Accident Severity Prediction using Machine Learning

Nischitha M[1], Pavankumar B Jois[2], Sheethal C J[3], Shreya S[4], Suman Jayakumar[5]

Student, Department of Information Science and Engineering[1,2,3,4]

Assistant Professor, Department of Information Science and Engineering[5]

Vidya Vikas Institute of Engineering & Technology, Mysuru, Karnataka, India

Affiliation to Visvesvaraya Technological University

Abstract: *Traffic accidents are among the most critical issues facing the world as they cause many deaths, injuries, and fatalities as well as economic losses every year. Accurate models to predict the traffic accident severity is a critical task for transportation systems. This investigation effort 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 and unsupervised machine learning algorithms are implemented on traffic accident data. The major objective is to discover the correlation between different types of the traffic accidents with the type of the injuries. The findings of this study indicate that the unsupervised learning techniques can be a promising tool for predicting the injury severity of traffic accidents.*

Keywords: Accidents, Supervised machine learning

I. INTRODUCTION

Traffic accidents are one of the biggest problems in the world. There are several ongoing experiments in trying to decrease them. Road traffic accidents have different underlying reasons behind. These reasons can be several different types of features. There are organizations focused on traffic accident understanding, and their proven strategies exist, which can reduce deaths. The primary purpose of using predictive analytic is to extracting enforceable insights from recent data. This technique can be used to recognize the samples, which may have gone unnoticed otherwise. Classification techniques are frequently used to release underlying relationships between features and pick out the essential predicting features. For example, in our data-set, there is some irregularity between accident severities, which are slight injuries that have so much larger number of samples than serious injuries. And the same numerical space occurs again with serious and fatal severities. So, there is a huge statistical difference with slightly versus fatal injuries. Consequential of this, deadly samples are rare and considered as random events that affect the prediction negatively. This kind of infrequent cases can be solved with sampling techniques. The experiment follows a proper methodology which at first pre processing the data with understanding and preparing it, construct the models, and model performance evaluation. Implementation results are delivered with key findings outlined. The ongoing parts will be mentioning the approaches followed in this project. Machine learning (ML) allows a computer system to learn and improve from past knowledge without directly programmed by the developer because it is hard to handle every situation on an object.

ML algorithms are generally based on mathematics and statistics. The main advantage of machine learning then a traditional software is that there is not a written code that shows the system how to make the decision to choose the right object between two different ones because it is hard to handle every situation on an object. Machine learning works for that. It mainly uses the given data to make intelligent decisions, give future predictions, or to detect anomalies. Today it is used in various applications such as autonomous cars, virtual assistants, search engine results, identifying objects, advertising, predictive analysis, etc. It can be categorized into three sub-fields called machine learning techniques are Supervised Learning, Unsupervised Learning, and Reinforcement Learning.

II. 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.

Solution: Proposed system is a real time application which is useful for government sector.It manages the traffic data such as accident type, injury types. It finds relationship between traffic accidents and injury types.

III. LITERATURE SURVEY

A. Data Mining Approach to Identify Key Factors of Traffic Injury Severity

Description: 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.

Methods Used: Classification and Regression Tree (CART).

Disadvantages: not suitable for real time application as it displays outputs graphically.

B. Analysis of Traffic Injury Severity: An Application of Non-Parametric Classification Tree Techniques

Description: The purpose of this study is to generate the CART model to identify relationships between injury severity and driver/vehicle characteristics, highway/environmental variables and crash variables.

Methods Used: Back-propagation Neural Network and Logistic Regression.

Disadvantages: Less parameters used, less accurate results, takes more time.

C. Mining Road Traffic Accident Data to Improve Safety in Dubai

Description: This work applied data mining technologies to link recorded accident, driver, and road factors to accident severity in Dubai, and generated a set of rules that could be used by the Dubai Police to improve safety.

Methods Used: Multi-layer Perceptron and Bayesian network

Disadvantages: suitable for traffic accidents prediction but not injuries, less efficient results.

D. Road Crash Proneness Prediction using Data Mining

Description: This work analyzed the crash and road data from the Queensland Department of Transport and Main Roads (Australia) that contained more than 42 000 records referring to years 2004 – 2007.

Methods Used: Decision trees with chi-square test, regression trees using f-test, logistic regression, neural networks and Bayesian.

