Analysis and Prediction of Road Accident Using Machine Learning

Ms. Pushpanjali M K[1] Sanjay Kumar S[2] Pavan Kashyap M V[3] Aditya P[4] Abhishek M R[5]
[1]*Assistant Professor, Department of Information Science and Engineering, East West Institute of Technology, Visvesvaraya Technological University, Karnataka*
[2,3,4,5]*Student, Department of Information Science and Engineering, East West Institute of Technology, Visvesvaraya Technological University, Karnataka*

***Abstract* - Traffic-related deaths and injuries have become a global phenomenon. Many countries around the world, including Thailand, are concerned about the rising prevalence of traffic fatalities and injuries. Road traffic accidents claimed the lives of 10,576 Thai citizens in 2016 [16]. Furthermore, according to the World Health Organization (WHO), 20 to 50 million people were injured non-fatally in 2016 [19]. Understanding the elements that influence road traffic accidents is a crucial component of road safety research. The purpose of this study is to provide an overview of elements that influence the severity of road traffic accidents, as well as to examine approaches that have been employed in past studies. According to a survey of the literature, the most commonly mentioned components that are found to be significant to the severity of road traffic accidents are vehicle speed and human traits. Vehicle types, weather, alcohol intake, driver weariness, and other factors have all been determined to be crucial.**

***Index Terms* - road accident analysis, risk factors, data analysis, Road safety.**

I.INTRODUCTION

Road accidents are without a doubt the most common cause of property damage. It's one of the most common reasons for fatalities. The causes for this are the severely congested highways and the drivers' relative freedom of movement. Accidents involving big cargo vehicles and even commercial vehicles colliding with public transportation, such as buses, are among the most deadly, claiming the lives of innocent people. There are numerous inventories in the automobile industry to create and manufacture safety features for autos, but traffic accidents continue to be a problem in both urban and rural locations. By constructing effective prediction models capable of automatically separating distinct unintentional occurrences, patterns involved in diverse conditions can be recognised. These clusters will aid in the prevention of accidents and the development of safety measures. Traffic accidents have a tremendous impact on society because of the high expenses of fatalities and injuries.

In recent years, academics have focused their efforts on determining the major impact of the severity of driver injuries caused by road accidents. The efficient use of accident records is contingent on a number of criteria, including data accuracy, data retention, and data analysis. Every year, more than 150,000 people are killed in automobile accidents in India. That's around 400 deaths each day, significantly more than industrialised motor markets like the United States, which had about 40,000 in 2001. Every year, over 1 million automobiles are added to traffic, resulting in an average of 1.2 million deaths and over 50 million injuries in road accidents around the

world. According to traffic studies, the number of road accidents and deaths due to laceration would rise.

on these assumptions, will reduce road accidents. It will be helpful to have an assumption system that is prepared with available data and new hazards. Highways are always a vulnerable place for these kinds of collisions that result in injuries and deaths. Rain, fog, and other meteorological conditions all play a role in amplifying the likelihood of accidents. Knowing the hotspots for accidents and the elements that contribute to them will assist to reduce them. It is necessary to provide quick emergency assistance even when casualties have happened, and this necessitates a thorough investigation of incidents.

Despite the existence of specified standards and highway codes, people's disregard for vehicle speed, vehicle condition, and their own recklessness in not wearing helmets has resulted in several accidents. If everyone had followed the rules, these accidents would not have turned fatal and cost the lives of innocent people. Road accident prevention is critical, and it will be bolstered by strict laws, technical and police controls, more rigorous training for drivers before issuing a driver's licence and raising public awareness of the importance of following the rules by imposing penalties and legalities on those who break them.

II. LITERATURE REVIEW

Many researchers looked on RTA from various angles and used various methods and algorithms. In India, Sonal et al. [2] utilised linear regression to study road incidents. The report offered a framework for analysing factors that influence the severity of accidents. Jayasudha et al. [31] looked at open data sources for traffic accidents as well as data mining tools and approaches for predicting the cause of accidents. They also employed visualisation to investigate traffic accidents in the United Kingdom, using multiple correspondence analysis to minimise the amount of variables in the graphic. The data is represented in two-dimensional graphs in this study.

