**TOPIC: PREDICTION AND ANALYSIS**

**AIM:**

To predict the accuracy of 4 algorithms on an iris dataset and infer the most significant one that is used in prediction of iris dataset class analysis.

**DATASET USED: Iris dataset**

**SOURCE CODE:**

import pandas

from pandas.plotting import scatter\_matrix

import matplotlib.pyplot as plt

from sklearn import model\_selection

from sklearn.metrics import classification\_report

from sklearn.metrics import confusion\_matrix

from sklearn.metrics import accuracy\_score

from sklearn.linear\_model import LogisticRegression

from sklearn.tree import DecisionTreeClassifier

from sklearn.neighbors import KNeighborsClassifier

from sklearn.discriminant\_analysis import LinearDiscriminantAnalysis

from sklearn.naive\_bayes import GaussianNB

from sklearn.svm import SVC

url = "https://raw.githubusercontent.com/jbrownlee/Datasets/master/iris.csv"

names = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'class']

dataset = pandas.read\_csv(url, names=names)

#shape

print(dataset.shape)

#head

print(dataset.head(20))

#description

print(dataset.describe())

#class distribution

print(dataset.groupby('class').size())

#data visualization

dataset.plot(kind='box',subplots=True, layout=(2,2), sharex=False, sharey=False)

plt.show()

scatter\_matrix(dataset)

plt.show()

#dataset spliting

array = dataset.values

X = array[:,0:4]

Y = array[:,4]

validation\_size = 0.30

seed = 6

X\_train, X\_validation, Y\_train, Y\_validation = model\_selection.train\_test\_split(X, Y, test\_size=validation\_size,random\_state=seed)

# Test options and evaluation metric

seed = 6

scoring = 'accuracy'

#algorithms

models = []

models.append(('LR', LogisticRegression()))

models.append(('LDA', LinearDiscriminantAnalysis()))

models.append(('KNN', KNeighborsClassifier()))

models.append(('CART', DecisionTreeClassifier()))

models.append(('NB', GaussianNB()))

models.append(('SVM', SVC()))

# evaluate each model in turn

results = []

names = []

for name, model in models:

kfold = model\_selection.KFold(n\_splits=10, random\_state=seed)

cv\_results = model\_selection.cross\_val\_score(model, X\_train, Y\_train, cv=kfold, scoring=scoring)

results.append(cv\_results)

names.append(name)

msg = "%s: %f (%f)" % (name, cv\_results.mean(), cv\_results.std())

print(msg)

# Compare Algorithms

fig = plt.figure()

fig.suptitle('Algorithm Comparison')

ax = fig.add\_subplot(111)

plt.boxplot(results)

ax.set\_xticklabels(names)

plt.show()

# Make predictions on validation dataset

knn = KNeighborsClassifier()

knn.fit(X\_train, Y\_train)

predictions = knn.predict(X\_validation)

print(accuracy\_score(Y\_validation, predictions))

print(confusion\_matrix(Y\_validation, predictions))

print(classification\_report(Y\_validation, predictions))

**DATA SET USED:**

Paramaters[sepal length, sepal width, petal length, petal width, Class]

