HUMBER INSTITUTE OF TECHNOLOGY AND ADVANCED LEARNING

(HUMBER COLLEGE)

BIA 5402 - OLA
Machine Learning 2:
Assignment 1

MANAGING SUPPLY CHAIN DELAYS AND RISKS

Problem Summary

Company SES Shipping Inc. analyzed their shipping data from last year (2018) and discovered that about 60% (refer to Fig 1.1) of their orders were delivered late to the customers. This caused a huge hit to the company both financially since they had to forfeit many of these deliveries as well as potential loss of customers due to their unsatisfactory service levels.

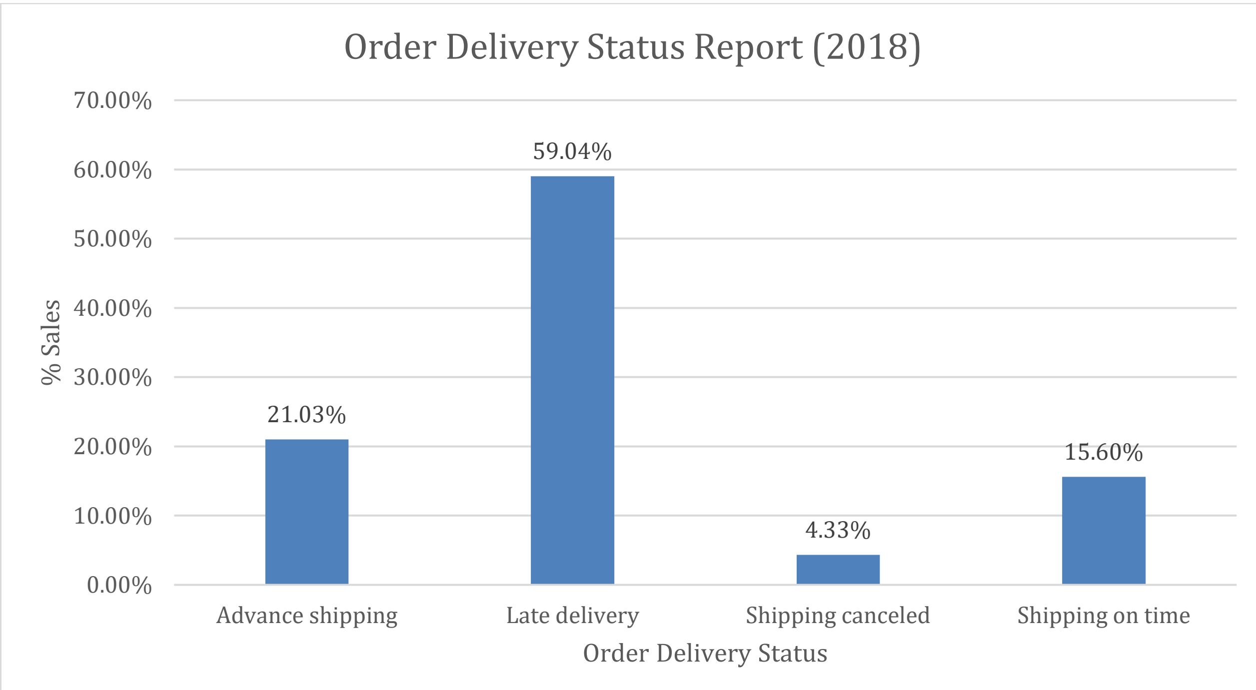

Fig. 1.1 Order Delivery Status Report

On further deep dive, their team realized that the root causes for these late deliveries are the following:

- No real-time/immediate reservations done on order placement.
    - The company does not hold inventory when an order is placed.
    - This can result in possible shortages or backordering during the processing of the order, which could lead to delays [2].
- Non-accurate promise dates
    - The lead time to manufacture the finished goods is calculated inaccurately, which could delay the shipping of the finished goods to the distribution center.

- The lead time to ship customer orders from the source to the destination has to be updated since the company went through a change in the supply chain network which included addition and removal of some facilities.
- Due to these inaccurate lead times, the delivery promise dates generated by their transportation system and sent to the customers were incorrect and led to delayed deliveries.

