

JADAVPUR UNIVERSITY

MACHINE LEARNING LABORATORY

ASSIGNMENT #2 - B) WINE DATASET

NAME - RITIK BAID

DEPARTMENT - INFORMATION TECHNOLOGY

ROLL NO - 001811001035

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

import pandas as pd
import numpy as np

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

col_name = ['Class','Alcohol','Malic acid','Ash','Alcalinity of ash','Magnesium','Total phenols','Flavanoids',
            'Nonflavanoid phenols','Proanthocyanins','Color intensity','Hue','OD280/OD315 of diluted wines','Proline']

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=0.3,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))

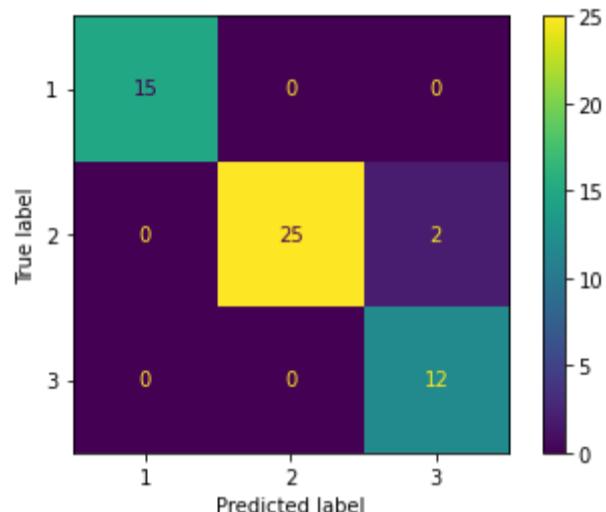
import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()
```

Confusion Matrix:

[[15 0 0]	[0 25 2]	[0 0 12]]
-------------	------------	-------------

Performance Evaluation				
	precision	recall	f1-score	support
1	1.00	1.00	1.00	15
2	1.00	0.93	0.96	27
3	0.86	1.00	0.92	12
accuracy			0.96	54
macro avg	0.95	0.98	0.96	54
weighted avg	0.97	0.96	0.96	54

Accuracy:
0.9629629629629629



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

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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(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:

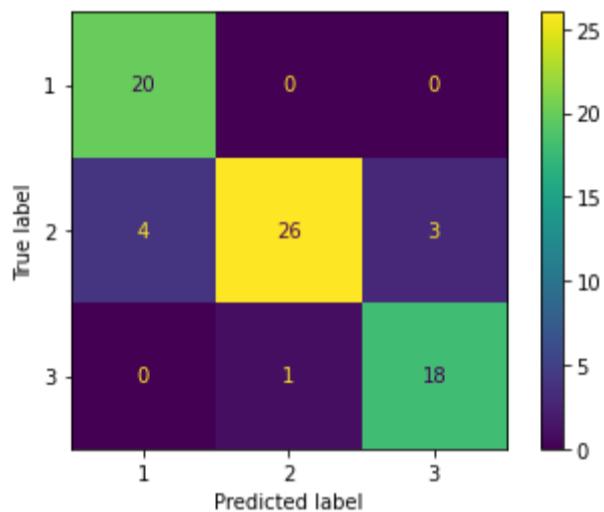
```
[[20  0  0]
 [ 4 26  3]
 [ 0  1 18]]
```

Performance Evaluation

	precision	recall	f1-score	support
1	0.83	1.00	0.91	20
2	0.96	0.79	0.87	33
3	0.86	0.95	0.90	19
accuracy			0.89	72
macro avg	0.88	0.91	0.89	72
weighted avg	0.90	0.89	0.89	72

Accuracy:

```
0.8888888888888888
```



In []:

```
# WINE DATASET
# Random Forest Classifier(Without Tuning)[50-50 split]

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
            'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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(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:

```

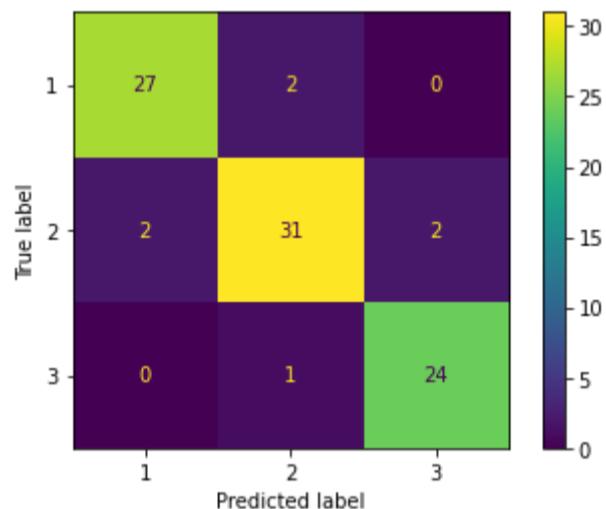
[[27  2  0]
 [ 2 31  2]
 [ 0  1 24]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
1	0.93	0.93	0.93	29
2	0.91	0.89	0.90	35
3	0.92	0.96	0.94	25
accuracy			0.92	89
macro avg	0.92	0.93	0.92	89
weighted avg	0.92	0.92	0.92	89

Accuracy:

