Draft Title

Last-Mile Delivery Time Prediction Using Real-Time and Historical Data: A Machine Learning Approach

---

Draft Research Questions

1. What are the key factors (e.g., traffic, parcel volume, time of day, route) that influence last-mile delivery time?
2. How do different machine learning models (Linear Regression, XGBoost, LSTM) compare in predicting delivery time windows?
3. Can combining real-time and historical data improve the accuracy of delivery time predictions compared to using historical data alone?

Methods and Approach

The aim of our dissertation is to build a predictive model that estimates last-mile parcel delivery time using both real-time and historical data. This is especially relevant in today's logistics industry, where delays in the final delivery stage can affect both operational efficiency and customer satisfaction.

To begin, we will explore existing literature on last-mile logistics, delivery time prediction, and machine learning approaches applied to similar problems. Our research will follow an *experimental research methodology*, where we will test and compare the performance of multiple machine learning models.

We plan to use a combination of open-source datasets and simulated data to build our dataset. We will extract features such as:

- Temporal: time of day, day of the week, holidays
- Spatial: delivery location, distance from depot
- Operational: parcel volume in a given area (parcel density)
- External: traffic congestion and weather conditions

For modelling, we will build and compare three types of predictive models:

1. Linear Regression – as a baseline for interpretability
2. XGBoost – a gradient boosting model known for high accuracy on structured data
3. LSTM (Long Short-Term Memory) – to capture time dependencies and trends

Each model's performance will be evaluated using standard regression metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score.

To support reproducibility and meet the department requirements, we will upload all code, experiments, and documentation to a public GitHub repository. Visualization and analysis will be done using Python libraries such as pandas, scikit-learn, TensorFlow/Keras, and Seaborn/Plotly.

This project allows us to apply machine learning in a real-world logistics scenario, while developing practical skills in data preprocessing, modelling, evaluation, and deployment. We believe the insights gained could be useful for logistics companies to better estimate delivery windows and optimize last-mile operations.

By:

- Kairav Roshan Sanghvi GH1039636
- Vijhisha Vinod Bhandary GH1039623