

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-06

PART: A) ASSIGNMENT NO: 06

Title: Data Analytics-III

1. Implement Simple Naïve Bayes classification algorithm using Python/R on iris.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 Naive Bayes Algorithm using Python for any open source dataset.

PREREQUISITE:

- Basic of Python Programming
- Concept of Join and Marginal Probability.

APPARATUS:

- Programming Language: Python.
- Dataset: Iris Dataset (<https://www.kaggle.com>)

ALGORITHM STEP: (Iris Dataset):

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

Step 2: Import the Iris dataset by calling URL.

Step 3: Initialize the data frame

Step 4: Perform Data Preprocessing

Step 5: Use Naive Bayes algorithm (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 concepts of Naïve Bayes classifier.
2. Consider the observation for the car theft scenario having 3 attributes color, Type and origin.

Example No	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Find the probability of car theft having scenarios Red SUV and Domestic.

3. Write python code for the preprocessing mentioned in step 4. And explain every step in detail.