ANALYSIS OF ROAD ACCIDENTS USING DATA MINING TECHNIQUES : A SURVEY

Dnymish Patil[1], Rohit Franklin[2], Sahil Deshmukh[3], Sarath Pillai[4], and Prof. Madhu Nashipudimath[5]

Department of Computer Engineering, Pillai College of Engineering, Navi Mumbai India - 410206

---***---

Abstract - *Road accidents are the main cause of death as well as serious injuries in the world. As a human being, everyone wants to avoid traffic accidents and stay safe. In order to stay safe, careful analysis of roadway traffic accident data is important to find out factors that are related to fatal, grievous injury, minor injuries, and non-injuries. The relationship between fatal rate and other attributes include combining weather conditions, road type, sunlight conditions, speed limit, drunk driver and so on are considered. Here, three data mining algorithms namely Apriori, Naïve-Bayes and K-Means are applied on the accident dataset to predict the accident severity. Apriori Algorithm is used for finding association rules between the attributes. Naïve-Bayes based approach is used for classifying how attributes are conditionally independent. K-means is used to form clusters and analyze them based on attributes. Comparison based on parameters is done to prove the efficiency of the various road accident detection techniques and approaches. By using this analysis, government/private agencies can take decisions in developing new roads and taking additional safety measures for the general public and awakening a sense of responsibility among road users.*

*Key Words***:** Road Accidents, Road conditions, Algorithms, Road accident detection systems.

1.INTRODUCTION

Road accident detection is considered to be the contemporary ever growing process focused primarily to reduce death. This study provides road accident detection techniques by analyzing the novel ideas. The analysis of these methods provides a better understanding of the steps involved in each process in a way of consequently increasing the scope for finding the efficient techniques to achieve maximum accurate performance. The comparison of the techniques used here, that is Apriori, Naive-Bayes and K-Means is carried out in terms of precision and recall. Environmental factors like roadway surface, weather, and light conditions do not strongly affect the fatal rate, while the human factors like being drunk or not, and the collision type, have a stronger effect on the fatality rate. From the clustering result we can see the states/regions which have a higher fatality rate, while some others lower. We should pay more attention when driving within these risky states/regions. Current system is manual where government sector makes use of this data and analyze it manually. Based on the analysis, they will take precautionary measures to reduce the number of accidents.

2. Literature Survey

This section reviews the research works carried out by different researchers that are related to the proposed work and provides a review of recent trends in motor vehicular accidents, factors influencing motor traffic safety and various methodologies commonly used in traffic safety studies. The literature review focuses primarily on the type of analysis available and the importance of determining and examining risk factors in general.

Liling Li, Sharad Shrestha and Gongzhu Hu [1] applied data processing algorithms on an outsized dataset for the analysis of road accidents. The link between fatal rate and different attributes together with collision manner, weather, surface condition, lightweight condition, and drunk driver were investigated. Association rules were discovered by Apriori rule, classification model was designed by Naive Bayes classifier and clusters were fashioned by straightforward K-means agglomeration rule.

Poojitha Shetty, S. P., Kashyap, S. V., & Madi, V [2] describes the way to mine frequent patterns inflicting road accidents from collected information. It finds associations among road accidents and predict the kind of accidents for existing also as for brand new roads. use of association and classification rules to find the patterns between road accidents and also as predicting road accidents for brand spanking new roads. Descriptive mining is applied on previous road accidents to mine frequent patterns in combos with alternative factors. Within the planned system, apriori rule is employed to predict the patterns of road accidents by analyzing previous road accidents

information. The results obtained from data processing approach will facilitate perceive the foremost important factors or usually continuance patterns. The generated pattern identifies the foremost dangerous roads in terms of road accidents and necessary measures will be taken to avoid accidents in those roads.

Atnafu, B., & Kaur, G [3] considered Deep Belief Network, supervised Latent Dirichlet Allocation, Support Vector Machine , Hybrid cluster, association rule mining, Random Forest, AdaBoostM1, Naïve mathematician, J48, PART, Preliminary real time autonomous accident detection system , Naive mathematician, C4.5, C&RT, RndTree,call list,rule induction, random tree, multi-class Support Vector Machine , Naive mathematician, J48, Random Forest algorithmic program, Apriori association rule mining, Decision Trees,Neural Networks, call tree and Support Vector Machines for analysis. The comparison is finished by the experimental results of the ways in terms of accuracy, precision, recall and F-measure. Attributes used square measure driver age, driver sex, vehicle category for result analysis.

