International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

# PREDICTING ACCIDENT SEVERITY USING MACHINE LEARNING

**Hemanth Kumar V[1], Ashwini M[2]**

*[1]Dept. of ISE, The National Institute of Engineering, Mysore*
*[2]Asst. Professor, Dept. of ISE, The National Institute of Engineering, Mysore, Karnataka, India*

---

**Abstract –** *According to World Health Organization (WHO), every year the lives of approximately 1.35 million people are cut short as a result of a road traffic crash. Between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability as a result of their injury. Road traffic injuries cause considerable economic losses to individuals, their families, and to nations as a whole. Road traffic crashes cost most countries 3% of their gross domestic product. These effects can be reduced to a considerable amount by using modern technologies. Technology like Machine Learning which is a sub-branch of Artificial Intelligence (AI) provides different types of techniques that can be applied to existing traffic accident data set. This study establishes a procedure to select a number of influential factors and to build up a model for identifying relationship between different types of accidents and different types of injuries. Among different techniques of Machine Learning, unsupervised learning is employed wherein a model is built and devoured with required dataset of particular place and* ***Eclat algorithm*** *identifies and displays the patterns among different types of accidents as well as different types of injuries which can be used by public, traffic departments and doctors for analysis purpose.*

***Key Words*****: WHO, Non-fatal injuries, Economic loss, Machine Learning, Unsupervised Learning, Patterns, Eclat.**

## 1. INTRODUCTION

Motorization has enhanced the lives of many individuals and societies, but the benefits have come with a price. Although the number of lives lost in road accidents in high-income countries indicate a downward trend in recent decades, for most of the world's population, the burden of road-traffic injury—in terms of societal and economic costs—is rising substantially. Injury and deaths due to road traffic accidents (RTA) are a major public health problem in developing countries where more than 85% of all deaths and 90% of disability-adjusted life years were lost from road traffic injuries. The costs of fatalities and injuries due to road traffic accidents (RTAs) have a tremendous impact on societal well-being and socioeconomic development. Road traffic crashes result in the deaths of approximately 1.35 million people around the world each year and leave between 20 and 50 million people with non-fatal injuries. In addition to the human suffering caused by road traffic injuries, they also incur a heavy economic burden on victims and their families, both through treatment costs for the injured and through loss of productivity of those killed or disabled. More broadly, road traffic injuries have a serious impact on national economies, costing countries 3% of their annual gross domestic product.

In recent years, many researchers studied the impact of influencing factors of traffic accidents, mainly focusing on people, cars, roads or the environment. Some researchers [8-9] studied on driver's behavior and analyzed the characteristics of the process during changing the lanes to identify dangerous driving behaviors. There are also studies [10] on the impact of road conditions on traffic accidents, they proposed a point that high and steep roadbed will undermine the traffic safety. Also other studies [11] focused on the impact of weather or dynamic traffic flow on accidents. However, most of these studies focused on a single factor (people, cars, roads, and the environment) on the impact of traffic accidents. Another study [1] focuses on different parameters that are affecting the road accidents. In most of the cases, some type of injuries occurs in a pair with other type of injuries like broken bones and brain injuries. In this study, we'll propose a model that uses **Eclat algorithm** to show different pattern among the accidents as well as these types of related injuries. With the development of data mining technology, a variety of data mining approaches can be used to study the pattern in the road accidents. Among them, the association rule mining can be used to analyze the relationship between the influencing factors of traffic accidents. Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness. The strong association rules can be used to find a pattern hidden in the accident data. In order to select interesting rules from the set of all possible rules, constraints on various measures of significance and interest are used. The best-known constraints are minimum thresholds on support and confidence.

The rest of the paper is organized into Related Work, Methodology, Conclusions and References.

