BDNS_End_Term_Final_Assignment_Gourab Saha (C23012)

Problem Statement:

The data scientists at BigMart have collected 2013 sales data for 1559 products across 10 stores in different cities. Also, certain attributes of each product and store have been defined. The aim is to build a predictive model and find out the sales of each product at a particular store.

Using this model, BigMart will try to understand the properties of products and stores which play a key role in increasing sales.

Variable Description:

- **Item_Weight**- Weight of product
- **Item_Visibility**- The % of the total display area of all products in a store allocated to the particular product
- **Item_MRP**- Maximum Retail Price (list price) of the product
- **Outlet_Establishment_Year**- The year in which the store was established
- **Outlet_Size**- The size of the store in terms of ground area covered
- **Item_Outlet_Sales**- Sales of the product in t particular store. This is the outcome variable to be predicted.

Models Implemented:

- ➢ Logistic Regression
- ➢ Naïve Bayes
- ➢ Decision Tree
- ➢ Random Forest

Conclusion:

- Random Forest is giving highest accuracy because it is ensemble of decision tree i.e. bagging ensemble.
- Logistic regression is giving least accuracy because in logistic regression decision boundary should be linear and in given datapoints they are non-linearly distributed.
- Decision Tree is almost giving Same accuracy as random forest as classes are less so it can easily divide them one class to other class.
- For further improvement some hyperparameter optimization can be done to improve accuracy.