- Incorrect order of stock allocations
    - Since automated reservations on the inventory is absent, it is possible that order B, which was placed after order A was placed, is fulfilled, and shipped before order A since order allocations are done manually. This could result in:
        - Order A having a shortage and hence, resulting in a backorder which could delay order delivery.
        - Customer cancelling the order due to experiencing extreme delay in their delivery.
        - Order A getting cancelled due to unavailability of stock.

Goal

To break even from the previous losses, the finance and analytics team of SES Shipping Inc. has recommended that the company deliver service levels of 24 hours for their A level customers, 48 hours for their B level customers and 72 hours for the C level customers. Hence, the company approached our business intelligence team to help them build a feed forward model that can predict whether an order item would get delayed or not, as soon as the order comes into the system. This model will help their team flag those orders, so that precautionary measures can be taken to avoid late deliveries and hence, reduce the overhead costs.

Data Acquisition

The project kick started with a discussion between the client and our BI team where the goal, which is the target service levels was agreed upon. Our BI team, then, requested the client for all the historical raw data regarding their past deliveries. The client provided the raw dataset spanning over the last five years having 180520 rows and 52 columns as attributes.

Analysis

In this project, our BI team used 18 attributes as input predictors and 1 target variable to predict Late delivery risk by using the Neural Network algorithm. The description of the used variables is given below:

Attribute Name	Data Type	Description
shipment_date	Date	It represents the date of Shipment.
Category_Id	Integer	Unique number against each category.
Customer_City	String	In which city the customers are placing orders.
Customer_Country	String	In which Country the customers are placing orders.
Customer_Segment	String	Indicates grouping of customers.
Customer_State	String	In which State the customers are placing orders.
Customer_Zipcode	String	Unique number against each location.
Department_Name	String	Name of each department
Latitude	Decimal	Distance of Equator (North or South)
Longitude	Decimal	Distance of Equator (East or West)
Market	String	Indicates the location of market.
Order_Price	Decimal	Represents the total amount of orders for each customer.
Order_Quantity	Integer	Indicates the total number of products ordered by each customer.
Sales	Decimal	Amount of sale according to each customer
Product_Card_ID	Integer	Unique Identification number for same type of card
Product_Category_Id	Integer	Unique Identification number for same type of products

Product_Price	Decimal	Unit price per product
Ship_Mod	String	Shipping mode as per class
Late_delivery_risk	Integer	Status of delivery risk

Table 1.1 Variable Description

- Analyzing the delivery status frequency in our supply chain – By analyzing and predicting late delivery risks, our business can work on the contributing factors to minimize the delays and number of cancelled orders as this effect the business profit.

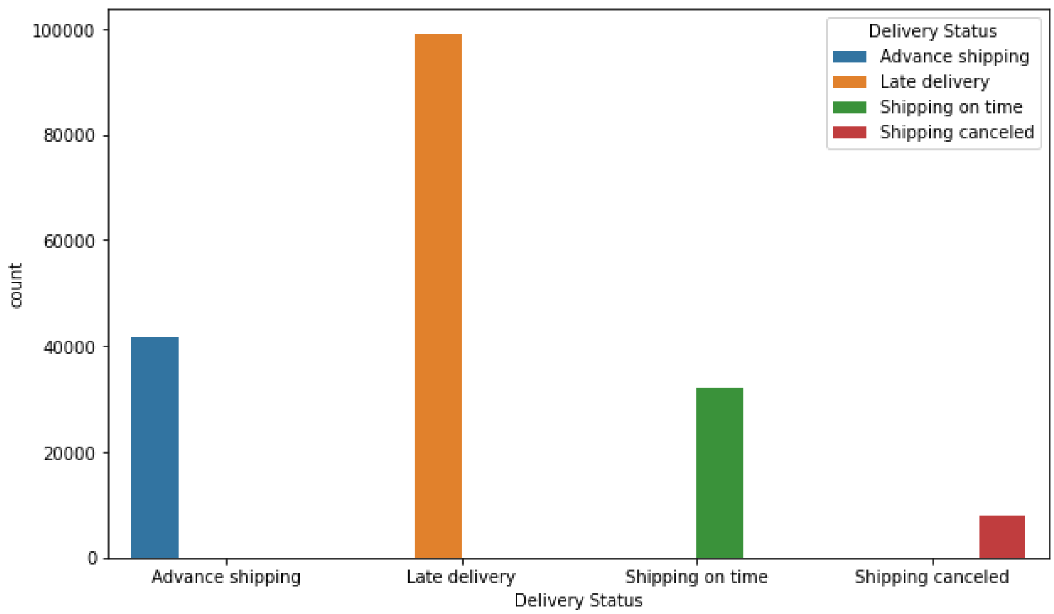

Fig 1.2 Graph for Delivery Status in our dataset

- Analyzing countries with highest fraud rates – By using the machine learning algorithms, we can train our model to predict whether an order ids fraud or not based on past order history of the customers and the country.

