

Predicting Relationship between Traffic Accidents with Related Injuries

Vidyashree N¹ and Pradeep Kumar H S²

¹⁻²The National Institute of Engineering, Department of Information Science & Engineering, Mysuru, India
Email: vidyashreenprabhu@gmail.com, pradeep@nie.ac.in

Abstract— Traffic accidents can be considered as a one of the most crucial problem because they causes many deaths, fatalities, injuries and economic losses. To predict traffic accident severity level a model is built in this inquiry to pick a collection of influential characteristics in order to construct a model for assessing the injury severity. Two separate unsupervised algorithms of machine learning are used to analyses traffic accident data. The major intension of this study is to analyse the relationship between three different patterns, namely, different forms of traffic accidents and different injuries types or between different types of the traffic accidents itself or else between different types of the injuries itself.

Index Terms— Road crashes, Injury and Patterns.

I. INTRODUCTION

In road accidents, there are different underlying reasons. Each reasons can be composed of several features of different types. With the primary purpose of extracting valuable perceptions from most recent data the predictive analytic is using in this model. To determine the underlying correlations between features and to identify the most important predictive characteristics techniques for classification are often employed. The same statistical space may exist with serious and fatal severities. As a result, there is a significant statistical gap between minor and deadly injuries. In contrast, lethal samples can be thought of as a rare and unpredictable event that has a negative impact on prediction. This type of infrequent instance can be resolved with the use of sampling techniques. The experiment follows a suitable methodology that starts with pre-treatment the data by comprehending and setting it, then building the models for evaluating the model performance. With key finding outlined implementation results are delivered. Technology allows a computer system to adapt and learn lessons information without being manually built by the developer, which is difficult to do because every condition on an element is tricky to manage.

A. Motivation

For helping traffic departments, doctors and public this model isbuilt has a real time applications with the major purpose is to identify the most common forms of traffic accidents as well as the various injuries associated with each type of accident in-order to reduce traffic accidents.The technology uses Association learning to detect patterns among traffic crashes and associated injury and intensity.

II. LITERATURE SURVEY

[1]The goal of this article is to recognize the most connected things that influence the degree of injuries sustained by drivers involved in traffic disasters on the highways for limiting those elements, best ways to ensure can be improved. In this article "Classification and Regression Tree (CART)" methods is used. The main drawback is that results are graphically shown, regression and classification models are not suited for practical applications.

[2]To use the CART model to find connections among symptom severities and vehicle properties, roadways elements, and related crash variables. Back propagation is a way of "rewinding of faults" in network of neural and logistic regression is essential when dealing with a data where the classes are more or less "discrete." The main drawback is that here, there are limited options and also it takes longer to establish a link between injury risk and motorist parameters.

[3]One of the primary goals of accident data analysis is to identify the major elements that contribute to road crashes. Because of the uniqueness of road fatalities data, investigation is difficult to come across uniqueness of data division is used. The described methodology uses the clustering method to segment data from road accidents. With this association rule mining is adapted to explore scenarios contributing towards occurrence of the entire data group as well as clusters found by this technique. Major information is produced when the results of K-means clustering and the Association rule are combined. The disadvantages are listed below

- Only categorization has been completed.
- Does not forecast the patterns of traffic collision injuries.

[4]Uses data mining techniques to relate documented accident, road and relevant elements to crash probability and came up with a group of guidelines for the police to improve safety. Methods can be used for a variety of tasks such as estimation, fault diagnosis, prognostics, automated insight, reasoning, time series prediction, and making decisions under uncertainty. Multi-layer Decoder wont to solve simple regression issues because it helps to urge a quick glimpse about deeper causes within the amazingly concepts of deep learning and also gather data about it. The Multi-layer Perceptron and Bayesian network methods are used. The disadvantage is that it produces inferior effective outcome.

III. PROPOSED SYSTEM

The recommended method is being developed as a real world applicant with the primary goal of reducing the number of road accidents and determining the degree of injuries, which will be beneficial to the government sector. Because traffic safety is such a vital element of our lives, it must be continually improved using all available possibilities and tools. The crash severity classifications into divided into fatal, serious, and minor using the data set. It will be explained in the suggested system how to utilize the obtained information on road accidents to track common patterns and critical elements affecting various sorts of road accidents and injuries. Furthermore, the technology will assist the transportation sector in effectively reducing traffic accidents and injuries. The numerous tasks done by Administrators and Doctors are depicted in Figure 1.

