Car Accidents Severity Prediction in the USA Using Machine Learning Techniques

1st Harshada Dhawade
Computer and Information Sciences
University of Wisconsin - Parkside
Kenosha, USA
dhawa001@rangers.uwp.edu

2nd Shweta Avachat
Computer and Information Sciences
University of Wisconsin - Parkside
Kenosha, USA
avach001@rangers.uwp.edu

3rd Dhanashree Mehta
Computer and Information Sciences
University of Wisconsin - Parkside
Kenosha, USA
mehta007@rangers.uwp.edu

I. Introduction

It is observed that the economic and societal impact of traffic accidents cost U.S. citizens hundreds of billions of dollars every year and a large part of losses is caused by a small number of serious accidents. Fatality rates per capita and vehicle miles traveled to provide a way of examining motor vehicle deaths relative to the population and amount of driving. However, many factors can affect these rates, including types of vehicles driven, travel speeds, rates of licensure, state traffic laws, emergency care capabilities, weather, and topography. Between January and June of 2021, it was reported that car fatalities increased by 16 percent from the same period last year, with areas as diverse as Texas and New York City reporting sharp increases. It is observed that if the trend continues for the rest of the year, nationwide deaths would reach the highest level since 2006. According to the National Highway Traffic Administration, car accidents happen every 60 seconds. That equates to about 5.25 million accidents across the nation yearly. Along with this, for a person who just met with a car accident may suffer from mental and emotional injuries and may include mental anguish, emotional distress, fear, anger, humiliation, anxiety, shock, embarrassment, random episodes of crying, loss of appetite, weight fluctuations, lack of energy, sexual dysfunction, mood swings, and sleep disturbances.

One of the two main approaches for dealing with traffic safety problems focuses on preventing potential unsafe road conditions from occurring in the first place. If we can identify the patterns of how these serious accidents happen based on the key factors, we might be able to implement well-informed actions and better allocate financial and human resources. Thus, with real-time traffic accidents data, a prediction solution can be created and in turn, a model might be able to further predict a few or more severe accidents in real-time.

The cost of road crashes varies in different economies. According to the WHO, crashes amount to approximately 1 percent of the gross domestic product (GDP). In middle-income countries, the cost is 1.5 percent of the GDP and in high-income countries, the cost is 2 percent of the GDP. Along with the social and technical aspects, we can see that the business is hampered. In 2019, according to the CDC, employers lost 39 billion dollars in work-related car accidents. Statistics highlight that motor vehicle accidents are one of the leading causes of death for workers across major industries. Whether related to their job or not, employees injured in a car accident are more likely to miss work. Along with this, other employees need to pitch in to help their injured co-workers, and thus potentially lowering overall productivity. We also see that businesses whose employees are involved in car accidents often see their health insurance premiums increase. Motor vehicle accidents might be the most significant threat to maintaining a healthy workforce, stable earnings, and market competitiveness. Therefore, employers have a vested interest in doing their part to help reduce the number of winter-related car accidents.

II. Literature Review

A study by B. Geyik and M. Kara [1] States that safety and accident issues are a global problem in the world. This study aims to establish models to predict the accident severity levels of traffic accident injury records for possible accidents by using some data mining classification methods. The dataset used for this work is named Stats19, which has the traffic accident data from 2010 to 2012 in the United Kingdom (UK), and it is collected by the UK government data service. The given database is classified into three different accident severity categories, those are fatal, serious, and slight. The used in this research are Multilayer Perceptron (MLP), Decision Tree classifier, and Random Forest classifier and Naive Bayes classifier. The extraction of the data from the dataset will make sense to compare and predict a level degree. The result of the MLP model shows an accuracy of 86.67 percent which is important for estimating accident costs, increasing safety, and determining a strategy and it aims to reduce injury levels.

Recently, researchers in [2] Investigated the risk factors that contribute to crash injury severity among elderly drivers. A car crash can cause serious and severe injuries that impact people every day. Those injuries could be especially damaging for elderly drivers of age 60 or more. The results

are accomplished by designing accurate machine learning-based predictive models. Naïve Bayesian (NB), Decision Tree (DT), Logistic Regression (LR), Light-GBM, and Random Forest (RF) models are proposed. The Dataset is taken from The Michigan Traffic Crash Facts (MTCF), provided by the Office of Highway Safety Planning in Michigan. This dataset includes traffic crash records in the period from 2010 to 2017. In data pre-processing missing values are recovered by using mean and mode because the dataset has fewer missing values. The highest accuracy is achieved by light-GBM models with an accuracy of up to 87.97 percent. Results have also confirmed that the most important features are the Age, and Traffic volume.

