**E-Commerce Shipping Classification Project**

By: - Saneel Tare, Saloni Sankhe, Mitali Gupta, Harsh Oswal

**Question/Problem Statement:**

Today’s world is an extremely opportunistic, yet complicated, time to be in E-Commerce. Competition is now on a global scale with millions of brands launching each year. As shopper expectations grow, the sellers need to find creative ways to go above and beyond to delight shoppers. With supply chain issues like labor shortages, port congestion, and freight and last-mile carrier costs increasing, it is very important to have trusted logistics partners to meet consumer demands and get goods shipped on time.

E-commerce shipping encompasses receiving and processing orders, picking and packing the purchased product at a warehouse, printing shipping labels, and even managing returns. All the moving pieces create a level of complexity that depends on the types and sizes of products you sell, to which regions you’ll deliver, the delivery options and shipping speeds you choose to offer and more.

Perhaps the most important issue that customers are concerned about is time. A company’s adherence to delivery dates is an essential factor that decides the success of the company in an industry. Those who cannot meet their delivery date have little chance of long-term success. Just as important as in major e-commerce companies, obedience to the delivery schedule is one of the most imperative factors influencing the success of the companies’ reliability, reputation, and prestige.

Shipping on time has major significance for E-Commerce companies. It helps in increasing the reputation of the company, broadens the relationship with the clients and hence enhances the customers’ confidence and their buying ability and raise the overall profit for the company.

**The goal of our project is to determine whether the products are shipped on time or not.**

**Data Description:**

We extracted [dataset](https://www.kaggle.com/datasets/prachi13/customer-analytics?resource=download) from Kaggle. The dataset contained 10999 rows and 12 columns: (ID, Warehouse block, Mode of shipment, Customer care calls, Customer rating, Cost of the product, Prior purchases, Product importance, Gender, Discount offered, Weight in gms and reached on time), with data type (integer, string).   
According to the Data, we found out that 59.7% of e-commerce shipping deliveries are late received by the customers (6.563 of 10.999 customers). Ship & Warehouse F has the highest frequency of delivery. But it looks the same based on the percentage. There is an assumption that lateness is influenced by other factors. Every product that gets a discount above 10 is confirmed Late. There is an assumption that this happens in specific months but needs further checking. Shipping delivery is confirmed late when the product weight is between 2-4 kg.

**Tools:**

**Programs**: Python and Jupyter Notebook.

**Libraries**: Pandas, NumPy, sklearn-Learn, sklearn-metrics, imblearn.under\_sampling, Statsmodels, Matplotlib and Seaborn.

**Plots**: Bar plot and pair plot.

By applying 5 different Machine Learning algorithms - Decision Tree, Logistic Regression, Random Forest, XGBoost and KNN we observe that we can predict the timely shipment with fairly good accuracy. We evaluate these different models with Accuracy, Precision, Recall, F1-Score and AUC Scores.

The following comparison can be observed:

Table

Description automatically generated