



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 naive 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.

बहुजन हिताय, बहुजन सुखाय।

Naive Bayes is a statistical classification technique based on the Bayes theorem & one of the simplest supervised learning algorithm. The naive bayes classifier is a quick, accurate and trustworthy method especially on large datasets.

simple formula of Bayes theorem.

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

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 occurring given that B is true.

$p(B|A)$: is the conditional probability of event B occurring given that A is true.

$p(A)$ & $p(B)$ are the probabilities of A & B occurring independently of one another.

What is Naive Bayes classification?

- The Naive Bayes classification algorithm is a probabilities classifier, and it belongs to supervised learning. It is based on probability 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 naive Bayes classification has the following different types:-

- Multinomial 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.

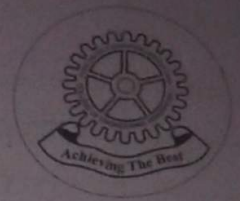
It only takes binary values. There may be multiple features, but each is



Pune District Education Association's College Of Engineering

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

Accredited by NAAC



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 gaussian distribution with no covariance between parameters.

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.preprocessing 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 warnings
warnings.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



Pune District Education Association's College Of Engineering

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

Accredited by NAAC



"(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

Pune District Education Association

point (classification-report (y-test, y-pred))

Q.1) Explain Bayes Theorem

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

①

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

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.