Accident Zone Detailed System Using Data Mining Technique

Pallavi s n[1], sumana k r[2],

[1]PG student, Department of MCA, NIE, Mysuru, India

[2]Assistant Professor ,Department of MCA ,NIE, Mysuru, India

---***--

Abstract - *Road accidents are a human tragedy. They involve high human suffering and monetary costs in terms of untimely deaths, injuries and loss of potential income. Road accident is one of the major issues in India. A variety of research has been done on data collected through police records covering a limited portion of highways. The analysis of such data can only reveal information regarding that portion only; but accidents are scattered not only on highways but also on local roads. We have used data mining techniques to analyze the data in which we first cluster the accident data and further association rule mining technique is applied to identify circumstances for each cluster. The results can be utilized to put some accident prevention.*

Key Words: *Data Mining; Road Accidents; Association Rule Mining.*

1. INTRODUCTION

Data Mining is a process that discovers the knowledge or hidden pattern from large databases. DM is known as one of the core processes of Knowledge Discovery in Database (KDD). It is the process that results in the discovery of new patterns in large data sets. It is a useful method at the intersection of artificial intelligence, machine learning, statistics, and database systems. It is the principle of picking out relevant information from data. It is usually used by business intelligence organizations, and financial analysts, to extract useful information from large data sets or databases DM is use to derive patterns and trends that exist in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems.

1.1 Motivation

The amount of data stored in computer files and databases is growing at a phenomenal rate. At the same time user of these data are expecting more sophisticated information from them. Data mining is the analysis step of the "knowledge Discovery in databases" process, or KDD). It is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

1.2 Problem Statement

Road and accidents are uncertain and unsure incidents. In today's world, traffic is increasing at a huge rate which leads to a large numbers of road accidents. The highway safety is being compromised and there are not enough safety factors by which we can analyse the traffic collisions before it happens. A method is proposed by which we can pre-process the accidental factors. Young drivers tend to be more daring and are unable to avoid a crush when they face one. They tend Knowledge Discovery Process. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD.

1.3 Objectives

DM is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics.

1.4 Scope

The goal of this technique is to find accurate patterns that were previously not known by us. So, the overall goal of the DM process is to extract information from a data set and transform it into an understandable structure for further use. Organizations like retail stores, hospitals, banks, and insurance companies currently using mining techniques.

2. LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy n company strength. Once these things r satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

2.1 Existing System

There are several major data mining techniques which have been developed and used in various data mining projects. In the proposed work, k –means performance will be enhanced by using hybrid approach for better result to show the effect of noise on the performance of various clustering techniques. Clustering may be applied on database using various approaches, based upon distance, density hierarchy and partition. Clustering is being widely used in many application including medical, finance etc. our purpose is to study how a particular clustering technique is responsive to the noise in the term of time. Apriori algorithm minimum support is needed to generate the large item set from candidate set in which not so required candidate item sets are pruned by utilizing user defined minimum support threshold. Moreover, in Apriori Algorithm, if the frequencies of items vary a great deal, we will encounter two problems. First of all, if minimum support is set too high, those rules that involve rare items will not be found. Secondly, to find rules that involve both frequent and rare items, min support has to be set very low. This may cause combination explosion because those frequent items will be associated with one another in all possible ways. So, Apriori is utilizing hit and trial method to find the required number of rules.

2.2 Proposed System

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

a. The size of cluster is fixed and the output of the first phase forms initial clusters. Here, the input array of elements is scanned and split up into sub –arrays, which represent the initial clusters.

b. 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. Furthermore, the data elements having less 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.

3. MODULES SPECIFICATIONS

There are 3 modules

- Admin
- User
- Reporter

Admin
- Admin will login using admin id and password.
- Admin will add the location based on city and area.

Reporter
- Reporter will register and login using id and password.
- Admin will assign the reporter to particular location [city and area].
- They can view the accidents and they will give Description prevention and reason. Why they met for the accident.
- Reporter will upload the accident data.

User
- User will register and login using user id and password.
- Once user login they can view the accident data uploaded [based on city and area] by reporter.
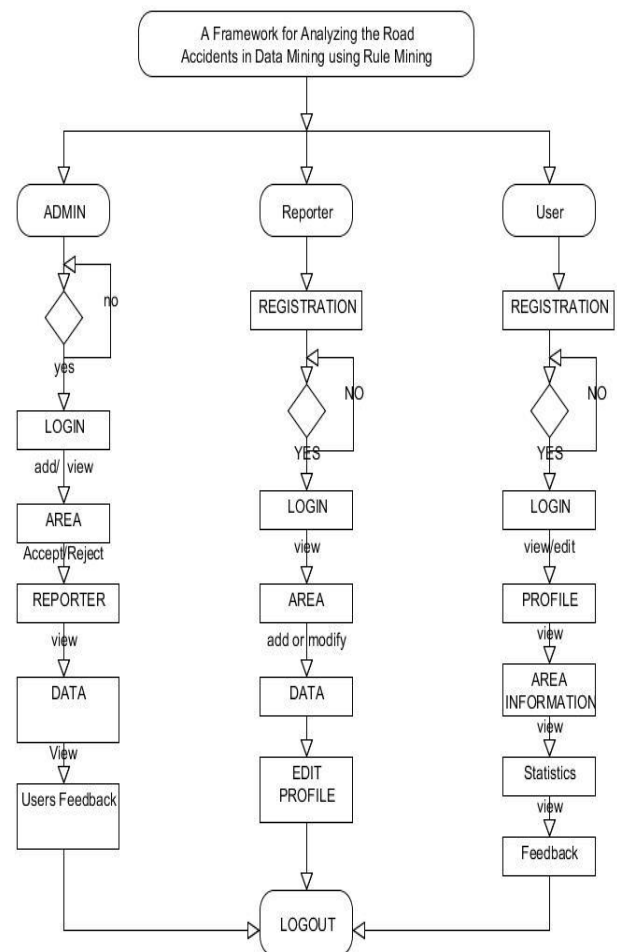- User can get know the accident details on particular area or city.

