International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 07 | July -2020 ISSN: 2582-3930

A Framework for Analyzing Road Accidents Using Machine Learning Paradigms

Priya Mandhyan[1] & M. Arvindhan[2]

[1]Priya Mandhyan School Of Computer Science and Engineering & Galgotias University
[2]M. Arvindhan School Of Computer Science and Engineering & Galgotias University

---

Abstract - Globally, amongst all groups, one among the leading explanation for deaths has been road accidents since a couple of years and are predicted to be seventh leading explanation for death by 2030 (WHO, 2020). It's entirely admissible and saddening to let citizens get killed in road accidents. As a result, to handle this sort of situation, an in depth analysis is required. Here, we are trying to find the transparent effect of road accidents and their severity level. In this research paper, we are going to analysis the strength of road accident and we will try to provide some beneficial suggestion. Machine learning algorithms will help us to extract the data and hidden patterns that will helps us to form the policies for the prevention from road accidents. During this research we are going to use clustering, k-means clustering and feature selection algorithm. Here we are using k-means clustering that helps to segment road accidents data and further using the feature selection for accurate predictions.

Key Words: Clustering; Feature Selection; K-means Clustering; Machine Learning; Road Accidents.

1.INTRODUCTION *(Size 11, Times New roman)*

This Road accidents are the foremost unfortunate and unforeseen thing occur to a road user .Road accidents data are the essential measure of safety with the assistance of which we will establish the size and therefore the nature of road safety problems. Therefore, accidents database is that the key think about the management of road safety. The measure things that help us to find the factors of road accidents is the data set. The country data are not accurate either they are lack of useful information or they have not strong format system. For the analysis, that data set must be accurate. So the first thing is strong format system and full information. Lack of information can cause the poor analysis of system which leads to wrong results.

Different parameters have different effect on road accidents. The most difficult thing for the analysis is the dissimilarity. Therefore, the segmentation is required. We can measure the dissimilarity with help of analysis of the given information and finding the connection between this dissimilarity that can help us to find the hidden pattern required for analysis of the road accidents.

Machine learning algorithms helps us to find the hidden pattern and extract the useful information. Clustering helps to group the data on the basis of similarities. Clustering can helps to segment the road accidents data and find the beneficial suggestions.

2. EXISTING WORK

Road accidents are not the problem of today. This problem has been increasing as the number of vehicles are increasing now a days. Here are some related work of researchers who had came up with new solutions to this problem.

In this research paper the author has used the data mining techniques. Firstly the author has pre- processed the data to find the locations with high frequencies of road accident and the analyzed these data to find the factors that are impacting on that locations. After that for using the K- means clustering algorithm the author divided the accident locations into k groups. Then, he has used the association rule to find out the connection between the different attributes with the help of which he exhibited that different locations have different accident frequencies. And with this approach the author was able to find some hidden information which can be used for prevention of road accidents (Kumar, 2016).

In this paper , the author has proposed the vision based method to analysis the road traffic data with the help of which he can learn the traffic pattern that help him to extract the velocity of vehicles and its distances. The author has used here the PPNN i.e., Parzen Probabilistic Neural Network for machine learning. He concluded that his proposed approach show approximately 85% accuracy in detecting special situation.(Elahi, 2014).

In this research paper, the author has analyzed the situation of occurrence of road accident with the help of machine learning algorithms . Here the author has used different algorithms of machine learning like CART, Naive bayes, ROC value etc. The author has concluded that by applying the CART algorithm he achieved 81.5% of accuracy. (Bulbul, 2016).

In this research paper, the author has done analysis to identity the strength of roach accidents in Bangladesh by using machine learning algorithms. Here the author had tried to find out the factors which have impact on road accidents and using that factors he tried to give us some beneficial suggestions regarding rod accidents. Here the author has used the four supervised machine learning techniques namely Decision Tree, K-Nearest Neighbour(KNN), Naive Bayes and AdaBoost. The author has used these algorithms to classify the severity of accidents into four categories that are Fatal, Grievous, Simple Injury and Motor Collision. Amongst these algorithms, AdaBoost algorithm achieved best performance. (M. F. Labib, 2019).

