

Road Accident Analysis using UK Dataset 2016

Chanchal Sharma, Mr. Anil Wadhwa

Department of Computer Science and Engineering, Rao Pahlad Singh Group of Institutions, Balana, Mohindergarh

Abstract: - Roadway traffic wellbeing is a noteworthy worry for transportation representing offices and in addition customary nationals. Information Mining is removing from concealed examples from colossal database. It is generally utilized as a part of an advertising, observation, misrepresentation location and logical revelation. In information mining, machine learning is mostly engaged as research which is consequently learnt to perceive complex examples and settle on keen choices in light of information. Globalization has influenced numerous nations. There has been an extreme increment in the financial exercises and utilization level, prompting extension of movement and transportation. The expansion in the vehicles, traffic prompt road accidents. Thinking about the significance of the road wellbeing, government is endeavoring to distinguish the reasons for road accidents to decrease the accidents level. The exponential increment in the accidents information is making it hard to break down the imperatives causing the road accidents. The paper portrays how to mine regular examples causing road accidents from gathered informational collection. We discover relationship among road accidents and anticipate the kind of accidents for existing and additionally for new roads. We make utilization of affiliation and order principles to find the examples between road accidents and also foresee road accidents for new roads.

Keywords: Data mining, Association rule, Classification rule, Apriori algorithm, Naïve Bayes algorithm

I. INTRODUCTION

There are a ton of vehicles driving on the roadway consistently, and traffic accidents could occur whenever anyplace. Some accident includes casualty, implies individuals kick the bucket in that accident. As person, we as a whole need to evade accident and remain safe. To discover how to drive more secure, information mining procedure could be connected on the traffic accident dataset to discover some significant data, consequently give driving recommendation. Accidents occurred because of the carelessness of driving vehicle on the roads. There are different reasons in charge of the accident like forsake of traffic manages however road conditions and the traffic are viewed as the one of prime reason for casualty and causality over the globe. These accidents happen because of dynamic plan and improvement of car ventures. A traffic crash occurs due specific reasons like crushes of two vehicles on road, strolling individual, creature, or some other regular deterrents. It could bring about damage, property harm, and demise. Traffic accident

examination required investigation of the different factor influencing behind them. In review it's seen that inexact 1.2 million passing and 50 million wounds assessed worldwide consistently. The inexact estimation of causality and wounds because of poor road foundation is a major test before the living creatures. The request to manage the issue, in computational science, we can receive information digging model for various situation. In any vehicle accident, it learns about the driver's conduct, road framework and potential outcomes of climate gauge that could be some place associated with various accident occurrences. The fundamental issue in the examination and investigation of accident information is its blend heterogeneous condition and information division which is utilized generally to beat accident issue. [2,5,7] Information Mining is a computational strategy to manage expansive and complex informational collection and these informational collections can be of ordinary, ostensible and blended. It is very simple to use in assortment of space have a place with science and administration; additionally, it could be utilized as a part of misrepresentation ID and numerous more logical cases and in accident seriousness issue. Segment of articles in a gathering of groups or in a homogeneous set is a major activity of information mining. Bunching is a strategy to segment questions in a comparative gathering. The k-implies calculation having a decent proficiency for bunching vast informational indexes yet confined in shaping groups for genuine word information while working just on numerical information since it helps in diminishing the cost work by modifying the importance of the groups [1,3]. Information mining procedure is perceived as dependable strategy for examination of traffic accident seriousness issue and discovering factors behind them. Harm like property, individuals because of road accident is bothersome. Happened that road accident episodes are more typical at specific places that can help in recognizing factors behind them. Control based multi portable charger framework is actualized in the proposed framework which can be utilized as a part of open spots like railroad stations, transport stands, doctor's facilities and parks and so forth.

II. GOALS

The essential goals of this investigation can be outlined as takes after:

1. To process the dataset.
2. To bunch the dataset in light of Parameters and apply k-implies for grouping and Bayesian calculation for order.

3. To examined the chart and Predict the Road Accidents in light of different parameters.
4. To create programmed logical apparatus which mines road savvy accident designs
5. To build up a Decision emotionally supportive network for overall population and for government which will anticipate the conceivable outcomes of accidents and spread mindfulness about the roads and take choice about redevelopment of roads

III. PROPOSED WORK

Roadway traffic safety is a major concern for transportation governing agencies as well as ordinary citizens so, for that purpose we are introducing an analytic tool in which Genetic algorithm will be used for classification. We are taking dataset of a country and analyzing that database month wise for one year, To find out which states are similar to each other considering fatal rate, and which states are safer or more risky to drive, clustering algorithm was performed on the fatal accidents dataset. Before applying the algorithms, the tuples with missing value in chosen attributes were removed. The proposed work is planned to be carried out in the following manner.

