Supply Chain Management

A Fast Moving Consumer Goods (FMCG) company entered into the instant noodles business two years back. Their higher management has noticed that there is a mismatch in the demand and supply. Where the demand is high, supply is pretty low and vice-versa which results in a loss in inventory cost and ultimately loss to the company. Hence, the higher management wants to optimize the supply quantity in each and every warehouse in the entire country

GOALS

The objective of this exercise is to build a model, using historical data that will determine an optimum weight of the product to be shipped each time from the respective warehouse.

1. Focus on all steps of data science (EDA, data processing, model, evaluation, charts)
2. Highlight any trend in data, deep insight, novel steps that you take
3. Highlight next steps and improvements.
4. Apply 5 to 6 machine learning algorithms and evaluate it

The objective of this exercise is to build a model, using historical data that will determine an optimum weight of the product to be shipped each time from the respective warehouse.

Dataset The dataset for this problem consists of various features such as warehouse ID, manager ID, zone, regional zone, number of refill requests received by the warehouse in the last 3 months, number of transport issues for the warehouse in the last 1 year, number of competitors in the market, number of retail shops who sell

noodles produced by the warehouse, whether the warehouse is owned by the company or it is on rent, number of distributors who work between warehouse and retail shops ,whether the warehouse is in a flood impacted area or not, whether the warehouse has proper electric supply along with some power backup, distance from the warehouse to production hub, number of workers in the warehouse, warehouse establishment year, storage issues reported by the warehouse in the last 3 months, whether the warehouse has temperature regulating machine indicator or not ,type of approval warehouse having been issued by government, number of times the warehouse faces the breakdown in the last 3 months, product weight, etc.

Data Science Process The following data science process was employed to build the model:

1.Project Overview

2.Data Exploration

3.Data Cleaning

4.Feature Engineering

5.Exploratory Data Analysis

6.Data Preprocessing

7.Model Building & Evaluation

8.Evaluation

9.Conclusion This project focuses on improving supply chain management using supervised regression analysis. It involves the following key steps:

Data preprocessing and cleaning to handle missing values and irrelevant features. Feature engineering to create new variables and enhance the dataset. Exploratory Data Analysis (EDA) to gain insights and understand the relationships between different features. Model building using various regression techniques such as Linear Regression, Decision Tree, Random Forest, AdaBoost, and Gradient Boosting. Evaluation of the regression models to determine their effectiveness in predicting and improving supply chain efficiency.

Data Exploration Exploring the dataset to understand the characteristics of the data is a crucial step in the project. This includes examining data statistics, visualizing key features, and identifying correlations.

Data Cleaning Data cleaning is essential to ensure that the dataset is free from missing values and irrelevant information. This step involves handling null values, removing unnecessary columns, and preparing the data for analysis.

Feature Engineering Feature engineering involves creating new variables or transforming existing ones to enhance the dataset's predictive power. In this project, features like warehouse age, demand, supply, and demand interaction have been created to better understand the supply chain dynamics.

Exploratory Data Analysis (EDA) The EDA process was conducted to get a better understanding of the dataset and the relationships between the different variables. EDA is a critical step in the project as it helps uncover valuable insights and relationships within the data. This includes visualizing data, identifying trends, and understanding how different factors affect the supply chain.

Data Preprocessing Data preprocessing involves preparing the data for modelling by encoding categorical variables, normalizing data, and making it ready for regression analysis.

Model Building and Evaluation Few models were built using the following machine learning algorithms:

Linear Regression

Decision Tree Regressor

Random Forest Regressor

Bagging Regressor

AdaBoost Regressor

Gradient Boosting Regressor

The models were fitted to the training data and evaluated using the testing data.

Evaluation The performance of the regression models is assessed using appropriate evaluation metrics. This step helps determine which model is the most effective in optimizing the supply chain.

Conclusion In conclusion, this project aims to enhance supply chain management through supervised regression analysis. By leveraging the power of data preprocessing, feature engineering, and regression modelling , we can uncover valuable insights and improve the efficiency of the supply chain. The project's findings and models will have practical applications in real-world supply chain optimization.