3. MATERIALS AND METHODS

There are mainly three types of machine learning algorithms. They are supervised learning, semi- supervised, unsupervised learning and reinforcement learning [8]. Among these three approaches unsupervised learning has been utilized in this

research paper. Here, the three popular machine learning algorithms has been used namely, Clustering, K-means Clustering and have Selection. Figure1. It represents the working processes of proposed model.

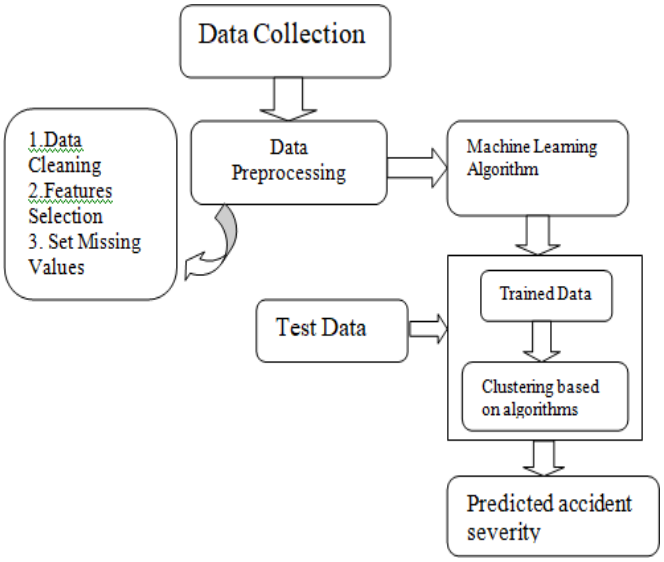


Fig -1: working processes of working model

3.1 Dataset Preparation

Extensive and accurate data records of road accidents are the foremost important got to recover performance by using machine learning algorithms. But getting an accurate and an ideal data set isn't a simple task, its quite challenging.

Therefore, to process the info supported the necessity we are using two following steps. They are:

3.1.1 Data collection

The most important aspect for any data analysis is the data. It is very important to collect the right kind of data. Special attention is needed for analyzing and understanding the content and structure of the data. This data set is a cleaned version of the Killed or Seriously Injured (KSI) traffic accident reports from the City of Toronto Police Open Data portal. It contains the information for all traffic accidents reported between 2007 and 2017. There is data on the time, location, and the type of incident with various attributes about the traffic conditions at the time of the incident.

3.1.2 Data Pre-processing

Data preprocessing is the important steps. It is the technique of data mining. Here we will try to remove the dimensionality of the data by encoding mechanism

The table below contains the attributes name and its description that are used in this research.

Table -1: Attributes name and its description.

Attribute	Attribute Description
ACCNUM	The total number of accidents.
YEAR	The year in which accidents took place.
MONTH	Month in which the accidents took place.
DAY	Day on which the accidents took place
HOUR	Hour in which the Accident (24hrs)
MINUTES	Minute in which the Accidents took place.
WEEKDAY	Weekday on which accidents took place (0 is Monday)
LATITUDE	Latitude
LONGITUDE	Longitude
Ward_Name	City Ward
Ward ID	City Ward ID
Hood_Name	Neighborhood Name

And many more attributes used in the dataset.

There are following steps taken in data pre-processing. They are:

3.1.2.1 Data cleaning

Data Cleaning is the process of clean or removing the noisy, incomplete and unwanted data. It helps us to modify and remove the data which is not useful or relevant for our analysis. This data is not important or helpful for analyzing because it may provide inaccurate result or it can hinder the process. There are different method for cleaning the data its depend upon how it is sorted.

Here Data is been cleaned by replacing blank value by NA and dropping columns with large amounts of missing value and by changing its data types like from int to string as per the analysis.