```
0.9213483146067416
```



In []:

```

# WINE DATASET
# Random Forest Classifier(Without Tuning)[40-60 split]

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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(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:

```

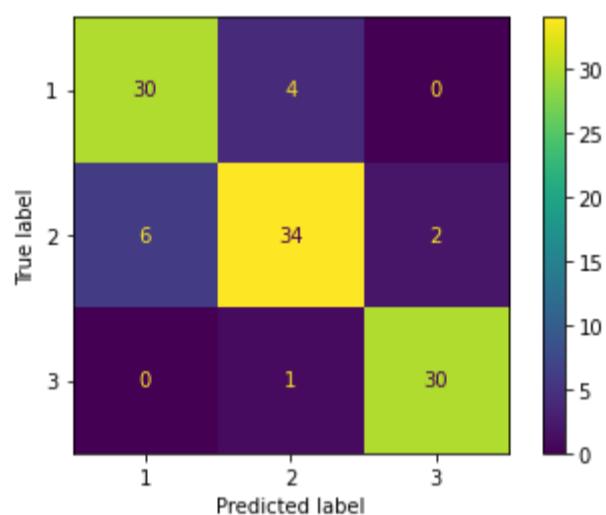
[[30  4  0]
 [ 6 34  2]
 [ 0  1 30]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
1	0.83	0.88	0.86	34
2	0.87	0.81	0.84	42
3	0.94	0.97	0.95	31
accuracy			0.88	107
macro avg	0.88	0.89	0.88	107
weighted avg	0.88	0.88	0.88	107

Accuracy:

```
0.8785046728971962
```



In []:

```

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

import pandas as pd
import numpy as np

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

col_name = ['Class','Alcohol','Malic acid','Ash','Alcalinity of ash','Magnesium','Total phenols','Flavanoids',
           'Nonflavanoid phenols','Proanthocyanins','Color intensity','Hue','OD280/OD315 of diluted wines','Proline']

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(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:

```

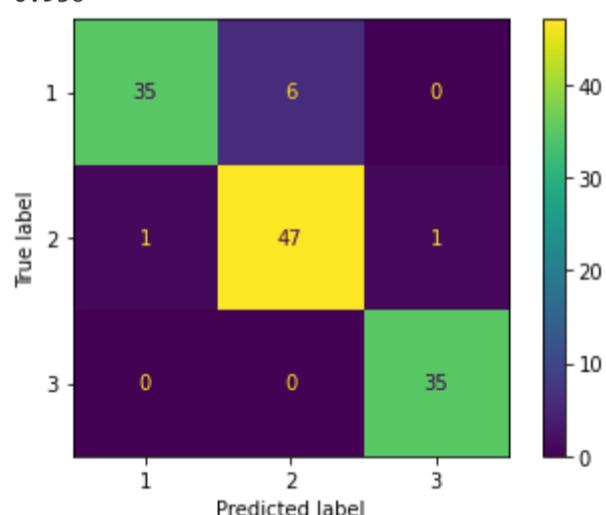
[[35  6  0]
 [ 1 47  1]
 [ 0  0 35]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
1	0.97	0.85	0.91	41
2	0.89	0.96	0.92	49
3	0.97	1.00	0.99	35
accuracy			0.94	125
macro avg	0.94	0.94	0.94	125
weighted avg	0.94	0.94	0.94	125

Accuracy:

0.936



In []:

```

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

import pandas as pd
import numpy as np

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

col_name = ['Class','Alcohol','Malic acid','Ash','Alcalinity of ash','Magnesium','Total phenols','Flavanoids',
           'Nonflavanoid phenols','Proanthocyanins','Color intensity','Hue','OD280/OD315 of diluted wines','Proline']

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:  52.9s
[Parallel(n_jobs=-1)]: Done 158 tasks    | elapsed:  3.6min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  7.0min finished

```

Confusion Matrix:

