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.

Some new products have to be added to BigMart for which business wants to predict the sales based on the sales analysis of the existing products.

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.

Please note that the data may have missing values as some stores might not report all the data due to technical glitches. Hence, it will be required to treat them accordingly.

Data

We have train ( sales analysis of existing products - 8523) and test (New products range - 5681) data set. You need to predict the sales for test data set.

|  |  |
| --- | --- |
| **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 particulat store. This is the outcome variable to be predicted. |

**Steps Involved**

1. Data Gathering
2. Data Cleaning (Filling Null Values, removing Garbage Data )
3. Feature Training
4. Research Of Market : Models and Algorithms ( With Efficiency)
5. Inferences – Checking Accuracy
6. Results
7. Improve Results – (Comparing Other Models on the basis of Output)

|  |  |  |
| --- | --- | --- |
| **Steps** | **Requirements** | **Action** |
| 1) Data Gathering | Relevent Data Sets, having sales details ,also require test data | have got 2013 BigMart datahaving sales details |
|  |  |  |
| 2) Data Cleaning (Filling Null Values, removing Garbage Data ) | Should have configured ML enabled with Python. | Downloaded Anaconda 3.x ( https://anaconda.org/) |
|  |  |  |
| 3) Feature Training | Should know Algorithms of Supervised and Unsupervised Learning. | Used DecisionTreeRegression- Linear Regression |
|  |  |  |
| 4) Research Of Market : Models and Algorithms ( With Efficiency) | Should have knowledge of ML models | Classification model |
|  |  |  |
| 5) Inferences – Checking Accuracy | Should Know functions of python | Used Libraries (Pandas, matplotlib,Sklearn,numpy ) for using functions. |
|  |  |  |
| 6) Results |  | got an output of sales prediction of new product with 3 attributes (item type,item id,outlet sales) |
|  |  |  |
| 7) Improve Results – (Comparing Other Models on the basis of Output) |  |  |