In a UK traffic accident database, Connor et al. [4] evaluated different forms of decision trees to predict death. CHAID trees, C5.0, and Bayes net were employed in the study. Once sample approaches are implemented, classification techniques can be used to anticipate uncommon events, according to the study. Staines et al. [5] looked at the link between the type of road, or driving conditions, and the racial disparity. Clustering was done using statistical tools and the K-Means technique. The study discovered a reversal association between the kind of road and the meteorological condition and the likelihood of an accident.

On heterogeneous accidents data (4570 records) from Haridwar, Uttarakhand, Kumar et al. [7] employed LCC and k-modes clustering techniques to investigate their impact on performance and accuracy. They came to the conclusion that both strategies functioned effectively in processing data with little variation in computational speed. The cumulative logistics model, neural networks model, and Bayesian networks model were compared by Jin and Deng [8] to evaluate the degree of traffic infractions on traffic violation data from Guangzhou 2015. The Bayesian networks model fared better in predicting the level of traffic offences, according to the researchers. The Bayesian networks model fared better in predicting the level of traffic offences, according to the researchers. This was determined by recording 70% for Bayesian networks, 47% for the cumulative logistics model, and 51% for the neural networks model.

Deep learning has shown considerable success in uncovering complicated structures in large and multi-dimensional data over the last few years. Deep learning has been employed as a state-of-the-art technique in object detection and classification from photos for the past five years. It's also utilised in things like natural language processing, crowd management, and intelligent transportation systems. Wu et al. [9] presented a variety of case studies in which deep learning models were employed to forecast traffic flow. Ma et al. [10] looked at a number of large-scale transportation systems that used a deep learning model to forecast network congestion. The deep learning approach is utilised to adjust the timings of traffic signals in a study by Li et al. [1 l].The traffic light timings are controlled using a deep learning approach. The timing of the signals is automatically modified using this method dependent on traffic flow and pedestrian availability.

Ren et al. [12] analysed the spatial and temporal patterns of accident data to offer a spatiotemporal correlation of traffic accidents. They built a model to anticipate accidents and risks using a recurrent neural network. Traffic accidents and their spatial-temporal patterns teach the model.There is no work in the literature that we are aware of that describes storing traffic accident data in a big data platform.

III.DATAMANAGEMENT

Figure 3.1. shows the layers of the data management. The traffic accident data table contains 275,697 traffic accidents along with other relational tables, such as, district, roads, light, weather, vehicle and causalities. The following subsections describes each layer in more detail.

A. Data Collection

Data collection is essential for assessing the true nature of data. From 2014 to 2016, open source data from the UK Road Transport Authority (RTA) was used in this study. The following attributes are present in each row of the accident table: (accident id, latitude,longitude,accident severity, number of vehicles, number of casualties, data, time,district, road number, road speedlimit, vehicle speed, weather condition, light condition, road surface condition, special condition, urban or rural area, vehicle type, vehicle manoeuvre, skidding and overturning, journey urpose of driver, sex driver, age driver, engine_ casuality_age, casuality_ severity, pedestrian location, pedestrian movement, car passengers,

The available data is not cleaned properly and contains lots of noisy empty fields.