Sanaa Elyassami [3] proposed machine learning-based models for analyzing the crash data, identifying the important risk factors, and predicting the injury severity of drivers. Dataset is provided by Maryland State Police and it only includes approved crash reports. It has three types of injury severities: the fatal crash, personal injury, and property damage crash. Missing values are recovered by using the list-wise deletion method. Three machine learning-based models were trained using Decision Tree (DT), Random Forest Tree (RFT), and Gradient Boosted Tree (GBT). All classifiers were implemented in the RapidMiner software platform. All three techniques are part of the software package. This experiment uses three different variants of decision trees for the classification. The experiment analyzes and compares the performance of each of the models. The GBT reported a maximum accuracy of 73.08 percent.

Mario Muñoz-Organero. [4] explored that the levels of stress while driving can cause an accident. Different types of sensors such as heart rate or skin conductivity have been used to measure stress. Research states how effectively upcoming stress levels can be predicted considering current stress levels, current driving behavior, and the shape of the road. Data is gathered from tracking four different drivers with three different car models and a motorbike and more than 220 test drives. They evaluated machine learning tools for scenarios like the Average Stress Values for the Next Minute and Regression Techniques. Cluster the average stress values for the next minute in 3 groups (low, medium, and high stress) and use classification techniques. They have also calculated the correlation between the upcoming stress signal and the real one. Results show that upcoming stress levels can be predicted correctly for a single user than the different users.

The authors of [5] compared the factors that are affected in the passenger car driver and truck driver crash severity. Crash data is collected by Police in Canada. Data is from 2007 to 2017. Two-vehicle crashes involving one truck and one passenger car are mined for modeling. Methods used to process data imbalance are over-sampling, under-sampling, hybrid method, and cost-sensitive learning method. Five classification models are used to test the performances of classifiers. Models are multinomial logistic regression, Naive Bayes, Classification and Regression Tree, support vector machine, and eXtreme Gradient Boosting (XGBoost).

REFERENCES

[1] B. Geyik and M. Kara, "Severity Prediction with Machine Learning Methods," 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2020, pp. 1-7, doi: 10.1109/HORA49412.2020.9152601.

[2] 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, doi: 10.1109/EIT48999.2020.9208259.

[3] S. Elyassami, Y. Hamid and T. Habuza, "Road Crashes Analysis and Prediction using Gradient Boosted and Random Forest Trees," 2020 6th IEEE Congress on Information Science and Technology (CiSt), 2020, pp. 520-525, doi: 10.1109/CiSt49399.2021.9357298.

[4] M. Muñoz-Organero and V. Corcoba-Magaña, "Predicting Upcoming Values of Stress While Driving," in IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 7, pp. 1802-1811, July 2017, doi: 10.1109/TITS.2016.2618424.

[5] B. Zhou et al., "Comparing Factors Affecting Injury Severity of Passenger Car and Truck Drivers," in IEEE Access, vol. 8, pp. 153849-153861, 2020, doi: 10.1109/ACCESS.2020.3018183.

[6] Z. Rezaei and H. Ebrahimpour-Komleh, "Prediction and Detection of car accidents in video by deep learning," 2021 5th International Conference on Pattern Recognition and Image Analysis (IPRIA), 2021, pp. 1-9, doi: 10.1109/IPRIA53572.2021.9483506.

[7] C. Parra, C. Ponce and S. F. Rodrigo, "Evaluating the Performance of Explainable Machine Learning Models in Traffic Accidents Prediction in California," 2020 39th International Conference of the Chilean Computer Science Society (SCCC), 2020, pp. 1-8, doi: 10.1109/SCCC51225.2020.9281196.

[8] M. Müller, X. Long, M. Botsch, D. Böhmländer and W. Utschick, "Real-Time Crash Severity Estimation with Machine Learning and 2D Mass-Spring-Damper Model," 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018, pp. 2036-2043, doi: 10.1109/ITSC.2018.8569471.