Toronto Metropolitan University
Data Science & Analytics
Graduate Studies

2023-2024 M.Sc. in Data Science and Analytics

Road Traffic Accident Prone Analysis

Student Name: Ankita Tiwari

Supervisor: Dr. Farid Shirazi

Objective: The objective of this study is to generate and evaluate machine learning models to identify and prioritize zones with a high likelihood of accidents using historical traffic accident data. The research aims to utilize different features related to the locations, conditions, and environmental factors of accidents to forecast areas with increased accident risk.

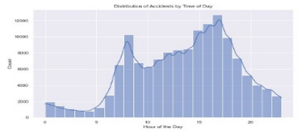
Background:

Road traffic accidents remain a significant global concern, leading to substantial human and economic losses. Despite various efforts to improve road safety, accidents continue to occur frequently, particularly in specific areas and under certain conditions. Understanding the factors such as weather, time of day, road types, and locations is essential to identifying high-risk situations and developing effective prevention strategies.

This study aims to analyze a comprehensive dataset of over 140,000 accident records to uncover patterns that contribute to traffic accidents. By examining the distribution of accidents and identifying key factors influencing their severity, the research seeks to provide actionable insights. These insights will guide targeted safety interventions, with the goal of reducing the frequency and severity of road traffic accidents.

Methodology:

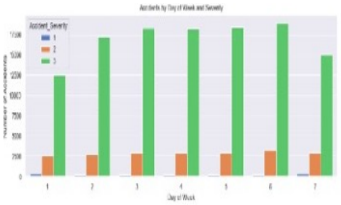
The project employed a thorough methodological approach, starting with Exploratory Data Analysis (EDA) to uncover patterns and relationships within the dataset. This analysis focused on identifying temporal trends, spatial distributions, weather impacts, and road condition influences on accident occurrences.

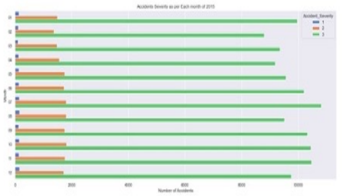


Key insights from the EDA informed the next phase, where various machine learning models—Logistic Regression, Decision Tree, Random Forest, AdaBoost, and Extra Trees Classifier—were trained to predict accident severity and identify high-risk zones. Each model was rigorously evaluated using metrics such as accuracy, F1-score, and the Jaccard Index, ensuring a comprehensive assessment of their predictive capabilities. This approach allowed for the identification of the most effective model, contributing to a deeper understanding of the factors influencing road traffic accidents.

Results:

The exploratory analysis revealed significant insights into accident patterns. Fridays and mid-afternoon hours exhibited the highest accident rates, with increased severity during summer months. Road surface defects and weather conditions were identified as major contributors to accidents.





The machine learning experiments demonstrated high predictive performance across all models, with test accuracies above 91% and F1-scores above 95%. The Logistic Regression, Decision Tree, and AdaBoost models, with test accuracies above 97%, were particularly effective, while the Random Forest and Extra Trees Classifier models, despite lower performance metrics, showed good balance between training and test performance, suggesting better generalization. The high Jaccard scores between the Random Forest, AdaBoost, and Extra Trees Classifier models indicate strong agreement in their predictions, enhancing confidence in these models' ability to accurately identify accident-prone zones. Given these results, the Random Forest or Extra Trees Classifier could be selected as the final model for identifying and ranking accident-prone zones, offering a good balance of performance and generalization, and capturing complex relationships in the data.

Conclusions: The project effectively developed and assessed machine learning models to predict road traffic accidents, identifying accident-prone zones, temporal patterns, and key environmental factors that can inform targeted interventions to reduce accidents. Future research could explore deep learning models, real-time risk assessment systems, and integration with autonomous vehicles to further advance road safety initiatives.