**METHODOLOGY AND METRICS**

In this section, we aim to describe the forecasting methods used in our study. All the models are trained and validated by 80/ 20 split. This section’s main goal is to list all the models that were applied before comparing the metrics that each of them produced. This section is further divided into 5 subsections: (1) Demand forecasting using Time Series Analysis;(2) Sales Forecasting: Regression ;(3) Multiclass Classification: Late Delivery Prediction ;(4) Fraud Detection; (5) Customer Segmentation using RFM Analysis to answer the research questions outlined in this paper.

**Demand forecasting using Time Series Analysis**

Demand forecasting is one of the main activities of supply chain management by means of which companies can forecast future customer behavior based on historical sales data. Accurate prognosis is critical to help in optimal inventory levels. This is while ensuring continuous product availability and reduction of the inventory holding and stock-out costs. We chose Seasonal AutoRegressive Integrated Moving Average (SARIMA) model for its advanced statistical approach which we applied to time-series data that has seasonality. The SARIMA model improves upon the ARIMA method by considering seasonality as a crucial factor and thus is best suited for those supply chain data with periodic and seasonal patterns in demand.  
Model Components: Seasonal Autoregressive (SAR) part: Contemplate some time and its seasonal changes. Seasonal Differencing (I): It assists in removing the seasonal orders’ variations and in stabilizing the mean of the series. Seasonal Moving Average (SMA) part: Improves the seasonal error components in the forecast are captured. On the SCM side, SARIMA models exploit the demand spikes and drops to plan strategically the procurement, production scheduling, and inventory management. Companies can therefore monitor when demand might be likely to increase or to decrease and accordingly adapt their safety stock levels to reflect the demand changes. We integrated SARIMA models to forecast demand and safety stock calculation as a test to the fact that ML and statistical models might be powerful methods to elevate decision making of supply chain. By virtue of making this type of models, supply chain managers have at their disposal a powerful tool that allows them to make evidence-based decisions in their attempts to cut down the supply chain costs, optimize the inventory levels, and improve the supply chain performance.

**Sales Forecasting: Regression**

In our study, we used 5 machine learning models: Linear Regression, Random Forest Regressor, Gradient Boosting Regressor, K-Neighbors Regressor, XGBoost Regressor, LightGBM Regressor and hyperparameter tuning, feature importance. In the realm of supply chain management, accurate forecasting of sales is the cornerstone based on which informed business decisions can be made. It secures that most companies may be able to forecast their production and carry out inventory properly as well as arrange their resources in a better way to be ready to respond in case of such uncertainties. Using regression models to forecast sales allows businesses to utilize the historical information in a way that would predict future sales, sweeping them onto a strategic business conduct.

* Linear Regression: This methodology of forecasting can be seen as a first step and the sales are predictably linearly related to independent variables. Such a characteristic as simplicity and evaluability makes it perfect for the business analysts to understand the results very simply and quickly act upon them.
* Random Forest Regressor: As a method of learning by groups, the Random Forest Regressor starts by generating a collection of the so-called decision trees and delivers results which serves to control over-fitting and give additional accuracy in a predictor. The solidity of this algorithm is a vital factor for processing multi-variable dataset as well as the ones intimately linked with the sales numbers.
* Gradient Boosting Regressor: In this model, the final prediction is made by fusing the results of the weak models, likely decision tree models, aiming to provide maximized prediction accuracy. It’s most impactful for non-linear datasets that are usual for sales data, which is not so easy to predict, for example.
* K-Neighbors Regressor: The regression performed by this K-Neighbors uses the historical data that are the ' k' most similar data points in finding the mean or median here. Discovering, measuring, and forecasting the sales trends are now made simple and detailed thanks to the data analysis that is based on market behaviors.
* XGBoost Regressor: XGBoost Regressor is a well-known decision-making technique that shows exceptional performance and speed. It is a scalable, end-to-end machine learning system, that has become very popular due to its contribution to competitions. The technology can deal with Big Data, it produces more precise forecasts that can affect the final outcomes of the business decisions singularly.

LightGBM Regressor**:** As a gradient boosting framework that uses tree-based learning algorithms, LightGBM is designed for distributed and efficient training, which is ideal for large datasets. Its application in sales forecasting allows for quick model training and prediction, providing businesses with timely insights.

**Hyperparameter tuning**  
The process of the hyperparameter search is to find out the values that will perform the best in terms of hyperparameters, that are the configurations given for learning. When it comes to the Random Forest regressor, a top-notch algorithm that is known to be one of the best performers when doing regression tasks, hyperparameter tuning can considerably enhance the accuracy of the overall sales prediction.

**Customer Segmentation using RFM analysis.**

Customer segmentation is a method of classifying clients to groups based on shared features. Based on RFM values (Recent, Frequency, and Monetary) of customers, the successful classification of company customers is divided into groups with comparable behaviors. This technique is widely used in marketing and customer relationship management. Each customer is given 3 different ratings based on the latest, frequency, and Amount of money spent (economic volatility). Scores are used within a range from 4 to 1. The most important group is given a four-point scale, while the others are given 4, 3, 2 and 1.

RFM analysis categorizes customers based on three key dimensions: RFM model assigns consumers into three types of groups include the following:

1. Recency (R): This dimension analyses how recently a customer has made a purchase. The customers who made a purchase in the last few days/weeks/months are considered to be more engaged with the product and can have a higher potential for repeating purchases.  
  
2. Frequency (F): Purchase frequency determines the customer’s shopping habit within a set timeframe. Loyal consumers, for the most part, are likely to be among the most consistent buyers. In fact, they are very critical to business operations.  
  
  
3. Monetary Value (M): This feature assesses the monetary value of a certain customer’s purchase. The consumers who are ready to pay hefty amounts for the product may aid in the development of the enterprise.