

Road Accidents Analysis Using Machine Learning

¹Ramya S., ²Abhijna, ³Alfiya S., ⁴Kokila B. V., ⁵M. R. Thejaswini

¹Assistant Professor, ^{2, 3, 4 & 5}Student

^{1,2,3,4,5} Department of Computer Science, The National Institute of Engineering, Mysuru, India

Email - ¹ramyas@nie.ac.in, ²abhijnahosad@gmail.com, ³alfiya9247@gmail.com, ³kokilavacharya@gmail.com, ³thejaswinitheju835@gmail.com

Abstract: Traffic accidents are one among the foremost critical issues facing the earth as they cause many deaths, injuries, and fatalities also as economic losses once a year. Accurate models to predict the traffic accident severity could also be a critical task for transportation systems. This investigation effort establishes models to select a gaggle of influential factors and to make up a model for classifying the severity of injuries. These models are formulated by various machine learning techniques. Supervised machine learning algorithms and unsupervised machine learning algorithms are implemented on traffic accident data. The main objective is to get the correlation between differing types of the traffic accidents with the sort of the injuries. The findings of this study indicate that the unsupervised learning techniques are often a promising tool for predicting severity of traffic accidents.

Key Words: Traffic accidents, Eclat Algorithm, Machine Learning.

1. INTRODUCTION:

Traffic accidents are one among the most important problems within the world. There are several ongoing experiments in trying to decrease them. Road traffic accidents have different underlying reasons behind. These reasons are often several differing types of features. There are organizations focused on traffic accident understanding, and their proven strategies exist, which may reduce deaths. the first purpose of using predictive analysis is to extract enforceable insights from recent data. this system is often wont to recognize the samples, which can have gone unnoticed otherwise. Classification techniques are frequently wont to release underlying relationships between features and detect the essential predicting features. for instance, in our dataset, there's some irregularity between accident severities, which are slight injuries that have such a lot larger number of samples than serious injuries. and therefore, the same numerical space occurs again with serious and fatal severities. So, there's an enormous statistical difference with slightly versus fatal injuries. Consequential of this, deadly samples are rare and thought of as random events that affect the prediction negatively. This type of infrequent cases is often resolved with sampling techniques. The experiment follows a correct methodology which initially pre-processing the information with understanding and preparing it, construct the models, and model performance evaluation. Implementation results are delivered with key findings outlined. the continued parts are getting to be mentioning the approaches followed during this project. Machine learning (ML) allows a computer system to seek out and improve from past knowledge without directly programmed by the developer because it's hard to handle every situation on an object.

ML algorithms are generally supported mathematics and statistics. The most advantage of machine learning than traditional software is that there's no written code that shows the system the way to make the choice to settle on the proper object between two different ones because it's hard to handle every situation on an object. Machine learning works for that. It mainly uses the given data to form intelligent decisions, give future predictions, or to detect anomalies. Today it's utilized in various applications like autonomous cars, virtual assistants, program results, identifying objects, advertising, predictive analysis, etc. It is often categorized into three subfields called machine learning techniques are Supervised Learning, Unsupervised Learning, and Semi-supervised Learning.

1.1. PROBLEM STATEMENT:

Discovering the associations among the traffic accidents and related injuries is the key factor in reducing the traffic accidents. Identification of injuries severity is a key factor for the proper treatment. As number of traffic accidents are increasing and injuries severity is a critical factor to identify. Public suffering from many major injuries even after many years of accidents.

2. LITERATURE REVIEW:

A knowledge Mining Approach to identify Key Factors of Traffic Injury Severity [1]. The purpose of this study is to identify the most important factors which affect injury severity of drivers involved in traffic crashes on these roads, so that by eliminating or controlling such factors an overall safety improvement can be accomplished. The

methods used are Classification and Regression Tree (CART). Disadvantage is that it's not suitable for real time application because it displays outputs graphically. Analysis of traffic injury severity: An application of non-parametric classification tree techniques [2]. The aim of this study is to get the CART model to spot relationships between injury severity and driver/vehicle characteristics, highway/environmental variables, and crash variables. The methods used are Back-propagation Neural Network and 2 Logistic Regression. Disadvantages are less parameters used, less accurate results and it takes more time.

