**SVKM’s NMIMS**

**Mukesh Patel School of Technology Management & Engineering**

Program: B.Tech\MBA.Tech\MBA.Tech AI

**Course: Machine Learning**

**Experiment No.07**

PART A

(PART A : TO BE REFFERED BY STUDENTS)

**A.1 Aim:** To implement Naïve Bayes algorithm for classification.

**A.2 Prerequisite:**

Python Programming, Pandas library, Numpy Library, MatplotLib, Seaborn Library

**A.3 Outcome:**

**After successful completion of this experiment students will be able to:**

1. Implement Naïve Bayes technique for the classification
2. Compare results of Naïve Bayes and KNN
3. Understand and infer results of different classification metrics

**A.4 Theory:**

**A.4.1 Naïve Bayes Classifier:**

* Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems.
* It is mainly used in *text classification* that includes a high-dimensional training dataset.
* Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
* **It is a probabilistic classifier, which means it predicts on the basis of the probability of an object**.
* Some popular examples of Naïve Bayes Algorithm are **spam filtration, Sentimental analysis, and classifying articles**.

**A.4.2 Bayes’ Theorem:**

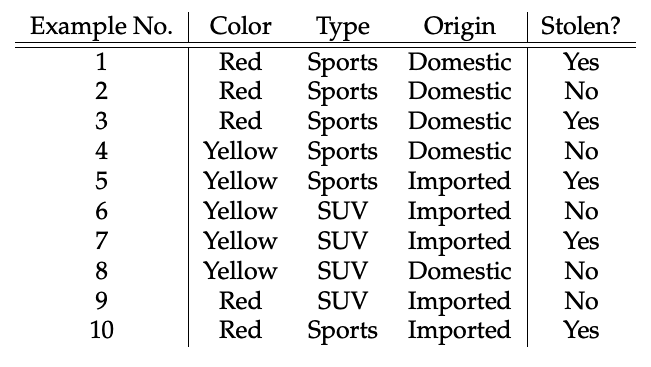
* Bayes' theorem is also known as Bayes' Rule or Bayes' law, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
* The formula for Bayes' theorem is given as:
* Naïve Bayes Classifier Algorithm
* Where,
* P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.
* P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.
* P(A) is Prior Probability: Probability of hypothesis before observing the evidence.
* P(B) is Marginal Probability: Probability of Evidence.

**A.4.3 Working of Naïve Bayes Classifier:**

* Convert the given dataset into frequency tables.
* Generate Likelihood table by finding the probabilities of given features.
* Now, use Bayes theorem to calculate the posterior probability.

**Tasks:**

**Task 1: For the given car data set, apply Naïve Bayes algorithm for the classification. Show all the steps of training phase and identify the class for test data point (color =Yellow, Type=Sports, Origin=Domestic). Solve answer on paper and upload the image.**



**Task 2:** **For the given adult.csv dataset, perform below operations.**

* **Upload data set into the dataframe**
* **Check the shape of the data set.**
* **Find out all the categorical columns from the data set**
* **Check if null values exist in all the categorical columns**
* **Identify the problems with “workclass”, “Occupation”,”native\_country” columns and rectify it.**
* **Explore numeric columns and check any null values exist for the numeric columns.**
* **Create a feature vector with x= all the columns except income and y=income**
* **Implement feature engineering for the train, test split data set:**
  + **Check the data types of columns of the input features of training data set**
  + **Identify categorical columns that has null values and fill them with most probable value in the data set**
  + **Repeat above step for the input features of test data set**
  + **Apply one hot encoding on all the categorical columns**
  + **Apply feature scaling using robust scaler**

**Task 3: Implement KNN algorithm on the given data set from Sklearn library with k=5**

**Task 4: Implement Naïve Bayes on the given data set.**

**Task 5: Compare confusion matrix for both the classifier**

**Task 6: Compare accuracy score of both the classifier.**

**Task 7: Draw ROC curve for comparing both the models**

PART B

(PART B : TO BE COMPLETED BY STUDENTS)

***(Students must submit the soft copy as per following segments within two hours of the practical.)***

|  |  |
| --- | --- |
| Roll No. I066 | Name: Srihari Thyagarajan |
| Class: I | Batch: B2 |
| Date of Experiment: 09/03/2023 | Date of Submission: 09/03/2023 |
| Grade : |  |

**B.1 Tasks**

**B.4 Conclusion:**

*(Students must write the conclusion in their own words.)*