3.1.2.2 Feature Selection

Working with sizable amount of features can affect the performance. Also, it's going to contain the danger of over fitting with the amount of features. Therefore, for relevant or appropriate outcomes the feature selection technique is used.

Feature selection is basically used to reduce the dimensionality. It allows machine learning algorithms to train the dataset faster.

4. PROPOSED MODEL

In this section we are going to explain how the dataset has been pre-processed and the algorithm used to get the result.

In this paper we are going to have the data set of the Killed or Seriously Injured (KSI) traffic accident reports from the City of Toronto between 2007 and 2017.

4.1 Data cleaning:

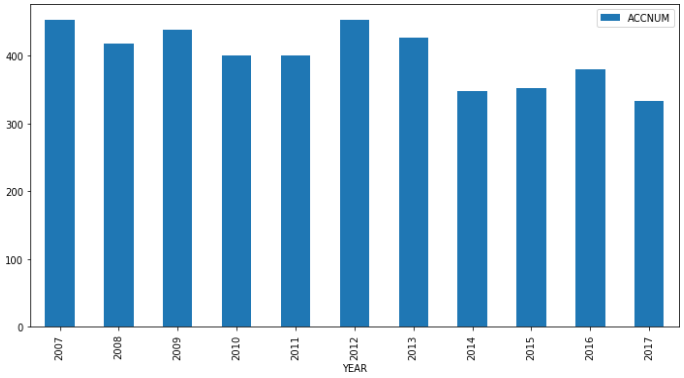
Usually the dataset will contain noisy, missing data values, which is going to affect the quality of the result. So we need to pre-process the raw set of data to improve the clustering

process. Here Data is been cleaned by replacing blank value by NA and dropping columns with large amounts of missing value and by changing its data type.

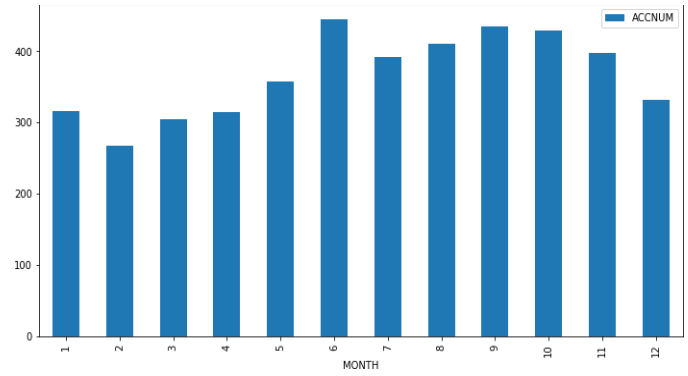
4.2 Data exploration

Here we will try to explore the data by analyzing the accident numbers against month and year.

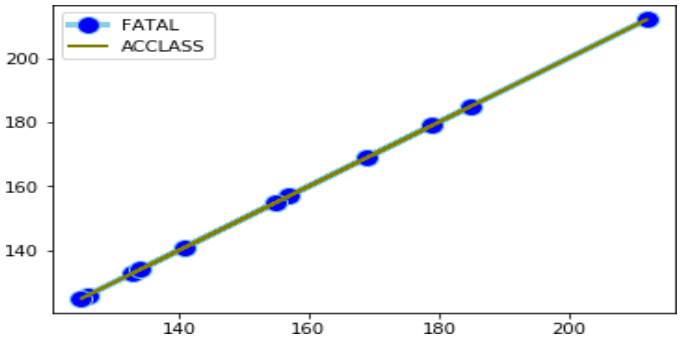
Total number of incidents has reduced slightly over the years.



From the below data, accidents happened more from June to October



Now we will look for the target column target column FATAL therefore we will analyze the ACCLASS VS Fatal VS injury .ACCLASS is columns that classified into 3 categories while fatal only show whether it is fatal or not.



4.3 Data Modeling:

Here we will analysis the data on the basis of categories like District category, Light Category, Visibility category and Road Surface Condition Category for applying the cluster analysis.

