



Road Accident Alert System using Data Mining

Sonal J. Mane¹, Mayuresh P. Patil², Sourabh C. Chauhan³, Omkar S. Raut⁴Assistant Professor¹, BE Student^{2,3,4}

Department of Information Technology

Bharati Vidyapeeth College of Engineering, Kharghar, Navi Mumbai, India

Abstract:

Roads are one of the substantive way of transportation, the rate of the number of vehicles is increasing proportionally day by day which gradually peaks the rate of accidents. In India, road accident is considered as one of the major cause leading to deaths and property loss. Reasons behind the accidents include over speeding, drunk driving, atmosphere conditions, defects in vehicle as well as poor conditions of road. The detail records of road accident that are available are not properly analyzed, no aiming governing body is available to gain knowledge from this data. The areas in which most number of accidents occur and the reason why they occur is not known in many cases. The current caution signs along the road are neglected by the travellers. In some areas signs are not present because of negligence in their maintenance. In order to give safe driving suggestions, careful analysis of road accident data is necessary to find out variables that are related to severity of accidents. Hence, we present a system which will alert the drivers about the most accident-prone areas present along the road they are driving. This will be based on the analysis of accidents occurred in the past along that road. System will give voice alerts to alert the drivers on basis of causes of accidents in that area and severity of that accident prone area. System will be used by highway authorities to view the accident statistics so that they can take precautionary measures to impede accident rate.

I. INTRODUCTION

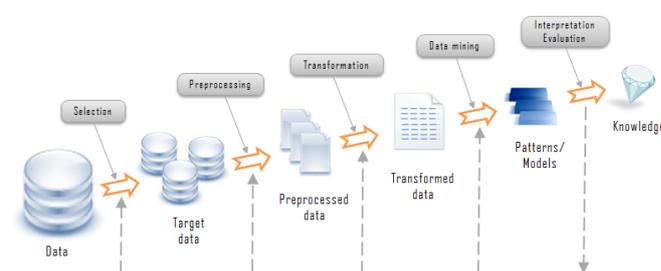
Roads are still the major way of transportation in India even if other modes like railways, airplanes etc. are available. Road network covers 75% of total transportation system. The Global status report on road safety 2013 estimates that more than 231 000 people are killed in road traffic crashes in India every year. Approximately half of all deaths on the country's roads are tagged to vulnerable road users. The statistics revealed a shocking report and if this continues, road accident death rate might reach 1.9 million people annually by 2020. To abate traffic incidents, it is not enough to just improve the road conditions, but also to control traffic accidents occurring by analyzing the cause and effect regulations. Some necessary steps should be taken by highway governing authorities. Significant help in this situation represents an identification of the key factors causing road accidents. Road traffic issues, moreover, has become the backbone for major injuries, deaths in recent times. The problem lies between negligence and the false approach towards the better analysis of traffic events. In India the accident rate per week is 1214 and supposed to be a negative factor with increasing road mishaps frequency. The liability of road accidents supposed to be varied due to various reasons such as drunk driver, natural cause (weather conditions), pedestrian crossings, brake failure, tire bursting, etc. It is at ease to find remedy if factors could be analyzed, grasped and act to take right measures. In order to control road casualties, it is not enough to just improve the road conditions, but also mandatory to examine the traffic event happened by the factors causing accident and hence following disciplinary traffic protocols. Application of suitable data mining methods on the collected dataset representing different situations on the roads and occurred accidents can help understand the most significant factors or often repeating patterns. Obtain results in the form of predictive models or generated rules can help relevant decision makers to identify the most accident prone plots and take precautionary measures to reduce the rate. In this paper we use a triggering mechanism; the driver will be alerted by pop-up notification of the accident prone spots on

his smartphone. We will also provide statistics about the place and a voice alert. This action would be real-time so that, the driver can take necessary actions to prevent the accident.

Data mining: Data mining is the process of analyzing hidden patterns of data according to different perspectives for categorization into useful information.