5.1,3.5,1.4,0.2,Iris-setosa

4.9,3.0,1.4,0.2,Iris-setosa

4.7,3.2,1.3,0.2,Iris-setosa

4.6,3.1,1.5,0.2,Iris-setosa

5.0,3.6,1.4,0.2,Iris-setosa

5.4,3.9,1.7,0.4,Iris-setosa

4.6,3.4,1.4,0.3,Iris-setosa

5.0,3.4,1.5,0.2,Iris-setosa

4.4,2.9,1.4,0.2,Iris-setosa

4.9,3.1,1.5,0.1,Iris-setosa

5.4,3.7,1.5,0.2,Iris-setosa

4.8,3.4,1.6,0.2,Iris-setosa

4.8,3.0,1.4,0.1,Iris-setosa

4.3,3.0,1.1,0.1,Iris-setosa

5.8,4.0,1.2,0.2,Iris-setosa

5.7,4.4,1.5,0.4,Iris-setosa

5.4,3.9,1.3,0.4,Iris-setosa

5.1,3.5,1.4,0.3,Iris-setosa

5.7,3.8,1.7,0.3,Iris-setosa

5.1,3.8,1.5,0.3,Iris-setosa

5.4,3.4,1.7,0.2,Iris-setosa

5.1,3.7,1.5,0.4,Iris-setosa

4.6,3.6,1.0,0.2,Iris-setosa

5.1,3.3,1.7,0.5,Iris-setosa

4.8,3.4,1.9,0.2,Iris-setosa

5.0,3.0,1.6,0.2,Iris-setosa

5.0,3.4,1.6,0.4,Iris-setosa

5.2,3.5,1.5,0.2,Iris-setosa

5.2,3.4,1.4,0.2,Iris-setosa

4.7,3.2,1.6,0.2,Iris-setosa

4.8,3.1,1.6,0.2,Iris-setosa

5.4,3.4,1.5,0.4,Iris-setosa

5.2,4.1,1.5,0.1,Iris-setosa

5.5,4.2,1.4,0.2,Iris-setosa

4.9,3.1,1.5,0.1,Iris-setosa

5.0,3.2,1.2,0.2,Iris-setosa

5.5,3.5,1.3,0.2,Iris-setosa

4.9,3.1,1.5,0.1,Iris-setosa

4.4,3.0,1.3,0.2,Iris-setosa

5.1,3.4,1.5,0.2,Iris-setosa

5.0,3.5,1.3,0.3,Iris-setosa

4.5,2.3,1.3,0.3,Iris-setosa

4.4,3.2,1.3,0.2,Iris-setosa

5.0,3.5,1.6,0.6,Iris-setosa

5.1,3.8,1.9,0.4,Iris-setosa

4.8,3.0,1.4,0.3,Iris-setosa

5.1,3.8,1.6,0.2,Iris-setosa

4.6,3.2,1.4,0.2,Iris-setosa

5.3,3.7,1.5,0.2,Iris-setosa

5.0,3.3,1.4,0.2,Iris-setosa

7.0,3.2,4.7,1.4,Iris-versicolor

6.4,3.2,4.5,1.5,Iris-versicolor

6.9,3.1,4.9,1.5,Iris-versicolor

5.5,2.3,4.0,1.3,Iris-versicolor

6.5,2.8,4.6,1.5,Iris-versicolor

5.7,2.8,4.5,1.3,Iris-versicolor

6.3,3.3,4.7,1.6,Iris-versicolor

4.9,2.4,3.3,1.0,Iris-versicolor

6.6,2.9,4.6,1.3,Iris-versicolor

5.2,2.7,3.9,1.4,Iris-versicolor

5.0,2.0,3.5,1.0,Iris-versicolor

5.9,3.0,4.2,1.5,Iris-versicolor

6.0,2.2,4.0,1.0,Iris-versicolor

6.1,2.9,4.7,1.4,Iris-versicolor

5.6,2.9,3.6,1.3,Iris-versicolor

6.7,3.1,4.4,1.4,Iris-versicolor

5.6,3.0,4.5,1.5,Iris-versicolor

5.8,2.7,4.1,1.0,Iris-versicolor

6.2,2.2,4.5,1.5,Iris-versicolor

5.6,2.5,3.9,1.1,Iris-versicolor

5.9,3.2,4.8,1.8,Iris-versicolor

6.1,2.8,4.0,1.3,Iris-versicolor

6.3,2.5,4.9,1.5,Iris-versicolor

6.1,2.8,4.7,1.2,Iris-versicolor

6.4,2.9,4.3,1.3,Iris-versicolor

6.6,3.0,4.4,1.4,Iris-versicolor

6.8,2.8,4.8,1.4,Iris-versicolor

6.7,3.0,5.0,1.7,Iris-versicolor

6.0,2.9,4.5,1.5,Iris-versicolor