## 2. RELATED WORK

The study "**Research on Automated Modeling Algorithm Using Association Rules for Traffic Accidents** "focuses on different type of factors responsible for accidents. In this paper, the traffic accidents of Shanghai Expressway from April to June 2014 were excavated using association rule mining which generated lots of frequent item sets. The strong rules hidden in these frequent item sets often uncover the association between influencing factors of accidents, which can be used to reduce the occurrence of accidents by

breaking them. The rules can also be used to probe usual scenes of accidents, and some corresponding security improvement measures can be taken to prevent the accidents, and ultimately improve the city's traffic safety level. General speaking, association rule mining can produce tons of weak rules, the study first designed a method to calculate minimal Support value of training parameters, and further put forward a way to extract strong rules automatically. The results of the experiments showed that these methods proposed in the paper are effective. Therefore, an automatic modeling algorithm using association rules was finally established to promote the effective application of association rule mining on intelligent transportation system.

Tibebe Beshah, Shawndra Hill in their work, applied data mining technologies to link recorded road characteristics to accident severity in Ethiopia, and developed a set of rules that could be used by the Ethiopian Traffic Agency to improve safety.

The study "**Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity**" establishes models to select a set of influential factors and to build up a model for classifying the severity of injuries. These models are formulated by various machine learning techniques. Supervised machine learning algorithms, such as AdaBoost, Logistic Regression (LR), Naive Bayes (NB), and Random Forests (RF) are implemented on traffic accident data. SMOTE algorithm is used to handle data imbalance. The findings of this study indicate that the RF model can be a promising tool for predicting the injury severity of traffic accidents. RF algorithm has shown better performance with 75.5% accuracy than LR with 74.5%, NB with 73.1%, and AdaBoost with 74.5% accuracy.

The paper "**Analysis on Traffic Accident Injury Level Using Classification**" presents some models to predict the severity of injury using some data mining algorithms. The study focused on collecting the real data from previous research and obtains the injury severity level of traffic accident data.

## 2. METHODOLGY

Figure 1 summarizes the different steps involved in the prediction of patterns among different types of accidents and as well as injuries.

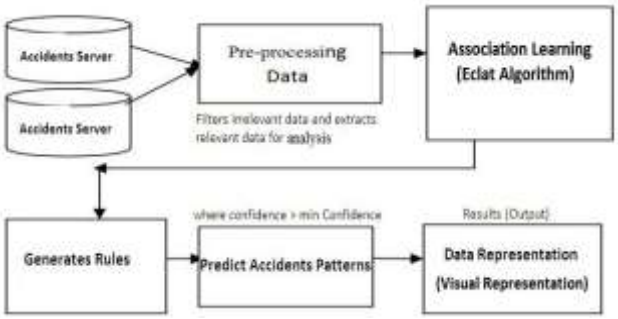


Figure 1: Architecture Diagram

### 1. Data Pre-processing

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Data preprocessing prepares raw data for further processing.

The dataset used in this study consists of several columns like year, Speed Limit, Weather Condition, Road Men at work etc. But we are only interested in accidents and injuries so all the irrelevant data are preprocessed (removed) and only relevant columns are taken into study.

### 2. Association Learning

Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness.

Several algorithms can be used to generating association rules. Some of popular algorithms are Apriori, Eclat and FP-Growth.

In this study we have used Eclat algorithm since it works efficiently for both small and large data sets, more efficient and scalable version of the Apriori algorithm.

### 2.1 Eclat Algorithm

The ECLAT algorithm stands for Equivalence Class Clustering and bottom-up Lattice Traversal. It is one of the popular methods of Association Rule mining. ECLAT algorithm works in a vertical manner just like the Depth-First Search of a graph.

When using the algorithm, we have considered two indicators to evaluate the rules, namely Support and Confidence. Support is the frequency of which frequent items appear in a transaction set. Assuming that there are frequent

item X and the transaction set D, then the Support of X presents the frequency of which X appears in D. For a rule like X => Y in which X is called Left-Hand-Side (LHS) and Y is called Right-Hand-Side (RHS), its Support is shown in Equation 1. The Confidence is the degree of trustworthiness of association rules, as shown in Equation 2.

