International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

DATA MINING TECHNIQUES FOR TRAFFIC MODELLING AND PREVENTION

Anupama K B[1], M R Pruthvish[1], Manogna H P[1], Monika Nag K J[2]

[1]*Dept. of ISE, The National Institute of Engineering, Mysore*
[2]*Asst. Professor, Dept. of ISE, The National Institute of Engineering, Mysore*

Abstract - *Road accident is one among the crucial areas of analysis. A variety of analysis has been done on information collected through police records covering a restricted portion of highways. The analysis of such information will solely reveal info relating to that portion solely, however accidents are scattered not only on highways but also on local roads. A distinct supply of road accident information is Emergency Management analysis Institute (EMRI) that serves and keeps track of each accident record on every form of road and cover info of entire state's road accidents. In this paper, we have used data mining techniques to analyze the information provided by EMRI in that we 1st cluster the accident information and more association rule mining technique is applied to spot circumstances within which an accident could occur for every cluster. The results are often used to place some accident prevention efforts within the areas known for various classes of accidents to beat the number of accidents.*

One of the key objectives in accident data analysis is to spot the main factors related to a road and traffic accident. The overall objective of this thesis is to attain the accuracy and determine the factors behind crashes or accident that might be useful to scale back accident magnitude in future and will be helpful to save several lives, deteriorate wealth destruction.

Key Words: **EMRI (Emergency Management Analysis Institute), Data Mining techniques, Association Rule Mining techniques, Accident Prevention, Factors behind Accidents, Wealth Destruction.**

1. INTRODUCTION

Road accidents are a major threat worldwide that continue to cause casualties, injuries, and fatalities on roadways on a daily basis, resulting in huge losses at the economic and social levels. According to World Health Organization records, road accidents have become one of the major causes of death worldwide. Accidents cause roughly 1.27 million annual deaths and between 20 and 50 million injuries. This global problem needs more attention to reduce the severity and the frequency of accident occurrence. The historical data about previous accidents represents a formidable opportunity for researchers to identify the most influential factors in such accidents, which in turn play a key role in finding appropriate solutions to mitigate this problem in the future. It is, however, a very challenging task to extract knowledge from these data as they are typically huge and high dimensional. In recent years, several data-mining techniques have been effectively used to extract useful knowledge from large data sets containing information about traffic accidents. Classification methods are among the most commonly used techniques in mining road traffic accidents, where the goal is building classifiers that are capable of predicting the severity of new accidents. These classifiers are built using training sets of data in which the severity of the accidents is known. Classification methods are categorized into three types: mathematical, probabilistic, and rule-based classifiers. The accuracy of a classifier depends heavily on the characteristic of the collected data, a fact that makes it almost impossible to recommend a certain classifier for a particular kind of problems. As a result, researchers tend to test multiple classifiers before deciding which one to use.

2. OBJECTIVES

- To provide an overview of the literature in road accident detection with various techniques and approaches implemented in them.
- To detect and develop effective road accidents mechanisms to save life.
- The key objectives in accident data analysis is to identify the main factors associated with a road and traffic accident.

3. EXISTING SYSTEM

- **A Real-Time Autonomous Highway Accident Detection Model Based on Big Data Processing and Computational Intelligence**

  —Due to increasing city population and growing wide variety of motor vehicles, traffic congestion is turning into a most important problem of the 21st century. One of the major reasons in the back of

traffic congestion is accidents which cannot only result in casualties and losses for the participants, however also in wasted and lost time for the others that are stuck behind the wheels. Early detection of an accident can store lives, offers faster avenue openings, as a result decreases wasted time and resources and increases efficiency. In this study, we advocate a preliminary real-time self-reliant accident-detection device based on computational Genius techniques. Istanbul city traffic-flow data from the year 2015 from quite a number sensor places are populated using massive records processing methodologies. The extracted features are then fed into a nearest neighbor model, a regression tree, and a feed-forward neural community model. For the output, the opportunity of an incidence of an accident is predicted. The outcomes point out that even although the quantity of false alarms dominates the actual accident cases, the machine can still furnish useful data that can be used for repute verification and early response to viable accidents.

- **Accident Prediction from Traffic Data using Hadoop author: Pallavi Dubey , Prof. Manaswini Panigrahi**

  —Accident prediction has been in trend to furnish warning signs previously than accidents happen. Traffic on highways is monitored and lots of data is processed each and every day to predict likelihood of accidents based totally surely on twin carriageway stipulations like avenue surface, slight on highway, turns etc. In this paper to predict accident based on special queries and approach this massive documents Hadoop has been used. It is placed that execution time is very a good deal less on Hadoop as in distinction to sequential techniques.