4.3.1 Cluster Analysis using K-means Clustering

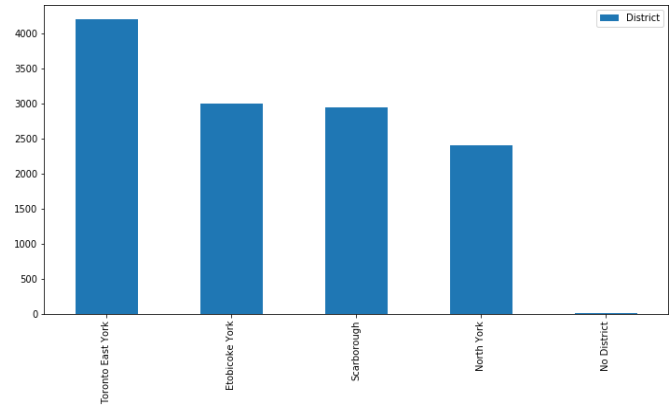
K-means clustering algorithm is the unsupervised machine learning algorithm that groups the data items into k cluster. It helps to solve the clustering problems in easy way. It help us to group the data on the bases of similarity as well as keeping it different as far as possible.

4.3.2 Feature Selection

Feature selection and cluster analysis was employed to reduce dimension and optimized for best modeling score From feature selection, column like accnum, year, month, hoodID or name, weekday, minute and hour has great impact to modeling score and relation to fatal injury. But accnum, minute and hour did not has a lot business sense in real situation and in this modeling, thus was not selecting for final modeling.

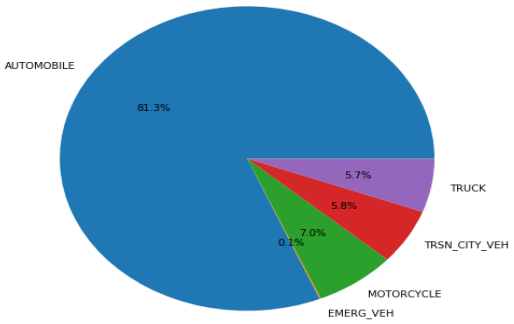
5. RESULT AND DISCUSSION

In this modeling, several graphs were used to explore relations inside the dataset. Column 'FATAL' was chosen as target output. Here we have taken different attributes for analysis and comparing the situations in which



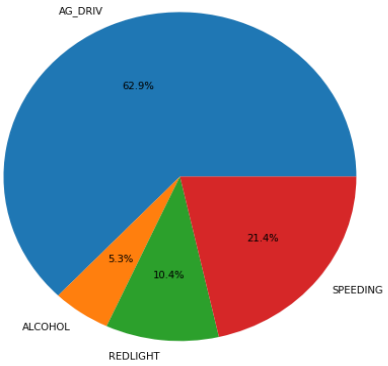
Region-wise, Toronto east region was reported to be the region with most accidents ending up in fatal injury as showed in figure.

Automobiles have been pretty consistent reason of accident over the years involving aggressive driving and pedestrians



Vechile type VS Accidents in Ontario in last 10 years(%age)

AG_DRIV (Aggressive and Distracted Driving) are the major cause of accidents (62.9%), speeding accounts for 21.4%, red-light is 10.4% and alcohol is 5.3 %.



Driving condition VS Accidents in Ontario in last 10 years(%age)

From k –means cluster method, 4 clusters was used to map the hoodname for final modeling to present which region result in more fatal injury and what is the cause behind it. In general, the results are not very significant that can demonstrate big reversal of the cause. But relative safer zone or dangerous area can be identified.

6. CONCLUSION

From this research we concluded that different attributes have different impact on road accidents, Like, using these datasets we found that tornado east region was reported more fatal to injury, automobiles is consistent reason of accidents involving aggressive driving and speeding.

In general, aggressive driving and speeding contribute greatly to fatal injury. But in dangerous neighborhood, dark light was found be to causing issues compared to other region.

