**Store Sales Prediction**

* Data Description

This is a Store Sales prediction data, and in this we have two data sets one is train.csv and other is test.csv.

Train.csv contains 8523 records and 12 columns, where Item\_Outlet\_Sale is Label column, whereas test.csv contains 5681 records and 11 columns.

**Directory**: Notebooks

**Python Script**: NA, all are ipynb files

**Work**: Contain all the jupyter notebook of data analysis.

-------------------------------------------------------------------------------------------------------

**Directory**: DataSet

**Python Script**: NA

**Work**: Contains the dataset

**Class**: NA

**Methods**: NA

-------------------------------------------------------------------------------------------------------

**Directory**: Logger

**Python Script**: Logging.py

**Work**: To log each and every activity of pipeline

**Class**: LogClass()

**Methods**: log

-------------------------------------------------------------------------------------------------------

**Directory**: DataLoader

**Python Script**: DataGetter.py

**Work**: Python script for loading the data from given source

**Class**: GetData()

**Methods**: load\_training\_data and load\_testing\_data

-------------------------------------------------------------------------------------------------------

**Directory**: DataFeatureClassification

**Python Script**: FeatureClassification.py

**Work**: Classifies the feature into categorical and numerical respectively.

**Class**: Features()

**Methods**: training\_numerical\_feature, training\_categorical\_features, testing\_numerical\_features and testing\_categorical\_features

-------------------------------------------------------------------------------------------------------

**Directory**: DataValidation

**Python Script**: DataValidation.py

**Work**: Validate the data on the basis of standard deviation, and missing values

**Class**: Validation()

**Methods**: training\_data\_stdzero, training\_data\_whole\_missing, training\_null\_values, testing\_data\_stdzero, testing\_data\_whole\_missing, testing\_null\_values

------------------------------------------------------------------------------------------------------

**Directory**: DataPreprocessing

**Python Script**: Preprocessing.py

**Work**: Pre-process or clean the data according to need.

**Class**: Preprocessing()

**Methods**: remove\_training\_duplicates, remove\_testing\_duplicates, trainig\_column\_cleaner, testing\_column\_cleaner.

-------------------------------------------------------------------------------------------------------

**Directory**: FeatureEngineering

**Python Script**: FeatureEngineering.py

**Work**: Apply feature engineering techniques on data.

**Class**: FeatureEngineering()

Methods: training\_categorical\_encoder, testing\_categorical\_encoder, training\_missing\_value\_imputation, testing\_missing\_value\_imputation, training\_outliers\_treatment, testing\_outleirs\_treatment, training\_feature\_selection, testinting\_feature\_selection

-------------------------------------------------------------------------------------------------------

**Directory**: ModelBuidling

**Python Script**: ModelBuilding.py

**Work**: To make ML model

**Class**: ModelBuilder()

**Methods**: random\_forest, xg\_boost, get\_best\_model

-------------------------------------------------------------------------------------------------------

**Directory**: Main

**Python Script**: Training.py

**Work**: To execute all the modules

**Class**: NA

**Methods**: NA

**Function**: training()

-------------------------------------------------------------------------------------------------------

**Directory**: Main

**Python Script**: app.py

**Work**: App for ML model

**Class**: NA

**Methods**: NA

**Function**: main(), predict\_sales()