41145

Page No.
Date

LP1 - DA3

* Title: Bigmart Sales Analysis

* Problem Statement:
For data comprising of transaction records of a sales store (8523 rows of 12 variables), predict the sales of a store.

* Objective:
To predict the sales for each item/product per store for a particular supermarket chain.

* S/w and H/w requirements:
i) 64 bit processor
ii) RAM
iii) Linux OS
iv) Python3

* Theory:
The BigMart Sales Analysis is a supervised Machine Learning, regression task, where an algorithm is expected to predict the sale price for a given product and store.

There are multiple influencing factors on the sales of a particular product, mainly the product itself and the type of store it is being sold at.
I) Store level hypothesis:
1) City type - Stores in urban areas should have higher sales due to high income households.

Page No.
Date

2) Population density :- Densely populated areas will have more sales.

3) Store capacity
4) Competitors
5) Establishment year.

II) Product level hypothesis:
1) Item Advertisement (visibility)
2) Item utility (type)
3) Price

Exploratory data analysis showed that-
i) Item visibility did not have a high correlation (+ve) as expected. It also had a lot of 0 values.
ii) No huge variations in sales due to item type either.
iii) Item weight & outlet size have values '0' or 'NaN'
iv) Item_fat_content contains varying values for 'Lowfat'
v) item_type can be converted to a more useful feature.

* Conclusion:
From this assignment, I was able to successfully analyze the dataset using the linear and decision tree regression models.