Assignment

- Title Bigmart sales Analysis
- Problem Statement:
  Bigmart sales analysis for data comprising of transaction records of a sales store. The data has 8523 rows and 12 cols. Predict the sales of the store
- Objective:
  To apply different regression techniques to find / predict the sales of a store
- Outcome:
  - To learn to preprocess tabular data
  - To apply different regression techniques.
- Theory:
  The data scientists of Bigm have collected 2013 sales data for 1559 products across 10 different stores in different cities

Different steps involved :-

- Data exploration:
  Looking at the categorical and continuous feature, summarize and make inferences about data

- Data cleaning:
  Inputing missing data values in the data such that zero or null Also checking for outliers

- Feature engineering:
  Modifying existing variables and creating new ones for analysis

- Model Building:
  Making predictive models on the data.

Since in the case of Bigmart analysis we have continuous values our target values (Item_Outlet_sales) this would come under regression problem

Algorithms :-

1) Linear regression :

It is a linear approach to modelling the relationships between a scalar response and one or more explanatory variables i.e independent variables

2) Random Forest :

Random Forest or random forest trees are an ensemble learning method for classification regression and other tasks. that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean / median prediction of individual trees.

Libraries used :-

- numpy
- pandas
- scikit learning

Splitting data into test and train data in ratio 30:70

- Analysis:

Algorithm	Validation score	Test score
Linear regression	1148.49	1277.805
Random forest regression	1138.185	1226.34
Evaluation Metric	Root mean squared errors	

- Conclusion:

We have thus build a machine learning model to predict outlet sales using big mart dataset.