Fig 3.1 architectural design

4. CONCLUSIONS

To overcome the above issues, in this paper, we proposed for an analysing accident patterns for different types of accidents on the road which makes use of Hybrid clustering and improved association rule mining algorithm. Comparing and analysing the results of proposed technique with K means clustering and Apriori algorithm on the basis of clustering time, accuracy and association rule mining time. The analysis on road accidents in this study will help to create awareness, guidelines and assist in informed decision making on road safety.

4. SCREENSHOTS
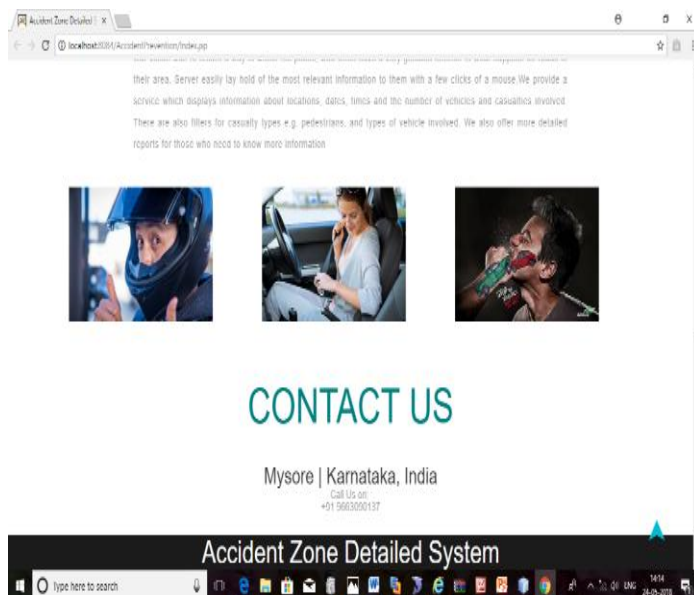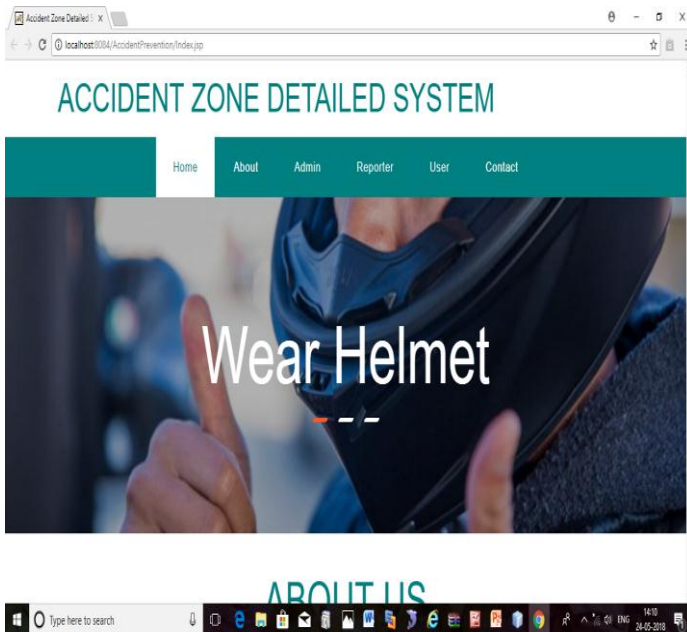
REFERENCES

[1] Seoung-hun Park and Young-guk Ha, "Large Imbalance Data Classification Based on Map Reduce for Traffic Accident Prediction", Eighth

International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, pp45-49,2014

[2] Evangels Gakis1, Dionysius Kehagias1 and Dimitrios Tzovaras1,"Mining Traffic Data for Road Incidents Detection", IEEE 17th international

Conference on Intelligent Transportation System (ITSC) October 8-11, 2014 Qingdao, Chine.

[3] An Shi, Zhang Tao, Zhang Xining, Wang Jian,"Evolution of Traffic Flow Analysis under Accidents on Highways Using Temporal Data Mining

", Fifth International conference On Intelligent System Design and Engineering Apllication,2014, pp-454-457.