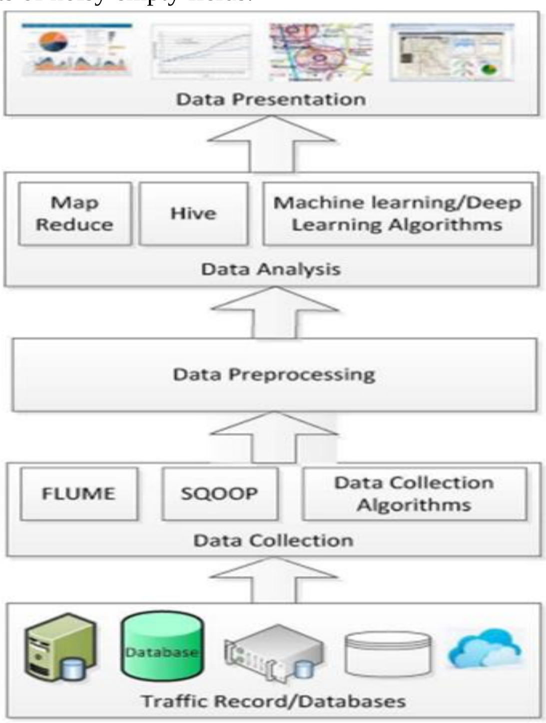

Fig 3.1: Data Management

B. Data Preprocessing

In data analysis applications, data preparation is the most time-consuming task. Because data is rarely available in a uniform format, work must be done to improve and transform it into the format necessary. Because the UK accident data received from many sources contains numerous errors, redundant, blank, non-relevant attributes, blank rows, and defective values, the data must be cleaned and reformatted before being used by various data mining methods. Some DM algorithms, for example, demand that all continuous variables be quantized into ranges, while others refuse to accept non-numeric values. Before going on to the following phases of the data mining process, the wrong or missing information must be repaired, destroyed, or accounted for during data cleansing.

C. Data Transformation

The process of transforming raw data into a set of new attributes that "make sense" when assessing the fundamental aspects of data is known as data transformation. In our scenario, we do a cleansing procedure on UK traffic data, where some values are turned into more precise attributes, such as the attribute "Full Date," which has been transformed into Year, Month, Date, Weekday, and WeekNo. The variable "Incident Time" has been converted to Hours and Minutes, and from Hours, the new attribute "Part of the Day" has been created, which divides the entire day into five zones: "Weekend," "formal," "Morning," "MidDay," and "Evening Rush." Data transformation will allow us to dig deeper into the data.

IV.PROPOSED SYSTEM OVERVIEW

The traffic accident data was kept in MS SQL server dump format, and retrieving the data took 10-15 minutes using a basic query. We moved the data from MS SQL server to the bigdata cluster to make the platform more efficient. Figure 4.1 depicts a high-level perspective of a suggested platform for analysing traffic accident data that we built.

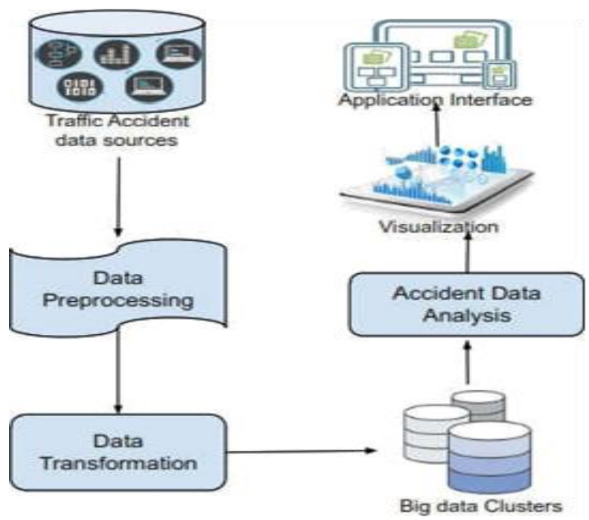

Figure 4.1: Overview of the Platform

The traffic accident data was kept in MS SQL server dump format, and retrieving the data took 10-15 minutes using a basic query. We moved the data from

MS SQL server to the bigdata cluster to make the platform more efficient. Figure 4.1 depicts a high-level perspective of a suggested platform for analysing traffic accident data that we built.
The goal of the study is to use a deep neural network to analyse data from traffic incidents. As a proof of concept, the proposed model will be used on the UK dataset. The model will then be applied to data from traffic accidents in Saudi Arabia.

