

# JADAVPUR UNIVERSITY

## MACHINE LEARNING LABORATORY

### ASSIGNMENT #2 - D) WISCONSIN BREAST CANCER DATASET

NAME - RITIK BAID

DEPARTMENT - INFORMATION TECHNOLOGY

ROLL NO - 001811001035

```
In [ ]:
# BREAST CANCER DATASET
# Random Forest Classifier(Without Tuning)[70-30 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data",header=None)

col_name = ['1','Class','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32']

df.columns = col_name

X = df.drop(['1','Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.7,test_size=0.3,random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier(n_estimators=20, random_state=0)
classifier.fit(X_train,y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

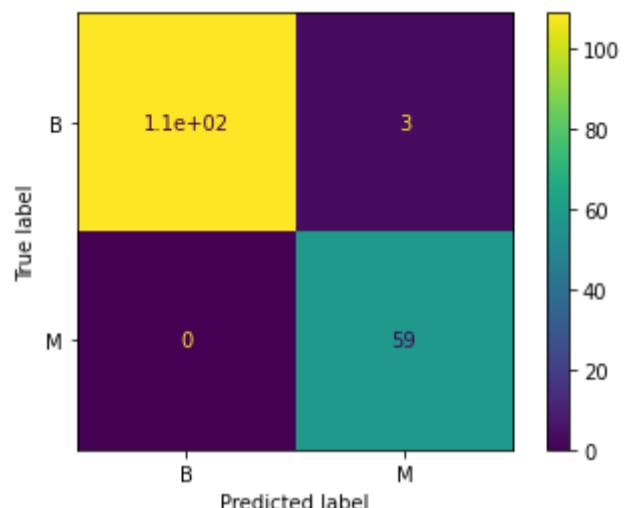
import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()
```

Confusion Matrix:  
[[109 3]  
 [ 0 59]]  
-----  
-----

Performance Evaluation

	precision	recall	f1-score	support
B	1.00	0.97	0.99	112
M	0.95	1.00	0.98	59
accuracy			0.98	171
macro avg	0.98	0.99	0.98	171
weighted avg	0.98	0.98	0.98	171

Accuracy:  
0.9824561403508771



```
In [ ]:
# BREAST CANCER DATASET
# Random Forest Classifier(Without Tuning)[60-40 split]
```

```
import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.6, test_size=0.4, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier(n_estimators=20, random_state=0)
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

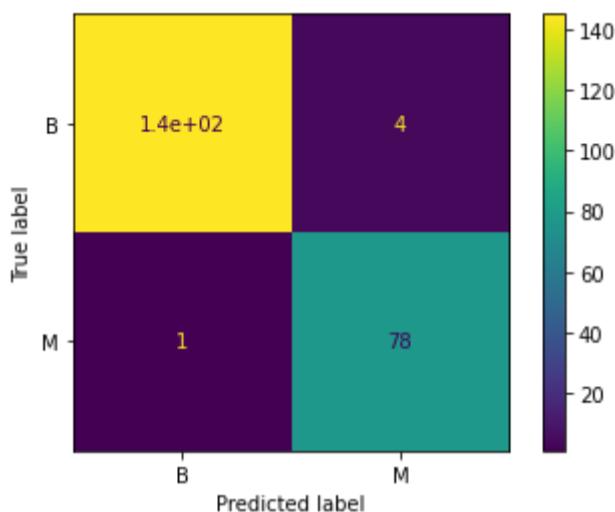
print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()
```

Confusion Matrix:

```
[[145  4]
 [ 1 78]]
-----
Performance Evaluation
      precision    recall   f1-score   support
      B       0.99     0.97     0.98     149
      M       0.95     0.99     0.97      79
accuracy                           0.98     228
macro avg       0.97     0.98     0.98     228
weighted avg     0.98     0.98     0.98     228
```

Accuracy:  
0.9780701754385965



```
In [ ]:
# BREAST CANCER DATASET
# Random Forest Classifier(Without Tuning)[50-50 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.5, test_size=0.5, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier(n_estimators=20, random_state=0)
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
```

```
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()
```

Confusion Matrix:

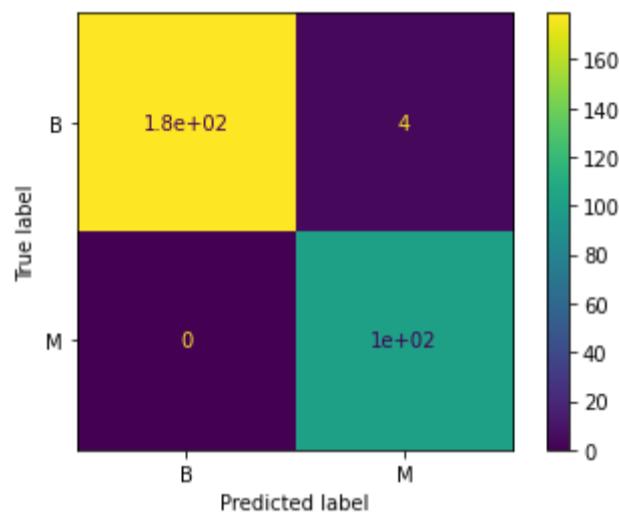
```
[[179  4]
 [ 0 102]]
```

Performance Evaluation

	precision	recall	f1-score	support
B	1.00	0.98	0.99	183
M	0.96	1.00	0.98	102
accuracy			0.99	285
macro avg	0.98	0.99	0.98	285
weighted avg	0.99	0.99	0.99	285

Accuracy:

```
0.9859649122807017
```



In [ ]:

```
# BREAST CANCER DATASET
# Random Forest Classifier(Without Tuning)[40-60 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1','Class','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32']

df.columns = col_name

X = df.drop(['1','Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.4,test_size=0.6,random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier(n_estimators=20, random_state=0)
classifier.fit(X_train,y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")
```

```

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

```

Confusion Matrix:

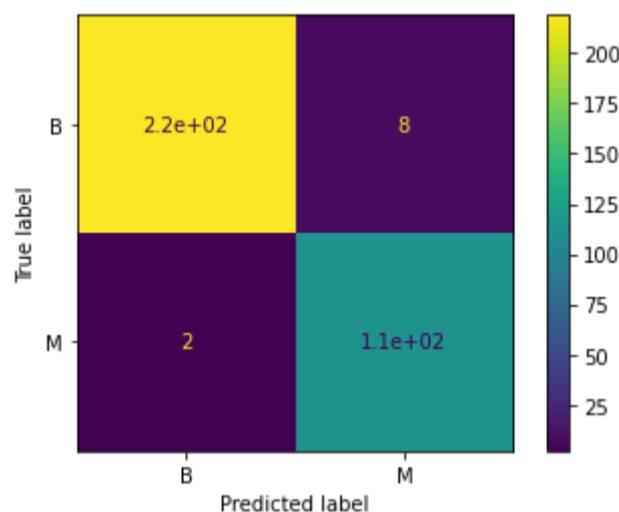
```
[[219  8]
 [ 2 113]]
```

-----  
Performance Evaluation

	precision	recall	f1-score	support
B	0.99	0.96	0.98	227
M	0.93	0.98	0.96	115
accuracy			0.97	342
macro avg	0.96	0.97	0.97	342
weighted avg	0.97	0.97	0.97	342

-----  
Accuracy:

```
0.9707602339181286
```



In [ ]:

```

# BREAST CANCER DATASET
# Random Forest Classifier(Without Tuning)[30-70 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.3, test_size=0.7, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier(n_estimators=20, random_state=0)
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")

```

```

print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

```

Confusion Matrix:

```

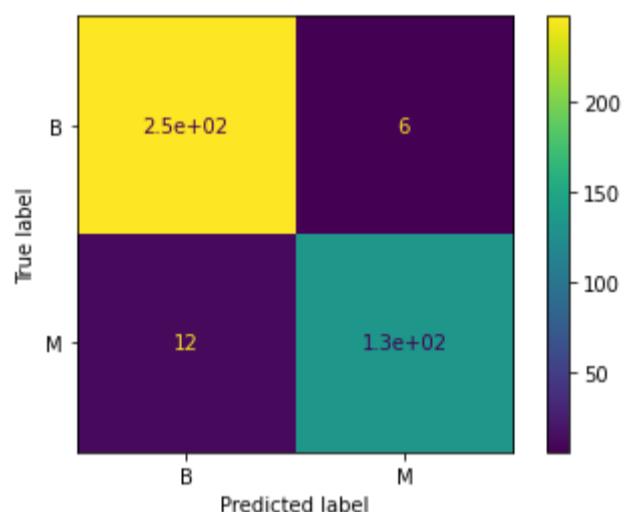
[[247  6]
 [ 12 134]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
B	0.95	0.98	0.96	253
M	0.96	0.92	0.94	146
accuracy			0.95	399
macro avg	0.96	0.95	0.95	399
weighted avg	0.95	0.95	0.95	399

Accuracy:

```
0.9548872180451128
```



In [ ]:

```

# BREAST CANCER DATASET
# Random Forest Classifier(With Tuning)[70-30 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7, test_size=0.30, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier()

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

```

```

#####
# Creating a set of important sample features

from sklearn.model_selection import RandomizedSearchCV
# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 200, stop = 2000, num = 10)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
max_depth.append(None)
# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10]
# Minimum number of samples required at each Leaf node
min_samples_leaf = [1, 2, 4]
# Method of selecting samples for training each tree
bootstrap = [True, False]
# Create the random grid
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}
pprint(random_grid)

#####

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = RandomForestClassifier()
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores
rf_random = RandomizedSearchCV(estimator = classifier, param_distributions = random_grid, n_iter = 100, cv = 3, verbose=2, random_
# Fit the random search model
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```

Parameters currently in use:

```

{'bootstrap': True,
 'ccp_alpha': 0.0,
 'class_weight': None,
 'criterion': 'gini',
 'max_depth': None,
 'max_features': 'auto',
 'max_leaf_nodes': None,
 'max_samples': None,
 'min_impurity_decrease': 0.0,
 'min_impurity_split': None,
 'min_samples_leaf': 1,
 'min_samples_split': 2,
 'min_weight_fraction_leaf': 0.0,
 'n_estimators': 100,
 'n_jobs': None,
 'oob_score': False,
 'random_state': None,
 'verbose': 0,
 'warm_start': False}
{'bootstrap': [True, False],
 'max_depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None],
 'max_features': ['auto', 'sqrt'],
 'min_samples_leaf': [1, 2, 4],
 'min_samples_split': [2, 5, 10],
 'n_estimators': [200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]}
Fitting 3 folds for each of 100 candidates, totalling 300 fits

```

```
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 2 concurrent workers.
[Parallel(n_jobs=-1)]: Done 37 tasks      | elapsed:   57.6s
```

```
[Parallel(n_jobs=-1)]: Done 158 tasks | elapsed: 3.9min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed: 7.3min finished
```

Confusion Matrix:

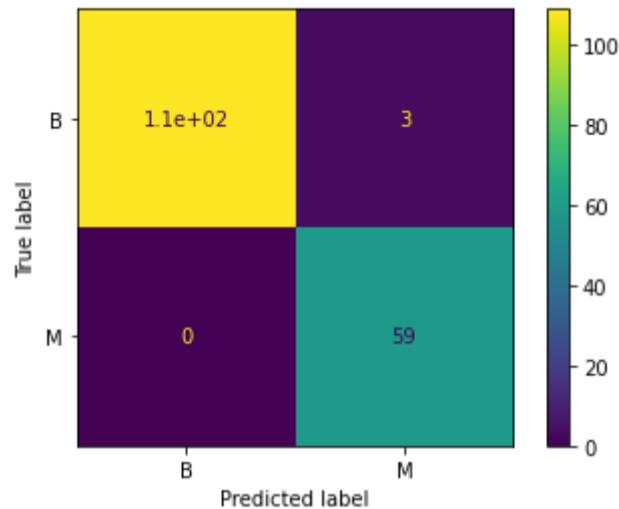
```
[[109  3]
 [ 0  59]]
```

Performance Evaluation

	precision	recall	f1-score	support
B	1.00	0.97	0.99	112
M	0.95	1.00	0.98	59
accuracy			0.98	171
macro avg	0.98	0.99	0.98	171
weighted avg	0.98	0.98	0.98	171

Accuracy:

```
0.9824561403508771
```



In [ ]:

```
# BREAST CANCER DATASET
# Random Forest Classifier(With Tuning)[60-40 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = [ '1','Class','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32']

df.columns = col_name

X = df.drop(['Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.6,test_size=0.4,random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier()

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

from sklearn.model_selection import RandomizedSearchCV
# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 200, stop = 2000, num = 10)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
max_depth.append(None)
```

```

# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10]
# Minimum number of samples required at each leaf node
min_samples_leaf = [1, 2, 4]
# Method of selecting samples for training each tree
bootstrap = [True, False]
# Create the random grid
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}
pprint(random_grid)

#####
# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = RandomForestClassifier()
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores
rf_random = RandomizedSearchCV(estimator = classifier, param_distributions = random_grid, n_iter = 100, cv = 3, verbose=2, random_
# Fit the random search model
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```

Parameters currently in use:

