CS5610 Advanced R for Data Science

Team Registration and Final Project Proposal

Q1) Title of your project proposal

Crash Severity Prediction

Q2) Team Member 1's Name

Ifrat Zaman

Q3) Team Member 2's Name

Asif Irfanullah Masum

Q4) Background and Motivation

Road traffic crashes are a significant public health issue and a major cause of death worldwide. According to some statistics, approximately 1.25 million people die and 50 million people are injured every year due to vehicular crashes [1]. The economic and social burden associated with these crashes is enormous, and the costs and consequences of the losses are significant. Therefore, it is essential to develop effective interventions to prevent or at least minimize crash-related fatalities and injuries.

To address this issue, several government agencies, including police, health departments, and education institutions, have implemented many strategies to improve road safety. A few examples of these interventions are designing safer infrastructure, integrating road safety features in transport planning, improving vehicle safety features and driver behavior. These strategies are formulated using road traffic crash data sourced from various organizations.

One such crash data was used to devise a machine learning model to predict real-time crashes in freeway work zones [2]. The study compared Convolutional Neural Network and Binary Logistic Regression using crash and traffic data from several freeways in the Los Angeles region. The Convolutional Neural Network displayed promising results with a global accuracy of 79.50% in predicting these crashes. This study shows that machine learning techniques are revolutionizing the way crash datasets are analyzed and interpreted, providing potential insights to improve road safety and inform the development of more effective interventions and policies to prevent future accidents.

One potential area of improvement in road safety is predicting freeway crash injury severity based on a number of environmental and human factors. Accurate predictions can assist first responders and medical personnels to prioritize the most seriously injured victims and provide them with the necessary medical care. These predictions can also be used by local transportation agencies to identify hazardous conditions and strategize accordingly to avoid severe road crash incidents. Therefore, the proposed project aims to develop a machine learning model that predicts freeway crash injury severity using a crash dataset. By predicting injury severity, this model will provide valuable insights to improve post-crash care for victims of road crashes and help develop accurate diagnosis and remedial measures for road traffic operational problems.

Q5) Project Objectives

The project is divided into two components:

i. Creating a machine learning prediction model that predicts road crash injury severity based on *environmental factors* such as weather condition, lighting condition, and road condition

ii. Creating a machine learning prediction model that predicts road crash injury severity based on *human factors* such as alcohol and drug consumption, and hazardous actions taken by the driver

The decision to divide the project into two separate components was based on the different end users of the models. For example, the first component can be used by local transportation agencies and police to take necessary precautions to make roads safer in suboptimal environmental conditions (like snowy road conditions and strong winds). The second component can be developed and deployed as a consumer application that predicts whether it is safe for an individual to drive based on a few Q&As within the application interface.

Q6) What Data?

The freeway crash dataset used for this project is obtained from Western Michigan University Transportation Research Center for Livable Communities (TCRLC). The data was provided by the Michigan State Police, Office of Highway Safety Planning, to WMU for research and learning purposes.

Q7) Design Overview

Machine learning techniques will be implemented to complete this project. Some of these techniques are listed below:

1. Data Cleaning and Preprocessing
2. Data Exploration (including visualization and correlation analysis)
3. Training machine learning models
4. Evaluating machine learning models

The initial choices of machine learning models for this project are Support Vector Machine and Multinomial Logistic Regression. However, model requirements might change and other models will be evaluated as time permits.

Q8) Schedule / timeline

Week 1: *Data Preprocessing and Exploration* - Clean the data using conventional data preprocessing techniques such as null value removal and exploratory data analysis with statistical and visualization tools. This will enable us to derive valuable insights and reveal hidden patterns in the data.

Week 2: *Feature Engineering and Model* - Selecting the most relevant predictors and employing them to create a robust model. We will also apply hyperparameter tuning techniques to optimize the model and ensure maximum accuracy on both training and testing data.

Week 3: *Comparison and Analysis of different models* - Comparison of different model's outputs to find the best model for the problem. Since the dataset is divided into smaller parts, we will infer how the model behaves and gain insights about the impact of different predictors. This will enable us to identify the key factors that influence our model.

Week 4: *Summary and Report* - Create a comprehensive report that summarizes our findings. This report will be made available on a Github Repository for easy reference and accessibility. It will include details about our approach, results, and the implications of our findings.

Week 5: *Conclusion* - Conclusion of the project and its outcomes. We will discuss how our findings can be applied to future data science projects, and highlight the strengths and limitations of our approach.

References

[1] Abdulhafedh, A. (2017) Road Traffic Crash Data: An Overview on Sources, Problems, and Collection Methods. *Journal of Transportation Technologies*, **7**, 206-219. doi: 10.4236/jtts.2017.72015.

[2] Wang, J., Song, H., Fu, T., Behan, M., Jie, L., He, Y., & Shangguan, Q. (2022). Crash prediction for freeway work zones in real time: A comparison between convolutional neural network and binary logistic regression model. *International Journal of Transportation Science and Technology*, *11*(3), 484–495. https://doi.org/10.1016/j.ijtst.2021.06.002