Rohit Kulkarni 41346

M T W T F S S
Page No.:
Date:
YOUVA

DA ASSIGNMENT 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 the store.

OBJECTIVES:

To predict the sales for each item (product) per store of a particular supermarket chain.

LEARNING OUTCOMES:

Identify the products which play a key role in the sales of the supermaket chain (best & worst performing) to enable proper strategies to be put in place to ensure the business success.

S/W & H/W REQUIREMENTS:

Python 3, Jupyter notebook, sklearn, matplotlib, seaborn, UNIX, LINUX Based OS, pandas, numpy, keyboard, mouse.

THEORY:

1] The BigMart Sales Analysis (Prediction) is a supervised machine learning, regression task, where an algorithm is expected to predict the sale price for a given product & store.

A more in depth analysis of the main factors -

Store level Hypothesis :

1. City type
2. Population density
3. Store capacity
4. Competitors
5. Establishment year

Product level Hypothesis :

1. Item advertisment (visibility)
2. Item utility (type)
3. Price

Exploratory Data Analysis showed that :

1] Item visibility did not have a high correlation (+ve) as expected. It also have a lot of 0 values

2] No huge variations due to Item_Type in sales either

3] Item_Weight & Outlet_size have 0 values or NaN.

4] Item_Fat_Content contains varying values for 'lowfat'

5] Item_Type can be converted to a more useful feature

M T W T F S S

Page No.: YOUVA

Date:

i) Linear Regression & Ridge Regression models were built to perform the actual prediction.

ii) Both models performed within the same range, giving a Root Mean squared Error of.

iii) Decision Tree Regression model was then built, resulting in an improved Root Mean Squared Error.

CONCLUSION:

The BigMart Sales was successfully predicted using the linear, ridge & decision tree regression models.