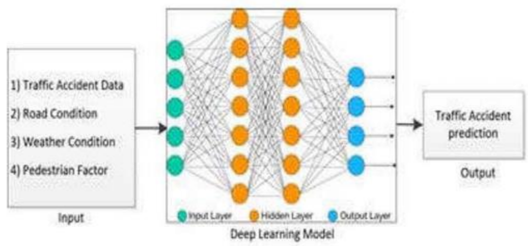

Figure 4.2: Flowchart of Traffic accident prediction
The proposed methodology would examine the spatial and temporal patterns of traffic accidents, producing a spatiotemporal correlation of data and identifying hotspots. The results of the accident analysis will be visualised in the visualisation section. The graphic conveys a clear picture of the suggested deep neural model's output.

V.CONCLUSION AND FUTURE WORK

In many nations, road accidents are the leading cause of death. Saudi Arabia has the largest number of accidents per year. Predicting traffic accidents and identifying hotspots on the road can help reduce the number of accidents in the future. In this paper, we provide a system for transferring traffic data to a big data platform.
We are using traffic data from the United Kingdom as a test case and developing a prototype based on deep learning to solve difficult computational models and learn several layers from raw data.

REFERENCES

[1] Mansuri FA, AI-Zalabani AH, Zalat MM, Qabshawi RI. Road safety and road traffic accidents in Saudi Arabia. A systematic review of existing evidence. Saudi Med J. 2015 ;36(4):418-424. doi:10.I5537/smj.2015.4. 10003.
[2] S. Sonal and S. Suman, "A framework for analysis of road accidents", 2018 Intemational Conference on Emerging Trend s and Innovations In Engineering And Technological Research, ICETIETR 2018, pp. 1-5, 2018.
[3] K. Jayasudha, "An overview of data mining in Road Traffic and Accident Analysis," Computer, vol. II, no. 4, pp. 32-37, 2009.
[4] O'Connor, A "An analysis ofthe predictive capability ofC5.0 and Chaid decision trees and Bayes net in the classification offatal traffic accidents in the UK". DIT 2015.
[5] A. T. Staines, "Environmental Conditions and Road Traffic Collisions in the UK," unpubl ished2016.
[6] S. Haynes, P. C. Estin, S. Lazarevski, M. Soosay, and A. L. Kor, "Data Analytics: Factors of Traffic Accidents in the UK," Conference Proceedings of 2019 10th International Conference on Dependable Systems , Services and Technologies, DESSERT 2019, pp. 120-126, 2019.
[7] Kumar S, Toshniwal D, Parida M, "A comparative analysis of heterogeneity in road accident data using data mining techniques.", Evol. Syst. 8(2), 147-155.
[8] J. Jin and Y. Deng, "A comparative study on traffic violation level prediction using different models,", 2017 4th Intemational Conference on Transportation Information and Safety (ICTIS), Banff, AB, 2017, pp. 1134-1139. doi: 1O.1109/ICTIS .2017.8047913.
[9] Y. Wu and H. Tan, "Short-term traffic flow forecasting with spatialtemporal correlation in a hybrid deep learning framework," arXiv preprint arXiv :1612 .01022, 2016.
[10] X. Ma, H. Yu, Y. Wang, and Y. Wang, "Large-scale transportation network congestion evolution prediction using deep learning theory, " PloS one, vol. 10, no. 3, p. eOI19044, 2015.
[11] L, Li, Y. Lv, and F.-Y. Wang, "Traffic signal timing via deep reinforcement learning," IEEE/CAA Journal of Automatica Sinica, vol. 3, no. 3, pp. 247-254,2016.
[12] Honglei Ren, You Song, Jingwen Wang, Yucheng Hu, and Jinzhi Lei. "A Deep Learning Approach to the Citywide Tra!c Accident Risk Prediction.", 21st International Conference on Intelligent Transportation Systems (ITSC), 2018, pp