```

{'bootstrap': True,
 'ccp_alpha': 0.0,
 'class_weight': None,
 'criterion': 'gini',
 'max_depth': None,
 'max_features': 'auto',
 'max_leaf_nodes': None,
 'max_samples': None,
 'min_impurity_decrease': 0.0,
 'min_impurity_split': None,
 'min_samples_leaf': 1,
 'min_samples_split': 2,
 'min_weight_fraction_leaf': 0.0,
 'n_estimators': 100,
 'n_jobs': None,
 'oob_score': False,
 'random_state': None,
 'verbose': 0,
 'warm_start': False}
{'bootstrap': [True, False],
 'max_depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None],
 'max_features': ['auto', 'sqrt'],
 'min_samples_leaf': [1, 2, 4],
 'min_samples_split': [2, 5, 10],
 'n_estimators': [200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]}
Fitting 3 folds for each of 100 candidates, totalling 300 fits

```

```

[Parallel(n_jobs=-1)]: Using backend LokyBackend with 2 concurrent workers.
[Parallel(n_jobs=-1)]: Done 37 tasks      | elapsed:   53.9s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.7min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  7.0min finished

```

Confusion Matrix:

```

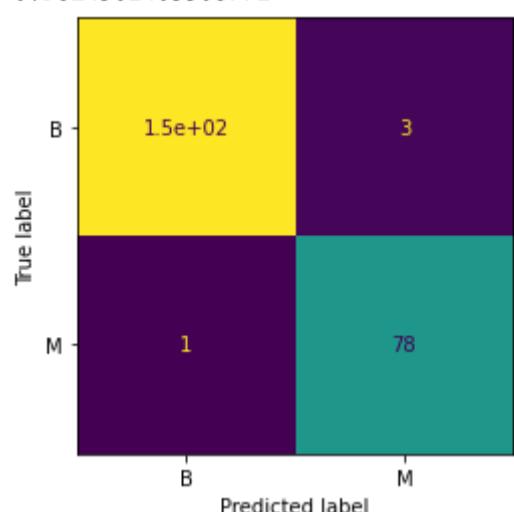
[[146  3]
 [ 1 78]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
B	0.99	0.98	0.99	149
M	0.96	0.99	0.97	79
accuracy		0.98		228

macro avg	0.98	0.98	0.98	228
weighted avg	0.98	0.98	0.98	228

-----  
Accuracy:  
0.9824561403508771



```
In [ ]:
# BREAST CANCER DATASET
# Random Forest Classifier(With Tuning)[50-50 split]
```

```
import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.5, test_size=0.5, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier()

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

from sklearn.model_selection import RandomizedSearchCV
# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 200, stop = 2000, num = 10)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
max_depth.append(None)
# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10]
# Minimum number of samples required at each Leaf node
min_samples_leaf = [1, 2, 4]
# Method of selecting samples for training each tree
bootstrap = [True, False]
# Create the random grid
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}
```

```

pprint(random_grid)

#####
# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = RandomForestClassifier()
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores
rf_random = RandomizedSearchCV(estimator = classifier, param_distributions = random_grid, n_iter = 100, cv = 3, verbose=2, random_
# Fit the random search model
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```

Parameters currently in use:

```

{'bootstrap': True,
 'ccp_alpha': 0.0,
 'class_weight': None,
 'criterion': 'gini',
 'max_depth': None,
 'max_features': 'auto',
 'max_leaf_nodes': None,
 'max_samples': None,
 'min_impurity_decrease': 0.0,
 'min_impurity_split': None,
 'min_samples_leaf': 1,
 'min_samples_split': 2,
 'min_weight_fraction_leaf': 0.0,
 'n_estimators': 100,
 'n_jobs': None,
 'oob_score': False,
 'random_state': None,
 'verbose': 0,
 'warm_start': False}
{'bootstrap': [True, False],
 'max_depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None],
 'max_features': ['auto', 'sqrt'],
 'min_samples_leaf': [1, 2, 4],
 'min_samples_split': [2, 5, 10],
 'n_estimators': [200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]}
Fitting 3 folds for each of 100 candidates, totalling 300 fits

```

```

[Parallel(n_jobs=-1)]: Using backend LokyBackend with 2 concurrent workers.
[Parallel(n_jobs=-1)]: Done  37 tasks      | elapsed:   51.5s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.5min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  6.7min finished

```

Confusion Matrix:

```

[[180  3]
 [ 2 100]]
-----
```

```

Performance Evaluation
      precision    recall  f1-score   support

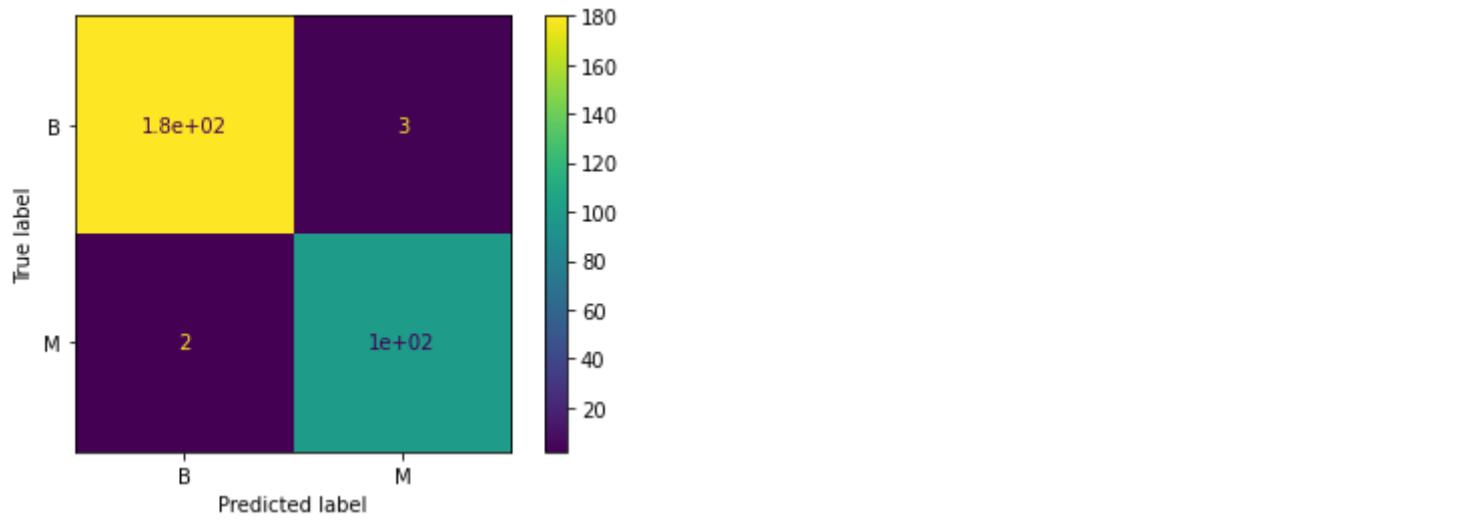
          B       0.99      0.98      0.99      183
          M       0.97      0.98      0.98      102

   accuracy                           0.98      285
  macro avg       0.98      0.98      0.98      285
weighted avg       0.98      0.98      0.98      285

```

```

-----
-----
Accuracy:
0.9824561403508771
```



```
In [ ]:
# BREAST CANCER DATASET
# Random Forest Classifier(With Tuning)[40-60 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.4, test_size=0.6, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier()

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

from sklearn.model_selection import RandomizedSearchCV
# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 200, stop = 2000, num = 10)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
max_depth.append(None)
# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10]
# Minimum number of samples required at each leaf node
min_samples_leaf = [1, 2, 4]
# Method of selecting samples for training each tree
bootstrap = [True, False]
# Create the random grid
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}
pprint(random_grid)

#####
# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = RandomForestClassifier()
```

```

# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores
rf_random = RandomizedSearchCV(estimator = classifier, param_distributions = random_grid, n_iter = 100, cv = 3, verbose=2, random_
# Fit the random search model
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```

Parameters currently in use:

```

{'bootstrap': True,
 'ccp_alpha': 0.0,
 'class_weight': None,
 'criterion': 'gini',
 'max_depth': None,
 'max_features': 'auto',
 'max_leaf_nodes': None,
 'max_samples': None,
 'min_impurity_decrease': 0.0,
 'min_impurity_split': None,
 'min_samples_leaf': 1,
 'min_samples_split': 2,
 'min_weight_fraction_leaf': 0.0,
 'n_estimators': 100,
 'n_jobs': None,
 'oob_score': False,
 'random_state': None,
 'verbose': 0,
 'warm_start': False}
{'bootstrap': [True, False],
 'max_depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None],
 'max_features': ['auto', 'sqrt'],
 'min_samples_leaf': [1, 2, 4],
 'min_samples_split': [2, 5, 10],
 'n_estimators': [200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]}
Fitting 3 folds for each of 100 candidates, totalling 300 fits

```

```

[Parallel(n_jobs=-1)]: Using backend LokyBackend with 2 concurrent workers.
[Parallel(n_jobs=-1)]: Done 37 tasks      | elapsed:  48.5s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.3min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  6.4min finished

```

Confusion Matrix:

```

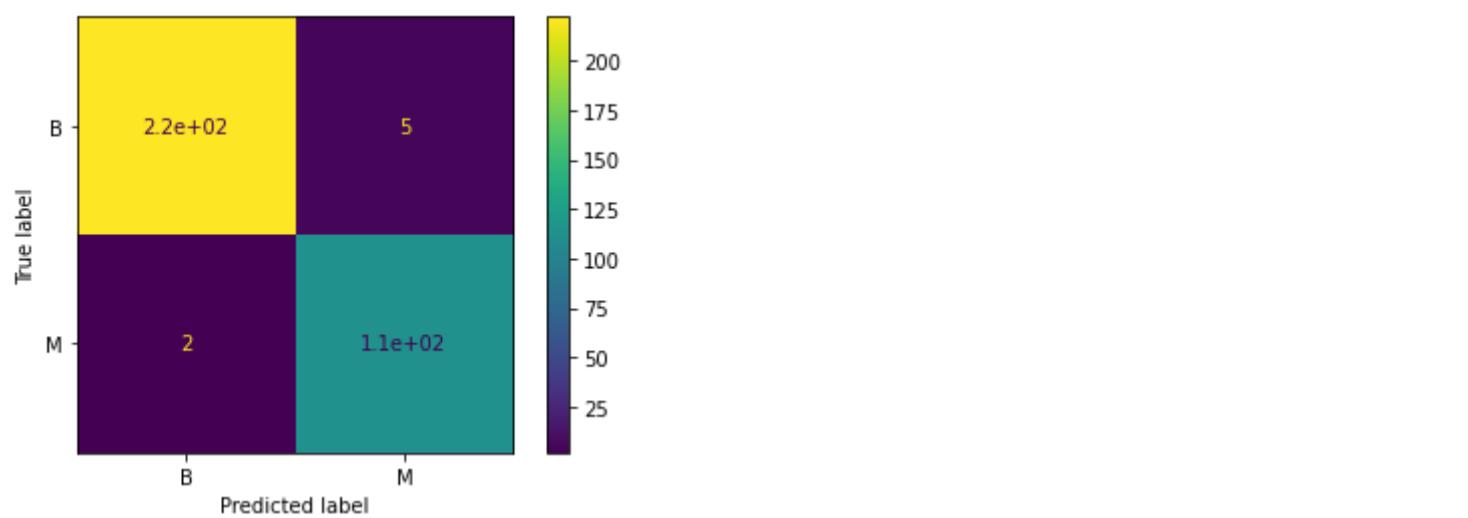
[[222  5]
 [ 2 113]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
B	0.99	0.98	0.98	227
M	0.96	0.98	0.97	115
accuracy			0.98	342
macro avg	0.97	0.98	0.98	342
weighted avg	0.98	0.98	0.98	342

Accuracy:

0.97953216374269



```
In [ ]:
# BREAST CANCER DATASET
# Random Forest Classifier(With Tuning)[30-70 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.3, test_size=0.7, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier()

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

from sklearn.model_selection import RandomizedSearchCV
# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 200, stop = 2000, num = 10)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
max_depth.append(None)
# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10]
# Minimum number of samples required at each Leaf node
min_samples_leaf = [1, 2, 4]
# Method of selecting samples for training each tree
bootstrap = [True, False]
# Create the random grid
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}
pprint(random_grid)

#####
# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = RandomForestClassifier()
```

```

# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores
rf_random = RandomizedSearchCV(estimator = classifier, param_distributions = random_grid, n_iter = 100, cv = 3, verbose=2, random_
# Fit the random search model
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```

Parameters currently in use:

```

{'bootstrap': True,
 'ccp_alpha': 0.0,
 'class_weight': None,
 'criterion': 'gini',
 'max_depth': None,
 'max_features': 'auto',
 'max_leaf_nodes': None,
 'max_samples': None,
 'min_impurity_decrease': 0.0,
 'min_impurity_split': None,
 'min_samples_leaf': 1,
 'min_samples_split': 2,
 'min_weight_fraction_leaf': 0.0,
 'n_estimators': 100,
 'n_jobs': None,
 'oob_score': False,
 'random_state': None,
 'verbose': 0,
 'warm_start': False}
{'bootstrap': [True, False],
 'max_depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None],
 'max_features': ['auto', 'sqrt'],
 'min_samples_leaf': [1, 2, 4],
 'min_samples_split': [2, 5, 10],
 'n_estimators': [200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]}
Fitting 3 folds for each of 100 candidates, totalling 300 fits

```

```

[Parallel(n_jobs=-1)]: Using backend LokyBackend with 2 concurrent workers.
[Parallel(n_jobs=-1)]: Done 37 tasks      | elapsed:  46.0s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.2min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  6.0min finished