5.7,2.6,3.5,1.0,Iris-versicolor

5.5,2.4,3.8,1.1,Iris-versicolor

5.5,2.4,3.7,1.0,Iris-versicolor

5.8,2.7,3.9,1.2,Iris-versicolor

6.0,2.7,5.1,1.6,Iris-versicolor

5.4,3.0,4.5,1.5,Iris-versicolor

6.0,3.4,4.5,1.6,Iris-versicolor

6.7,3.1,4.7,1.5,Iris-versicolor

6.3,2.3,4.4,1.3,Iris-versicolor

5.6,3.0,4.1,1.3,Iris-versicolor

5.5,2.5,4.0,1.3,Iris-versicolor

5.5,2.6,4.4,1.2,Iris-versicolor

6.1,3.0,4.6,1.4,Iris-versicolor

5.8,2.6,4.0,1.2,Iris-versicolor

5.0,2.3,3.3,1.0,Iris-versicolor

5.6,2.7,4.2,1.3,Iris-versicolor

5.7,3.0,4.2,1.2,Iris-versicolor

5.7,2.9,4.2,1.3,Iris-versicolor

6.2,2.9,4.3,1.3,Iris-versicolor

5.1,2.5,3.0,1.1,Iris-versicolor

5.7,2.8,4.1,1.3,Iris-versicolor

6.3,3.3,6.0,2.5,Iris-virginica

5.8,2.7,5.1,1.9,Iris-virginica

7.1,3.0,5.9,2.1,Iris-virginica

6.3,2.9,5.6,1.8,Iris-virginica

6.5,3.0,5.8,2.2,Iris-virginica

7.6,3.0,6.6,2.1,Iris-virginica

4.9,2.5,4.5,1.7,Iris-virginica

7.3,2.9,6.3,1.8,Iris-virginica

6.7,2.5,5.8,1.8,Iris-virginica

7.2,3.6,6.1,2.5,Iris-virginica

6.5,3.2,5.1,2.0,Iris-virginica

6.4,2.7,5.3,1.9,Iris-virginica

6.8,3.0,5.5,2.1,Iris-virginica

5.7,2.5,5.0,2.0,Iris-virginica

5.8,2.8,5.1,2.4,Iris-virginica

6.4,3.2,5.3,2.3,Iris-virginica

6.5,3.0,5.5,1.8,Iris-virginica

7.7,3.8,6.7,2.2,Iris-virginica

7.7,2.6,6.9,2.3,Iris-virginica

6.0,2.2,5.0,1.5,Iris-virginica

6.9,3.2,5.7,2.3,Iris-virginica

5.6,2.8,4.9,2.0,Iris-virginica

7.7,2.8,6.7,2.0,Iris-virginica

6.3,2.7,4.9,1.8,Iris-virginica

6.7,3.3,5.7,2.1,Iris-virginica

7.2,3.2,6.0,1.8,Iris-virginica

6.2,2.8,4.8,1.8,Iris-virginica

6.1,3.0,4.9,1.8,Iris-virginica

6.4,2.8,5.6,2.1,Iris-virginica

7.2,3.0,5.8,1.6,Iris-virginica

7.4,2.8,6.1,1.9,Iris-virginica

7.9,3.8,6.4,2.0,Iris-virginica

6.4,2.8,5.6,2.2,Iris-virginica

6.3,2.8,5.1,1.5,Iris-virginica

6.1,2.6,5.6,1.4,Iris-virginica

7.7,3.0,6.1,2.3,Iris-virginica

6.3,3.4,5.6,2.4,Iris-virginica

6.4,3.1,5.5,1.8,Iris-virginica

6.0,3.0,4.8,1.8,Iris-virginica

6.9,3.1,5.4,2.1,Iris-virginica

6.7,3.1,5.6,2.4,Iris-virginica

6.9,3.1,5.1,2.3,Iris-virginica

5.8,2.7,5.1,1.9,Iris-virginica

6.8,3.2,5.9,2.3,Iris-virginica

6.7,3.3,5.7,2.5,Iris-virginica

6.7,3.0,5.2,2.3,Iris-virginica

6.3,2.5,5.0,1.9,Iris-virginica

6.5,3.0,5.2,2.0,Iris-virginica

6.2,3.4,5.4,2.3,Iris-virginica

5.9,3.0,5.1,1.8,Iris-virginica

**EXECUTION AND OUTPUT SCREENSHOTS**

===== RESTART: C:\Users\Harini\Desktop\predection\prediction test.py =====

Printing the datasets:

(150, 5)

sepal-length sepal-width ... petal-width class

0 5.1 3.5 ... 0.2 Iris-setosa

1 4.9 3.0 ... 0.2 Iris-setosa

2 4.7 3.2 ... 0.2 Iris-setosa

3 4.6 3.1 ... 0.2 Iris-setosa

4 5.0 3.6 ... 0.2 Iris-setosa

5 5.4 3.9 ... 0.4 Iris-setosa

6 4.6 3.4 ... 0.3 Iris-setosa

7 5.0 3.4 ... 0.2 Iris-setosa

8 4.4 2.9 ... 0.2 Iris-setosa

9 4.9 3.1 ... 0.1 Iris-setosa

10 5.4 3.7 ... 0.2 Iris-setosa

11 4.8 3.4 ... 0.2 Iris-setosa

12 4.8 3.0 ... 0.1 Iris-setosa

13 4.3 3.0 ... 0.1 Iris-setosa

14 5.8 4.0 ... 0.2 Iris-setosa

15 5.7 4.4 ... 0.4 Iris-setosa

16 5.4 3.9 ... 0.4 Iris-setosa

17 5.1 3.5 ... 0.3 Iris-setosa

18 5.7 3.8 ... 0.3 Iris-setosa

19 5.1 3.8 ... 0.3 Iris-setosa

Detailed description of the dataset:

[20 rows x 5 columns]