Beshah, T., & Hill, S. [4] proposed that road traffic accidents (RTAs) are a significant public health concern, leading to associate calculable one.2 million deaths and fifty million injuries worldwide every year. Within the developing world, RTAs are among the leading reasons behind death and injury; Abyssinia above all experiences the very best rate of such accidents. The results of this study might be utilized by the individual stakeholders to market road safety. Whereas the ways are easy, the results of this work may have an incredible impact on the well-being of Ethiopian civilians. The algorithms used are Decision Tree (J48), Naive Bayes, K-Nearest Neighbors.

Durga Karthik, P. Karthikeyan, S.Kalaivani & K.Vijayarekha [5] proposed that road accidents are a major cause of death and were analysed based on data mining algorithms such as disabilities. It uses Naive Bayes, Random Forest and J48. The aim of the traffic accident analysis for a region is to investigate the cause for accidents and to determine dangerous locations in a region. Multivariate analysis of traffic accidents data is critical to identify major causes for fatal accidents.

Through literature survey, it was clearly observed that road accident analysis is tremendously important to cut back road accidents which motivated us to pick up this project. The target of our project aimed to beat all the shortcomings of the previous road accident analysis system.In our project,we are going to use Apriori, Naïve-Bayes and K-Means Algorithms to overcome the road accident issues.

2.1 Summary of Literature Survey

The summary of methods used in literature is given in Table 1.

Table 1 Summary of literature survey

Literature	Apriori	Naive Bayes	K-means
Analysis of Road Traffic Fatal Accidents using Data Mining Techniques	Yes	Yes	Yes
Analysis of Road Accidents using Data Mining Techniques	Yes	No	No
Survey on Analysis and Prediction of Road Traffic Accident Severity Levels using Data Mining Technique	No	Yes	Yes
...ining Road Traffic Accident ...ata to Improve Safety	No	Yes	Yes
Identifying Efficient Road Safety Prediction Model Using Data Mining Classifiers	No	Yes	No

3. Proposed Work

In order to predict the pattern of recent road accidents, associate association and classification data processing techniques are an area unit used, namely, Apriori, Naïve Thomas Bayes classifier and Kmeans, that area unit extremely scalable . although we have a tendency to area unit functioning on an information set with various records with some attributes, this classifier will yield best results.

There are unit models that assign category labels to downside instances, that area unit depicted as vectors of feature values, and also the category labels area unit drawn from some finite set. The info is collected from police stations that area unit restricted to a region.

3.1 System Architecture

The system architecture is given in Figure 1. Each block is described in this Section.
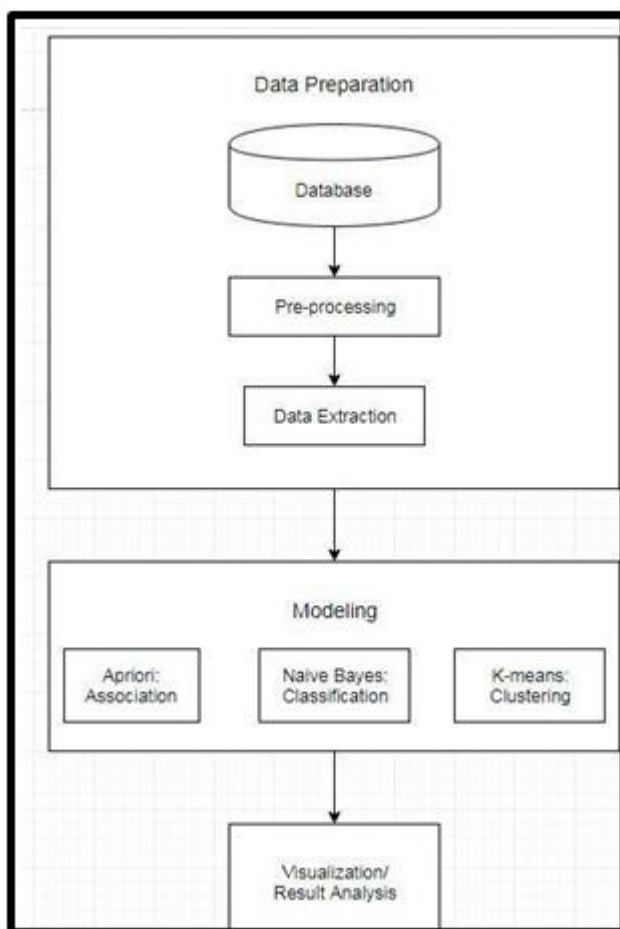

Fig. 1 Proposed system architecture

A. Data Preparation :

Data preparation is performed before each model construction. All records with missing value in the chosen attributes are removed. All numerical values are converted to nominal value according to the data dictionary in the attached user guide. Fatal rate is calculated and binned into two categories: High and Low. Preprocessing is done to remove unnecessary words from the dataset and make the

raw data for implementing algorithms on it.