The data mining process is commonly defined with the stages:

- (1) Selection
- (2) Pre-processing
- (3) Transformation
- (4) Data mining
- (5) Interpretation/evaluation.



II. LITERATURE REVIEW

Road mishaps are nowadays one of the concerned issues, many works related to road accidents analysis and extraction process of event information is recorded. In many countries, road accidents data are studied in order to explore the factors leading to road accidents. Researchers and authorities seek to identify patterns and relationships between risk factors and injury severity levels. Among the important attributes that are usually studied in road accident data are: driver characteristics, vehicle characteristics, road condition variables, and weather and environmental factors. Jayasudha proposed system that analyzed the traffic accident using data mining technique that could possibly reduce the fatality rate. They used web data for generating self-organizing map for pattern analysis. It could classify information and provide warning as an audio or video.

They also identified that the fatality rate increased to 27% due to rash driving or over speeding. Kumar performed partition based clustering and density based clustering to categorize the similar accidents. Based on this categorization nature data k-modes algorithm were used. To find correlation between various set of attributes association rule mining were used. They observed that certain factors such as accident location, time, weather conditions, vehicle type, age of driver and gender have impact on the complete road event occurred. Chapter 2 6 Lilingli, sharad shrestha, gongzhu hu proposed a system where preprocessing is performed before the data is converted into csv format. Later the data is analysed by analysing tool Weka. Naive Bayes classification is applied to classify the cleaned data. A group of similar clusters are formed with the help of K means algorithm. It is easy to explore the basic characteristics of the fatal accidents. Reasons are defined with proper observation. Another study was conducted using CART, TreeNet, and Random Forest to understand car accidents data in Ethiopia. The main goal of the study was to identify the role of road users in car accidents. A total of 14,254 records and 12 attributes were used to train the classification models. Many attributes were found to be important in predicting injury cases, including: pedestrian movement during the accident, occupation of victims, and victim's age. The average accuracies of the CART, TreeNet, and RandomForest are 84.5 %, 98.94%, and 86.59%, respectively. It was observed that the trained models performed better in predicting non-injury risk of an accident. Isra Al-turaiki, Maryam Aloumi and Khulood Alghamdi proposed a system where data mining is performed in order to understand the factors leading to car accidents severity in Riyadh. They were provided with 85,834 records in spreadsheet format for accidents that occurred between October 2014 to October 2015. There are 83,605 records for accidents with no injuries, 1,808 records for accidents that resulted in injuries, and 421 for accidents that led to death. Three classification techniques are used: CHAID, J48, and Naive Bayes. The performance of all the obtained models is evaluated and compared.

III. SYSTEM ARCHITECTURE

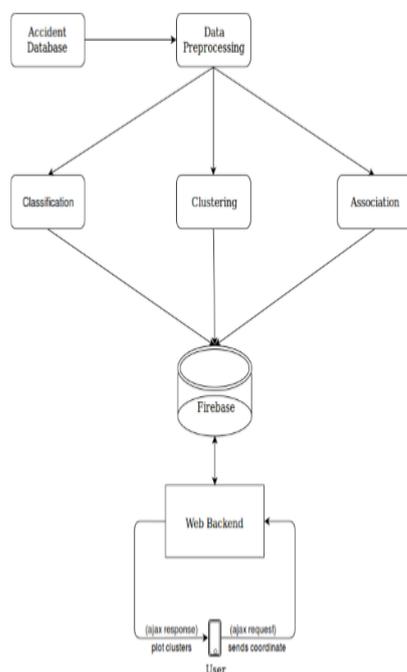


Fig 3.1 System Architecture

Data mining

Data mining is the process of analyzing hidden patterns of data according to different perspectives for categorization into useful information. This data mining process contains following processes:

- **Classification:** Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. Once rules are obtained by training the model, these rules can be used to predict the class of data.
- **Clustering:** Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics.
- **Association:** Association is a data mining function that discovers the probability of the co-occurrence of items in a collection. The relationships between co-occurring items are expressed as association rules. Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships.

