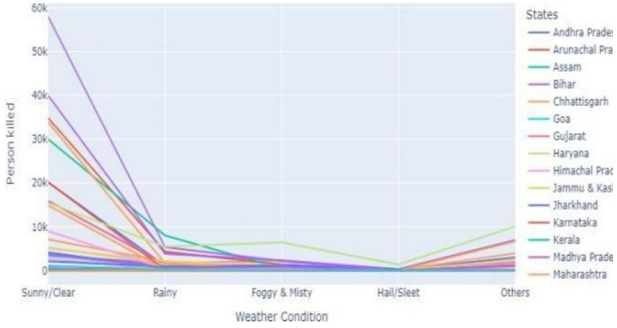


Figure 5: Person killed onthe basis of Weather Condition

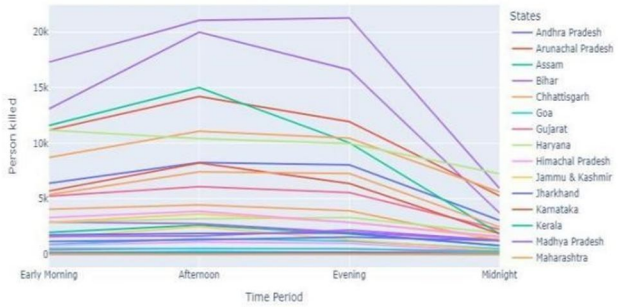


Figure 6: Person killed on the basis of Time Period

c) And lastly, in Karnataka 4588 people were killed by "hitting from back" and in Tamil Nadu 3507 pedestrian were killed. While, Uttar Pradesh records the most hit and run death cases nearly 3300 in the year 2017.

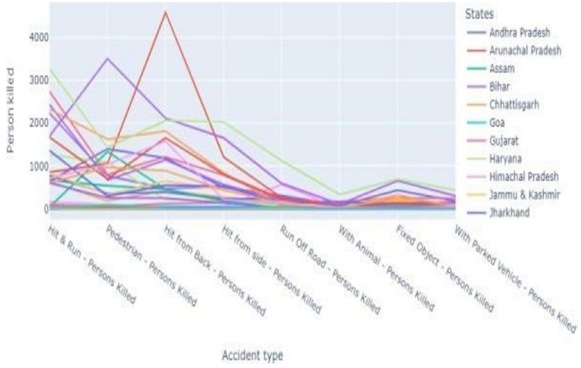


Figure 7: Person killed on the basis of accident type

d) If chances of any 2 features among Midnight, Overnight, exceeding lawful speed, driving on wrong side, it from side is true then probability of Accident is 0.4.