```

Confusion Matrix:

```

[[248  5]
 [ 7 139]]
-----
```

```

Performance Evaluation
      precision    recall  f1-score   support

        B       0.97      0.98      0.98      253
        M       0.97      0.95      0.96      146

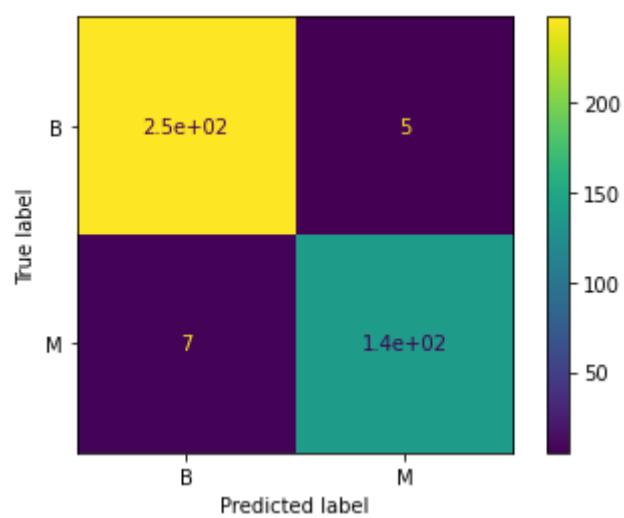
   accuracy                           0.97      399
  macro avg       0.97      0.97      0.97      399
weighted avg       0.97      0.97      0.97      399

```

```

-----
```

Accuracy:  
0.9699248120300752



```
In [ ]: #####
```

```
In [ ]: # BREAST CANCER DATASET
# Multi Layer Perceptron(Without Tuning)[70-30 split]
```

```
import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7, test_size=0.30, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification using MLP
from sklearn.neural_network import MLPClassifier

classifier = MLPClassifier()
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (200) reached and the optimization hasn't converged yet.
```

```
% self.max_iter, ConvergenceWarning)
```

```
Confusion Matrix:
```

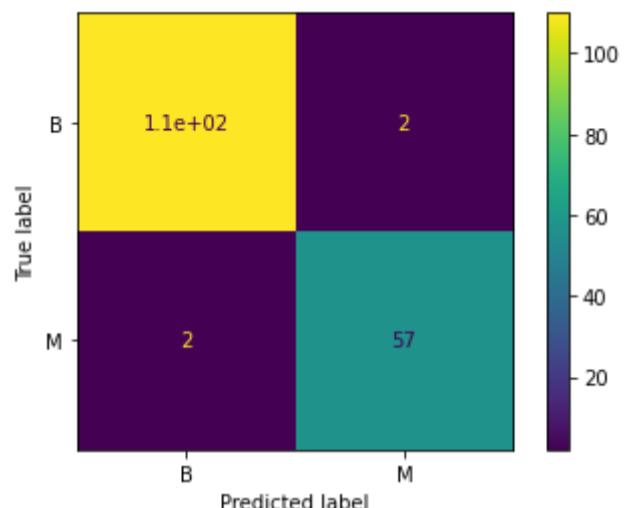
```
[[110  2]
 [ 2 57]]
```

```
-----
```

```
Performance Evaluation
```

	precision	recall	f1-score	support
B	0.98	0.98	0.98	112
M	0.97	0.97	0.97	59
accuracy			0.98	171
macro avg	0.97	0.97	0.97	171
weighted avg	0.98	0.98	0.98	171

Accuracy:  
0.9766081871345029



```
In [ ]:
# BREAST CANCER DATASET
# Multi Layer Perceptron(Without Tuning)[60-40 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.6, test_size=0.4, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification using MLP
from sklearn.neural_network import MLPClassifier

classifier = MLPClassifier()
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (200) reached and the optimization hasn't converged yet.
```

```
% self.max_iter, ConvergenceWarning)
```

```
Confusion Matrix:
```

```
[[147  2]
 [ 2  77]]
```

```
-----
```

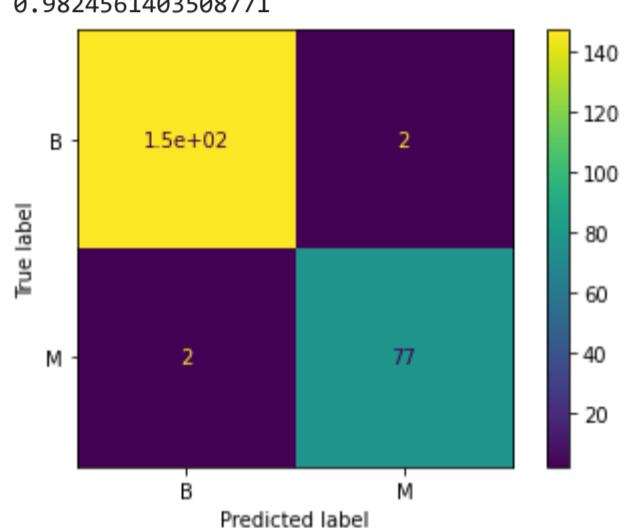
```
Performance Evaluation
```

	precision	recall	f1-score	support
B	0.99	0.99	0.99	149
M	0.97	0.97	0.97	79
accuracy			0.98	228
macro avg	0.98	0.98	0.98	228
weighted avg	0.98	0.98	0.98	228

```
-----
```

```
Accuracy:
```

```
0.9824561403508771
```



```
In [ ]:
```

```
# BREAST CANCER DATASET
# Multi Layer Perceptron(Without Tuning)[50-50 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.5, test_size=0.5, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification using MLP
from sklearn.neural_network import MLPClassifier

classifier = MLPClassifier()
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")
```

```

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

/usr/local/lib/python3.7/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (200) reached and the optimization hasn't converged yet.
  % self.max_iter, ConvergenceWarning)
Confusion Matrix:
[[178  5]
 [ 2 100]]
-----
Performance Evaluation
      precision    recall   f1-score   support
      B       0.99     0.97     0.98     183
      M       0.95     0.98     0.97     102
      accuracy           0.98     285
      macro avg       0.97     0.98     0.97     285
  weighted avg       0.98     0.98     0.98     285
-----
Accuracy:
0.9754385964912281


```

```

In [ ]:
# BREAST CANCER DATASET
# Multi Layer Perceptron(Without Tuning)[40-60 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.4, test_size=0.6, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification using MLP
from sklearn.neural_network import MLPClassifier

classifier = MLPClassifier()
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")

```

```

print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

```

/usr/local/lib/python3.7/dist-packages/sklearn/neural\_network/\_multilayer\_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (200) reached and the optimization hasn't converged yet.

% self.max\_iter, ConvergenceWarning)

Confusion Matrix:

```

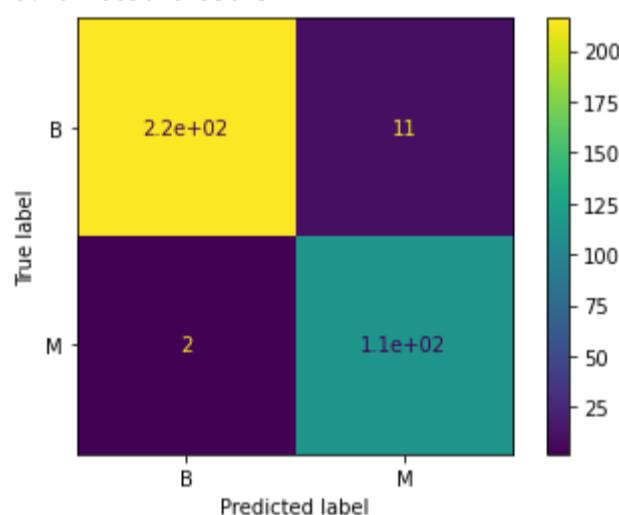
[[216 11]
 [ 2 113]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
B	0.99	0.95	0.97	227
M	0.91	0.98	0.95	115
accuracy			0.96	342
macro avg	0.95	0.97	0.96	342
weighted avg	0.96	0.96	0.96	342

Accuracy:

0.9619883040935673



In [ ]:

```

# BREAST CANCER DATASET
# Multi Layer Perceptron(Without Tuning)[30-70 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = [ '1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32' ]

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.3, test_size=0.7, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification using MLP
from sklearn.neural_network import MLPClassifier

classifier = MLPClassifier()
classifier.fit(X_train, y_train)

```

```

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

```

/usr/local/lib/python3.7/dist-packages/sklearn/neural\_network/\_multilayer\_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (200) reached and the optimization hasn't converged yet.  
% self.max\_iter, ConvergenceWarning)

Confusion Matrix:

```

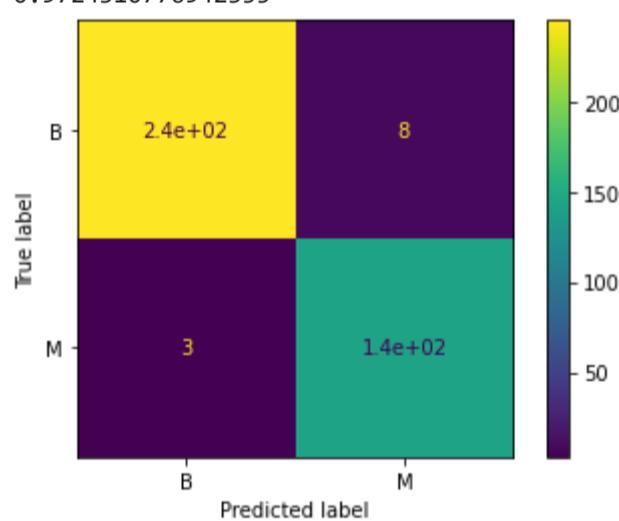
[[245  8]
 [ 3 143]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
B	0.99	0.97	0.98	253
M	0.95	0.98	0.96	146
accuracy			0.97	399
macro avg	0.97	0.97	0.97	399
weighted avg	0.97	0.97	0.97	399

Accuracy:

0.9724310776942355



In [ ]:

```

# BREAST CANCER DATASET
# Multi Layer Perceptron(With Tuning)[70-30 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7, test_size=0.3, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

```

```

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.neural_network import MLPClassifier
classifier = MLPClassifier(max_iter=100)

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

parameter_space = {
    'hidden_layer_sizes': [(50,50,50), (50,100,50), (100,)],
    'activation': ['tanh', 'relu'],
    'solver': ['sgd', 'adam'],
    'alpha': [0.0001, 0.05],
    'learning_rate': ['constant','adaptive'],
}
pprint(parameter_space)

#####

from sklearn.model_selection import GridSearchCV

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = MLPClassifier(max_iter=100)
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores

rf_random = GridSearchCV(classifier, parameter_space, n_jobs=-1, cv=3)
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```

Parameters currently in use:

```

{'activation': 'relu',
 'alpha': 0.0001,
 'batch_size': 'auto',
 'beta_1': 0.9,
 'beta_2': 0.999,
 'early_stopping': False,
 'epsilon': 1e-08,
 'hidden_layer_sizes': (100,),
 'learning_rate': 'constant',
 'learning_rate_init': 0.001,
 'max_fun': 15000,
 'max_iter': 100,
 'momentum': 0.9,
 'n_iter_no_change': 10,
 'nesterovs_momentum': True,
 'power_t': 0.5,
 'random_state': None,
 'shuffle': True,
 'solver': 'adam',
 'tol': 0.0001,
 'validation_fraction': 0.1,
 'verbose': False,
 'warm_start': False}
{'activation': ['tanh', 'relu'],
 'alpha': [0.0001, 0.05],
 'hidden_layer_sizes': [(50, 50, 50), (50, 100, 50), (100,)]}

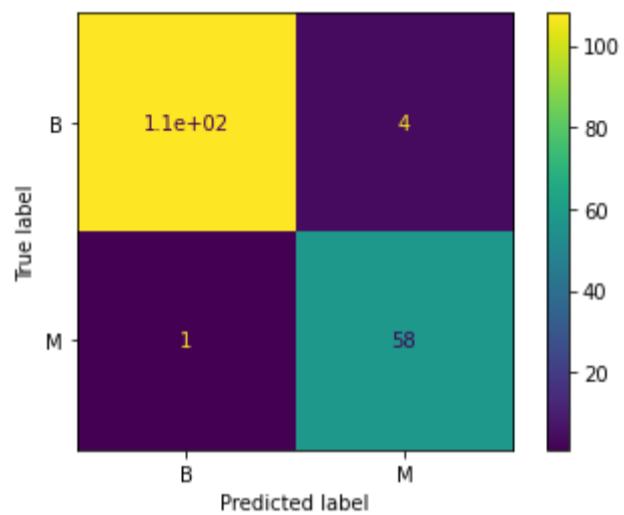
```

```

'learning_rate': ['constant', 'adaptive'],
'solver': ['sgd', 'adam']}
/usr/local/lib/python3.7/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (100) reached and the optimization hasn't converged yet.
  % self.max_iter, ConvergenceWarning)
Confusion Matrix:
[[108  4]
 [ 1 58]]
-----
Performance Evaluation
      precision    recall   f1-score   support
B       0.99     0.96     0.98     112
M       0.94     0.98     0.96      59
accuracy                           0.97    171
macro avg       0.96     0.97     0.97    171
weighted avg    0.97     0.97     0.97    171

```

Accuracy:  
0.9707602339181286



```
In [ ]:
# BREAST CANCER DATASET
# Multi Layer Perceptron(With Tuning)[60-40 split]
```

```

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.6, test_size=0.4, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.neural_network import MLPClassifier
classifier = MLPClassifier(max_iter=100)

