**PG-DBDA CDAC Mumbai**

**Subject: Practical Machine Learning**

**Marks: 40 Time: 2hrs.**

**Problem Statement 1:**

Amit wants to purchase office space. He does a detailed survey of the offices and corporate complexes in the area and tries to quantify a lot of factors, such as the distance of the offices from residential and other commercial areas, schools, and workplaces; the reputation of the construction companies, and builders involved in constructing the apartments; the distance of the offices from highways, freeways, and important roads; the facilities around the office space and so on.

Each of these factories was quantified, normalized, and mapped to values on a scale of 0 to 1. Amit then makes a table. Each row in the table corresponds to Amit's observations for a particular house. If Amit has observed and noted F features, the row contains F values separated by a single space, followed by the office-space price in dollars/square-foot. If Amit makes observations for H houses, his observation table has (F+1) columns and H rows, and a total of (F+1) \* H entries.

Amit does several such surveys and provides you with the tabulated data. At the end of these tables are some rows which have just F columns (the price per square foot is missing). Your task is to predict these prices. F can be any integer number between 1 and 5, both inclusive.

There is one important observation that Amit has made. The prices per square foot, are (approximately) a polynomial function of the features in the observation table. This polynomial always has an order less than 4

**Dataset:** data.csv

**Input values in the dataset:**

The first line contains F and N. Over here, F is the number of observed features. N is the number of rows for which features as well as price per square foot.

**Output values in the prediction:**

1 = 180.38

2 = 1312.07

3 = 440.13

4 = 343.72

[20 Marks, 60 minutes]

**Problem Statement 2:**

A dataset collected in a cosmetics shop showing details of customers and whether or not they responded to a special offer to buy a new lipstick is shown in the table below. Use this dataset to build a decision tree, with Buys as the target variable, to help in buying lip-sticks in the future. Find the root node of the decision tree. According to the decision tree, you have made from the previous Training data set, what is the decision for the test data: [Age < 21, Income = Low, Gender = Female, Marital Status = Married]? Write the concluding statement for the implemented application.

[20 Marks, 60 minutes]

**Dataset:**

