A Perspective Analysis of Traffic Accident using Data Mining Techniques

Mrs.Vandana Munde[1] Dheeraj Singh[2] Rahul Singh[3] Sharad Singh[4]
[2, 3, 4] UG-Student
[1]AP-IT, TCET MUMBAI
[2, 3, 4] TCET MUMBAI

Abstract--- Nowadays vehicles are increasing day by day in town and cities roads. It is a well-known problem to manage traffics on the roads in towns and cities. A lot of accident occurs on the road due to careless driving and Technical faults in vehicles. The main problem of traffic authorities is to manage the traffic on the road for the smooth functioning of vehicles that can reduce the accident and violation on the road. There is a tremendous demand from traffic authorities to develop a system that can helps to avoid the accident and keep the accident report data and also maintain the accident report data. The main objective of this paper is to model a Traffic Accident Reporting System (TARS) through data mining using Clustering technique to solve the above problem.

I. INTRODUCTION

Data mining is the extraction of hidden predictive information from large databases and it is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. According to a Gartner HPC Research Note, "Due to data capture, transmission and storage, large-systems users have to implement new techniques. They innovative ways to mine the after-market value of their vast stores of detail data, employing MPP [massively parallel processing] systems to create new sources of business advantage".

According to the World Health Organization (WHO) [18], India is leading in the annual reported number of traffic deaths worldwide. In Global Status Report on Road Safety, the WHO revealed that India leads with 105,000 traffic deaths in a year, when compare to China with over 96,000 deaths on road. The survey was conducted in 178 countries, as per the survey 300 Indians die on roads every day. There are two million people have disabilities caused from a traffic accident. This survey is based on data collection for 2006 (Data collection began from March2008 and completed in September 2008).A traffic collision occurs when a road vehicle collides with another vehicle, pedestrian, animal, or geographical or architectural obstacle. It can result in injury, property damage, and death. Road accidents have been the major cause of injuries and fatalities in worldwide for the last few decades. It is estimated that the amount of data stored in the world's database grows every twenty months at a rate of 100%. This fact shows that we are getting more and more exploded by data/ information and yet ravenous for knowledge. Data mining is a useful tool to address the need for sifting useful information such as hidden patterns from database.

II. PROBLEM DEFINITION

Reducing the number of traffic accidents remains one of the greatest challenges facing many societies around the world. The cost of traffic accident on society and individuals is very high. Loss of life, disability and suffering are but a few of the impacts of traffic accidents. On average a higher proportion of Indian drivers are involved in road accidents compared to their relative population among licensed drivers.

A study of the reasons behind traffic accidents revealed four main factors: factors related to driving (the human factor); vehicle-related factors (physical environmental factors); mechanical factors, and socio-economic factors whether factor and Drinking driving. According to the World Health Organization (WHO), India is leading in the annual reported number of traffic deaths world-wide. In Global Status Report on Road Safety, the WHO revealed that India leads with 105,000 traffic deaths in a year, when compare to India with over96, 000 deaths on road. The survey was conducted in 178 countries, as per the survey 300 Indians die on roads every day. There are two million people have disabilities cause from a traffic accident. This survey is based on data collection for 2012(Data collection began from March 2008 and completed in September 2012).A traffic collision occurs when a road vehicle collides with another vehicle, pedestrian, animal, or geographical or architectural obstacle. It can result in injury, property damage, and death. Road accidents have been the major cause of injuries and fatalities in worldwide for the last few decades. It is estimated that the amount of data stored in the world's database grows every twenty months at a rate of 100%. This fact shows that we are getting more and more exploded by data/ information and yet ravenous for knowledge. Data mining is a useful tool to address the need for sifting useful information such as hidden patterns from databases.

III. PREVIOUS WORKS

IV. THE USE OF DATA MINING IN TRAFFIC ACCIDENT

There are many cities in our country which has a large Population compared to their extension. Naturally these People need a good transport system to cope with their needs like going to work, go shopping, etc. It should be done through city roads. The growth of population and the need of transportation system in one amount of traffic information in roads and cities hand and the addition of transport vehicles on the other hand, we need a good city Management all over the country. The addition of transport facilities in a town involves high financial and chronological expenses. These problems show the need of correct traffic management .As the traffic had bad effect on the air it has also undesired effects on human lives in different aspects, so it is necessary to investigate the ways to control and overcome the difficulties in this field. One of the most important problems of traffic is taking a lot of time, so we can prevent ton desired traffic effects. Today is the age of information increasing in any field; there are large databases.

V. K –MEANS ALGORITHM

A. 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. This results in a partitioning of the data space into Voronoi cells.The problem is computationally difficult (NP-hard); however, there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum.The algorithm for mixtures of distributions via an iterative refinement approach employed by both algorithms. Additionally, they both use cluster centers to model the data; however, k-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes'-Means clustering generates a specific number of disjoint, flat (non-hierarchical) clusters. It is well suited to generating globular clusters The K-Means method is numerical, unsupervised, non-deterministic and iterative.

B. K-Means Algorithm Properties

There are always K clusters. There is always at least one item in each cluster.

The clusters are non-hierarchical and they do not overlap. Every member of a cluster is closer to its cluster than any other cluster because closeness does not always involve the 'centre' of clusters.

C. The K-Means Algorithm Process

The dataset is partitioned into K clusters and the data points are randomly assigned to the clusters resulting in clusters that have roughly the same number of data points. For each data point: Calculate the distance from the data point to each cluster. If the data point is closest to its own cluster, leave it where it is. If the data point is not closest to its own cluster, move it into the closest cluster. Repeat the above step until a complete pass through all the data points results in no data point moving from one cluster to another. At this point the clusters are stable and the clustering process ends.The choice of initial partition can greatly affect the final clusters that result, in terms of inter-cluster and intra cluster distances and cohesion.

D. K-Means Clustering Advantages and Disadvantages

– K-Means Advantages:

1. If variables are huge, then K-Means most of the times computationally faster than hierarchical clustering, if we keep k small..
2. K-Means produce tight clusters than hierarchical clustering, especially if the clusters are globular.

– K-Means Disadvantages:

1. Difficult to analyze K-Value.
2. With global cluster, it won't work well.
3. Different initial partitions can result in different final clusters.
4. It does not work with clusters (in the original data) of Different size and Different density

VI. IMPLEMENTATION

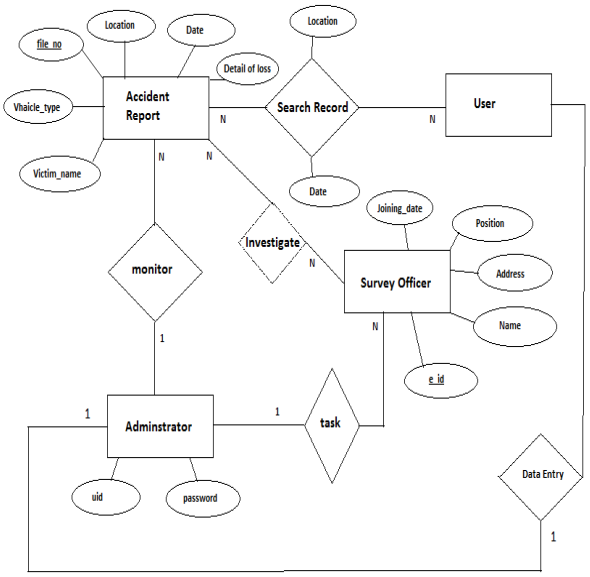

Fig. 1:

VII. FUTURE SCOPE

This project can be used for future references for the number of the accidents happened in past years. Hence, with the help of this Project, the analysis can be done and therefore preventive measures can be taken. It can help the government to keep track of records of the accidents, causes of accident, vehicle number, vehicle owner's name and address.

This project will also help us to add news record and delete old data .The viewer or user can also make their own account for viewing the site .you can view the data about causality .Our system will provide the graphical view of the accidents with respect to the data entered into the system according to the period .This system will provide the solutions as accidents causes. So that with the help of this system government can take the necessary actions according accidents cases.

1. Accurate Location of accident
2. GPS integration
3. Government ID Authentication for user Data
4. Advanced Filter technique Accident Solution prediction

VIII. RESULT

We can search any previous information about any accident, which we want to search so this project will help us to gather information about any past accident with correct date, time and vehicle owner .This project will also help us to add news record and delete old data. The viewer or user can also make their own account for viewing the site .you can view the data about causality .Our system will provide the graphical view of the accidents with respect to the data entered into the system according to the period .This system will provide the solutions as accidents causes .So that with the help of this system government can take the necessary actions according accidents cases.

IX. CONCLUSION

The aim of this project is to detect the causes of accidents. The dataset for the study contains traffic accident records of

the year 2013 produced by the transport department of government of India and investigates the performance of J48, The classification accuracy on the test result reveals for the following three cases such as accident, vehicle and casualty .Random Forest outperforms than other classification algorithms instead of selecting all the attributes for classification. Genetic Algorithm is used for feature selection to reduce the dimensionality of the dataset. In this work, we extended the research to three different cases such as Accident, Casualty and Vehicle for finding the cause of accident and the severity of accident. With the current data it is possible to recognize the risky road segments and the road user groups responsible for accidents in certain environments. However, it is not possible to find out very strict details for enhancing road construction plans from this data. More detailed location specific information from accident locations and situations are needed.

REFERENCES

[1] Suchita Borkar and K.Rajeswari. 1997 Attributes Selection for Predicting Students' Academic Performance using Education Data Mining and Artificial Neural Network.
[2] R. R. Kabra, R. S. Bichkar 2011,Performance Prediction of Engineering Students using Decision Trees.
[3] V.Ramesh,P.Parkavi, K.Ramar ,2013. Predicting Student Performance: A Statistical and Data Mining Approach.
[4] Singh, Randhir. An Empirical Study of Applications of Data Mining Techniques for Predicting Student Performance in Higher Education, 2013. Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.
[5] C. Romero, S. Ventura, "Educational data mining: A survey from 1995 to 2005", Expert system with applications 33(2007), 135-146.
[6] http://www.saedsayad.com/oner.htm
[7] http://en.wikipedia.org/wiki/Weka.
[8] Jiawei Han and Micheline Kamber, "datamining Concepts and Techniques", Elsevier Second Edition.
[9] http://www.soc.napier.ac.uk/~peter/vldb/dm/node8.html