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

parameter_space = {
    'hidden_layer_sizes': [(50,50,50), (50,100,50), (100,)],
    'activation': ['tanh', 'relu'],
    'solver': ['sgd', 'adam'],
    'alpha': [0.0001, 0.05],
    'learning_rate': ['constant','adaptive'],
}
```

```

}

pprint(parameter_space)

#####
# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = MLPClassifier(max_iter=100)
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores

rf_random = GridSearchCV(classifier, parameter_space, n_jobs=-1, cv=3)
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```

Parameters currently in use:

```

{'activation': 'relu',
'alpha': 0.0001,
'batch_size': 'auto',
'beta_1': 0.9,
'beta_2': 0.999,
'early_stopping': False,
'epsilon': 1e-08,
'hidden_layer_sizes': (100,),
'learning_rate': 'constant',
'learning_rate_init': 0.001,
'max_fun': 15000,
'max_iter': 100,
'momentum': 0.9,
'n_iter_no_change': 10,
'nesterovs_momentum': True,
'power_t': 0.5,
'random_state': None,
'shuffle': True,
'solver': 'adam',
'tol': 0.0001,
'verification_fraction': 0.1,
'verbose': False,
'warm_start': False}
{'activation': ['tanh', 'relu'],
'alpha': [0.0001, 0.05],
'hidden_layer_sizes': [(50, 50, 50), (50, 100, 50), (100,)],
'learning_rate': ['constant', 'adaptive'],
'solver': ['sgd', 'adam']}

```

/usr/local/lib/python3.7/dist-packages/sklearn/neural\_network/\_multilayer\_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (100) reached and the optimization hasn't converged yet.  
 % self.max\_iter, ConvergenceWarning)

Confusion Matrix:

```

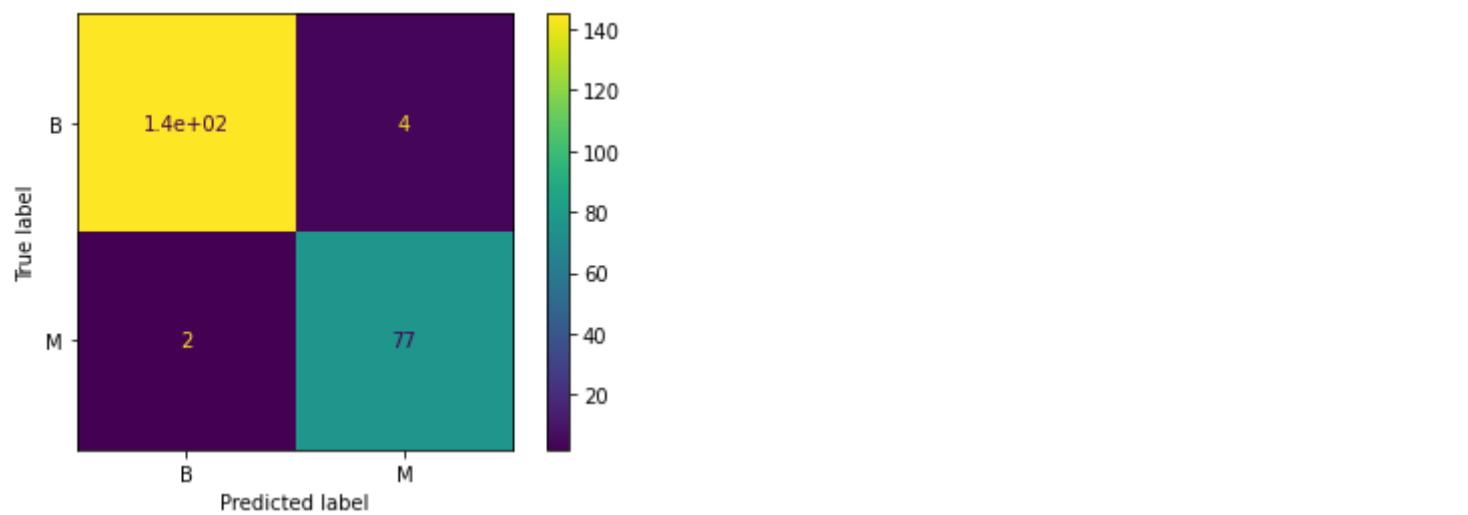
[[145  4]
 [ 2 77]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
B	0.99	0.97	0.98	149
M	0.95	0.97	0.96	79
accuracy			0.97	228
macro avg	0.97	0.97	0.97	228
weighted avg	0.97	0.97	0.97	228

Accuracy:

0.9736842105263158



```
In [ ]:
# BREAST CANCER DATASET
# Multi Layer Perceptron(With Tuning)[50-50 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.5, test_size=0.5, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.neural_network import MLPClassifier
classifier = MLPClassifier(max_iter=100)

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

parameter_space = {
    'hidden_layer_sizes': [(50,50,50), (50,100,50), (100,)],
    'activation': ['tanh', 'relu'],
    'solver': ['sgd', 'adam'],
    'alpha': [0.0001, 0.05],
    'learning_rate': ['constant','adaptive'],
}
pprint(parameter_space)

#####
from sklearn.model_selection import GridSearchCV

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = MLPClassifier(max_iter=100)
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores

rf_random = GridSearchCV(classifier, parameter_space, n_jobs=-1, cv=3)
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))
```

```

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```

Parameters currently in use:

```

{'activation': 'relu',
 'alpha': 0.0001,
 'batch_size': 'auto',
 'beta_1': 0.9,
 'beta_2': 0.999,
 'early_stopping': False,
 'epsilon': 1e-08,
 'hidden_layer_sizes': (100,),
 'learning_rate': 'constant',
 'learning_rate_init': 0.001,
 'max_fun': 15000,
 'max_iter': 100,
 'momentum': 0.9,
 'n_iter_no_change': 10,
 'nesterovs_momentum': True,
 'power_t': 0.5,
 'random_state': None,
 'shuffle': True,
 'solver': 'adam',
 'tol': 0.0001,
 'validation_fraction': 0.1,
 'verbose': False,
 'warm_start': False}
{'activation': ['tanh', 'relu'],
 'alpha': [0.0001, 0.05],
 'hidden_layer_sizes': [(50, 50, 50), (50, 100, 50), (100,)],
 'learning_rate': ['constant', 'adaptive'],
 'solver': ['sgd', 'adam']}

```

```
/usr/local/lib/python3.7/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (100) reached and the optimization hasn't converged yet.
  % self.max_iter, ConvergenceWarning)
```

Confusion Matrix:

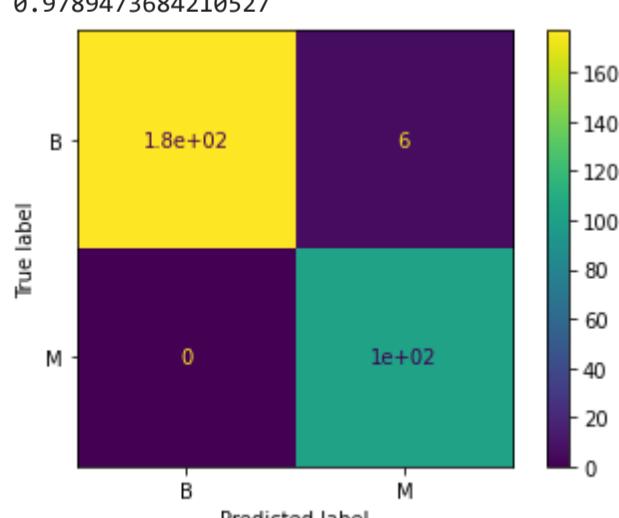
```
[[177  6]
 [ 0 102]]
```

-----

		precision	recall	f1-score	support
	B	1.00	0.97	0.98	183
	M	0.94	1.00	0.97	102
accuracy				0.98	285
macro avg		0.97	0.98	0.98	285
weighted avg		0.98	0.98	0.98	285

-----

Accuracy:  
0.9789473684210527



In [ ]:

```
# BREAST CANCER DATASET
# Multi Layer Perceptron(With Tuning)[40-60 split]

import pandas as pd
import numpy as np
```

```

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.4, test_size=0.6, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.neural_network import MLPClassifier
classifier = MLPClassifier(max_iter=100)

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

parameter_space = {
    'hidden_layer_sizes': [(50,50,50), (50,100,50), (100,)],
    'activation': ['tanh', 'relu'],
    'solver': ['sgd', 'adam'],
    'alpha': [0.0001, 0.05],
    'learning_rate': ['constant','adaptive'],
}
pprint(parameter_space)

#####

from sklearn.model_selection import GridSearchCV

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = MLPClassifier(max_iter=100)
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores

rf_random = GridSearchCV(classifier, parameter_space, n_jobs=-1, cv=3)
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```

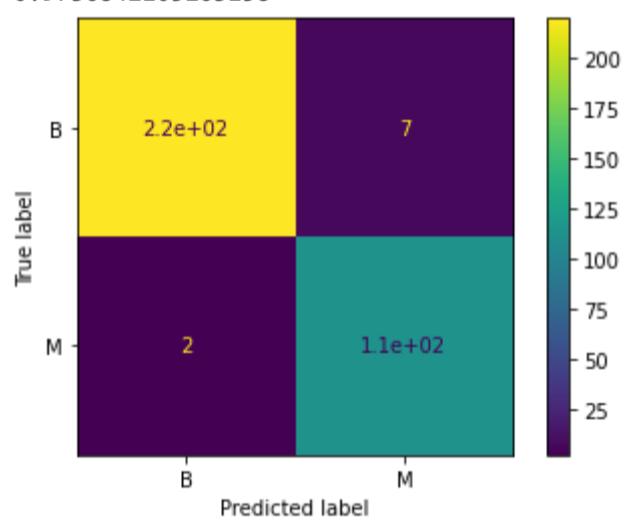
Parameters currently in use:

```
{'activation': 'relu',
 'alpha': 0.0001,
 'batch_size': 'auto',
```

```

'beta_1': 0.9,
'beta_2': 0.999,
'early_stopping': False,
'epsilon': 1e-08,
'hidden_layer_sizes': (100,),
'learning_rate': 'constant',
'learning_rate_init': 0.001,
'max_fun': 15000,
'max_iter': 100,
'momentum': 0.9,
'n_iter_no_change': 10,
'nesterovs_momentum': True,
'power_t': 0.5,
'random_state': None,
'shuffle': True,
'solver': 'adam',
'tol': 0.0001,
'verlbose': False,
'warm_start': False}
{'activation': ['tanh', 'relu'],
'alpha': [0.0001, 0.05],
'hidden_layer_sizes': [(50, 50, 50), (50, 100, 50), (100,)],
'learning_rate': ['constant', 'adaptive'],
'solver': ['sgd', 'adam']}
/usr/local/lib/python3.7/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (100) reached and the optimization hasn't converged yet.
  % self.max_iter, ConvergenceWarning)
Confusion Matrix:
[[220  7]
 [ 2 113]]
-----
-----
Performance Evaluation
      precision    recall   f1-score   support
      B       0.99     0.97     0.98     227
      M       0.94     0.98     0.96     115
      accuracy           0.97     342
      macro avg       0.97     0.98     0.97     342
  weighted avg       0.97     0.97     0.97     342
-----
-----
Accuracy:
0.9736842105263158

```



```

In [ ]:
# BREAST CANCER DATASET
# Multi Layer Perceptron(With Tuning)[30-70 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1','Class','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19','20','21','22','23','24','25','26','27','28','29','30','31','32']

df.columns = col_name

X = df.drop(['1','Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.3,test_size=0.7,random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

```

```

# Classification
from sklearn.neural_network import MLPClassifier
classifier = MLPClassifier(max_iter=100)

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

parameter_space = {
    'hidden_layer_sizes': [(50,50,50), (50,100,50), (100,)],
    'activation': ['tanh', 'relu'],
    'solver': ['sgd', 'adam'],
    'alpha': [0.0001, 0.05],
    'learning_rate': ['constant','adaptive'],
}
pprint(parameter_space)

#####

from sklearn.model_selection import GridSearchCV

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = MLPClassifier(max_iter=100)
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores

rf_random = GridSearchCV(classifier, parameter_space, n_jobs=-1, cv=3)
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```

Parameters currently in use:

```

{'activation': 'relu',
 'alpha': 0.0001,
 'batch_size': 'auto',
 'beta_1': 0.9,
 'beta_2': 0.999,
 'early_stopping': False,
 'epsilon': 1e-08,
 'hidden_layer_sizes': (100,),
 'learning_rate': 'constant',
 'learning_rate_init': 0.001,
 'max_fun': 15000,
 'max_iter': 100,
 'momentum': 0.9,
 'n_iter_no_change': 10,
 'nesterovs_momentum': True,
 'power_t': 0.5,
 'random_state': None,
 'shuffle': True,
 'solver': 'adam',
 'tol': 0.0001,
 'validation_fraction': 0.1,
 'verbose': False,
 'warm_start': False}
{'activation': ['tanh', 'relu'],
 'alpha': [0.0001, 0.05],
 'hidden_layer_sizes': [(50, 50, 50), (50, 100, 50), (100,)],
 'learning_rate': ['constant', 'adaptive'],
 'solver': ['sgd', 'adam']}

```

```
/usr/local/lib/python3.7/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (100) reached and the optimization hasn't converged yet.  
% self.max_iter, ConvergenceWarning)
```

```
Confusion Matrix:
```