- **Analysis and Prediction of Bangalore Traffic South Road Accidents author: Ramya v**

  —Data mining is the approach of inspecting records from exclusive views and summarizing it into beneficial facts - statistics that can be used to enlarge revenue, cut fees or both. The speedy proliferation of Global Position Service (GPS) gadgets and mounting variety of site visitors monitoring structures employed through means of municipalities have opened the door for most excellent site traffic manipulate and custom-made route planning. But the complexity of site visitors accident evaluation has brought many difficulties to web site visitors administration and decision-making. This paper gives necessary points about how avenue accidents and site visitors information can be analyzed and used to predict the threat of an accident to occur. To start with, the evaluation has been finished on the Bangalore town visitors thinking about the 5 site visitors stations of the south area of the metropolis. There are a lot of lookup about the traffic and transportation systems; on the other hand most of them are focused on infrastructure development, enhancements on bodily infrastructures, etc. Meanwhile, some research focus on some unique shrewd transportation machine components such as subjects of accident prevention, traffic float estimation, match detection, route optimization etc. There are additionally some lookup primarily based absolutely on image and video processing techniques that have in mind on data acquired by using CCTV cameras. The benefit of these structures is the possibility of searching at exclusive types of motors and their behaviors.

4. PROPOSED SYSTEM

The proposed system, datamining approach for traffic accident modelling and prediction is web and android based definitely application. Road collisions are the second leading reason behind the death for people, this utility goals to identify the important elements that make a contribution to the severity of such accidents. The proposed website consists of an Admin module who is responsible for widespread functioning of the model. Admin will login and add accident vicinity with town and area, he is additionally accountable for assigning journalists to particular town and area. The reporters will view the accident small print delivered through the admin and supply description about rationale for incidence of the accident and preventive measures and they can moreover add accident statistics in form of image. Users can register and login the use of their android software and view the accident small print with the resource of identifying on precise city or area.

The research involves exploring various data classification
1. Collection of raw data and then apply filtering techniques to make that raw data into structured format: Filtering techniques like replace Missing Value Filter

2. Enhanced K Means Clustering algorithm 1. The size of cluster is fixed. Here, the input array of elements is scanned and split up into sub –arrays, which represent the initial clusters. 2. The cluster size varies and the output of this phase is the finalized clusters. Initial clusters are inputs for this phase. The centered of these initial clusters are computed first, on the basis of which distance from other data elements are calculated. The data elements having less than or equal distance remains in the same cluster otherwise they are moved to appropriate clusters. The entire process continues until no changes in the clusters are detected.

There are 3 modules:

1. Admin
2. User
3. Reporter

Admin:

- Admin will login the usage of admin identification and password.
- Admin will add the location primarily based on metropolis and area.

User:

- User will register and login using the consumer identification and password.
- Once a person login they can view the accident records uploaded [based on the metropolis and area] through reporter.
- User can get recognize accident details in particular area or city.

Reporter:

- Reporter will register and login with identification and password.
- Admin will assign the reporter to specific vicinity [city and area].
- They can view the accidents and they will provide Description about prevention and reason. Why they met for the accident.
- Reporter will upload the accident data.

5. IMPLEMENTATION

The Proposed Application consists of Web and Android implementations for one of a kind modules. The Admin and reporter Modules of the software is applied as a web utility running in desktops or laptops. The frontend of the internet web page is developed using HTML5, CSS, Bootstrap and JavaScript with NetBeans IDE and Java Servlet for backend Programming. MySQL Database in backend to keep facts and User module is developed in android the use of XML as frontend, for designing Android Studio IDE is utilized and MySQL as Backend. For Server Side Development, Apache is chosen as net server.