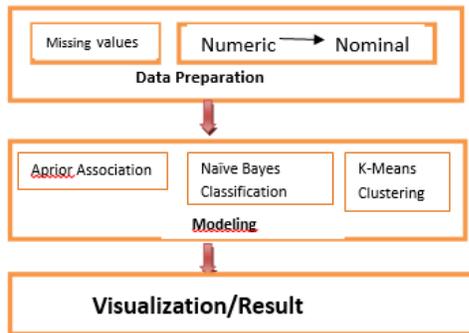


Fig.1: Architectural Design

Data preparation was performed before each model construction. All records with missing value (usually represented by 99 in the dataset) in the chosen attributes were removed. All numerical values were converted to nominal value according to the data dictionary in attached user guide.

Algorithms:

1. Association Rule

To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion.

Data mining uses many different techniques and algorithms to discover the relationship in large amount of data. It is considered one of the most important tool in information technology in the previous decades.

Association rule mining algorithm is a popular methodology to identify the significant relations between the data stored in large database and also plays a very important role in frequent

itemset mining. A classical association rule mining method is the Apriori algorithm who main task is to find frequent itemsetss, which is the method we use to analyze the roadway traffic data.

Before applying the algorithms, the tuples with missing value in chosen attributes were removed, the numerical values were converted to nominal values. The clean data was stored in CSV format and ready to be analyzed by the data analyzing tool. The clean data for association rule mining and classification contains number of tuples, 5 condition attributes, and 1 decision attribute After applying Apriori algorithm with minimum support and minimum confidence in analyzing tool, association rules with fatal rate at the right side as decision were generated.

Rule Measures: Support and Confidence

- Confidence($X \rightarrow Y$) = #tuples containing both X & Y / #tuples containing X = $\Pr(Y/X) = \Pr(X \cup Y) / \Pr(X)$
- Support($X \rightarrow Y$) = #tuples containing both X & Y / total number of tuples = $\Pr(X \cup Y)$

2. Naïve Bayes Classification

Classification in data mining methodology aims at constructing a model (classifier) from a training data set that can be used to classify records of unknown class labels. The Naive Bayes technique is one of the very basic probability-based methods for classification that is based on the Bayes' hypothesis with the presumption of independence between each pair of variables.

Naive Bayes classifier was built on the cleaned data. The Naive Bayes Classifier shows that the fatal rate does not strongly depend on the given attributes, although they are considered feature in comparison to other attributes in the dataset.

On the basis of these formulas the naïve bayes classification table will be created.

$$Tp = \frac{\text{Number of accidents per event}}{\text{Total number of accidents}}$$

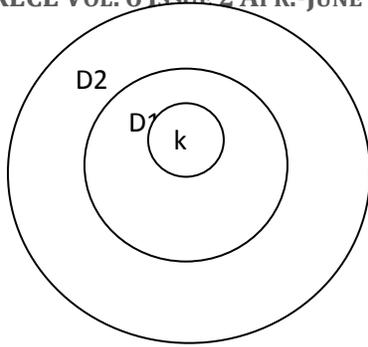
$$Fp = \frac{\text{Number of false accidents per event}}{\text{Total number of accidents}}$$

$$\text{Precision} = \frac{| \{ \text{relevant documents} \} \cap \{ \text{retrieved documents} \} |}{| \{ \text{retrieved documents} \} |}$$

$$\text{Recall} = \frac{| \{ \text{relevant documents} \} \cap \{ \text{retrieved documents} \} |}{| \{ \text{relevant documents} \} |}$$

3. K-means clustering

To find out which states are similar to each other considering fatal rate, and which states are safer or more risky to drive, clustering algorithm was performed on the fatal accident dataset. With the fatal accident and the population dataset, fatalities per million people in the state was calculated.



K= Parameter (i.e. weather condition)

For given k cluster evaluation will be done as (M)

$$\text{Mean cal} = \frac{\sum \text{all entities}}{\text{Cluster mean}(k)}$$

Recluster till two consecutive (M) has some values.

Work Flow

In the modeling We first calculated several statistics from the dataset to show the basic characteristics of the fatal accidents. We then applied association rule mining, clustering, and Naive Bayse classification to find relationships among the attributes and the patterns.

The results of our analysis include association rules among the variables, clustering of states in the USA on their populations and number of fatal accidents, and classification of the regions as being high or low risk of fatal accident.