Firestore

Firestore is a cloud-hosted real-time NoSQL database. It makes use of WebSocket which is much faster than regular HTTP connections and provides an API that allows developers to store and sync data across multiple clients. The data is stored in JSON format, synced across connected devices in milliseconds, and available even if the app goes offline.

IV. METHODOLOGY

Driver will initiate the application at the start of his journey. GPS enabled phone will send the location of user to maps API. Application will be frequently sending the location. As the driver enters inside range of 100 meters of a blackspot, application will trigger a notification with voice alert to make driver attentive. It will also include major reason behind the accident occurred in that area and severity of accidents according to the number of accidents occurred. Also, highway patrolling authorities will be able to see the statistics via a web front-end. This system is divided into following parts-

- Data mining on the previous accident data
- Handling of the rules on web back-end and alert application.

Data mining: The historical accident data will be first preprocessed for handling of missing values and formatting of data. This is then categorized according to the season, time zone, type of vehicle, and severity and is visualized into graphs and pie charts. Clustering is then performed on this to get density wise clusters to get accident-prone areas called "blackspots". Mean values of these clusters are plotted on maps. Naive bayes classification algorithm is applied to training dataset to generate set of rules that will classify data into different classes of severities. This model when provided with input result the class. Apriori algorithm will be used to find relationship between different attributes that will generate interesting patterns which can be used to decrease accident rate by taking preventive measures.

Back-end: All data and rules derived from data mining process are stored in Firestore. Data available is in excellent

format and will have to be converted into JSON since JSON is much faster to parse. User location and time will be fetched via Android app. This location and time will trigger one of the rules stored in database and display blackspots according to it. This process will be a realtime process.

Front-end: An android application will be developed for end-user (driver). This application will use firebase as a back-end and will retrieve and display the clusters. A voice alert and pop-up message will be provided alongside.

V. CONCLUSION

Road issues are moreover increasing concerns which cannot be neglected. A certain kind of system is needed to improve the current road protocols and hence give a kind of assurance to avoid the mishaps. Prioritization of road events analysis is mandatory to have certain kind of reign over the overall chaos of the road events. Based on observation of traffic incidents, it is at ease to determine factors causing accidents. As system is working in interest for both drivers and traffic authority, it helps both to impede the road events and thereby saving property loss. Records are real hence it would make more real time implementation to enhance the system and therefore would definitely benefit both kind of users.

VI. REFERENCES

[1]. Liling Li, Sharad Shrestha, Gongzhu Hu. "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques" IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA) 2017

[2]. Isra Al-Turaiki, Maryam Aloumi, Nour Aloumi, and Khulood Alghamdi. "Modeling Traffic Accidents in Saudi Arabia using Classification Techniques" 4th Saudi International Conference on Information Technology (Big Data Analysis) (KACSTIT) 2016

[3]. Rui Tian and Zhaosheng Yan, Kg Maolei Zhang. "Method of Road Traffic Accidents Causes Analysis Based on Data Mining" International Conference on Computational Intelligence and Software Engineering 2010

[4]. František Babic, Karin Zuskáová. "Descriptive and Predictive Mining on Road Accidents Data" IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMI) 2016

[5]. Data mining - https://en.wikipedia.org/wiki/Data_mining

[6]. Nikam S. S. A Comparative Study of Classification Techniques in Data Mining Algorithms. Orient. J. Comp. Sci. and Technol; 8(1). Available from: <http://www.computerscijournal.org/?p=1592>

[7]. K. M. A. Patel and P. Thakral, "The best clustering algorithms in data mining," 2016 International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, 2016,

[8]. Association rules: <http://search.businessanalytics.techtarget.com/definition/association-rules-in-data-mining>

[9]. FIREBASE - <https://firebase.google.com/>

[10]. Android Application Development - [https:// developer.android.com/](https://developer.android.com/)

[11]. Technology" 2015 Seventh International Conference on Measuring Technology and Mechatronics Automation.