Toronto Transportation

Students in Data Science and Statistics

Introduction

The TTC serves as a lifeline for Toronto, offering millions of daily trips across streetcars, subways, and buses. Covering key routes from Union to Finch and Pearson Airport to the Toronto Zoo, the network plays a crucial role in urban mobility. However, recurring delays across various routes hinder the TTC's ability to deliver consistent service. This reliability issue undermines public confidence and complicates efforts to promote sustainable transportation, crucial for reducing the city's carbon footprint by encouraging fewer car journeys. Addressing these challenges is vital for the TTC to fulfill its mandate of providing an incentive to alternative transit solutions.

Challenge

The goal of this project is to develop a solution to forecast the occurrence, location, and duration of delays within the TTC network. By accurately predicting when and where delays are most likely to occur, as well as estimating their length, the model will provide valuable insights for optimizing transit operations and enhancing the reliability of service across the city. Additionally, the model can be integrated into real-time applications, allowing for proactive resource management by the TTC.

Usage

This model has practical applications for both commuters and businesses. Most riders have faced delays, leading to disrupted plans or missed appointments. With this model, commuters can anticipate delays and make informed travel decisions, whether adjusting their schedule or choosing alternative routes. The model use to provide real-time delay forecasts and helping users avoid the frustration of unforeseen disruptions. For businesses, the predictive insights offer opportunities to target advertisements more effectively, provide alternative transportation options during delays, and attract customers through enhanced service offerings. For example, marketing strategies can be adjusted based on delay hotspots, turning potential disruptions into strategic advantages.

Deliverables

- **Visualizations:** A series of detailed visualizations that highlight key patterns and trends in the TTC delay data, including geographical hotspots, peak times for delays, and the frequency of delays. These visualizations should effectively communicate findings to both technical and non-technical audiences.
- **Summary of EDA Findings:** A comprehensive report summarizing the insights gained from EDA, covering correlations between variables, significant trends, and any anomalies identified in the data. This report should also provide context for how these findings influence the modeling approach.
- **Final Model:** A fully developed and trained machine learning model, accompanied by a performance evaluation and other relevant metrics as well as key leading factors. The model should also include an explanation of the model selection process and any hyperparameter tuning performed.
- **Documentation:** Comprehensive documentation covering the entire project, including data sources, methodologies, model architecture, and usage instructions. This should allow for easy replication and further development of the project by others.

Award Categories

- **Best Visualizations:** This prize goes to the team that produces the most insightful and impactful visualizations from the TTC delay data. Winning visualizations should clearly depict trends such as delay hotspots, peak times, and recurring issues, making complex data accessible and useful for decision-makers and the public. Engaging design and clarity of communication will be key in this category.
- **Best Business Insights:** This prize goes to the team that extracts the most actionable insights from the data. These insights might include identifying how businesses could adjust their marketing or operations based on predicted delays or how TTC authorities could optimize route schedules to reduce delays. Teams that provide clear, data-driven recommendations with potential real-world applications will excel here.
- **Best Model:** This prize goes to the team that builds the most accurate, reliable, and scalable machine learning model to predict delays. Evaluation will be based on metrics like accuracy, precision, recall, and overall performance in predicting delay occurrence, location, and duration. Bonus points for creativity in feature engineering and using advanced techniques like ensemble methods or neural networks.