Mining Road traffic accident data to reinforce safety in Dubai [3]. This work applied data processing technologies to link recorded accident, driver, and road factors to accident severity in Dubai, and generated a set of rules that might be used by the Dubai Police to reinforce safety. The methods used are Multi-layer Perceptron and Bayesian network. Disadvantages are they are suitable for traffic accidents prediction but not injuries and less efficient results.

Road Crash Proneness Prediction using data processing [4]. This work analysed the crash and road data from the Queensland Department of Transport and Main Roads (Australia) that contained quite 42 000 records pertaining to years 2004 – 2007. The methods used are Decision trees with chi-square test, regression trees using ftest, logistic regression, neural networks, and Bayesian. Disadvantages are uses data mining technique, requires more amount of data.

2.1. EXISTING SYSTEM:

Current system is manual where government sector makes use of ledger data and analyse the data manually, based on the analysis they will take the precautionary measures to reduce the number of traffic accidents and related injuries. We also get many tools and software to maintain traffic accidents and related injuries, these tools just collect the data stores in sever but no analysis is done. Current system is a manual process where they maintain traffic accidents data and injuries severity using books or ledgers. Drawbacks of Existing System are Manual process, Time Consuming, Expensive, Store’s traffic accidents and retrieves the same, no extraction of useful information, Lack of user satisfaction, Less Efficient and Doesn’t discovers associations among traffic accidents and related injuries.

2.2. PROPOSED SYSTEM:

Proposed system is a real time application which is beneficial for state sector to scale back the amount of traffic accidents and to spot injuries. Traffic safety represents a crucial a part of our lives, so it's necessary to continuously improve within all possible and available opportunities and resources. Proposed system describes one possibility of the way to use the collected data about traffic accidents to mine frequent patterns and important factors causing differing types of accidents and related injuries. Thus, proposed system goal is to seek out the frequent traffic accident types, seek out different injuries associated with accident type and also 1. System aim is to scale back the amount of traffic accidents. 2. System will identify the foremost occurring accident type and related injuries. 3. System makes use of “Association Learning” to get the patterns between traffic accidents and related injuries.

3. SYSTEM DESIGN AND IMPLEMENTATION:

3.1. System Design: Fig. 1 describes the system architecture.

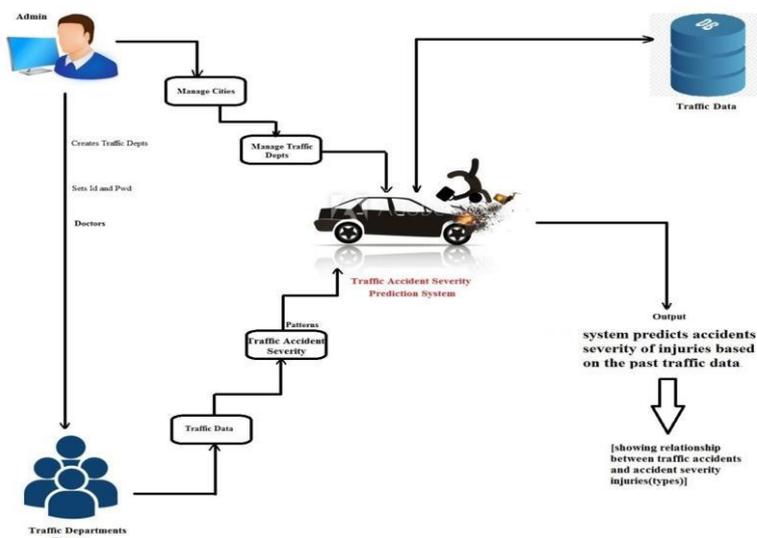


Figure 1 System Architecture

Application mainly consists three actors:

- Administrator: Administrator is a one who maintains the entire system. Administrator is responsible for member creation and uploading the necessary data for processing.
- Member: Member is a service receiver and responsible for uploading the traffic accidents data into the server.
- Visitor: Visitor is a user who visits the application. Visitor can view the most occurring accidents in a particular road of a city and take a look at its corresponding graph to take precautionary measures .

3.2. Implementation:

Application uses three-tier architecture which is a client-server architecture. It contains three tiers namely Data tier (SQL Server Stored Procedures), Business tier (C# Classes) and Presentation tier (ASP.NET Web Forms, Web User Controls, ASP.NET Master Pages). Three-tier architecture is a software design pattern and a well-established software architecture.

Step 1: Data Collection: We are performing on real time application; we build a replacement which contains data servers (used to store data). Data collection means collecting data from different sources. Data includes Accidents, road name, city name and Injuries.

Step 2: Data Preparation: Here data from servers extracted and analysed. Complete data is extracted and analysed where we remove irrelevant data and retain data required for processing. consistent with the project only accidents and injuries are required to get outputs.

Step 3: Specify Constraints

SUPPORT COUNT : The relationship between the entire number of transactions containing that item (A) with the entire number of transactions in data set.

CONFIDENCE : Confidence of item set is defined as total number of transactions containing the item set to the entire number of transactions containing LHS.

Step 4: Association Rules Mining (Eclat Algorithm): Association (or relation) is perhaps the higher known and most familiar and simple data processing technique. Here, we make an easy correlation between two or more items, often of an equivalent type to spot patterns.

For example, Market-basket analysis, where we track people's buying habits, we'd identify that a customer always buys cream once they buy strawberries, and thus suggest that subsequent time they buy strawberries they could also want to shop for cream. We use Eclat algorithm to process e commerce data and to seek out the patterns. Here we generate patterns associated with traffic accidents and injuries. Eclat algorithm is chosen due to Quicker results, works fine for little data set also as huge dataset, one scan of database is enough and also works fine for multiple constraints.

Step 5: Patterns Prediction

Here system predicts the connection between frequent traffic accidents with injury types. Machine learning may be a process of studying a system supported data. Machine learning may be a part of data science where we use machine learning algorithms to process data.

3.3. Unsupervised Learning:

A Descriptive model is employed for tasks that might enjoy the insight gained from summarizing data in new and interesting ways. There are not any predefined labels in unsupervised learning technique. The goal is to explore the info and find some structure with in. Unsupervised learning works well on transactional data.

Descriptive model developed using clustering techniques and association learning techniques. we've many efficient algorithms like "Eclat algorithm", "AIT algorithm", "SFIT algorithm", "STEM Algorithm", "FP Growth algorithm", "K Means algorithm", "Fuzzy C Means algorithm" etc....

In the project we use "Eclat algorithm" to seek out the connection between traffic accidents and injuries. Eclat algorithm is one among the efficient algorithms and takes less time for processing. This algorithm works fine for little datasets also as large datasets.

4. RESULTS:.

Figure 2, describes the most frequently occurring accidents in a particular road of a city which is accessible by the visitor.

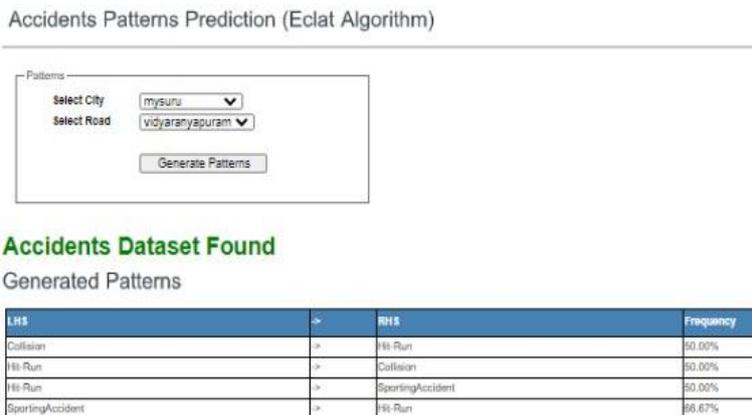


Figure 2: Frequently occurring Accident Patterns

Figure 3, describes the graph generated for the most frequently occurring accidents of a road in a city.

