



Big Mart Store Sales



Objective

Nowadays, shopping malls and Big Marts keep track of individual item sales data in order to forecast future client demand and adjust inventory management. In a data warehouse, these data stores hold a significant amount of consumer information and particular item details. By mining the data store from the data warehouse, more anomalies and common patterns can be discovered.

Benefits :

- [] Keep the Stock which is more in demand.
- [] Give better insites of customer base
- [] Helps Easy flow and for managing resources

Data Validation and Data Transformation :

Name Validation - Data Validation is used to validate the data. checking the accuracy and quality of source data before training a new model version.

Number of Columns – Validation of number of columns present in the files.

Name of Columns - The name of the columns is validated and should be the same as given in the schema file.

Data type of columns - The data type of columns is given in the schema file. It is validated when we insert the files.

Null values in columns - If any of the columns in a file have all the values as NULL or missing, we discard such a file

Model Selection –

After the clusters are created, we find the best model for each cluster. By using Algorithms "Random Forest", "Linear Regression", "AdaBoost", "Gradient Boost", "SVM" and "XGBoost". For each cluster both the hyper tuned algorithms are used. We calculate the AUC scores for both models and select the model with the best score. Similarly, the model is selected for each cluster. All the models for every cluster are saved for use in prediction

Prediction:

The testing files are shared in the batches and we perform the same Validation operations ,data transformation and data insertion on them.

The accumulated data iis exported in csv format for prediction

We perform data pre-processing techniques on it.

Based on the cluster number respective model is loaded and is used to predict the data for that cluster.

Once the prediction is done for all the clusters. The predictions are saved in csv format and shared.



Thank
you