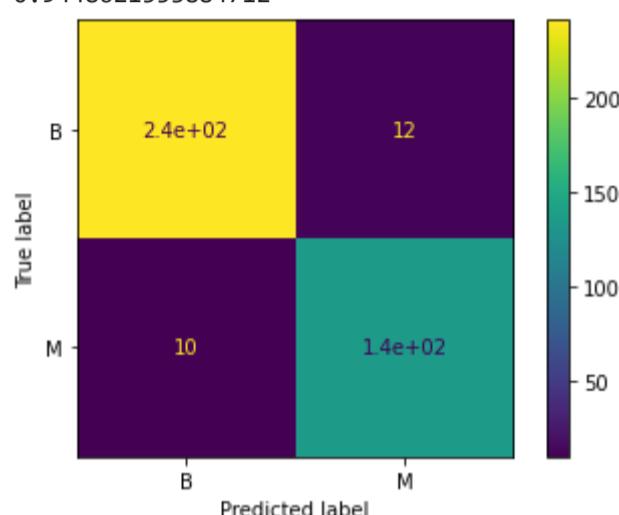
```
[[241 12]  
 [ 10 136]]
```

```
-----  
Performance Evaluation
```

	precision	recall	f1-score	support
B	0.96	0.95	0.96	253
M	0.92	0.93	0.93	146
accuracy			0.94	399
macro avg	0.94	0.94	0.94	399
weighted avg	0.95	0.94	0.94	399

```
-----  
Accuracy:
```

```
0.9448621553884712
```



```
In [ ]:
```

```
#####
#####
```

```
In [ ]:
```

```
# BREAST CANCER DATASET  
# SVM(Without Tuning)[70-30 split]
```

```
import pandas as pd  
import numpy as np  
  
# Dataset Preparation  
df = pd.read_csv("wdbc.data",header=None)  
  
col_name = ['1','Class','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19'  
           , '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']  
  
df.columns = col_name  
  
X = df.drop(['1','Class'], axis=1)  
y = df['Class']  
  
from sklearn.model_selection import train_test_split  
  
X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.7,test_size=0.3,random_state=10)  
  
# Feature Scaling  
from sklearn.preprocessing import StandardScaler  
  
sc = StandardScaler()  
X_train = sc.fit_transform(X_train)  
X_test = sc.transform(X_test)  
  
# Classification  
from sklearn.svm import SVC  
  
classifier = SVC()  
classifier.fit(X_train,y_train)  
  
y_pred = classifier.predict(X_test)  
  
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score  
  
print("Confusion Matrix:")  
print(confusion_matrix(y_test, y_pred))  
  
print("-----")  
print("-----")  
  
print("Performance Evaluation")  
print(classification_report(y_test, y_pred))
```

```

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

```

Confusion Matrix:

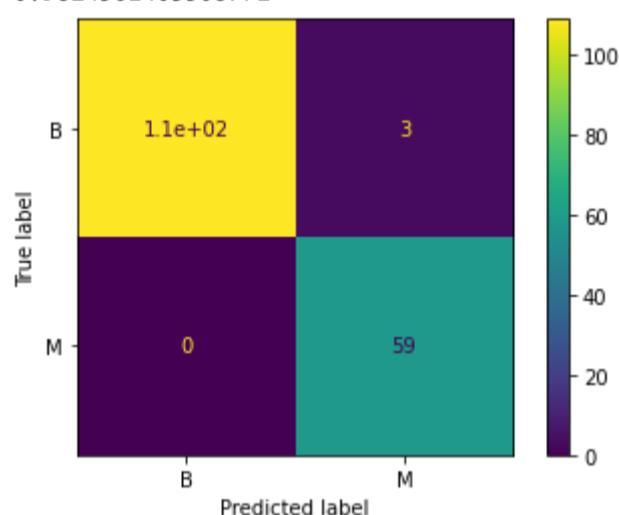
```
[[109  3]
 [ 0  59]]
```

Performance Evaluation

	precision	recall	f1-score	support
B	1.00	0.97	0.99	112
M	0.95	1.00	0.98	59
accuracy			0.98	171
macro avg	0.98	0.99	0.98	171
weighted avg	0.98	0.98	0.98	171

Accuracy:

```
0.9824561403508771
```



In [ ]:

```

# BREAST CANCER DATASET
# SVM(Without Tuning)[60-40 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.6, test_size=0.4, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.svm import SVC

classifier = SVC()
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")

```

```

print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

```

Confusion Matrix:

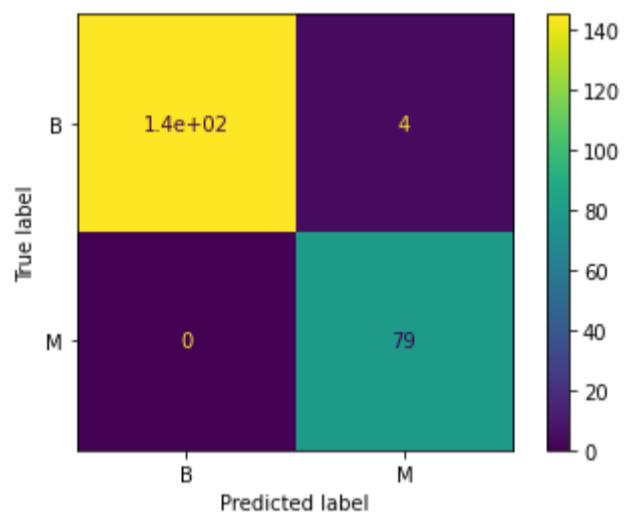
```
[[145  4]
 [ 0  79]]
```

Performance Evaluation

	precision	recall	f1-score	support
B	1.00	0.97	0.99	149
M	0.95	1.00	0.98	79
accuracy			0.98	228
macro avg	0.98	0.99	0.98	228
weighted avg	0.98	0.98	0.98	228

Accuracy:

```
0.9824561403508771
```



In [ ]:

```

# BREAST CANCER DATASET
# SVM(Without Tuning)[50-50 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.5, test_size=0.5, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.svm import SVC

classifier = SVC()
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

```

```

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

```

Confusion Matrix:

```
[[177  6]
 [ 0 102]]
```

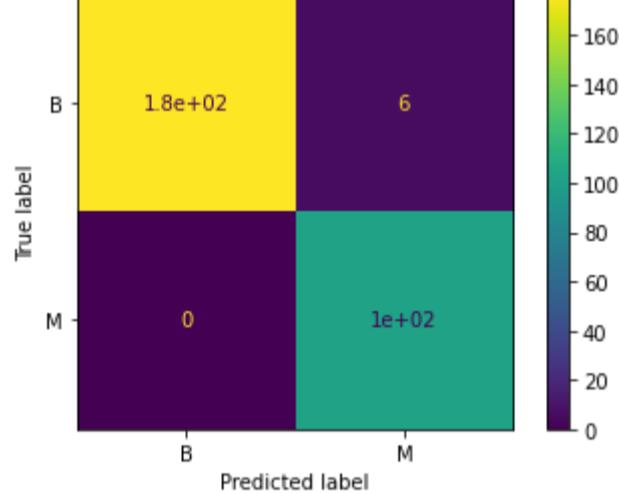
-----

	precision	recall	f1-score	support
B	1.00	0.97	0.98	183
M	0.94	1.00	0.97	102
accuracy			0.98	285
macro avg	0.97	0.98	0.98	285
weighted avg	0.98	0.98	0.98	285

-----

Accuracy:

```
0.9789473684210527
```



In [ ]:

```

# BREAST CANCER DATASET
# SVM(Without Tuning)[40-60 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.4, test_size=0.6, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.svm import SVC

```

```

classifier = SVC()
classifier.fit(X_train,y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

```

Confusion Matrix:

```

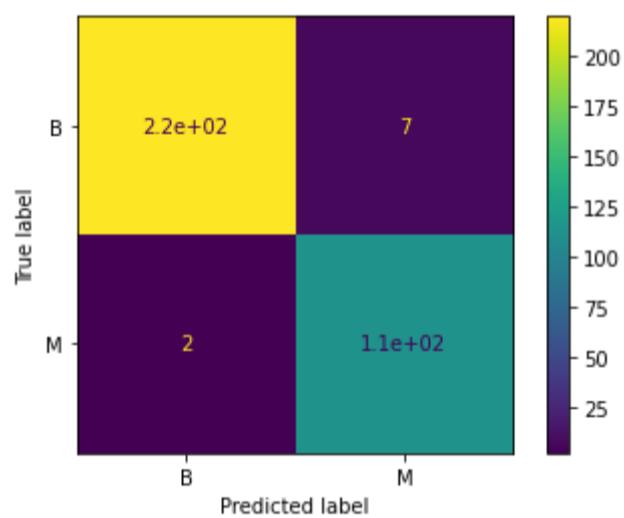
[[220  7]
 [ 2 113]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
B	0.99	0.97	0.98	227
M	0.94	0.98	0.96	115
accuracy			0.97	342
macro avg	0.97	0.98	0.97	342
weighted avg	0.97	0.97	0.97	342

Accuracy:

```
0.9736842105263158
```



In [ ]:

```

# BREAST CANCER DATASET
# SVM(Without Tuning)[30-70 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data",header=None)

col_name = ['1','Class','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32']

df.columns = col_name

X = df.drop(['1','Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.3,test_size=0.7,random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

```

```

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.svm import SVC

classifier = SVC()
classifier.fit(X_train,y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

```

Confusion Matrix:

```

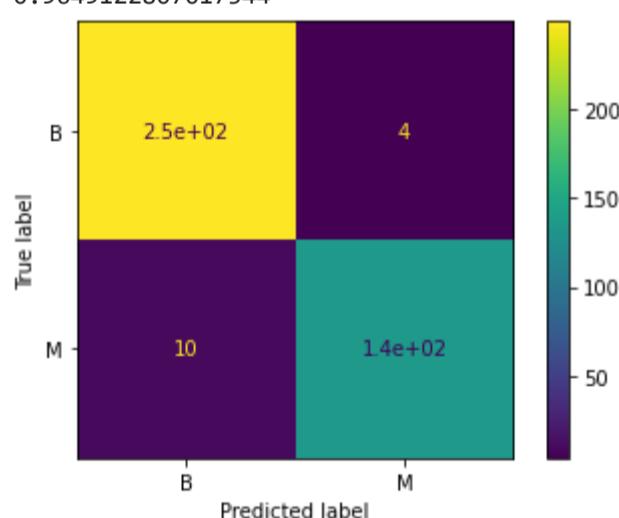
[[249  4]
 [ 10 136]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
B	0.96	0.98	0.97	253
M	0.97	0.93	0.95	146
accuracy			0.96	399
macro avg	0.97	0.96	0.96	399
weighted avg	0.97	0.96	0.96	399

Accuracy:

0.9649122807017544



In [ ]:

```

# BREAST CANCER DATASET
# SVM(With Tuning)[70-30 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data",header=None)

col_name = ['1','Class','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32']

df.columns = col_name

X = df.drop(['1','Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

```

```

X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.7,test_size=0.3,random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.svm import SVC

classifier = SVC()

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

param_grid = {'C': [0.1, 1, 10, 100], 'gamma': [1, 0.1, 0.01, 0.001], 'kernel': ['rbf', 'poly', 'sigmoid']}
pprint(param_grid)

#####

from sklearn.model_selection import GridSearchCV

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = SVC()
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores

rf_random = GridSearchCV(SVC(), param_grid, refit=True, verbose=2)
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```

Parameters currently in use:

```

{'C': 1.0,
 'break_ties': False,
 'cache_size': 200,
 'class_weight': None,
 'coef0': 0.0,
 'decision_function_shape': 'ovr',
 'degree': 3,
 'gamma': 'scale',
 'kernel': 'rbf',
 'max_iter': -1,
 'probability': False,
 'random_state': None,
 'shrinking': True,
 'tol': 0.001,
 'verbose': False}
{'C': [0.1, 1, 10, 100],
 'gamma': [1, 0.1, 0.01, 0.001],
 'kernel': ['rbf', 'poly', 'sigmoid']}
Fitting 5 folds for each of 48 candidates, totalling 240 fits
[CV] C=0.1, gamma=1, kernel=rbf .....
[CV] ..... C=0.1, gamma=1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=1, kernel=rbf .....

```









### Confusion Matrix:

