

CAPSTONE TWO - PROJECT PROPOSAL

E-Commerce Delivery Performance Prediction

Dataset Link: [E-Commerce Shipping Data \(Kaggle\)](#)

Problem Statement

On-time delivery is a critical success factor for e-commerce businesses. Delayed deliveries lead to poor customer experiences, negative reviews, and higher churn rates. Predicting whether an order will arrive on time allows businesses to proactively manage logistics, allocate resources, and improve customer satisfaction.

Context

E-commerce platforms manage millions of deliveries across multiple locations. Factors such as shipping method, product type, distance, and warehouse processing times can impact whether an order arrives on time. Predicting delivery performance ahead of time allows companies to flag high-risk orders and take preventive measures such as faster shipping or adjusting logistics partners.

Criteria for Success

- Develop a classification model to predict whether an order will be delivered on time.
- Achieve at least 75% accuracy with strong recall (minimizing false negatives, orders incorrectly predicted as on time).
- Provide insights into key drivers of delays (e.g., distance, shipment mode, warehouse processing).
- Translate findings into business recommendations for logistics and supply chain teams.

Scope of Solution Space

- **Included:** Classification modeling using features like product weight, shipping mode, warehouse block, customer location, and order value.

CAPSTONE TWO - PROJECT PROPOSAL

- **Excluded:** Real time logistics tracking and external data (e.g., weather, traffic), as they are not included in the dataset.
- **Approach:** Apply ML classification models such as Logistic Regression, Random Forest, and XGBoost. Evaluate using ROC-AUC, precision, recall, and F1 score.

Constraints

- Potential **class imbalance** (majority of deliveries on time, fewer delays).
- Dataset limited to order-level features without external supply chain data.
- Missing or noisy entries may require cleaning and preprocessing.

Stakeholders

- **Logistics & Supply Chain Teams:** Use predictions to allocate resources and reduce late deliveries.
- **Customer Service Teams:** Proactively communicate with customers when delays are likely.
- **Executives & Management:** Improve customer satisfaction metrics and reduce churn caused by delays.
- **E-Commerce Retailers:** Strengthen customer trust and loyalty through timely delivery performance.

Data Source

- **Dataset:** E-Commerce Shipping Data (Kaggle)
- Contains order-level details such as customer demographics, product weight, shipment mode, warehouse block, distance, and whether the order reached on time.
- Rich in numerical and categorical features, well-suited for classification modelling.