sepal-length sepal-width petal-length petal-width

count 150.000000 150.000000 150.000000 150.000000

mean 5.843333 3.054000 3.758667 1.198667

std 0.828066 0.433594 1.764420 0.763161

min 4.300000 2.000000 1.000000 0.100000

25% 5.100000 2.800000 1.600000 0.300000

50% 5.800000 3.000000 4.350000 1.300000

75% 6.400000 3.300000 5.100000 1.800000

max 7.900000 4.400000 6.900000 2.500000

class

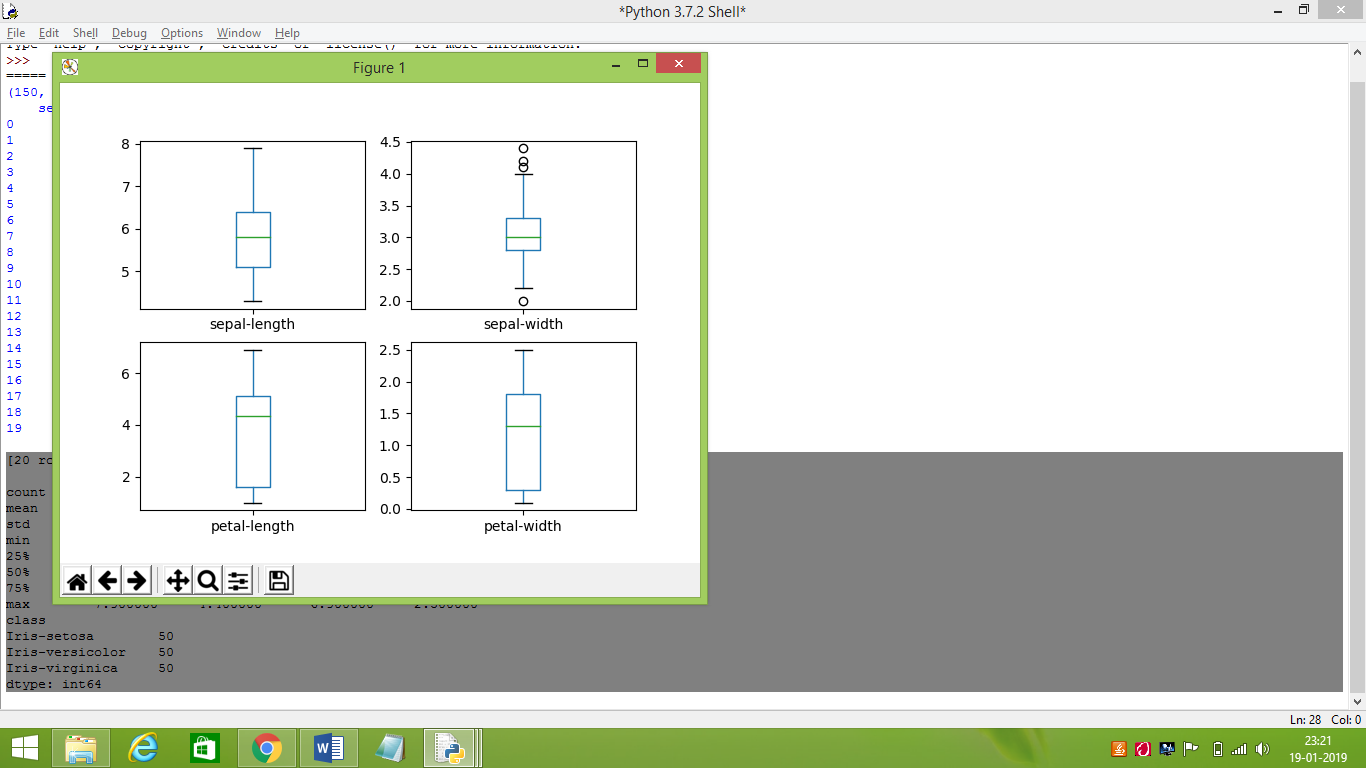
Iris-setosa 50

Iris-versicolor 50

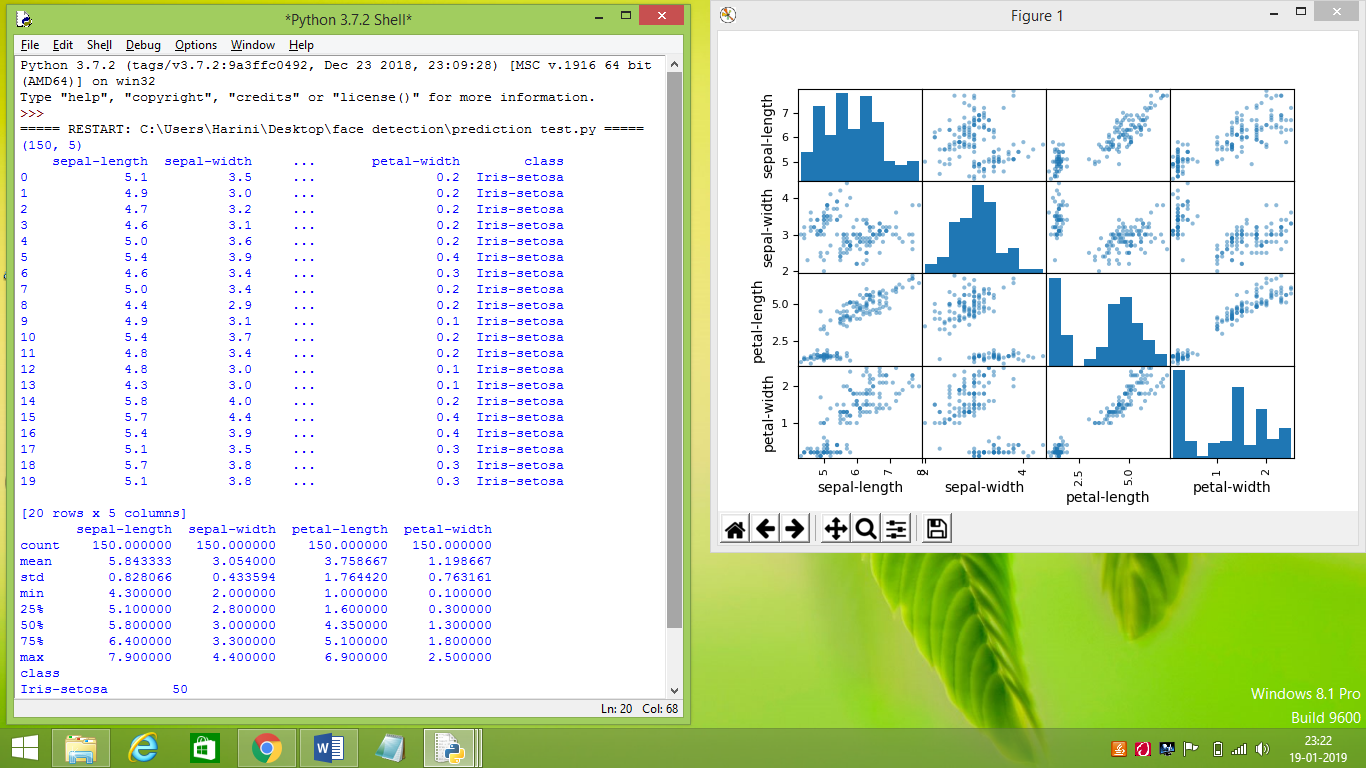
Iris-virginica 50

dtype: int64

Graphical representation of the 4 parameters within a given range used to classify the flowers and predict the class .



Plotting all the 3 classes of flowers and comparing them by a graph where the x and y ordinates are the initial parameters



Printing and graphically representing the efficiency of algorithms.

