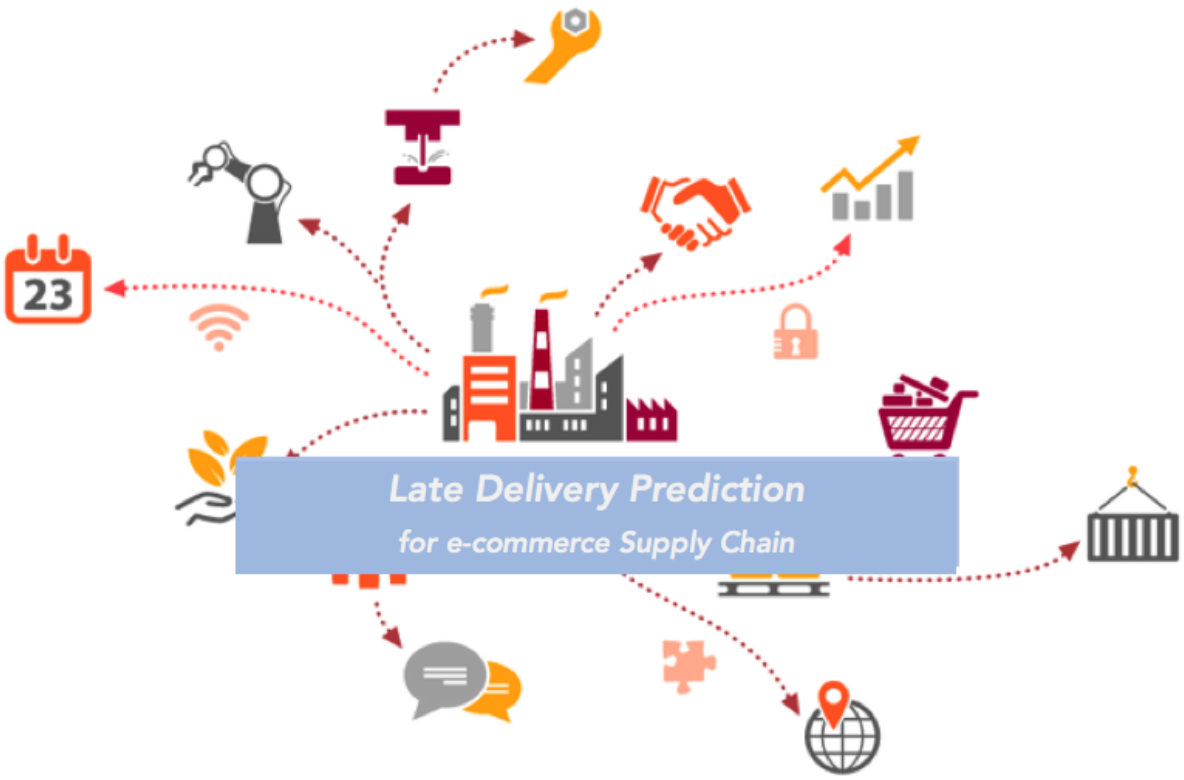

Late Delivery Prediction

For e-commerce supply chain

The Problem

Our target business is an e-commerce supply chain business. The business has two major concerns in its business operations:

1. It is facing high customer churn of unsatisfied customers due to late deliveries
2. It is facing losses due to fraudulent transactions

The business has decided to consult a data analytics firm (us) and wants to figure out:

- Any insights they can find on their business in terms of their product catalog and customer base.
- Which customers can they target immediately to reduce churn and boost sales.
- Do any patterns emerge from their customer base and transactional data that can point to patterns in late deliveries and fraudulent transactions.
- A prediction model that can predict late deliveries sing the order/sale data before they occur.

The Opportunity (Benefit to the customer)

Customer satisfaction is the bread and butter of ecommerce businesses. It is the repeat business that is most beneficial to e-commerce businesses and leads to maximum profit since there are no new acquisition costs involved in repeat business.

Detecting late deliveries before they occur can greatly benefit the supply chain business by rerouting and taking proactive steps to ensure on-time delivery and can also set appropriate customer expectations in-terms of delivery time so that customers feel dissatisfied.

Fraudulent transactions directly impact the company's profitability. This is simply money lost for the organization and it is in the best interest of the organization to curtail fraudulent transactions before they occur.

The DataSet

The data used is the transactions of all sales by product from January 2015 to January 2018 for a reputed ecommerce business. The data is published by DataCo and can be downloaded from :

https://data.mendeley.com/datasets/8gx2fvg2k6/5

The Solution

In order to solve the problems that our customer is facing, we will cover three major steps in our analysis:

1. Exploratory data analysis to detect trends in sales, product pricing and segments, markets and regions, and insights related to late delivery and fraudulent transactions
2. Customer segmentation analysis using RFM technique providing a systematic approach to customer loyalty programs.
3. A reliable machine Learning model that the company can deploy to detect late deliveries and improve customer satisfaction.

We will explore various industry standard classification algorithms, perform hyper-parameter tuning and compare their performance before choosing the best one for our model. We will also look at the feature importance to gain insight into our model.