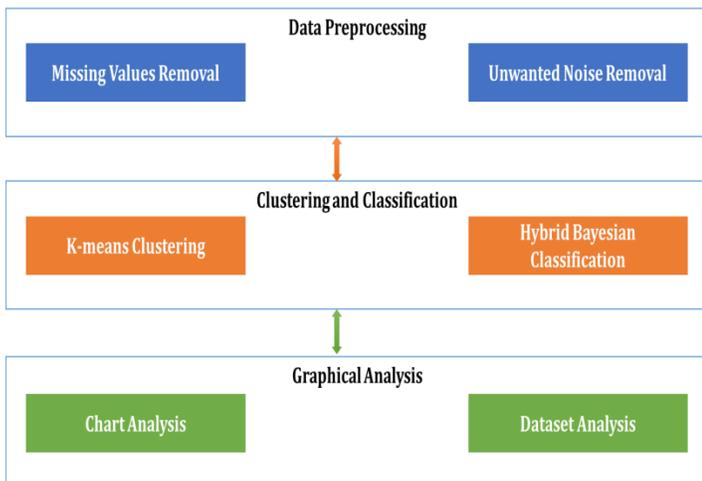


Fig 2. Work Flow

Clustering is a process of collection of objects which are similar between them while dissimilar objects belong to other clusters. A clustering technique is used to obtain a partition of N objects using a suitable measure such as resemblance function as a distance measure ‘d’.

K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. The algorithm has a loose relationship to the k-nearest neighbor classifier, a popular machine learning technique for classification that is often confused with k-means because of the k in the name. One can apply the 1-nearest neighbor classifier on the cluster centers obtained by k-means to classify new data into the existing clusters.

