

Store-Sales-Prediction

Problem Statement -

You are opening a new Store at a particular location. Now, Given the Store Location, Area, Size and other params. Predict the overall revenue/Sale generation of the Store.

Dataset Details- The data has 8523 rows of 12 variables.

Dataset Description -

Variable - Description

Item_Identifier- Unique product ID

Item_Weight- Weight of product

Item_Fat_Content - Whether the product is low fat or not

Item_Visibility - The % of total display area of all products in a store allocated to the particular product

Item_Type - The category to which the product belongs

Item_MRP - Maximum Retail Price (list price) of the product

Outlet_Identifier - Unique store ID

Outlet_Establishment_Year- The year in which store was established

Outlet_Size - The size of the store in terms of ground area covered

Outlet_Location_Type- The type of city in which the store is located

Outlet_Type- Whether the outlet is just a grocery store or some sort of supermarket

Item_Outlet_Sales - Sales of the product in the particular store. This is the outcome variable to be predicted.

Dataset Link :

https://drive.google.com/drive/folders/1-WqRLkzYFJJMe-e_QeVUAMqpvesZ7cft?usp=sharing

Steps to be followed:-

Data Preprocessing: Clean the data by handling missing values, outliers, and any inconsistencies. Convert categorical variables into numerical representations using techniques like one-hot encoding or label encoding.

Feature Engineering: Create additional features that could have an impact on sales. For instance, you could extract information like month, year, season, or even lagged sales from previous periods.

Data Splitting: Divide your dataset into training, validation, and test sets. The split ratios mentioned in the previous answer (70-80% for training, 10-15% for validation, and the rest for testing) can also be applied here.

Choose a Model: Start with simple regression models such as linear regression. Depending on the complexity of the data, you could explore more advanced regression algorithms like decision trees, random forests, gradient boosting, or even neural networks.

Model Training: Train your chosen model(s) on the training dataset. Tune hyperparameters to optimize the performance. Libraries like Scikit-Learn can be very helpful.

Model Evaluation: Evaluate the model using the validation set. Common metrics for regression problems include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Adjust the model and hyperparameters based on the evaluation results.

Fine-tuning and Validation: Iterate through model training and evaluation, adjusting parameters and making improvements based on validation results.