Disadvantages: uses data mining technique, requires more amount data.
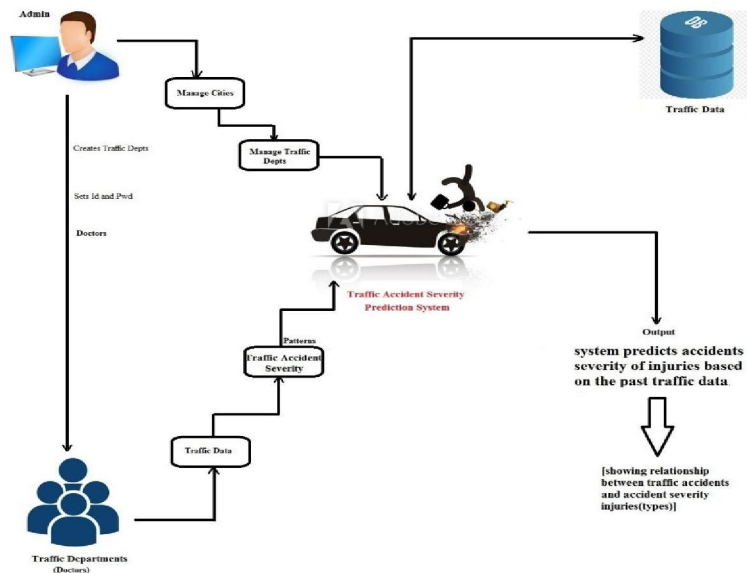
IV. PROPOSED SYSTEM

Fig 1: Proposed system for Accident severity prediction

Proposed system is a real time application which is useful for government sector to reduce the number of traffic accidents and to identify injuries severity. Traffic safety represents an important part of our lives, so it is necessary to continuously improve within all possible and available opportunities and resources. The data-set was classified into three accident severity categories, which are fatal, serious, and slight. Proposed system describes one possibility of how to use the collected data about traffic accidents to mine frequent patterns and important factors causing different types of accidents and related injuries. System helps traffic sector to reduce the traffic accidents and related injuries.

V. METHODOLOGY

ML concerns with construction and study of system that can learn from data. For example, ML can be used in E-mail message to learn how to distinguish between spam and inbox messages. There are there types of Machine learning(ML), they are:

1. Supervised Machine Learning: Here we have labels and the input is past examples.
2. Unsupervised Machine Learning: Extraction of patterns without labels.
3. Semi-Supervised Machine Learning: Mixture of both Supervised and Unsupervised Machine Learning.

5.1 Unsupervised Learning

A Descriptive model is used for tasks that would benefit from the insight gained from summarizing data in new and interesting ways. There are no predefined labels in unsupervised learning technique.The goal is to explore the data and find some structure with in. Unsupervised learning works well on transactional data.

Descriptive model developed using clustering techniques and association learning techniques. We have many efficient algorithms such as "*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 find the relationship between traffic accidents and injuries. Eclat algorithm is one of the efficient algorithm and takes less time for data processing. This algorithm works fine for small data-sets as well as large data-sets.

A. Accidents and Injuries Pattern Prediction Process

Step 1: Data Collection

We are working on real time application, we build a new application which contains data servers (used to store data). Data collection means collecting data from different sources. Data includes Year, Speed Limit, Weather-condition, School-zone, Humps, hospital zone, road type, men at work, Accidents and Injuries.

Step 2: Data Preparation

Here data from servers extracted and analyzed. Complete data extracted and analyzed where we remove irrelevant data and retain data required for processing. According to the project only accidents and injuries are required to generate outputs.

Step 3: Specify Constraints

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

CONFIDENCE**:** Confidence of item set defined as total number of transaction containing the item set to the total number of transaction containing LHS.

Step 4: Association Rules Mining (Eclat Algorithm)

Association (or relation) is probably the better known and most familiar and straightforward data mining technique. Here, we make a simple correlation between two or more items, often of the same type to identify patterns. For example, Market-basket analysis, where we track people's buying habits, we might identify that a customer always buys cream when they buy strawberries, and therefore suggest that the next time that they buy strawberries they might also want to buy cream. We use eclat algorithm to process e commerce data and to find the patterns. Here we generate patterns related to traffic accidents and injuries.

Eclat algorithm is selected because of the following reasons.

1. Quicker Results (takes less time for Prediction)
2. Works fine for small data set as well as Huge data set.
3. One scan of Database is Enough.
4. Works fine for multiple constraints.

Step 5: Patterns Prediction

Here system predicts the relationship between frequent traffic accidents with injury types. Machine learning is a process of studying a system based on data. Machine learning is a part of data science where we use machine learning algorithms to process data.

VI. EXPERIMENTAL RESULTS

Patterns generated (showing relationship b/w accidents and injury types)

- Single Car Accident → Skull & Maxillofacial
- Sporting Accident → Spine Fracture
- Drunkn Drive → Back and Spinal Cord Trauma
- Hit-Run → Broken Bones, Brain Injuries
- Hit-Run →Brain Injuries

VII. CONCLUSION

Road safety represents an important part of our lives, so it is necessary to continuously improve within all possible and available opportunities and resources. Descriptive or predictive mining applied on historical data about occurred accidents in combination with other important information as weather or road conditions creates an interesting alternative with potentially useful and helpful outcomes for all involved stakeholders. These factors motivated the creation of this work to analyze available data samples describing road accidents in UK representing a quite large amount of data which required the use of relatively new method in-memory data processing in this domain.

VIII. FUTURE ENHANCEMENTS

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 take precautionary

REFERENCES

[1]. WHO | Road traffic injuries, 2017. WHO.
[2]. Michigan State Police, Michigan Traffic Crash Decade-At-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 onaccident 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 thefort 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.