CSE4062

Introduction to Data Science and Analytics

Spring 2024

GROUP-1

“Predicting Revenue Trends Using Machine Learning”

Group Members:

Duygu Yasinoğlu-150122982-Computer Science Engineering-
duyguyasinoglu@gmail.com

Erdem Pehlivanlar-150119639-Computer Science Engineering -
erdemphl@gmail.com

Ulaş Kaba-150320512-Industrial Engineering

ulaskaba@marun.edu.tr

Yunus Emre Dar-150321823- Industrial Engineering-
yunusdar@marun.edu.tr

Muhammed Emin Bardakçı-150319057- Industrial Engineering -
eminbrdk4@gmail.com

Deniz Yücetepe-150620027-Chemical Engineering-
denizyucetepe@marun.edu.tr

Mustafa Özgür Hocaoğlu-150120058- Computer Science Engineering -
mustafaozgurh@gmail.com

Project Description

Our project aims to analyze a dataset containing information about orders, including various attributes such as product name, price and similar attributes.The primary objective is to predict revenue trends, which involves understanding how different variables influence revenue generation over time. This analysis can provide valuable insights for decision-making, such as identifying factors that contribute most significantly to revenue growth or decline. The main objective of the project is to develop a predictive model that can forecast revenue trends based on the DataCo dataset. By analyzing historical data and identifying patterns, the goal is to create a reliable model that can anticipate future revenue fluctuations accurately. Additionally, the project aims to uncover insights into the factors driving revenue changes, enabling stakeholders to make informed business decisions and possibly optimize strategies for revenue maximization.

Our group consists of three students from computer engineering, three students from industrial engineering and a student from chemical engineering. The dataset is named “DataCoSupplyChainDataset”, therefore the topic aligns well with the field of industrial engineering due to several reasons. Choosing a dataset related to order information and revenue prediction is relevant to the field of industrial engineering because it aligns with key areas of focus such as supply chain optimization, operations management, forecasting and planning, performance measurement, and system optimization. Analyzing such data can provide valuable insights and support informed decision-making to drive efficiency, profitability, and competitiveness all of which are crucial for optimization and maximizing revenue.

DATASET

DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS*

This dataset has 63 columns/attributes and 180K rows/instances. There are nominal, text and boolean type attributes.

Important Features:

- **Order date (DateOrders):** Numeric
- **Product Category Id:** Nominal
- **Order Item Quantity:** Numeric
- **Product Price:** Numeric
- **Order Item Discount:** Numeric
- **Customer Segment:** Nominal
- **Delivery Status:** Text
- **Late Delivery Risk:** Boolean
- **Product Status:** Boolean
- **Category Name:** Nominal (**Target Attribute**)

*https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis