Forecasting Sourcing Costs - Data Analysis and Modeling

Description:

This project uses historical data on product types, manufacturers, area codes, sourcing channels, product sizes, and sourcing costs to forecast sourcing costs. The task involves data preprocessing, exploratory data analysis, outlier detection, and modelling 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.

Approach:

- Data Preprocessing: Check for missing values, visualise data distributions, and handle outliers.
- Modelling:
    - 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) and Mean Squared Error (MSE) to evaluate model performance.

Insights:

- The ARIMA and LSTM models provided valuable insights into forecasting sourcing costs based on historical data.
- The models demonstrated the ability to capture data patterns and trends, enabling accurate sourcing cost predictions.

Conclusion:

This project focused on forecasting sourcing costs using historical data, involving data preprocessing, exploratory data analysis, outlier detection, and modelling with ARIMA and LSTM models. The dataset provided insights into various attributes influencing costs, split into training and test sets for evaluation. The ARIMA and LSTM models were utilised for forecasting, with evaluation based on Mean Absolute Error (MAE) and Mean Squared Error (MSE). The project successfully addressed cost forecasting through a comprehensive analysis, highlighting the models' ability to capture trends and patterns accurately. Further optimisation and feature exploration could enhance forecasting accuracy and adaptability to changing data dynamics.