Steps:-

1.Gather Data : The data preparation process begins with finding the right data. This can come from an existing data catalog or can be added when needed.

2.Discover and assess data : When grouping the info, it's vital to get every dataset. This step is concerning planning to recognize the info and understanding what needs to be done before the info becomes helpful in an exceedingly explicit context.

3.Cleanse and validate data : Improvement up the info is historically the foremost time intense a part of the info preparation method, however it's crucial for removing faulty knowledge and filling in gaps. Once knowledge has been cleaned, it should be valid by testing for errors within the knowledge preparation method up to the current purpose. Typically times, a blunder within the system can become apparent throughout this step and can get to be resolved before moving forward.

4.Transform and enrich information : Remodeling data is that method of changing the format or worth entries so as to succeed in a well-defined outcome, or to create the info simply understood by a wider audience. Enriching information refers to adding and connecting information with alternative connected info to supply deeper insights.

5.Store data : Once prepared, the data can be stored or channeled into a third party application—such as a business intelligence tool—clearing the way for processing and analysis to take place.

B. Modeling

Modeling is the process of making a knowledge model for the information to be stored in a very Database. This data model may be a conceptual representation of knowledge objects, the associations between different data objects and therefore the rules. Data modeling helps within the visual representation of knowledge and enforces business rules, regulatory compliance , and the government policies on the information. Data Models ensure consistency in naming conventions, default values, semantics, security while ensuring quality of the information. Data model emphasizes on what data is required and the way it should be organized rather than what operations have to be

performed on the information. Data Model is like architect's building plan which helps to create a conceptual model and set the connection between data items. We first calculate several statistics from the dataset to indicate the essential characteristics of the fatal accidents. We then apply Apriori, Naïve-Bayes and K-Means to search out relationships among the attributes and therefore, the patterns.

4. Requirement Analysis

Hardware and software details The experiment setup is carried out on a computer system which has different hardware and software specifications.

The hardware used are Processor Intel Core i7, HDD 1 TB, RAM 8 GB.

The software used are Operating Language - Windows 10, Programming Language - Java, Front-end - JTattoo, JFrame, JFreeChart, Database - MySQL and Back-end - Weka Tool, SQL.

5. Conclusion

The aim of this study was to show the application of data mining techniques in the field of accident investigation. It was done by reviewing various papers. We are currently enhancing it by considering several issues; variation in crash occurrence may have some consequence for traffic safety measures in several areas. The modeling will be to combine road related factors with driver information for better prediction, and to find interactions between the different attributes. The analysis of these methods provides a better understanding of the steps involved in each technique to achieve accurate results. The comparison of the techniques used here, that is Apriori, Naive-Bayes and K-Means is carried out. Environmental factors like roadway surface, weather, and light conditions do not strongly affect the fatal rate, while the human factors like being drunk or not, and the collision type, have a stronger effect on the fatality rate. It will be useful to the authorities as well as to the entire society for awareness. So the implementation will be done to analyse the road accidents.

Acknowledgment

References

[1] Li, L., Shrestha, S., & Hu, G. (2017, June). Analysis of road traffic fatal accidents using data mining techniques. In *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)* (pp. 363-370). IEEE.

[2] Poojitha Shetty, S. P., Kashyap, S. V., & Madi, V. (2017). Analysis of road accidents using data mining techniques.

[3] Atnafu, B., & Kaur, G. (2017). Survey on Analysis and Prediction of Road Traffic Accident Severity Levels using Data Mining Techniques in Maharashtra, India. *International Journal of Current Engineering and Technology*, *7*, 1974-1978.

[4] Beshah, T., & Hill, S. (2010, March). Mining road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia. In *2010 AAAI Spring Symposium Series*.

[5] Durga Karthik, P. Karthikeyan, S.Kalaivani & K.Vijayarekha (2019). Identifying Efficient Road Safety Prediction Model Using Data Mining Classifier.