```
[[108    4]
 [ 0   59]]
```

Performance Evaluation

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

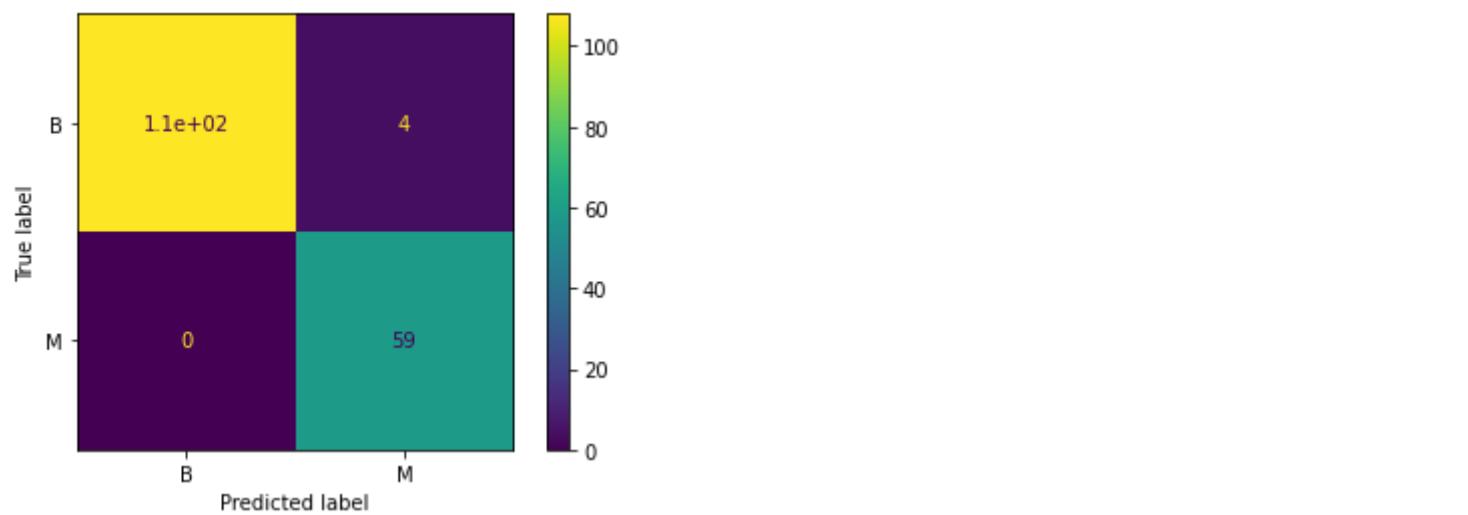
B	1.00	0.96	0.98	112
M	0.94	1.00	0.97	59

accuracy			0.98	171
macro avg	0.97	0.98	0.97	171
weighted avg	0.98	0.98	0.98	171

Accuracy:

Accuracy:  
0.9766081871345029

[Parallel](n\_jobs=1)]: Done 240 out of 240 | elapsed: 1.7s finished



```
In [ ]:
# BREAST CANCER DATASET
# SVM(With Tuning)[60-40 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.6, test_size=0.4, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.svm import SVC

classifier = SVC()

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

param_grid = {'C': [0.1, 1, 10, 100], 'gamma': [1, 0.1, 0.01, 0.001], 'kernel': ['rbf', 'poly', 'sigmoid']}
pprint(param_grid)

#####

from sklearn.model_selection import GridSearchCV

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = SVC()
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores

rf_random = GridSearchCV(SVC(), param_grid, refit=True, verbose=2)
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")
```









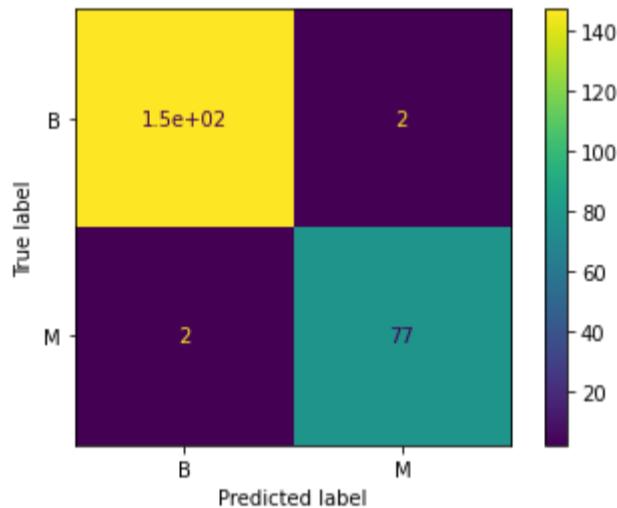


```
[CV] ..... C=100, gamma=0.001, kernel=rbf, total= 0.0s
[CV] C=100, gamma=0.001, kernel=poly .....
[CV] ..... C=100, gamma=0.001, kernel=poly, total= 0.0s
[CV] C=100, gamma=0.001, kernel=poly .....
[CV] ..... C=100, gamma=0.001, kernel=poly, total= 0.0s
[CV] C=100, gamma=0.001, kernel=poly .....
[CV] ..... C=100, gamma=0.001, kernel=poly, total= 0.0s
[CV] C=100, gamma=0.001, kernel=poly .....
[CV] ..... C=100, gamma=0.001, kernel=poly, total= 0.0s
[CV] C=100, gamma=0.001, kernel=sigmoid .....
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
[CV] C=100, gamma=0.001, kernel=sigmoid .....
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
[CV] C=100, gamma=0.001, kernel=sigmoid .....
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
[CV] C=100, gamma=0.001, kernel=sigmoid .....
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
[CV] C=100, gamma=0.001, kernel=sigmoid .....
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
Confusion Matrix:
[[147  2]
 [ 2 77]]
```

Performance Evaluation		precision	recall	f1-score	support
B		0.99	0.99	0.99	149
M		0.97	0.97	0.97	79
		accuracy		0.98	228
macro avg		0.98	0.98	0.98	228
weighted avg		0.98	0.98	0.98	228

Accuracy:  
0.9824561403508771

```
[Parallel(n_jobs=1)]: Done 240 out of 240 | elapsed: 1.4s finished
```



```
In [ ]:
# BREAST CANCER DATASET
# SVM(With Tuning)[50-50 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.5, test_size=0.5, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.svm import SVC

classifier = SVC()
```

```
#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

param_grid = {'C': [0.1, 1, 10, 100], 'gamma': [1, 0.1, 0.01, 0.001], 'kernel': ['rbf', 'poly', 'sigmoid']}

pprint(param_grid)

#####

from sklearn.model_selection import GridSearchCV

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = SVC()
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores

rf_random = GridSearchCV(SVC(), param_grid, refit=True, verbose=2)
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()
```

### Parameters currently in use









### Confusion Matrix:

```
[[178  5]
 [ 0 102]]
```

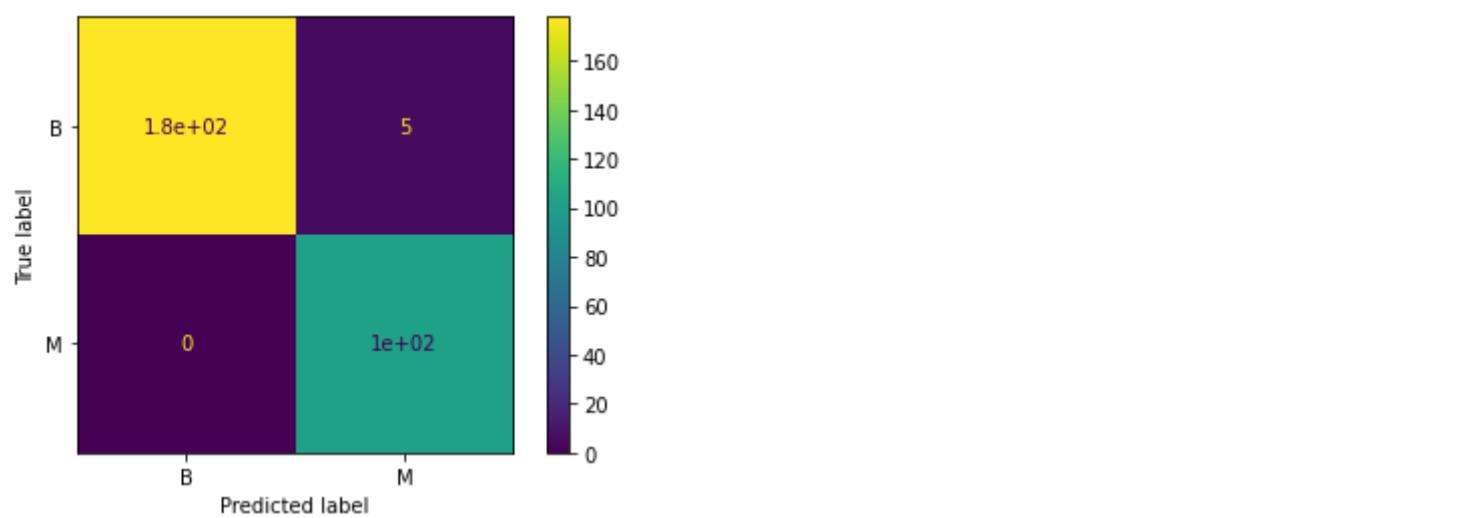
Performance Evaluation

```
[Parallel(n_jobs=1)]: Done 240 out of 240 | elapsed: 1.2s finished
```

precision	recall	f1-score	support
-----------	--------	----------	---------

B	1.00	0.97	0.99	183
M	0.95	1.00	0.98	102
accuracy			0.98	285
macro avg	0.98	0.99	0.98	285
weighted avg	0.98	0.98	0.98	285

Accuracy:  
0.9824561403508771



```
In [ ]:
# BREAST CANCER DATASET
# SVM(With Tuning)[40-60 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1','Class','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32']

df.columns = col_name

X = df.drop(['1','Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.4,test_size=0.6,random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.svm import SVC

classifier = SVC()

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

param_grid = {'C': [0.1, 1, 10, 100], 'gamma': [1, 0.1, 0.01, 0.001], 'kernel': ['rbf', 'poly', 'sigmoid']}
pprint(param_grid)

#####

from sklearn.model_selection import GridSearchCV

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = SVC()
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores

rf_random = GridSearchCV(SVC(), param_grid, refit=True, verbose=2)
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")
```

```
print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-" * 50)
print("-" * 50)

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()
```

Parameters currently in use:







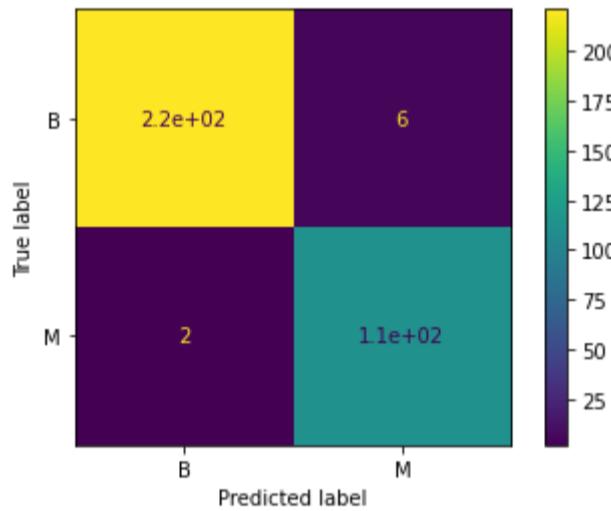


```
[CV] ..... C=100, gamma=0.001, kernel=rbf, total= 0.0s
[CV] C=100, gamma=0.001, kernel=poly .....
[CV] ..... C=100, gamma=0.001, kernel=poly, total= 0.0s
[CV] C=100, gamma=0.001, kernel=poly .....
[CV] ..... C=100, gamma=0.001, kernel=poly, total= 0.0s
[CV] C=100, gamma=0.001, kernel=poly .....
[CV] ..... C=100, gamma=0.001, kernel=poly, total= 0.0s
[CV] C=100, gamma=0.001, kernel=poly .....
[CV] ..... C=100, gamma=0.001, kernel=poly, total= 0.0s
[CV] C=100, gamma=0.001, kernel=sigmoid .....
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
[CV] C=100, gamma=0.001, kernel=sigmoid .....
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
[CV] C=100, gamma=0.001, kernel=sigmoid .....
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
[CV] C=100, gamma=0.001, kernel=sigmoid .....
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
[CV] C=100, gamma=0.001, kernel=sigmoid .....
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
Confusion Matrix:
[[221  6]
 [ 2 113]]
```

Performance Evaluation		precision	recall	f1-score	support
	B	0.99	0.97	0.98	227
	M	0.95	0.98	0.97	115
accuracy				0.98	342
macro avg		0.97	0.98	0.97	342
weighted avg		0.98	0.98	0.98	342

Accuracy:  
0.9766081871345029

```
[Parallel(n_jobs=1)]: Done 240 out of 240 | elapsed: 1.0s finished
```



```
In [ ]:
# BREAST CANCER DATASET
# SVM(With Tuning)[30-70 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.3, test_size=0.7, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Classification
from sklearn.svm import SVC

classifier = SVC()
```

```
#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

param_grid = {'C': [0.1, 1, 10, 100], 'gamma': [1, 0.1, 0.01, 0.001], 'kernel': ['rbf', 'poly', 'sigmoid']}

pprint(param_grid)

#####

from sklearn.model_selection import GridSearchCV

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = SVC()
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores

rf_random = GridSearchCV(SVC(), param_grid, refit=True, verbose=2)
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()
```

### Parameters currently in use:









[CV] .....  
Confusion Matrix:

```
[[250  3]
 [ 5 141]]
```

---

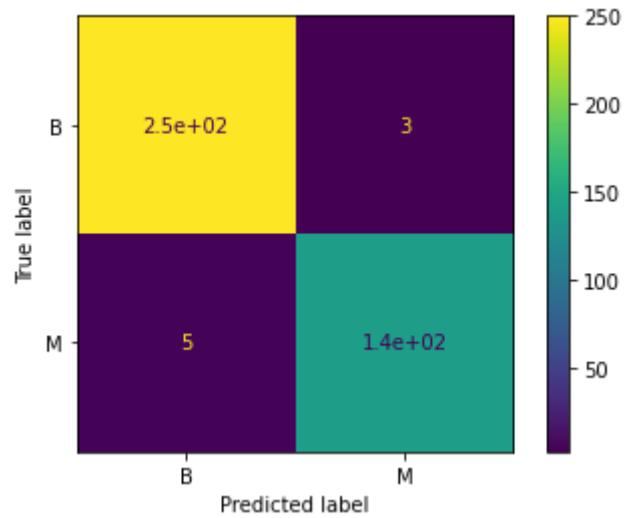
#### Performance Evaluation

Performance Evaluation		precision	recall	f1-score	support
B	0.98	0.99	0.98	253	
M	0.98	0.97	0.97	146	
accuracy				0.98	399
macro avg	0.98	0.98	0.98	399	
weighted avg	0.98	0.98	0.98	399	

Accuracy:

Accuracy:  
0.9799498746867168

[Parallel](n\_jobs=1): Done 240 out of 240 | elapsed: 0.8s finished



