MES Wadia College of Engineering Pune-01

Department of Computer Engineering

Name of Student:	**Class:**
Semester/Year:	**Roll No:**
Date of Performance:	**Date of Submission:**
Examined By:	**Experiment No: Part A-05**

PART: A) ASSIGNMENT NO: 05

Title: Data Analytics-II

1. Implement logistic regression using Python/R to perform classification on Social_Network_Ads.csv dataset.

2. Compute Confusion matrix to find TP, FP, TN, FN, Accuracy, Error rate, Precision, Recall on the given dataset.

OBJECTIVES:

Students should be able to data analysis using logistic regression using Python for any open source dataset.

PREREQUISITE:

- Basic of Python Programming
- Concept of Regression.

APPRATUS:

- Programming Language: Python.
- Dataset: Boston Dataset (https://www.kaggle.com)

ALGORITHM STEP: (Boston Dataset):

Step 1: Import libraries and create alias for Pandas, Numpy and Matplotlib

Step 2: Import the Social_Media_Adv Dataset

Step 3: Initialize the data frame

Step 4: Perform Data Preprocessing

Step 5: Use Logistic regression(Train the Machine) to Create Model

Step 6: Predict the y_pred for all values of train_x and test_x

Step 7: Evaluate the performance of Model for train_y and test_y

Step 8: Calculate the required evaluation parameters

CONCLUSION:

QUESTIONS:

1. Explain Logistic Regression in details
2. Differentiate between Linear and Logistic Regression
3. Consider the binary classification task with two classes positive and negative. Find out TP,TP, FP, TN, FN, Accuracy, Error rate, Precision, Recall

N = 165	Predicted YES	Predicted NO
Actual YES	TP = 150	FN = 10
Actual NO	FP = 20	TN = 100

4. Comment on whether the model is best fit or not based on the calculated values.
5. Write python code for the preprocessing mentioned in step 4. And explain every step in detail.