

Big Mart Sales Prediction Project Proposal Report

Group name: ML CREW

Members:

Aneeq Arif(cs211174)
Saud Ahmed(cs211170)
Haider Ali(cs211057)
Gaitree wadhwani(cs211086)

Problem Statement:

The goal of this project is to predict the sales of products in various Big Mart outlets. By leveraging historical sales data, we aim to develop a model that accurately forecasts future sales, assisting Big Mart in inventory management, demand planning, and strategic decision-making.

Significance of the Problem in the Pakistani Context:

In Pakistan, the retail sector is a significant contributor to the economy, providing employment and driving consumer spending. However, inefficiencies in inventory management and demand forecasting can lead to overstocking or stockouts, impacting profitability and customer satisfaction. Accurate sales predictions can help Big Mart optimize inventory levels, reduce waste, and improve supply chain efficiency, thus enhancing the overall shopping experience and contributing to economic stability.

Scope of the Problem:

This project will focus on predicting sales for a variety of products across multiple Big Mart outlets. While the primary focus is on improving sales forecasting accuracy, the scope includes:

- Analysis of factors influencing sales such as seasonality, holidays, and regional preferences.
- Identifying trends and patterns in sales data.
- Addressing challenges related to data quality and missing values.
- Limiting the scope to historical data provided, without real-time data integration.

Dataset Information:

We will use a publicly available dataset from Big Mart, which includes information on product sales across various outlets. The dataset comprises features such as item identifier, item weight, item visibility, item type, outlet identifier, outlet establishment year, outlet size, and sales. This dataset is chosen because it provides a comprehensive overview of factors affecting sales and is commonly used in academic and professional research for sales prediction.

Machine Learning Approach and Expected Algorithms:

To predict sales, we will employ the following machine learning techniques and algorithms:

- **Linear Regression**: To establish a baseline model and understand linear relationships between features.
- **Decision Trees**: For capturing non-linear relationships and interactions between features.
- **Random Forests**: To improve predictive accuracy and handle overfitting.

Methodology to Solve the Problem:

- Data Collection and Exploration**:
 - Collect and load the dataset.
 - Perform exploratory data analysis (EDA) to understand the data distribution, identify missing values, and detect outliers.
 - Visualize data to uncover trends and patterns.
- Data Preprocessing**:
 - Handle missing values through imputation or removal.
 - Encode categorical variables using techniques such as one-hot encoding.
 - Normalize or standardize numerical features if necessary.
- Feature Engineering**:
 - Create new features based on domain knowledge (e.g., interaction terms, date-related features).
 - Select important features using techniques like feature importance from tree-based models.
- Model Training**:
 - Split the data into training and validation sets.
 - Train multiple models (Linear Regression, Decision Trees, Random Forests, GBM, XGBoost).
 - Tune hyperparameters using techniques like grid search or random search.
- Model Evaluation**:
 - Evaluate model performance using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared.

- Compare model performance and select the best-performing model.

6. **Model Deployment**:

- Prepare the final model for deployment.
- Develop a user interface or API for easy access to the prediction model.
- Document the model and provide guidelines for its use.

7. **Monitoring and Maintenance**:

- Implement monitoring to ensure the model remains accurate over time.
- Update the model periodically with new data to maintain its predictive power.

Relevant Industry and Partner Name: Provide the name of the industry, organization, and the contact person you are consulting with.

Partner Name: Sara Sameer

Her linkedin : <https://www.linkedin.com/in/sarasameer/>

Relevant Industry : Retail Industry.