Fig 1.3 Graph for Top 5 Order Countries with Highest Fraud Rate

Fig 1.4 Graph for Top Customer Countries with Highest Fraud Rate

- Stock re-allocation and scheduling – We can train our model to frequently stock or schedule delivery for fast selling products to make sure we have enough inventory for products, avoiding the late deliveries.

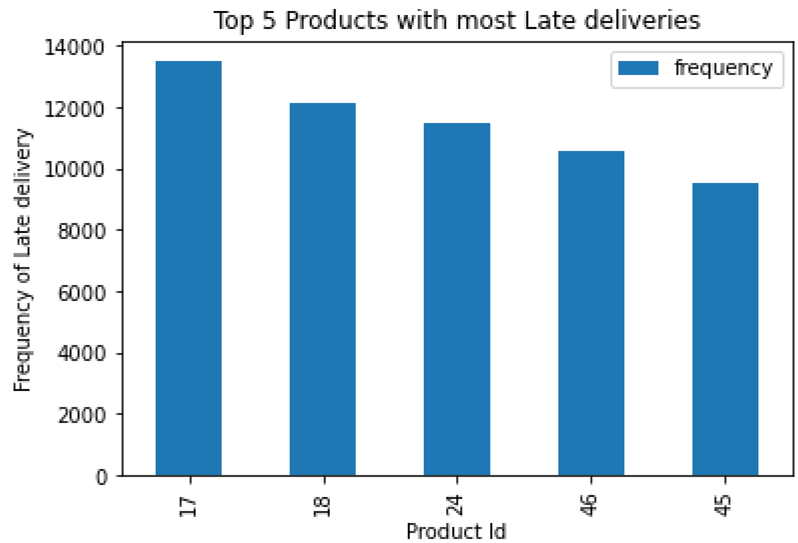

Fig 1.5 Graph for top 5 products with most late deliveries

- Warehouses for supply efficiency – We can use the analysis to gain insights about countries with highest number of orders and we can open warehouse in those countries to minimize the shipment time.