Experimental Graph Based Results

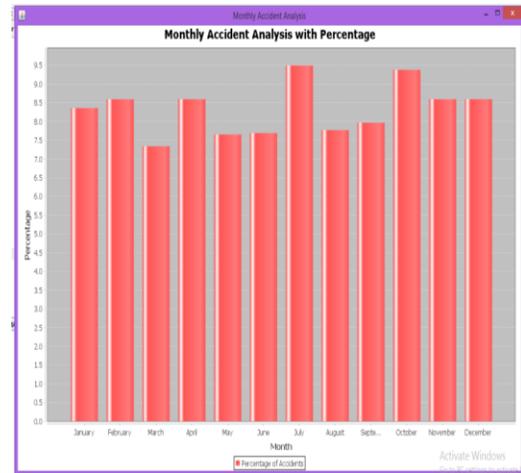


Fig 3. Monthly Analysis for number of Accidents

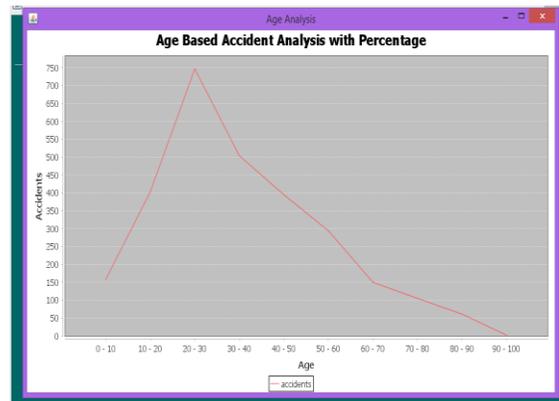


Fig 4. Age Based Analysis

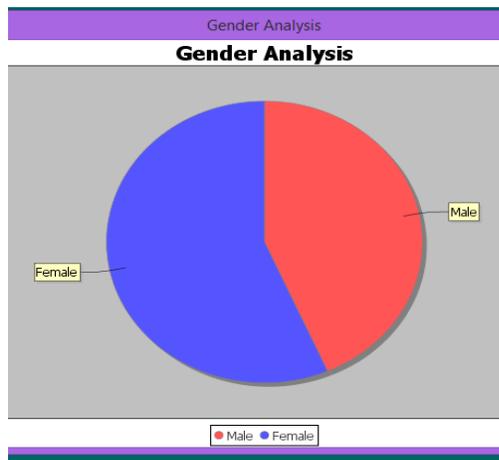


Fig 5. Gender Based Analysis

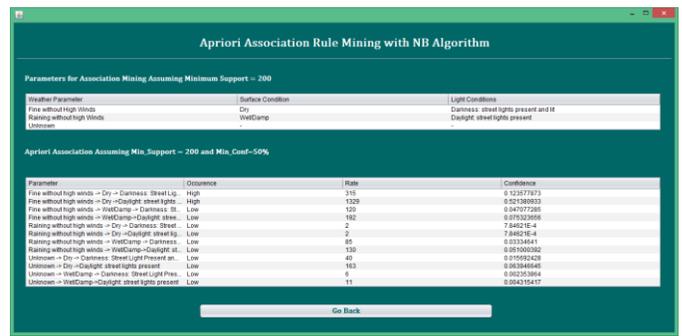


Fig 8: Support and Confidence Results

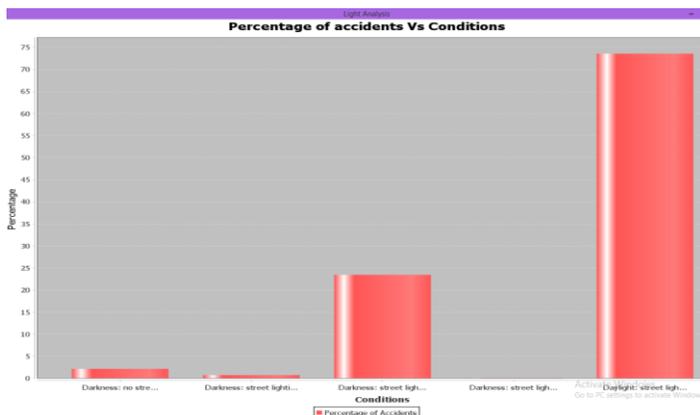


Fig 6. Number of Accidents categorized by Light Conditions

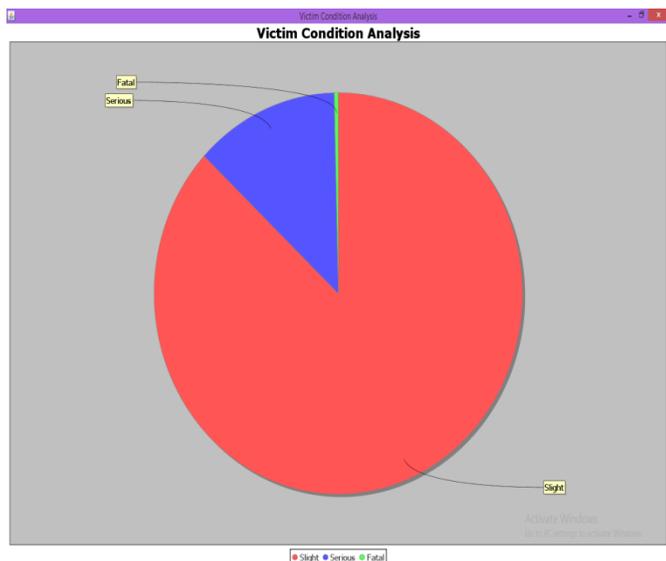


Fig 7. Victim Condition Analysis

IV. CONCLUSION

In this paper, we have collected multiple researchers' works together in single document and discussed about the contribution towards impact of road and traffic accident on human life and society. This highlights the number of approaches used to avoid the accident happened in various countries and cities. The paper also discussing about various data mining techniques which is proved supporting to resolve traffic accident severity problem and conclude which one could be optimal technique in road traffic accident scenario. The brief discussion will also help us to find better mining technique in this kind of problem.

V. REFERENCES

- [1]. Zhexue Huang, "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values", Data Mining and Knowledge Discovery 2, 283–304 (1998).
- [2]. Sachin Kumar and Durga Toshniwal, "A data mining framework to analyse road accident data", Journal of Big Data (2015) 2:26 DOI 10.1186/s40537-015-0035-y.
- [3]. S. Krishnaveni and Dr. M. Hemalatha, "A perspective analysis of Traffic Accident Using Data Mining Techniques", International Journal of Computer Application.
- [4]. Olutayo V.A and Eludire A.A, "Traffic Accident Analysis Using Decision Trees and Neural Networks", IJ. Information Technology and Computer Science, 2014, 02, 22-28 Published Online January 2014 in MECS (<http://www.mecs-press.org/>) DOI: 10.5815/ijitcs. 2014.02.03.
- [5]. K. Geetha and C. Vaishnavi, "Analysis on Traffic Accident Injury Level Using Classification", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 2, February 2015, ISSN: 2277 128X.
- [6]. Sachin Kumar and Durga Toshniwal, "A data mining approach to characterize road accident locations", J. Mod. Transport. (2016) 24(1):62–72 DOI 10.1007/s40534-016-0095-5.
- [7]. Tibebe Beshah, Shawndra Hill, "Mining Road Traffic Accident Data to Improve Safety: Role of Road-related Factors on Accident Severity in Ethiopia"
- [8]. Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993. [9] K. Jayasudha and Dr. C. Chandrasekar, "An overview of Data Mining in Road Traffic and Accident Analysis", Journal of Computer Applications, Vol – II, No.4, Oct – Dec 2009.

- [9]. Miao Chong, Ajith Abraham and Marcin Paprzycki, "Traffic Accident Analysis Using Machine Learning Paradigms", *Informatica* 29 (2005) 89–98.
- [10]. M. Sowmya and Dr.P. Ponmuthuramalingam, "Analyzing the Road Traffic and Accidents with Classification Techniques", *International Journal of Computer Trends and Technology (IJCTT)* – volume 5 number 4 –Nov 2013.