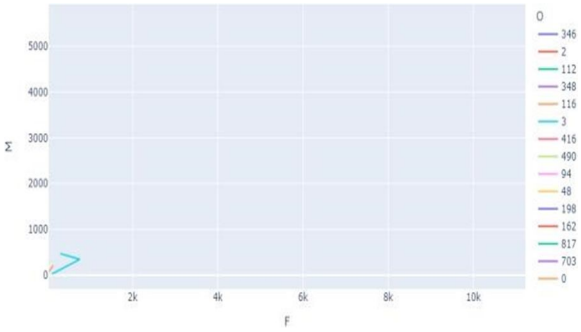


Figure 8: ANOVA defining correlation between the parameters

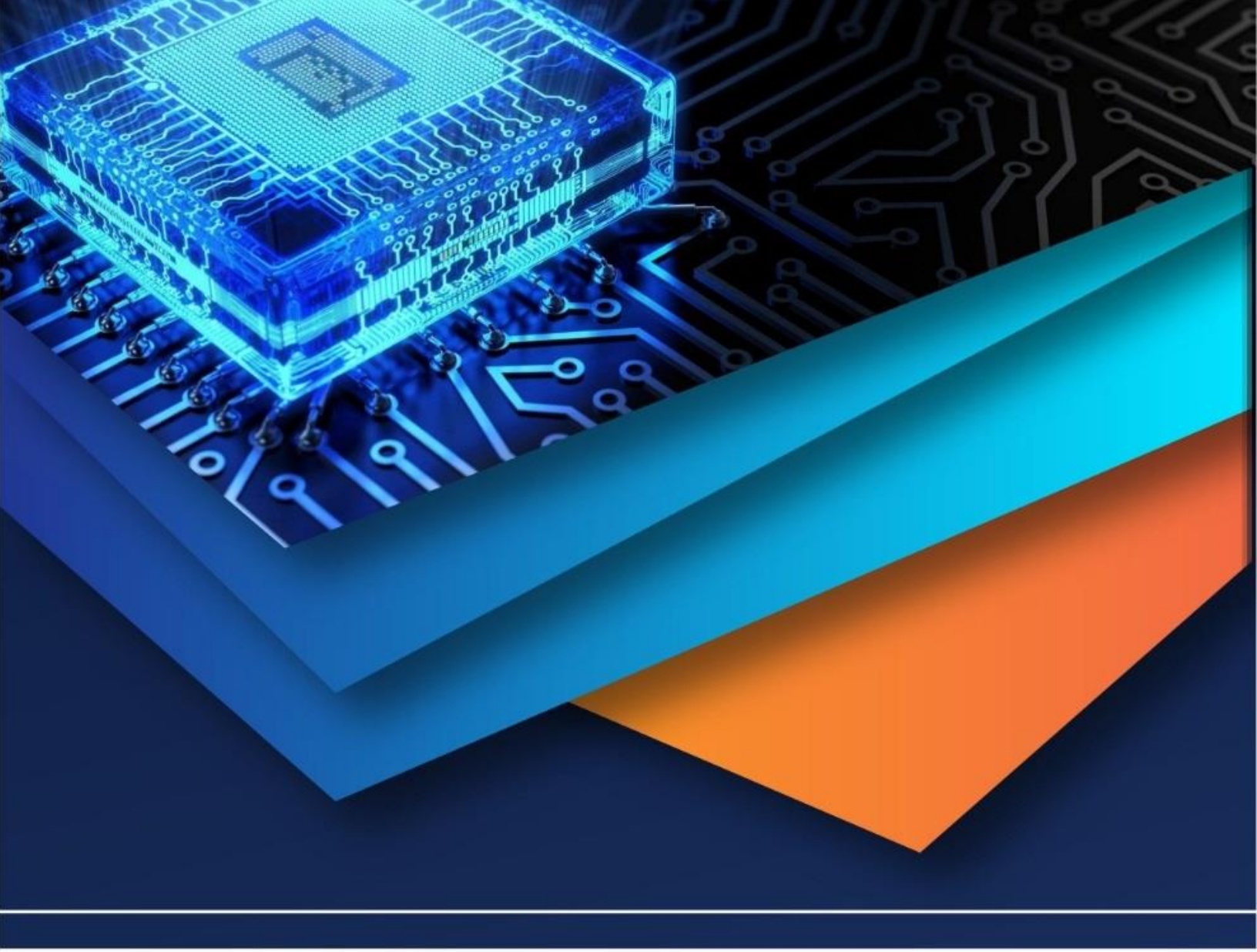
V. CONCLUSION

This model of Road Accident Analysis gives us many factors, controlling these factors can leads to decrease in accident rate. The Model uses Exploratory Data Analysis which give pattern in Dataset. There are two different Dataset i.e. state wise and year-wise.[4] When EDA is applied on Year-wise Dataset, it shows that the accident rate increases as per the year. When EDA is applied on State wise Dataset, it shows total accidents in all the state is 1300 average. Tamil Nadu is the state having more accident rates and Uttar Pradesh, Madhya Pradesh, Maharashtra and Karnataka are those states where most fatal accidents occurred in year 2017 due to exceeding lawful speed. Using ANOVA, we have seen that

Midnight, Overnight, exceeding lawful speed, driving on wrong side, hit from side, which have high impact on Fatal Accident rates. So, these features are chosen for finding the Probability of Accidents. User is asked to enter all the 5 parameters value as 1 or 0. If any 2 are true then there is 0.4 probability of Accident. This algorithm gives 95% efficiency and its performance is 1 min. This is the result of this Road Accident Model.

REFERENCES

[1] Chitnis S.D., P Gokhale, Statistical Analysis and Classifier Accuracy Improvements Models for Road Accident Issues on National Highways in India, Design Engineering, Issue 7, ISSN No. 0011-9342, Pages 55765591, 2021.
[2] Sahil Dabhade, Sai Mahale, Avinash Chitalkar, Pushkar Gawhad, Vicky Pagare, Road Accident Analysis and Prediction using Machine Learning, International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 23219653; Volume 8 Issue I, Jan 2020
[3] Ram Prasanth T, Spanglar Diaz V, Surendran N, Udhayavel V, Dr. C. Anand, Mrs. N. Vasuki, A Statistical Analysis of Road Accident Data-Using Data Mining Techniques, ISSN -2278-0181, Volume 6, Issue 08, 2018
[4] Road Accidents Data of Government of India, 2018 ,https://data.gov.in/catalo g/road- accidents-india2018?filters%5Bfield_catalog_r eference
[5] ROAD SAFETY FACTS https://www.asirt.org/safetravel/roadcrashes/
[6] Inderpreet kaur, Ashish Kumar Luhach, Pooja, PROPOSING K-MODE BASEDMETHODOLOGY FOR ROAD ACCIDENTSWITH IMPROVED APRIORI, International Journal of Engineering Applied Sciences and Technology, Vol. 2, Issue 5, ISSN No. 2455-2143, Pages 66-74, 2017
[7] Cloud analysis data of IBM on Road Accidentshttps://www.ibm.com/cloud/learn/explor at ory-data-analysis
[8] Gurchetan1000 — A Simple Introduction to ANOVA (with applications in Excel)January 15, 2018 -https://www.analyticsvidhya.com/blog/2018/01/anova-analysis-of-variance/
[9] Ali Moslah Aljofey, Khalil Alwagih, Analysis of Accident Times for Highway Locations Using K-Means Clustering and Decision Rules Extracted from Decision Trees, International Journal of Computer Applications Technology and Research Volume 7- Issue 01, 001011 , 2018, ISSN:-2319-8656
[10] Md. Farhan Labib, Ahmed Sady Rifat, Md. Mosabbir Hossain, Amit Kumar Das, Faria Nawrine, Road Accident Analysis and Prediction of Accident Severity by using Machine Learning in Bangladesh, International Conference on Smart computing and communiction, Issue 7, 2019.
[11] Muthusamy A P, Rajendra M, Ramesh K, Sivaprakash P, A Review on Road Traffic and Related Factors International Journal of Applied Engineering Research, Vol. 10, Issue 7, ISSN No. 0973-4562, Pages 2817728183, 2015.
[12] Tessa K. Anderson, Kernel density estimation and K-means clustering to profile roadaccident hotspots, Accident Analysis and Prevention Elsevier, 41 (2009) 359–364.
[13] Shanthi Selvaraj, Geetha Ramani, Classification of seating position specific patterns in road traffic accident data through data mining techniques, Second International Conference on Computer Applications, Pondicherry.
[14] Sachin Kumar and Durga Toshniwal, A data mining framework to analyse road accident data, Kumar and Toshniwal Journal of Big Data (2015) 2:26 DOI10.1186/s40537-015-0035-y



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)