```

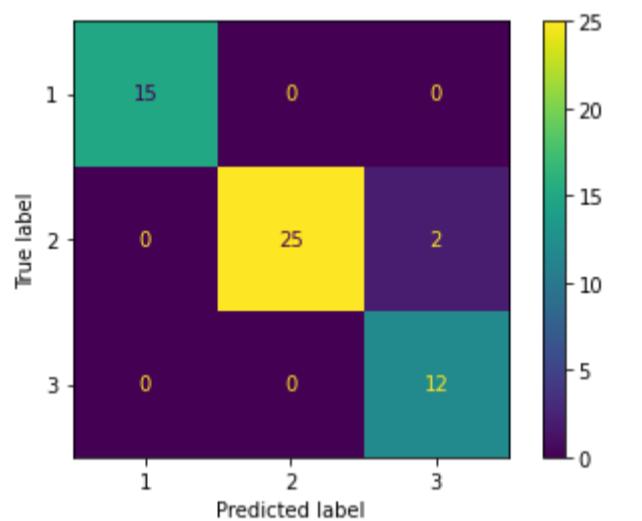
[[15  0  0]
 [ 0 25  2]
 [ 0  0 12]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
1	1.00	1.00	1.00	15
2	1.00	0.93	0.96	27
3	0.86	1.00	0.92	12
accuracy			0.96	54
macro avg	0.95	0.98	0.96	54
weighted avg	0.97	0.96	0.96	54

Accuracy:

```
0.9629629629629629
```



In []:

```

# WINE DATASET
# Random Forest Classifier(With Tuning)[60-40 split]

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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:

```

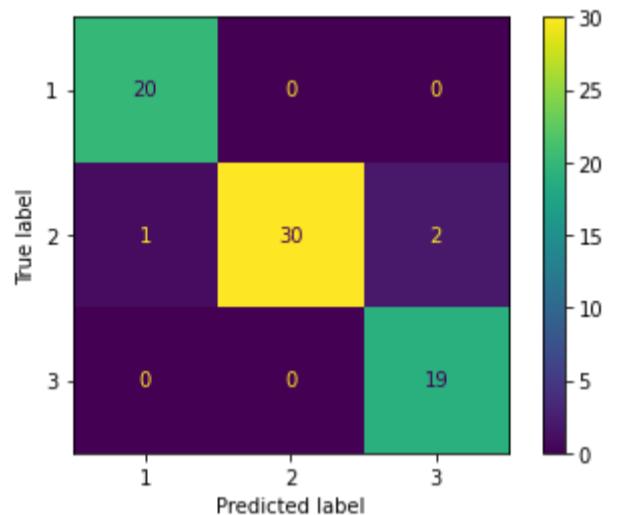
{'bootstrap': True,
 'ccp_alpha': 0.0,
 'class_weight': None,
 'criterion': 'gini',
 'max_depth': None,
 'max_features': 'auto',
 'max_leaf_nodes': None,
 'max_samples': None,
 'min_impurity_decrease': 0.0,
 'min_impurity_split': None,
 'min_samples_leaf': 1,
 'min_samples_split': 2,
 'min_weight_fraction_leaf': 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': 100}

```

```
'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:  52.2s
[Parallel(n_jobs=-1)]: Done 158 tasks    | elapsed:  3.6min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  7.0min finished
Confusion Matrix:
[[20  0  0]
 [ 1 30  2]
 [ 0  0 19]]
```

```
Performance Evaluation
      precision    recall   f1-score   support
      1          0.95     1.00     0.98      20
      2          1.00     0.91     0.95      33
      3          0.90     1.00     0.95      19
      accuracy                           0.96      72
   macro avg       0.95     0.97     0.96      72
weighted avg       0.96     0.96     0.96      72
```

Accuracy:
0.958333333333334



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

import pandas as pd
import numpy as np

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

col_name = ['Class','Alcohol','Malic acid','Ash','Alcalinity of ash','Magnesium','Total phenols','Flavanoids',
           'Nonflavanoid phenols','Proanthocyanins','Color intensity','Hue','OD280/OD315 of diluted wines','Proline']

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:   52.4s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.6min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  7.0min finished

```

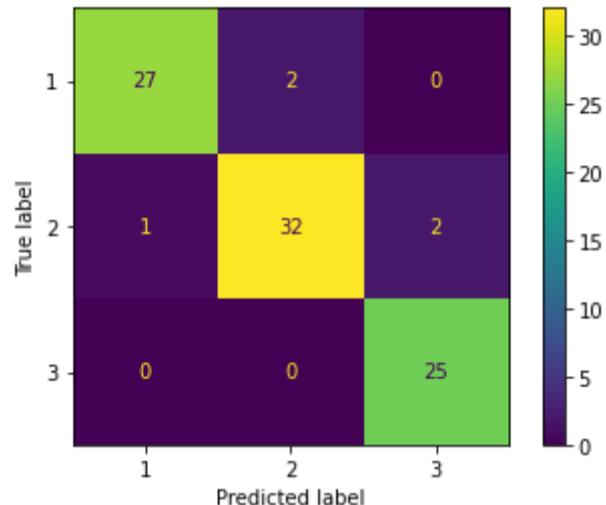
Confusion Matrix:

```

[[27  2  0]
 [ 1 32  2]
 [ 0  0 25]]
-----
```

