Sales Forecasting and Late Delivery Prediction
Documentation Date: 07/02/2023
Author: Jake Meyer

Project Description
The focus for this project was centered around Supply Chain Analytics.
The two main focuses were Sales Forecasting and Late Delivery Prediction
with data from DataCo from Mendeley Data. The CRISP-DM methodology was
followed to highlight challenges and determine the best models
fit to the data. The Sales Forecasting problem was treated as a
regression issue and the Late Delivery Prediction was
treated as a classification problem.

Table of Contents

    Supporting Files
    Project Environment Overview
    Data Preprocessing
    Model Training
    Model Evaluation
    Report an Issue
    Project References

Supporting Files
The dataset used for this analysis can be found through the link below:
[DataCo Data Set](https://data.mendeley.com/datasets/8gx2fvg2k6/5)

Project Environment Overview
The project was completed in Jupyter Notebook (through Anaconda Navigator)
via Python. All relevant libraries are called out in the Import Necessary
Libraries section of the code.

Data Preprocessing
Univariate, Multivariate, and additional Exploratory Data Analysis (EDA)
was performed in depth on the data set. All columns irrelevant to
Sales Forecasting or Late Delivery prediction were removed initially.
In addition, customer names, emails, and passwords were removed from
the analysis to retain data privacy. Please be aware of this privacy concern
within the data set if an alternative analysis is performed. The data
preparation for the model included features that were strongly correlated
with the target variables of Sales and Late Deliver Risk. The data was
separated into Training and Test subsets.

Model Training
Models used for Sales Forecasting included Linear Regression,
Lasso Regression, Decision Tree Regressor, and Random Forest Regressor.
Models used for Late Delivery Prediction included Logistic Regression,
Random Forest Classifier, and Decision Tree Classifier.

Model Evaluation
Evaluation metrics were compared for models within each respective problem.
For Sales Forecasting, Mean Absolute Error (MAE), Mean Squared Error (MSE),
R-squared Train, and R-squared Test metrics were used. The primary metric was
R-squared Test. For Late Delivery Prediction, Accuracy Test, Accuracy Train,
Precision, Recall, and F1-Score were used. The primary metric for the
classification problem was Accuracy Test.

Report an Issue
In the event of an error or major concerns, please reach out to my email
via meyerjake@gmail.com.

Project References

 AI Monks (n.d.). Top 10 Analytics Projects in Operations and Supply Chain.
Resources. Retrieved June 10, 2023, from https://aimonks.com/supply-chain-analytics-top-10-analytics-projects/
 Ariwala, P. (2022, September 30).
 9 Ways Machine Learning Can Transform Supply Chain Management.
 Artificial Intelligence and Machine Learning. Retrieved June 10, 2023,
 from https://marutitech.com/machine-learning-in-supply-chain/

Constante, F., Silva, F., & Pereira, A. (2019, March 12). Datasets.
DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS. Retrieved June 10, 2023,
from https://data.mendeley.com/datasets/8gx2fvg2k6/5

Diaz, J., Marcolis, C., Washburn, R. (2023, March 31). CDW.
How the Modern Data Platform Fuels Success. Retrieved June 19, 2023,
from https://www.cdw.com/content/cdw/en/articles/dataanalytics/how-the-modern-data-platform-fuels-success.html?cm_ven=acquirgy&cm_cat=bing&cm_pla=SLG+Analy

Hotz, N. (2023, January 19). What is CRISP DM?. Data Science Process Alliance.
Retrieved June 16, 2023, from https://www.datascience-pm.com/crisp-dm-2/

IBM (2021, August 17). CRISP-DM Help Overview. Documentation.
Retrieved June 10, 2023, from https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview

Tiwari, S. (2019, March 12). DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS.
About Dataset. Retrieved June 10, 2023, from https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis?resource=downloa

Ying, K. (2021, September 16). Data Science Ethics – What Could Go Wrong and
How to Avoid It. Data Science. Retrieved June 10, 2023,
from https://www.freecodecamp.org/news/the-ethics-of-data-science/