Support(X=>Y) = Support (XUY) (1)
Confidence(X=>Y)=P(Y|X)=Support(XUY)/Support(X) (2)

In the implementation of algorithm, we have set minimum support count as 2 and confidence as 80%. The data from dataset contains different types of information such as types of accidents like Hit n run, Over Speed, Inexperience etc. and different types of injuries like Broken Bones, Brain Injuries, Rib Fracture, Spine Fractures etc. This information serves as input to the Eclat algorithm as shown in figure.

## 3. Generating Rules and Pattern Prediction

The input above mentioned are processed by the algorithm and based on the constraints the weak rules are rejected and algorithm generates strong association rules through several iterations. These strong association rules are then selected for confidence of 80%, thus results in different types of pattern among different types of accidents and injuries.

## 4. Result Analysis

The proposed model gives the pattern among different types of accidents and as well as injuries like the relation between brain injuries and broken bones or like relation between over speed and Hit & Run etc. In this study we've set minimum support count to 2 and confidence as 80%.Depending on the dataset these parameters can be set to different level and hence more precise results can be obtained.

## 3. CONCLUSION

In this paper, we used the association rules to analyze the relationship between the influencing factors of traffic accidents and proposed a model which provides the different patterns among accidents and injuries. In contrast with the previously published work of the authors, here we focused not only on the type of accidents but also on the different injuries that occur together more frequently. The results help in studying the accidents severity. The results of this study could be used by the respective stakeholders to promote road safety. While the methods are simple, the results of this work could have tremendous impact. The next step in the modeling will be to combine road-related factors with driver information for better predictions, and to find interactions between the different attributes.

## 4. ACKNOWLEDGEMENT

We are proud and grateful to have been given an opportunity to present an idea such as ours. To our principal, Dr.Rohini Nagapadma, we are extremely grateful for her kind permission to carry out our work. Our sincere thanks to the Department of Information Science and Engineering, NIE, Mysuru.

I'm indebted to Ms.Ashwini M, our project guide for her constant support and valuable insights to the project.

## 5. REFERENCES

[1] Zhen Gao, Ruifeng Pan, Xuesong Wang, Rongjie Yu, "Research on Automated Modeling Algorithm Using Association Rules for Traffic Accidents", 2018 IEEE International Conference on Big Data and Smart Computing.

[2] Tibebe Beshah, Shawndra Hill, "Mining Road Traffic Accident Data to Improve Safety: Role of Road-related Factors on Accident Severity in Ethiopia".

[3] Rabia Emhamed Al, Keneth Morgan Kwayu, Maha Reda Alkasisbeh, Abdulbaset Ali Frefer, "Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity", 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT).

[4] K. Geetha, C. Vaishnavi,"Analysis on Traffic Accident Injury Level Using Classification", Volume 5, Issue 2, February 2015, ISSN: 2277 128X, International Journal of Advanced Research in Computer Science and Software Engineering.

[5] Vichika Iragavarapu, P.E., Dominique Lord, Ph.D, P.Eng., Kay Fitzpatrick, Ph.D., P.E.," ANALYSIS OF INJURY SEVERITY IN PEDESTRIAN CRASHES USING CLASSIFICATION REGRESSION TREES". Submitted to the Transportation Research Board 94th Annual Meeting January 11-15, 2015, Washington D.C.

[6] World Health Organization (WHO).

[7] Worrall R D, Bullen A G R. AN EMPIRICAL ANALYSIS OF LANE

[8] CHANGING ON MULTILANE HIGHWAYS[J]. Highway Research Record, 1970(303).

[9] Xu H, Cheng G, Pei Y. Study on effect of lane-changing behavioral characteristic to capacity [J]. Sciencepaper Online, 2010.

[10] Yulong P, Ji M, Research on countermeasures for road condition causes of traffic accidents [J]. China Journal of Highway and Transport, 2003, 16(4):77-82.

[11] Das S, Sun X. Investigating the Pattern of Traffic Crashes Under Rainy Weather by Association Rules in Data Mining [J]. Transportation Research Board Annual Meeting, 2014.