```
In [ ]: #####
```

```
In [ ]:
# BREAST CANCER DATASET
# Random Forest Classifier(Without Tuning)[70-30 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7, test_size=0.3, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Finding the important parameters that contribute to most of the variance in the data.

import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA

pca_test = PCA(n_components=30)
pca_test.fit(X_train)
sns.set(style='whitegrid')
plt.plot(np.cumsum(pca_test.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance')
plt.axvline(width=4, color='r', linestyle = '--', x=10, ymin=0, ymax=1)
display(plt.show())

# So we can see that we have 10 important parameters

pca = PCA(n_components=10)
pca.fit(X_train)
X_train = pca.transform(X_train)
X_test = pca.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier(n_estimators=20, random_state=0)
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
```

```

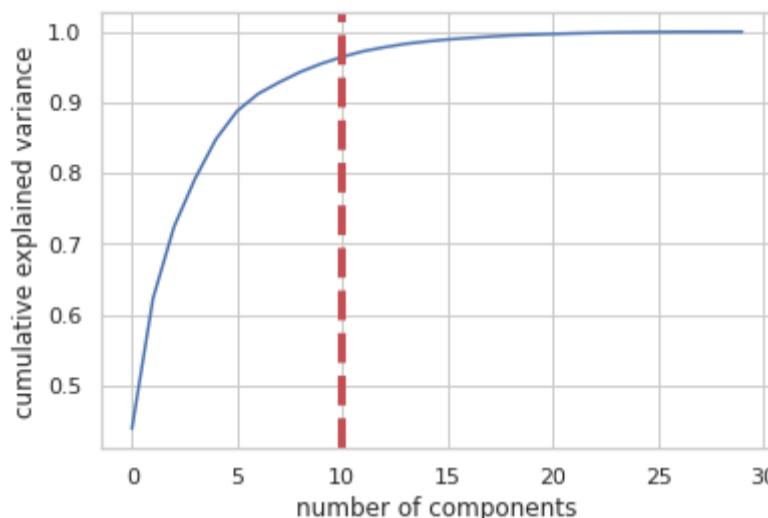
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

```



None

Confusion Matrix:

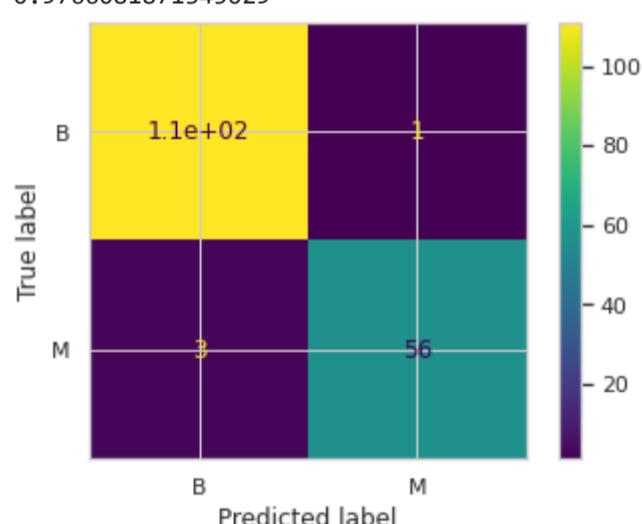
```
[[111  1]
 [ 3  56]]
```

Performance Evaluation

	precision	recall	f1-score	support
B	0.97	0.99	0.98	112
M	0.98	0.95	0.97	59
accuracy			0.98	171
macro avg	0.98	0.97	0.97	171
weighted avg	0.98	0.98	0.98	171

Accuracy:

0.9766081871345029



In [ ]:

```

# BREAST CANCER DATASET
# Random Forest Classifier(With Tuning)[70-30 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7, test_size=0.3, random_state=10)

# Feature Scaling

```

```

from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Finding the important parameters that contribute to most of the variance in the data.

import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA

pca_test = PCA(n_components=30)
pca_test.fit(X_train)
sns.set(style='whitegrid')
plt.plot(np.cumsum(pca_test.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance')
plt.axvline(linewidth=4, color='r', linestyle = '--', x=10, ymin=0, ymax=1)
display(plt.show())

# So we can see that we have 10 important parameters

pca = PCA(n_components=10)
pca.fit(X_train)
X_train = pca.transform(X_train)
X_test = pca.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier()

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

from sklearn.model_selection import RandomizedSearchCV
# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 200, stop = 2000, num = 10)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
max_depth.append(None)
# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10]
# Minimum number of samples required at each Leaf node
min_samples_leaf = [1, 2, 4]
# Method of selecting samples for training each tree
bootstrap = [True, False]
# Create the random grid
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}
pprint(random_grid)

#####

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = RandomForestClassifier()
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores
rf_random = RandomizedSearchCV(estimator = classifier, param_distributions = random_grid, n_iter = 100, cv = 3, verbose=2, random_
# Fit the random search model
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

```

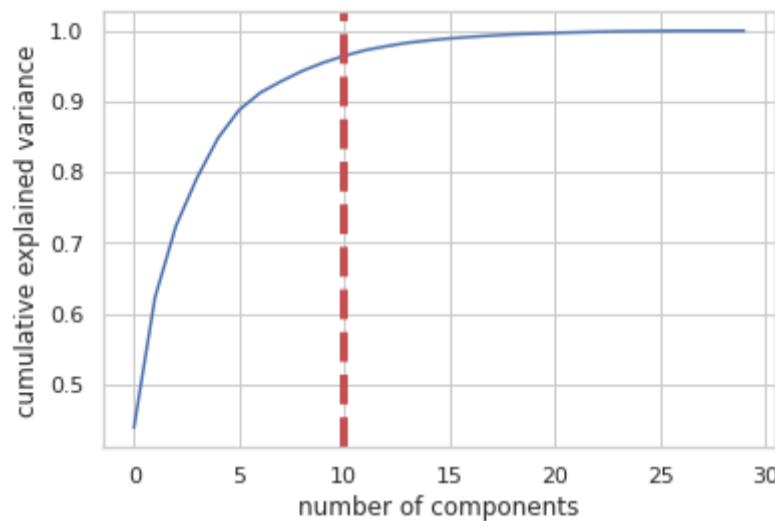
```

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```



None

Parameters currently in use:

```

{'bootstrap': True,
 'ccp_alpha': 0.0,
 'class_weight': None,
 'criterion': 'gini',
 'max_depth': None,
 'max_features': 'auto',
 'max_leaf_nodes': None,
 'max_samples': None,
 'min_impurity_decrease': 0.0,
 'min_impurity_split': None,
 'min_samples_leaf': 1,
 'min_samples_split': 2,
 'min_weight_fraction_leaf': 0.0,
 'n_estimators': 100,
 'n_jobs': None,
 'oob_score': False,
 'random_state': None,
 'verbose': 0,
 'warm_start': False}
{'bootstrap': [True, False],
 'max_depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None],
 'max_features': ['auto', 'sqrt'],
 'min_samples_leaf': [1, 2, 4],
 'min_samples_split': [2, 5, 10],
 'n_estimators': [200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]}

```

Fitting 3 folds for each of 100 candidates, totalling 300 fits

```

[Parallel(n_jobs=-1)]: Using backend LokyBackend with 2 concurrent workers.
[Parallel(n_jobs=-1)]: Done  37 tasks      | elapsed:   55.6s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.7min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  7.1min finished

```

Confusion Matrix:

```

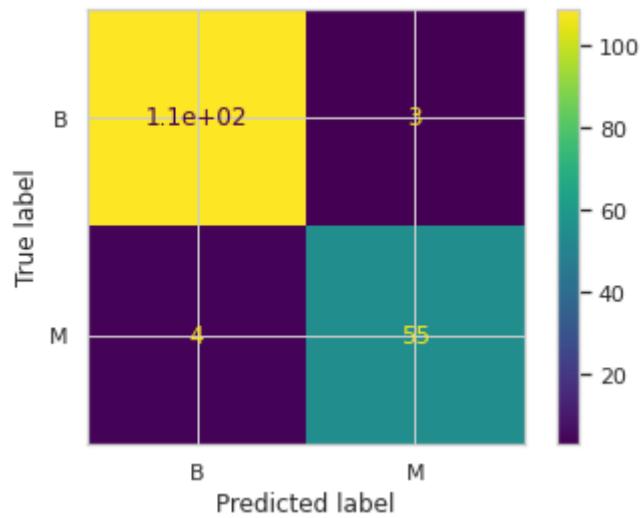
[[109  3]
 [ 4 55]]
-----
```

Performance Evaluation		precision	recall	f1-score	support
B		0.96	0.97	0.97	112
M		0.95	0.93	0.94	59
accuracy				0.96	171
macro avg		0.96	0.95	0.95	171
weighted avg		0.96	0.96	0.96	171

```

-----
```

Accuracy:  
0.9590643274853801



```
In [ ]:
# BREAST CANCER DATASET
# Multi Layer Perceptron(Without Tuning)[70-30 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1','Class','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32']

df.columns = col_name

X = df.drop(['1','Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.7,test_size=0.30,random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Finding the important parameters that contribute to most of the variance in the data.

import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA

pca_test = PCA(n_components=30)
pca_test.fit(X_train)
sns.set(style='whitegrid')
plt.plot(np.cumsum(pca_test.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance')
plt.axvline(linewidth=4, color='r', linestyle = '--', x=10, ymin=0, ymax=1)
display(plt.show())

# So we can see that we have 10 important parameters

pca = PCA(n_components=10)
pca.fit(X_train)
X_train = pca.transform(X_train)
X_test = pca.transform(X_test)

# Classification using MLP
from sklearn.neural_network import MLPClassifier

classifier = MLPClassifier()
classifier.fit(X_train,y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))
```

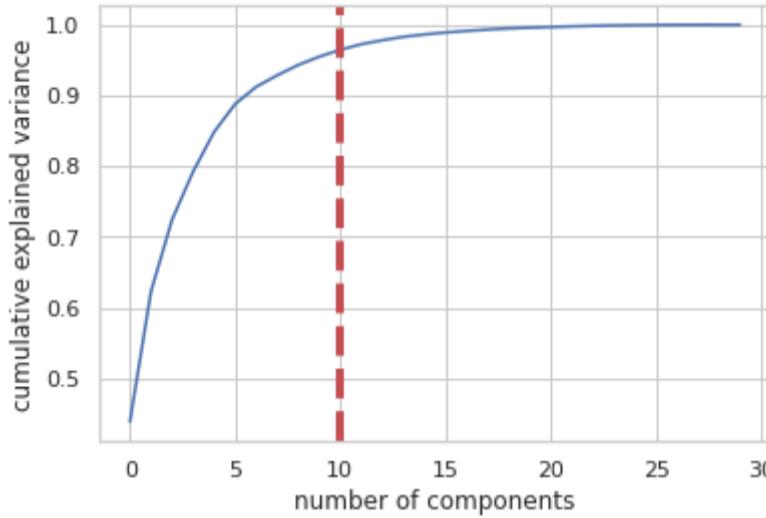
```

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

```



None

```
/usr/local/lib/python3.7/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (200) reached and the optimization hasn't converged yet.
  % self.max_iter, ConvergenceWarning)
```

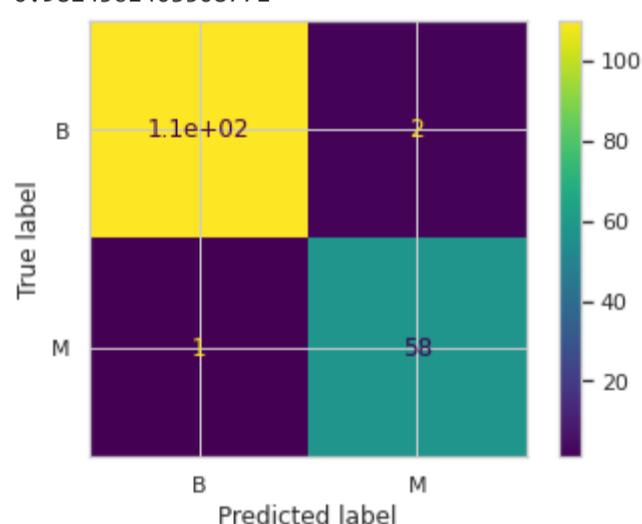
Confusion Matrix:

```
[[110  2]
 [ 1 58]]
```

Performance Evaluation		precision	recall	f1-score	support
	B	0.99	0.98	0.99	112
	M	0.97	0.98	0.97	59
accuracy				0.98	171
macro avg		0.98	0.98	0.98	171
weighted avg		0.98	0.98	0.98	171

Accuracy:

```
0.9824561403508771
```



In [ ]:

```

# BREAST CANCER DATASET
# Multi Layer Perceptron(With Tuning)[70-30 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
           '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7, test_size=0.3, random_state=10)

