Forecasting Sourcing Costs - Data Analysis and Modeling

Description:

This project focuses on forecasting sourcing costs using historical data on product types, manufacturers, area codes, sourcing channels, product sizes, and sourcing costs. The task involves data preprocessing, exploratory data analysis, outlier detection, and modeling using ARIMA and LSTM models.

Dataset:

The dataset includes information on ProductType, Manufacturer, Area Code, Sourcing Channel, Product Size, Product Type, Month of Sourcing, and Sourcing Cost. The dataset is split into training and test sets for model evaluation.

Files:

- Training_Dataset.csv: Contains the training data for model development.
- Test_Dataset.csv: Contains the test data for model evaluation.

Approach:

- Data Preprocessing: Check for missing values, visualise data distributions, and handle outliers.
- Modeling:
    - ARIMA Model: Implement an ARIMA model for time series forecasting.
    - LSTM Model: Develop an LSTM model for sequence prediction.
- Evaluation: Calculate Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) to evaluate model performance.