ACKNOWLEDGEMENT

I am whole heartedly thankful to my guide Prof. M. Arvindhan for his assistance & valuable suggestions on this study. His advices and support has, throughout, been inspirational. I am extremely thankful to my parents as their support during the development of this study was unfathomable. I would also like to express my very great appreciation to my special friend Aryan Singh for his abysmal support & adulation in this period of time.

REFERENCES

1. A. Hébert, T. G. (2019). High-Resolution Road Vehicle Collision Prediction for the City of Montreal. 2019 IEEE International Conference on Big Data (Big Data) (pp. 1804-1813). Los Angeles, CA, USA: IEEE.
2. Arvindhan, M. (2019). CLUSTERING ALGORITHM NETWORKS TEST COST SENSITIVE FOR. ICTACT JOURNAL ON IMAGE AND VIDEO PROCESSING , 2098-2102.
3. B. Kumeda, F. Z. (2019). Classification of Road Traffic Accident Data Using Machine Learning Algorithms. 2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN) , 682-687.
4. Bulbul, H. &. (2016). Analysis for Status of the Road Accident Occurance and Determination of the Risk of Accident by Machine Learning in Istanbul. 15th IEEE International Conference on Machine Learning and Applications , 426-430.
5. E. Bruehning, D. O. (2005). 30 years in-depth accident studies for improving traffic safety. Zeitschrift fuer Verkehrssicherheit , 175-181.
6. Elahi, M. M. (2014). Computer Vison Based Road Traffic Accident and Anomaly Detection in the Context of Bangladesh. INTERNATIONAL CONFERENCE ON INFORMATICS, ELECTRONICS & VISION 2014 , 1-6.
7. G. Parathasarathy, T. R. (2019). Using hybrid Data Mining algorithm for Analysing road accidents Data Set. 2019 3rd International Conference on Computing and Communications Technologies (ICCCT) , 7-13.
8. Kasbe, A. V. (2017). A review on road accident data analysis using data mining techniques. International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS) (pp. 1-5). Coimbatore: IEEE.
9. Krause, S. &. (2019). New Insights into Road Accident Analysis through the Use of Text Mining Methods. 2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). , 1-6.
10. Kumar, S. &. (2016). A data mining approach to characterize road accident locations. Journal of modern transportation , 62-72.
11. M Arvindhan, A. A. (2019). Scheming an Proficient Auto Scaling Technique for Minimizing Response Time in Load Balancing on. International Conference on Advances in Engineering Science Management & Technology (ICAESMT) , 1-8.
12. M. F. Labib, A. S. (2019). Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh. 2019 7th International Conference on Smart Computing & Communications (ICSCC) (pp. 1-5). Sarawak: IEEE.
13. Pereira, F. &.-A. (2013). Text analysis in incident duration prediction. Transportation Research Part C: Emerging Technologies. , 177-192.
14. Prakash, M. A. (2020, january). Data Mining Security over DDOS Attacks. ICTACT Journal on Soft Computing , 2061-2065.
15. R. E. AlMamlook, K. M. (2019). Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity. 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT). (pp. 272-276). Amman, Jordan: IEEE.
16. Sanjai, P. R. (2018, jan 10). the Economics time. (economics times) Retrieved march 24, 2020, from economics india: https://economictimes.indiatimes.com/news/politics-and-nation/400-deaths-a-day-are-forcing-india-to-take-car-safety-seriously/articleshow/62439700.cms
17. Suman, S. S. (2018). A Framework for Analysis of Road Accidents. 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR) , 1-5.
18. T. Beshah, D. E. (2012). Learning the Classification of Traffic Accident Types. 2012 Fourth International Conference on Intelligent Networking and Collaborative Systems. (pp. 463-468). Bucharest: IEEE.
19. Vijayarani, E. S. (2017). Analysis of road accidents in India using data mining classification algorithms. 2017 International Conference on Inventive Computing and Informatics (ICICI) (pp. 1122-1126). Coimbatore, India: IEEE.
20. WHO. (2020, FEB 7). world health organization. Retrieved march 31, 2020, from who.int: https://www.who.int/gho/road_safety/mortality/number_text/en/