Pune District Education Association's

College Of Engineering

Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

DSBDAL

Assignment No:-06

* problem statement :- Data Analytics III

1) Implement simple Navie Bayes classification algorithm using python /R on iris. csv. dataset.

2) comput confusion matrix to find TP, FP, TN, FN, Accuracy, Error rate, precision, Recall on the given dataset.

बहुजन हिताय, बहुजन सुखाय।

Navie Bayes is a stastical classification technique based on the Bayes theorem & one of the simplest supervised learning algorithm. The navie bayes classifier is a quick, accurate and trustworthy method espically on large datasets.

simple formula of Bayes theoram.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Where P(A) & P(B) are two independent events & P(B) is not equal to zero.

P(A|B) : is the conditional probability of event A occuring given that B is true.

P(B|A): is the conditional probablity of event B occuring given that A is true.

P(A) & P(B) are the probabilities of A & B occuring independently of one another.

What is Navie Bayes classification?

- The Navie Bayes classification algorithm is a probabilities classifier, and it belongs to suspperuised learning. It is based on probabilitity model that incorporate strong independent assumption.
Therefore they are considered naive.

- Another assumption made by the naive bayes classifier is that all the predictors have an equal effect on the outcome.

- The navie bayes classification has the following different types:-

• Multionominal Naive Bayes method:-

It is a common Bayesian learning approach in natural language processing.
Using the Bayes theorem, the program estimates the tag of a text, such as an email.
It assesses the likelihood of each tag for a given sample.

• Bernouli Naive Bayes:-

It is a part of the family of Naive Bayes.
It is only take binary values. There a may multiple features, but each is

assumed to be a binary-valued variable.

• Gaussian Naive Bayes:-

It is a variant of Naive Bayes that follows gaussian normal distribution and support continuous data.
To build a simple model using gaussian naive bayes, we assume the data is characterized by a guassian distribution with no covariance betn parameter.
This model may be fit by calculating the mean and standard deviation of the points within each label.

* conclusion:-
In above problem statement we use iris flower dataset to implement simple naive bayes classification algorithm.
use sepal length, sepal width, petal length & petal width as input & class is as output.

* csv File / Dataset - Iris_dataset
Required libraries
import pandas as pd
import numpy as np

```
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.datasets import load_iris
from sklearn.processing import standardscaler
from sklearn.model_selection import train_test_split.
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import confusion_matrix.
  accuracy_score, classification_report,
precision score, recall score, f1_score.
import warning
warning.filterwarnings ("ignore")
%matplotlib inline.
```

* functions used

```
iris = load_iris()
iris.keys()
x.head()
x.shape , y.shape
x.info()
y.info()
x.describe()
```

Data preparation
model Building
Evaluation

```
cm= confusion_matrix()
plot= confusion_matrix()
```

calculate Accuracy score

"(Accuracy score is {accuracy_score (y-test, y-pred)}")

calculate Error rate
(" Error rate is {1- accuracy_score (y-test, y-pred)}")

calculate precision score
("precision score is {precision_score (y-test y_pred , average = '')}")

calculate Recall score
("Recall score is {Recall_score (y_test, y_pred , average = ' ' })}

classification Report

print (classification_report (y_test, y_pred)

Q.1) Explain Bayes Theorem

→ $$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{——— ①}$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \quad \text{——— ②}$$

From ① & ②

$$P(A|B) \cdot P(B) = P(A \cap B) \quad — ③$$

$$P(B|A) \cdot P(A) = P(B \cap A) \quad — ④$$

from eqⁿ ③ & ④

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \rightarrow ⑤$$

posterior $= P(A|B)$ probability of the hypothesis given that.

$P(B|A)$ = likelihood probability

$P(A)$ - prior probability of hypothesis.

$P(B)$ - marginal probability of evidence.