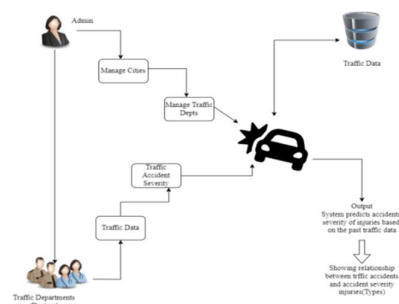


Figure 1 Working of the system

IV. METHODOLOGY

Machine Learning is primarily concerned with the design and analysis of machine to learn from information. Supervised, Semi-supervised and Unsupervised machine learning are the different types of this learning. However, both "Apriori", "Eclat" Unsupervised techniques are used in this research. The primary goal of using Unsupervised Machine Learning is to extract meaningful patterns from the data. Figure 2 depicts the operation of algorithms using a flowchart diagram.

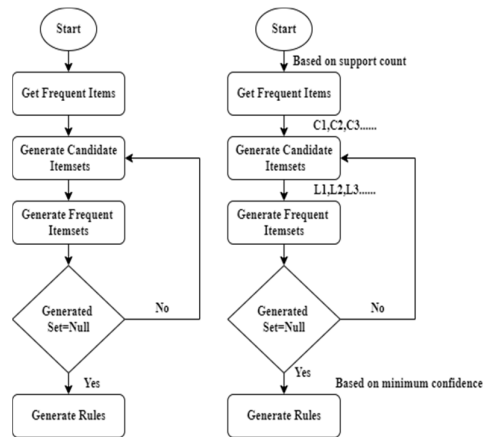


Figure 2 Flow chart of Apriori and Eclat algorithms

A. Unsupervised Learning

For tasks that would benefit from the perception gained through summarizing data in fresh and fascinating ways, a descriptive model is utilized. There are no specified labels in this technique. The purpose is to understand information and see whether there is any structure in it. This learning works effectively with synchronous data. The Apriori and Eclat algorithms are applied in this project to determine the association between injuries and accidents.

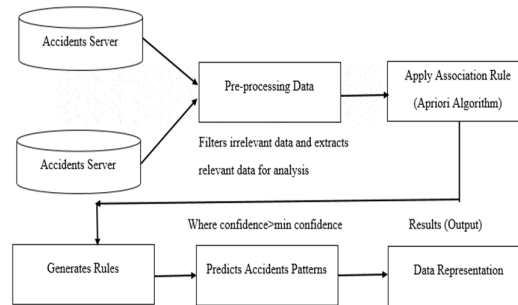


Figure 3 System Diagram

Figure 3 summarizes the overall functionalities of the system with the different components. The steps carried out by the system in order to predict injuries and accident patterns are as follows

Step 1: Data Collection

Data collection is the process of gathering information from various sources. Year, Speed Limit, Weather Conditions, Hospital Zone, Road Type, Men at Work, School Zone, Humps, Accident, and Injury are among the details included in the data.

Step 2: Preparation of data

Data will be retrieved and processed from servers by eliminating irrelevant data.

Step 3: Define the Limitations

Support count-The support count is a correlation between total number of transactions that contain that element (A) and the total volume of transactions in the set of data.

Confidence-The total volume of transaction holding the items set divided by the total amount of subject to regulations LHS is the reliability of the frequent item sets.

Step 4: Establishing Association Rules

A basic correlation between two or more elements will be built to discover patterns of the same sort. These algorithms, however, will be applied in this research to generate patterns relating to road injuries and accidents.

Step 5: Prediction of Patterns

The association between frequent road accidents and injuries types will be predicted using a system developed in this paper. Machine learning is a data-driven approach to understanding a system and it also a branch of data science that employs machine learning algorithms to efficiently handle data.

