Team No. 5

Professor: Dr. Amir Jafari

Topic: Supply Chain Analytics

Date: 1st November 2021

1. What problem did you select and why did you select it?

   One of the principal components of the supply chain pipeline is the timely delivery of the order. Not only does it affect the profit margins for large companies like Amazon, but it has a substantial impact on smaller e-commerce companies. Streamlining the process helps them build a better customer experience by providing customers with reliable expected delivery dates. Hence, we analyze historical order data from the company DataCo Global and build a machine learning model to predict the expected number of days required to complete the delivery.

2. What database/dataset will you use? Does it need to be cleaned?

   Link: DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS

   The dataset does not contain any missing data, but specific preprocessing steps such as feature selection, feature engineering, one-hot encoding are necessary. Data transformations for various models are needed.

3. What data mining algorithm will you use? Will it be a standard form, or will you have to customize it?

   a. Decision Tree & Random Forest

   b. Ensemble Average

4. What packages will you use to implement the model? Why?
    a. Pandas : Importing data and manipulation
    b. Sklearn: For ML model building, training and prediction
    c. Matplotlib and Seaborn: For plotting and graphs
5. What reference materials will you use to obtain sufficient background on applying the chosen model to the specific problem that you selected?
    a. Data mining class, Sklearn documentation
    b. The Elements of Statistical Learning
    c. sklearn.tree.DecisionTreeClassifier — scikit-learn 1.0.1 documentation
    d. https://scikit-learn.org/stable/modules/tree.html
6. How will you judge the performance of your results? What metrics will you use?

    The result will be measured on the basis of classification error metrics such as:
    a. Roc_auc_score
    b. Recall and Precision
    c. F1_score
    d. Accuracy
7. Provide a rough schedule for completing the project.

    Topic Proposal: 11/1/2021

    EDA Analysis: 11/7/2021

    Data Preprocessing: 11/20/2021

    Model Selection: 11/25/2021

    Dashboard creation: 11/30/2021

    Final Project Report: 12/5/2021