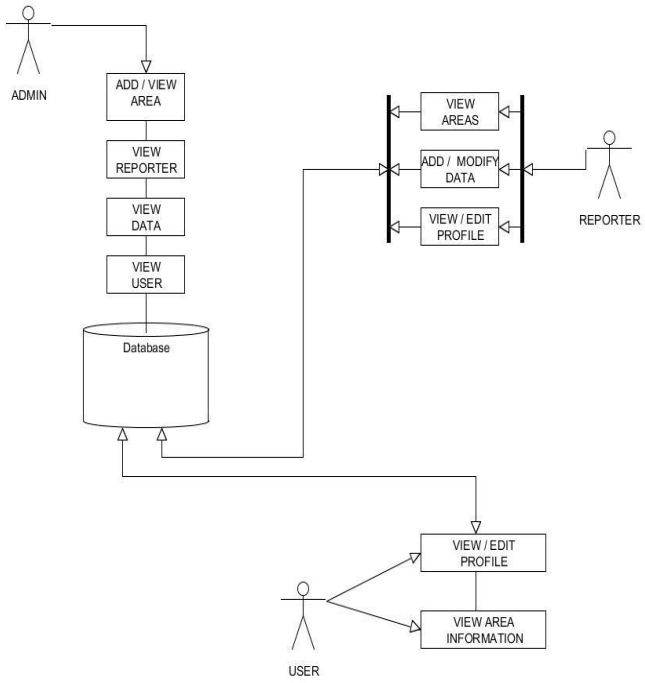

Fig 1 – System Architecture

5.1 Algorithm

Given a dataset of n data points x1, x2, …, xn such that each data point is in R d , the problem of finding the minimum variance clustering of the dataset into k clusters is that of finding k points {mj} (j=1, 2, …, k) in R d such that $1/n\sum$ i=1 to n[Min d2(X2,Mj)] is minimized, where d(xi , mj) indicates the Euclidean distance between xi and mj . The points {mj} (j=1, 2, …,k) are called as cluster centroids. The problem in eq.(1) is to search k cluster centroids, such that the average squared Euclidean distance (mean squared error, MSE) between a data point and its nearest cluster centroid is minimized. The k-means algorithm gives an easy method to implement solution to eq.(1). The reason for the popularity of k-means algorithm are its simplicity and ease of implementation, scalability, speed of convergence and adaptability to sparse data. The k-means algorithm can be thought of as a gradient descent procedure, which begins at starting cluster centroids, and repetitively updates these centroids to decrease the objective function in eq.(1). The k-means algorithm always converge to a local minimum. The specific local minimum found depends on the starting cluster centroids. The k-means updates cluster centroids till local minimum is discovered. Before the k-means algorithm

converges, distance and centroid calculations are done while loops are executed a number of times, say i, where the positive integer i is known as the number of k-means iterations. The precise value of i varies depending on the initial starting cluster centroids even on the same data set. So the computational time complexity of the algorithm is O(nki), where n is the total number of objects in the dataset, k is the required number of clusters we identified and i is the number of iterations, k≤n, i.

6. CONCLUSION

Although the proposed approach is quite sufficient to uncover reasonable information from the selected data set, the results remain at very general level as source data does not contain other accident related information such as the speed of vehicle at the time of accident, weather, road surface condition. Hence in future we can provide the data with more number of attributes so that we can reveal more information using our approach.

REFERENCES

[1] Abellan, J., Lopez, G., & De Ona, J. (2013). Analysis of traffic accident severity using Decision Rules via Decision Trees. Expert Systems with Applications, 40, 6047–6054.
[2] Chang, L.-Y., & Chien, J.-T. (2013). Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. Safety Science, 51(1), 17–22.
[3] Chang, L. Y., & Wang, H. W. (2006). Analysis of traffic injury severity: An application of nonparametric classification tree techniques. Accident; Analysis and Prevention, 38, 1019–1027.
[4] Chen, W. H., & Jovanis, P. P. (2000). Method for identifying factors contributing to driverinjury severity in traffic crashes. Transportation Research Record, 1717, 1–9.
[5] De Ona, J., Mujalli, R. O., & Calvo, F. J. (2011). Analysis of traf ~ fic accident injury severity on Spanish rural highways using Bayesian networks. Accident Analysis & Prevention, 43, 402–411.
[6] De Ona, J., L ~ opez, G., & Abellan, J. (2013). Extracting decision rules from police accident reports through decision trees. Accident Analysis & Prevention, 50, 1151–1160. [7]. Delen, D., Sharda, R., & Bessonov, M. (2006). Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accident Analysis and Prevention, 38, 434–444.
[7] Gregoriades, A. (2007). Towards a user-centred road safety management method based on road traffic simulation. In Proceedings of the 39th Conference on Winter Simulation: 40 years! The Best is Yet to come (pp. 1905–1914). Washington, DC: IEEE. [9]. Holmes, G., Donkin, A., & Witten, I. H. (1994). Weka: A machine learning workbench. In Proceedings of the 2nd Australia and New Zealand Conference on Intelligent Information Systems (pp. 357–361). Brisbane, Australia: IEEE.