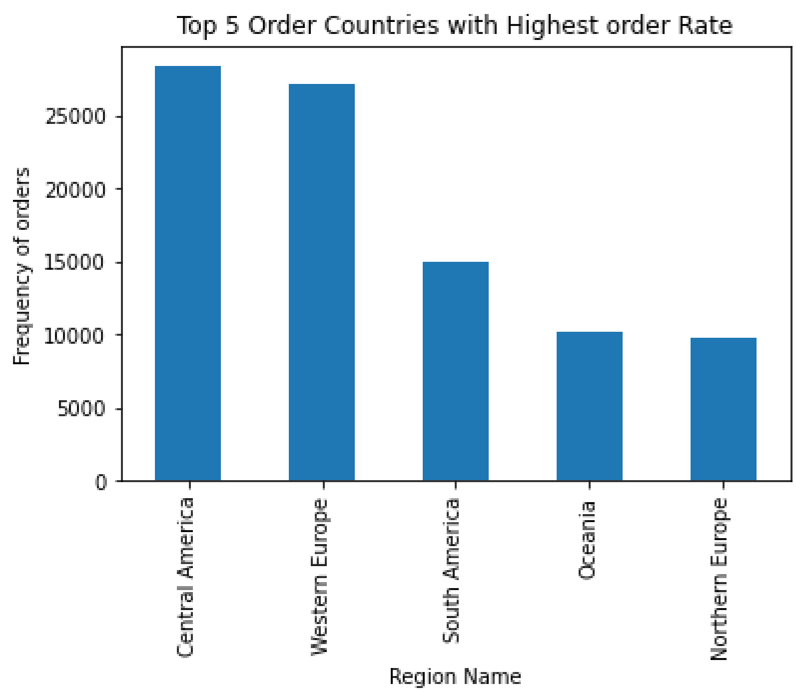

Fig 1.6 Graph for top 5 order countries with highest order rate

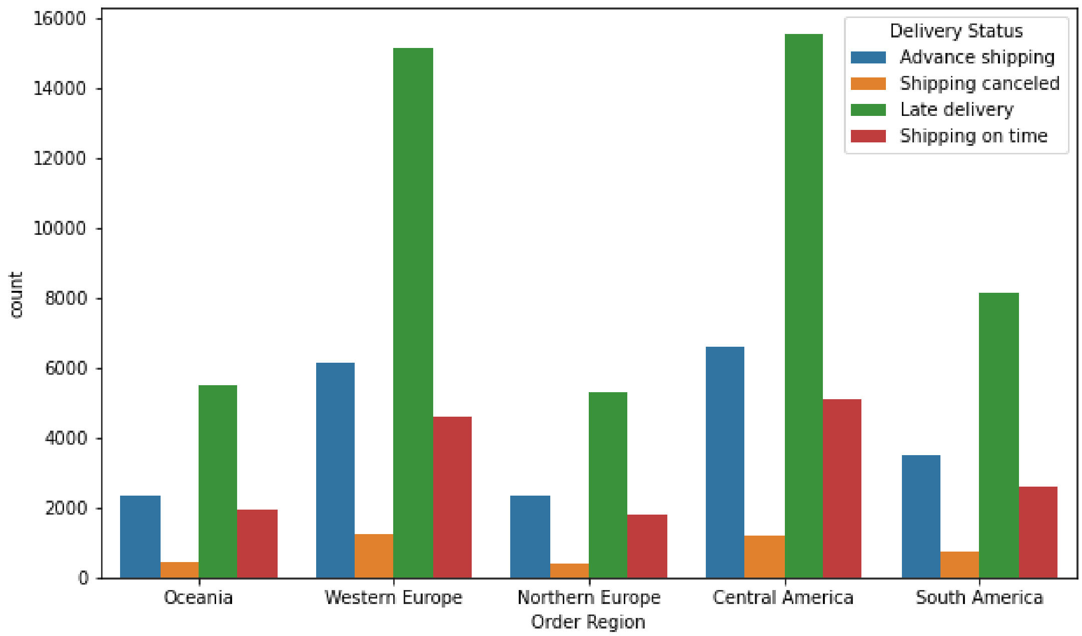

Fig 1.7 Delivery status for countries with highest order rate

Confusion matrix: Confusion matrix in machine learning is used to understand the performance of classification models. There are four terminologies which are related to the confusion matrix. They are as follows. True Positive represents the Number of cases predicted positive and are truly positive. False Positive represents the Number of cases predicted negative and are truly negative. True negative represents the Number of cases predicted positive and are truly negative. False negative represents the Number of cases predicted negative and are truly positive.

We got following values for Confusion matrix for Neural Network model:

array ([[83,81457],

[102, 98874]],

Here,True positive predictions: 98874

True negative predictions:83

False positive predictions: 81457

False negative predictions: 102

Algorithm

Working on the dataset to predict the late delivery risk, we are using a neural network model where we have divided our dataset in a 60-40% ratio for training and testing purposes, using 3 hidden layers. The objective for choosing 3 hidden layers is to balance between complexity and interpretability of the dataset. The model gets enough capacity to learn complex relationships in the data, while still being computationally efficient [3].

Model Comparison and Conclusion

Hence, our team created a multi-layer feed forward neural network model that was able to train and predict the late delivery risk with a certainty of 54.82% using a neural network, whereas, while using the logistic regression model accuracy to predict the delays came out to be 65.21% accurate, slightly higher than the accuracy of neural network.

With this information, the client team can flag the orders having the delivery risk and ensure:

- A reservation for this order is created in the system.
- An accurate promise date is calculated by the system and sent to the customer.
- The order is allocated systemically and communicated to the respective facility.

References

[1] Tiwari, S. (2019, December 5). *Dataco Smart Supply Chain for Big Data Analysis*. Kaggle. Retrieved January 31, 2023, from https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis

[2] Panova, Y., & Hilletofth, P. (2018). Managing supply chain risks and delays in construction project. Industrial Management & Data Systems, 118(7), 1413-1431.

[3] Uzair, M., & Jamil, N. (2020, November). Effects of hidden layers on the efficiency of neural networks. In 2020 IEEE 23rd international multitopic conference (INMIC) (pp. 1-6). IEEE.