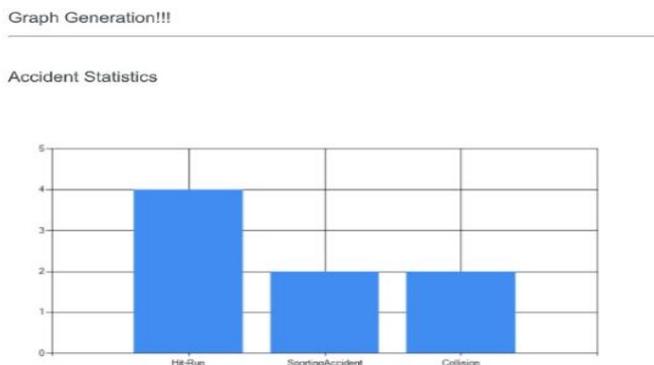


Figure 3: Graph Generation

Figure 4, describes the association among the traffic accidents and related injuries which is accessible to the member.



Figure 4: Pattern Prediction

5. CONCLUSION AND FUTURE WORK::

Road safety represents a crucial a part of our lives, so it's necessary to continuously improve within all possible and available opportunities and resources. Descriptive or predictive mining applied on historical data about occurred accidents together with other important information as weather or road conditions creates a stimulating alternative with potentially useful and helpful outcomes for all involved 4 stakeholders. These factors motivated the creation of this work to analyse available data samples describing road accidents in UK representing a quite great deal of knowledge which required the utilization of relatively new method in memory processing during this domain.

Future Work:

- We can add public Notifications which helps public.
- We can add query module for the interaction between administrator and member.
- We can predict reasons for accidents which helps traffic departments to require precautionary measures.

REFERENCES:

1. WHO | Road traffic injuries, 2017. WHO.
2. Michigan State Police, Michigan Traffic Crash DecadeAt-A-Glance, 2018.
3. 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.
4. T. Dejene, A. Ajith, V. Snášel, and P. Krömer, "Knowledge discovery from road traffic accident data in Ethiopia: data quality, ensembling and trend analysis for improving road safety", Neural Network World, vol.22, no. 3, 2012, pp. 215–244.
5. S. Krishnaveni and M. Hemalatha, "A perspective analysis of traffic accident using data mining techniques," International Journal of Computer Applications, vol. 23, no. 7, 2011, pp.40-48.
6. T. Beshah and S. Hill, "Mining Road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia," AAAI Spring Symposium, 2010.
7. G. Chen, Z. Zhang, R. Qian, R. A. Tarefder, and Z. Tian, "Investigating Driver Injury Severity Patterns in Rollover Crashes Using Support Vector Machine Models," Accident Analysis and Prevention, vol. 90, 2016, pp. 128–139.
8. S. Nazneen, M. Rezapour, and K. Ksaibati, "Determining causal factors of severe crashes on the fort peck Indian reservation," Journal of Advanced Transportation' Montana, 2018, pp. 1-8.
9. S. Dissanayake and U. Roy, "Crash Severity Analysis of Single Vehicle Run-off-Road Crashes," Journal of Transportation Technologies, vol. 4, 2014, pp. 1-10.
10. D. Shinstine, S. Wulff, and K. Ksaibati, "Factors associated with crash severity on rural roadways in Wyoming," Journal of Traffic and Transportation Engineering, vol. 3, no. 4, August 2016, pp. 308-323.
11. M. I. Ratnayake, "Effectiveness of seat belts in reducing Injuries: a different approach based on KABCO injury severity scale," Midwest Transportation Consortium, Iowa, 2006.
12. N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, "SMOTE: synthetic minority oversampling technique," Journal of Artificial Intelligence Research, vol. 16, 2002, pp. 321-357.
13. L. Lei and X.-D. Wang, "Improved ad boost ensemble approach based on loss function," Journal of Computer Applications, vol. 32, no. 10, 2013, pp.2916–2919.
14. G. James, D. Witten, T. Hastie, and R. Tibshirani, 2013. "An introduction to statistical learning with applications in R," Springer Texts in Statistics Series, New York, 2013.
15. Dong. C, Shao. C, Li. J, and Xiong. Z, "An improved deep learning model for traffic crash prediction," Journal of Advanced Transportation, 2018, pp.1-13.