

Assignment OA-3

Title :- Bigmart sales Analysis.

Problem statement :-

For data comprising of transaction records of a sales store the data has 8523 rows of 12 variables. Predict the sales of store.

Objective :-

Learn Sales Analysis.

Outcomes :-

To analyze sales dataset and make prediction.

Theory :-

The data scientist at big Mart have collected 2013 sales data for 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. Big Mart will try to understand the properties of products and stores which play a key role in increasing sales.

The dataset consist of following attributes-

variable	Description
Item-identifier	Unique product ID.
Item_weight	weight of product
Item-fat_content	whether product is low fat or not
Item_visibility	The % of total display area of all products in a store allocated to particular product. The category to which product belongs.
Item-MRP	Max retail price of an item.
outlet-identifier	unique store id
outlet_size	Ground area of store
outlet_location_type	outlet is grocery / supermarket
Item_outlet_sales	sales of product in particular store

Valuation Metric :-

Root mean square error value

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\text{prediction} - \text{actual}_i)^2}{N}}$$

The problem can be explored in following steps.

1) Hypothesis generation :-

Understanding the problem better by brainstorming possible factors that can impact outcome

2) Data exploration :- Looking at categories and continuous feature summaries & making inferences about data

3) Data cleaning :- Imputing missing values in data and checking outliers

4) Feature Engineering :- Modifying existing variable and creating new ones for analysis

5] Model Building:-

Making predictive models on data

Different sets of hypothesis are created for stores and products:-

store level hypothesis include:-

- 1] City type
- 2] Population density
- 3] Capacity
- 4] Competitions
- 5] Marketing
- 6] Location
- 7] Customer behavior
- 8] Ambulance

Product level hypothesis include

- 1] Brand
- 2] Packaging
- 3] Utility
- 4] Display area
- 5] Visibility
- 6] Advertising
- 7] Promotion offers

Conclusion 2-

In this way big mart sales prediction is done using linear regression, RF & decision tree classifier.