V. RESULTS AND DISCUSSION

The proposed system is developed like a real time application which is used to predict the patterns using both Apriori and Eclat algorithm respectively. In addition to that the comparative examination of both algorithms by its performance is shown in Figures 6.1 and 6.2 using a bar chart. By this figures it clearly shows that Eclat algorithm predicts the patterns at very less time compared to the Apriori algorithm. Hence Eclat algorithm more efficient compared to Apriori algorithm. Following snapshots shows the result of proposed work:

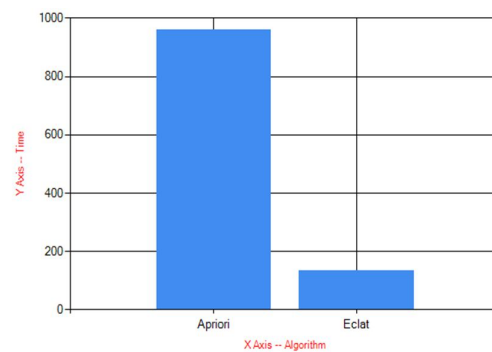
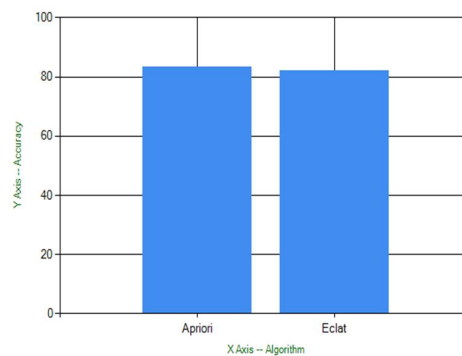


Fig: 6.1 Comparative analysis of Apriori and Eclat algorithm [Efficiency] Fig:

Graph Representation (Algorithm Vs Accuracy)!!!



6.2 Comparative analysis of Apriori and Eclat algorithm [Accuracy]

V. CONCLUSION

Road safety is frequently seen as an important aspect of our daily life. As a result, ongoing improvement is essential among all viable and suitable chances with exist resources. With predictive/descriptive mining, along with other key data such as weather or rather traffic conditions, gives exciting alternative with major beneficial and useful conclusions for available stakeholders. With these circumstances prompts towards the development of a strategy to evaluate accessible data sets documenting road accidents within the United Kingdom, which contained a huge amount of data and necessitated the uses of a relatively modern approach of in-memory execution in this area.

REFERENCES

- [1] [1] Kashani, Ali & Shariat, Afshin & Ranjbari, Andishe. (2011). A Data Mining Approach to Identify Key Factors of Traffic Injury Severity.
- [2] [2] Chang, Li-Yen, and Hsiu-Wen Wang. "Analysis of traffic injury severity: An application of non-parametric classification tree techniques."
- [3] [3] P. A. Nandurge and N. V. Dharwadkar, "Analyzing road accident data using machine learning paradigms," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017, pp. 604-610.
- [4] [4] Araar, Abdelaziz and Amira A. El Tayeb. "MINING ROAD TRAFFIC ACCIDENT DATA TO IMPROVE SAFETY IN DUBAI." (2013).
- [5] [5] D. Dinesh, "A Novel Multi-Model Machine Learning Approach to Real-Time Road Accident Prediction and Driving Behavior Analysis," 2021 International Symposium on Computer Science and Intelligent Controls (ISCSIC), 2021, pp. 67-72
- [6] [6] M. Evin, J. Taillard, H. L. De la Fuente, E. Galy and C. Berthelon, "Detection of functional state after alcohol consumption by classification and machine learning technics," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018, pp. 2237-2242
- [7] [7] A. Najah, M. Abuzwidah and D. Khalil, "The Impact of the Rear Seat Belt Use on Traffic Safety in the UAE," 2020 Advances in Science and Engineering Technology International Conferences (ASET), 2020, pp. 1-6.
- [8] [8] R. E. Al Mamlook, T. Z. Abdulhameed, R. Hasan, H. I. Al-Shaikhli, I. Mohammed and S. Tabatabai, "Utilizing Machine Learning Models to Predict the Car Crash Injury Severity among Elderly Drivers," 2020 IEEE International Conference on Electro Information Technology (EIT), 2020, pp. 105-111.
- [9] [9] M. Chong, A. Abraham, M. Paprzycki, "Traffic accident data mining using machine learning paradigms", Fourth International Conference on Intelligent Systems Design and Applications (ISDA'04), Hungary, 2004, pp. 415- 420.