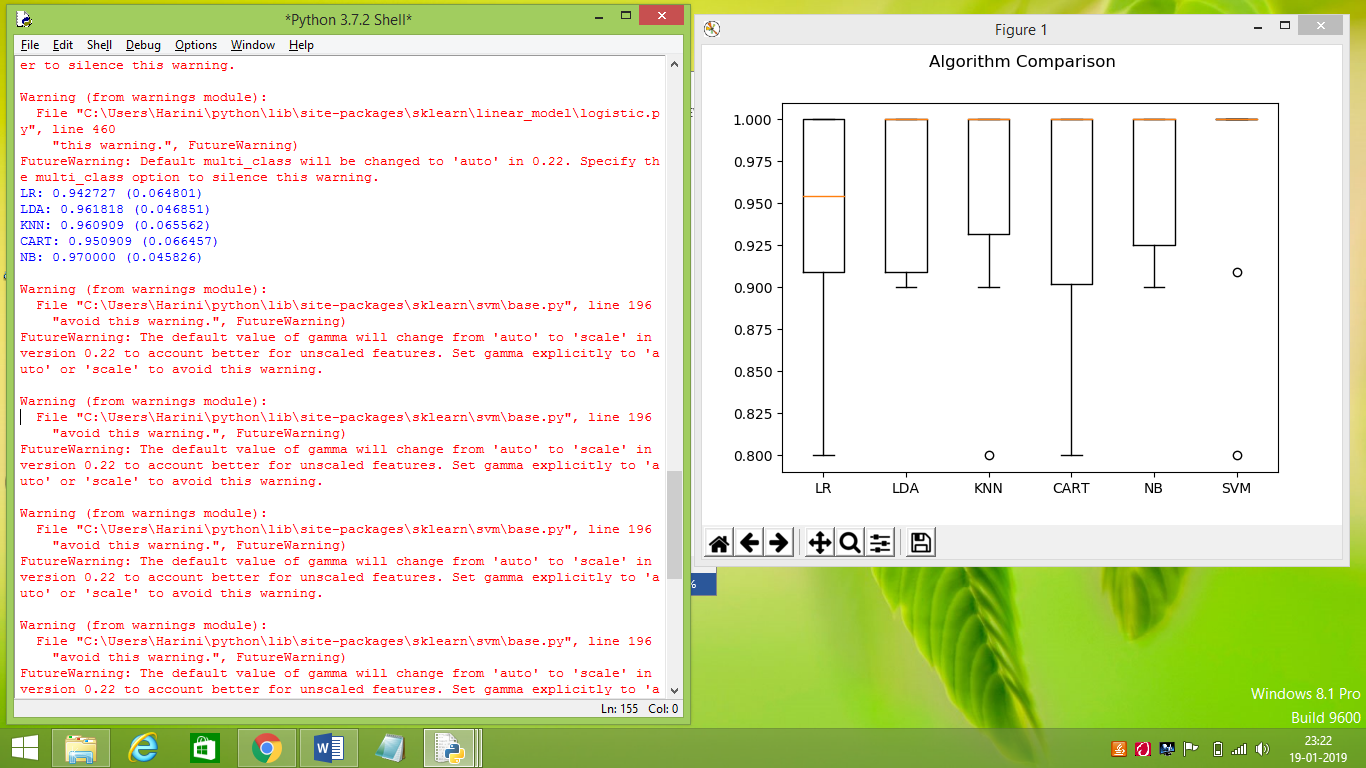
LR: 0.942727 (0.064801)

LDA: 0.961818 (0.046851)

KNN: 0.960909 (0.065562)

CART: 0.950909 (0.066457)

NB: 0.970000 (0.045826)



SVM: 0.970909 (0.063089)

0.9777777777777777

[[15 0 0]

[ 0 15 0]

[ 0 1 14]]

precision recall f1-score support

Iris-setosa 1.00 1.00 1.00 15

Iris-versicolor 0.94 1.00 0.97 15

Iris-virginica 1.00 0.93 0.97 15

micro avg 0.98 0.98 0.98 45

macro avg 0.98 0.98 0.98 45

weighted avg 0.98 0.98 0.98 45

**INFERENCE:**

From the above analysis it is inferred that LDA algorithm is efficient in preventing the class of iris flower with its sepal and petal length,width.