```

```

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Finding the important parameters that contribute to most of the variance in the data.

import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA

pca_test = PCA(n_components=30)
pca_test.fit(X_train)
sns.set(style='whitegrid')
plt.plot(np.cumsum(pca_test.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance')
plt.axvline(linewidth=4, color='r', linestyle = '--', x=10, ymin=0, ymax=1)
display(plt.show())

# So we can see that we have 10 important parameters

pca = PCA(n_components=10)
pca.fit(X_train)
X_train = pca.transform(X_train)
X_test = pca.transform(X_test)

# Classification
from sklearn.neural_network import MLPClassifier
classifier = MLPClassifier(max_iter=100)

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

parameter_space = {
    'hidden_layer_sizes': [(50,50,50), (50,100,50), (100,)],
    'activation': ['tanh', 'relu'],
    'solver': ['sgd', 'adam'],
    'alpha': [0.0001, 0.05],
    'learning_rate': ['constant','adaptive'],
}
pprint(parameter_space)

#####

from sklearn.model_selection import GridSearchCV

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = MLPClassifier(max_iter=100)
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores

rf_random = GridSearchCV(classifier, parameter_space, n_jobs=-1, cv=3)
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

print("Performance Evaluation")
print(classification_report(y_test, y_pred))

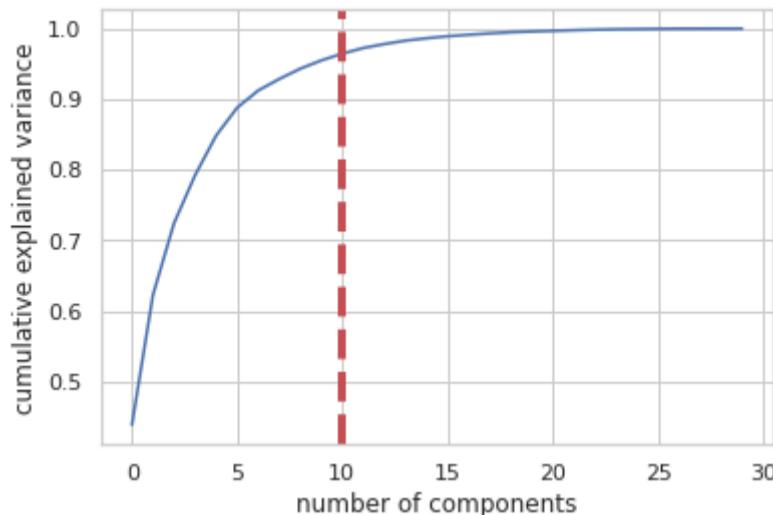
print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix

```

```
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()
```



None

Parameters currently in use:

```
{'activation': 'relu',
 'alpha': 0.0001,
 'batch_size': 'auto',
 'beta_1': 0.9,
 'beta_2': 0.999,
 'early_stopping': False,
 'epsilon': 1e-08,
 'hidden_layer_sizes': (100,),
 'learning_rate': 'constant',
 'learning_rate_init': 0.001,
 'max_fun': 15000,
 'max_iter': 100,
 'momentum': 0.9,
 'n_iter_no_change': 10,
 'nesterovs_momentum': True,
 'power_t': 0.5,
 'random_state': None,
 'shuffle': True,
 'solver': 'adam',
 'tol': 0.0001,
 'validation_fraction': 0.1,
 'verbose': False,
 'warm_start': False}
{'activation': ['tanh', 'relu'],
 'alpha': [0.0001, 0.05],
 'hidden_layer_sizes': [(50, 50, 50), (50, 100, 50), (100,)],
 'learning_rate': ['constant', 'adaptive'],
 'solver': ['sgd', 'adam']}
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (100) reached and the optimization hasn't converged yet.
  % self.max_iter, ConvergenceWarning)
```

Confusion Matrix:

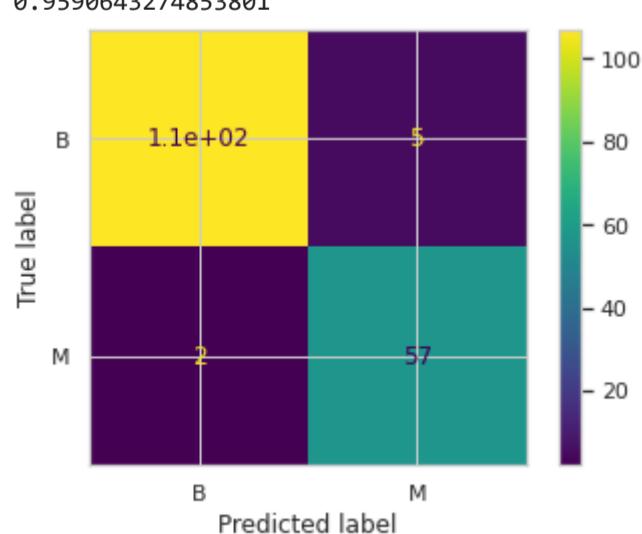
```
[[107  5]
 [ 2 57]]
```

Performance Evaluation

	precision	recall	f1-score	support
B	0.98	0.96	0.97	112
M	0.92	0.97	0.94	59
accuracy			0.96	171
macro avg	0.95	0.96	0.96	171
weighted avg	0.96	0.96	0.96	171

Accuracy:

```
0.9590643274853801
```



In [ ]:

```
# BREAST CANCER DATASET
# SVM(Without Tuning)[70-30 split]
```

```
import pandas as pd
```

```

import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7, test_size=0.3, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Finding the important parameters that contribute to most of the variance in the data.

import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA

pca_test = PCA(n_components=30)
pca_test.fit(X_train)
sns.set(style='whitegrid')
plt.plot(np.cumsum(pca_test.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance')
plt.axvline(linewidth=4, color='r', linestyle = '--', x=10, ymin=0, ymax=1)
display(plt.show())

# So we can see that we have 10 important parameters

pca = PCA(n_components=10)
pca.fit(X_train)
X_train = pca.transform(X_train)
X_test = pca.transform(X_test)

# Classification
from sklearn.svm import SVC

classifier = SVC()
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

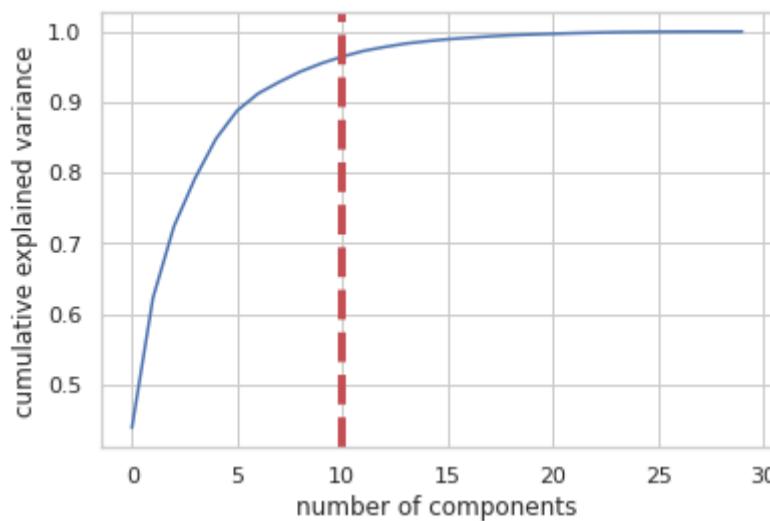
print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()

```

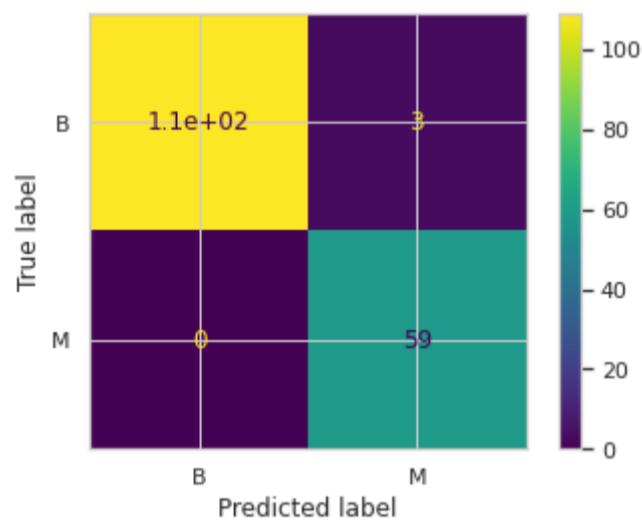


None  
Confusion Matrix:  
[[109 3]  
 [ 0 59]]

Performance Evaluation

	precision	recall	f1-score	support
B	1.00	0.97	0.99	112
M	0.95	1.00	0.98	59
accuracy			0.98	171
macro avg	0.98	0.99	0.98	171
weighted avg	0.98	0.98	0.98	171

Accuracy:  
0.9824561403508771



```
In [ ]:
# BREAST CANCER DATASET
# SVM(With Tuning)[70-30 split]

import pandas as pd
import numpy as np

# Dataset Preparation
df = pd.read_csv("wdbc.data", header=None)

col_name = ['1', 'Class', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
           '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32']

df.columns = col_name

X = df.drop(['1', 'Class'], axis=1)
y = df['Class']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7, test_size=0.3, random_state=10)

# Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Finding the important parameters that contribute to most of the variance in the data.

import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA

pca_test = PCA(n_components=30)
```

```

pca_test.fit(X_train)
sns.set(style='whitegrid')
plt.plot(np.cumsum(pca_test.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance')
plt.axvline(linewidth=4, color='r', linestyle = '--', x=10, ymin=0, ymax=1)
display(plt.show())

# So we can see that we have 10 important parameters

pca = PCA(n_components=10)
pca.fit(X_train)
X_train = pca.transform(X_train)
X_test = pca.transform(X_test)

# Classification
from sklearn.svm import SVC

classifier = SVC()

#####
# Showing all the parameters

from pprint import pprint
# Look at parameters used by our current forest
print('Parameters currently in use:\n')
pprint(classifier.get_params())

#####
# Creating a set of important sample features

param_grid = {'C': [0.1, 1, 10, 100], 'gamma': [1, 0.1, 0.01, 0.001], 'kernel': ['rbf', 'poly', 'sigmoid']}
pprint(param_grid)

#####

from sklearn.model_selection import GridSearchCV

# Use the random grid to search for best hyperparameters
# First create the base model to tune
classifier = SVC()
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores

rf_random = GridSearchCV(SVC(), param_grid, refit=True, verbose=2)
rf_random.fit(X_train, y_train)

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("-----")
print("-----")

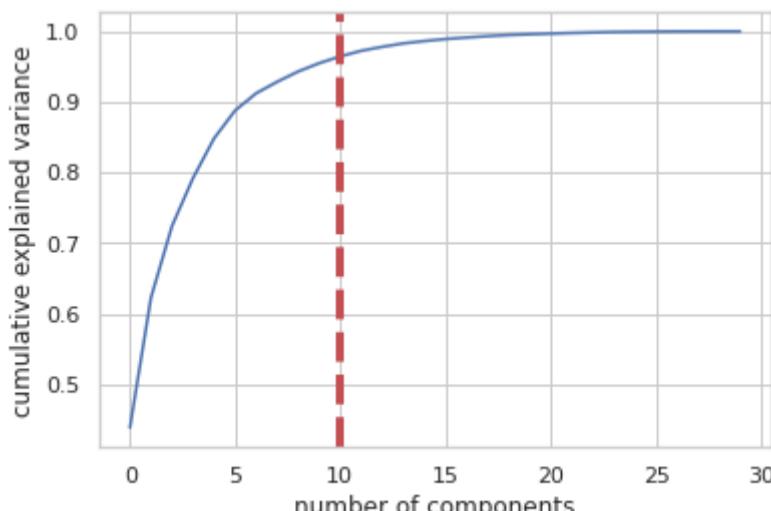
print("Performance Evaluation")
print(classification_report(y_test, y_pred))

print("-----")
print("-----")

print("Accuracy:")
print(accuracy_score(y_test, y_pred))

import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(rf_random, X_test, y_test)
plt.show()

```



None

Parameters currently in use:







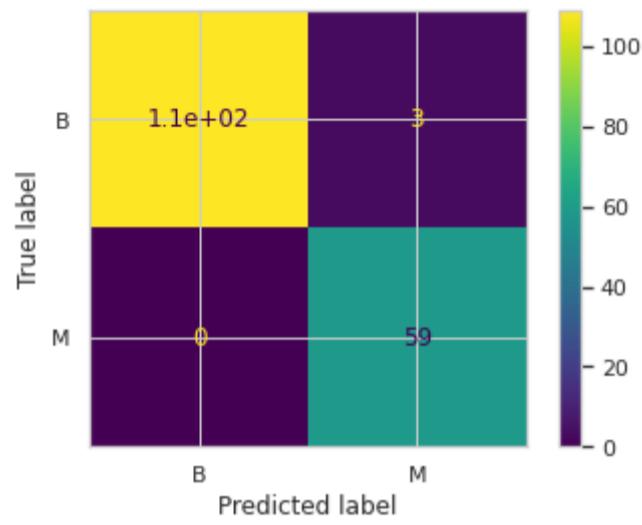


```
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
[CV] C=100, gamma=0.001, kernel=sigmoid .....
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
Confusion Matrix:
[[109  3]
 [ 0  59]]
```

```
-----  
Performance Evaluation  
precision    recall   f1-score   support  
  
      B       1.00      0.97      0.99      112  
      M       0.95      1.00      0.98      59  
  
accuracy                           0.98      171  
macro avg       0.98      0.99      0.98      171  
weighted avg     0.98      0.98      0.98      171
```

```
-----  
Accuracy:  
0.9824561403508771
```

```
[Parallel(n_jobs=1)]: Done 240 out of 240 | elapsed: 1.4s finished
```



```
In [ ]:
```