ASSIGNMENT C3

Title: BigMart Sales Analysis

Problem statement /definition:
For data comprising of transaction records of a sales store, predict the sales of the store in future. The data has 8523 rows of 12 variables.

Objectives:
- To understand sales prediction.
- To understand use of various data vizuation tools to draw necessary conclusions.
- To implement various predictive algorithms on given dataset.

Outcome:
We will be able to:
- Perform sales prediction.
- Learn about various data visualization tools.
- Lean implementation of various predictive algorithms.

Hardware and software requirements:
- Os : Fedora 20 /Ubuntu (64-bit)
- RAM : 4 GB
- HDD : 500GB
- Jupyter Notebook
- Python libraries.

Theory:

- There are two types of data analysis that can be used for extracting models describing important classes or to predict future trends.
- These forms are as follows:
  1) Classification
  2) Prediction

- Classification model predicts categorical classes labels & prediction model predicts continuous valued function

- Prediction:
  - Suppose the modelling manager needs to predict how much a given customer will spend during a sales at his company.
  - In this example, we are bothered to predict a numeric value therefore the data analytics task is an example of numeric prediction.
  - In this case a model or a predictor will be constructed that predicts a continuous valued function or ordered value.

→ Logistic Regression:
- It is a statistical analysis method used to predict a data value based on prior observation of data set.
- It predicts a dependent variable by analysing the relationship between one or more existing independent variable

In logistic regression, we seek to find vector β of parameter in the following equation that the minimizes the cost function.

$$\log_i^t(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = B_0 + B_1 x_i + \ldots B_k x_b^i$$

Algorithm:

1. Hypothesis Generation:
   Understand the problem better by brainstorming possible factors that can impact outcome

2. Data Exploration:
   Looking at categorical and continuous feature summarizes and making interface about data

3. Data cleaning:
   Imputing missing values of dataset and checking for outliers.

4. Feature engineering:
   Modifying existing variables & creating new ones for analysis

5. Model Building:
   Making predictive models on given dataset.

Test cases & analysis

Input.
test.csv
train.csv.

Output
- Graphics for drawing necessary conclusions
- Predictive model's output, predicting future sales.

Conclusion:

Hence, we have successfully predicted the future sales of BigMart Sales dataset using various predictive models.