Performance Evaluation				
	precision	recall	f1-score	support
1	0.96	0.93	0.95	29
2	0.94	0.91	0.93	35
3	0.93	1.00	0.96	25
accuracy			0.94	89
macro avg	0.94	0.95	0.95	89
weighted avg	0.94	0.94	0.94	89

Accuracy:
0.9438202247191011



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

import pandas as pd
import numpy as np

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

col_name = ['Class','Alcohol','Malic acid','Ash','Alcalinity of ash','Magnesium','Total phenols','Flavanoids',
           'Nonflavanoid phenols','Proanthocyanins','Color intensity','Hue','OD280/OD315 of diluted wines','Proline']

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:  52.0s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.6min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  6.9min finished

```

Confusion Matrix:

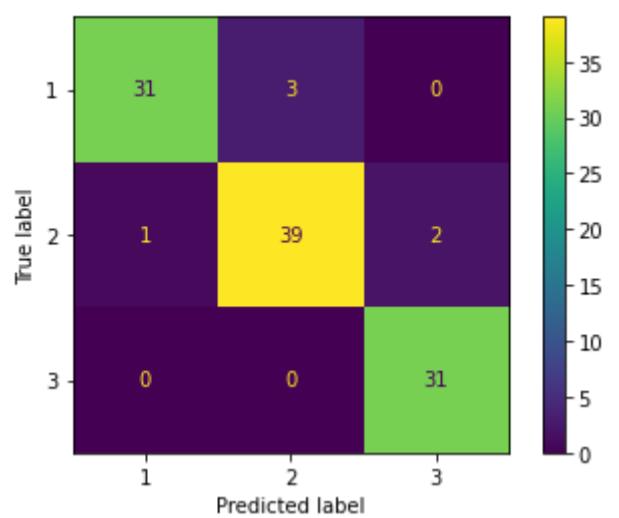
```

[[31  3  0]
 [ 1 39  2]
 [ 0  0 31]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
1	0.97	0.91	0.94	34
2	0.93	0.93	0.93	42
3	0.94	1.00	0.97	31
accuracy			0.94	107
macro avg	0.95	0.95	0.95	107
weighted avg	0.94	0.94	0.94	107

Accuracy:
0.9439252336448598



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

import pandas as pd
import numpy as np

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

col_name = ['Class','Alcohol','Malic acid','Ash','Alcalinity of ash','Magnesium','Total phenols','Flavanoids',
           'Nonflavanoid phenols','Proanthocyanins','Color intensity','Hue','OD280/OD315 of diluted wines','Proline']

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.,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:   51.3s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.5min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  6.8min finished

```

Confusion Matrix:

```

[[35  6  0]
 [ 1 46  2]
 [ 0  1 34]]
-----
```

```

Performance Evaluation
      precision    recall  f1-score   support

          1       0.97      0.85      0.91       41
          2       0.87      0.94      0.90       49
          3       0.94      0.97      0.96       35

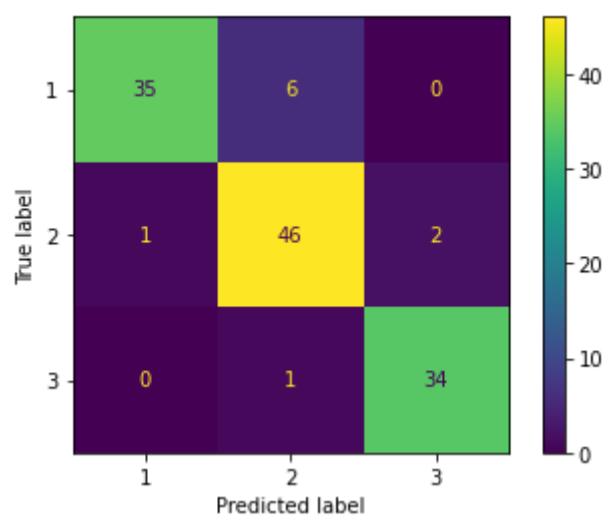
   accuracy                           0.92      125
  macro avg       0.93      0.92      0.92      125
weighted avg       0.92      0.92      0.92      125

```

```

-----
```

Accuracy:
0.92



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

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

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
            'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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 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)
```

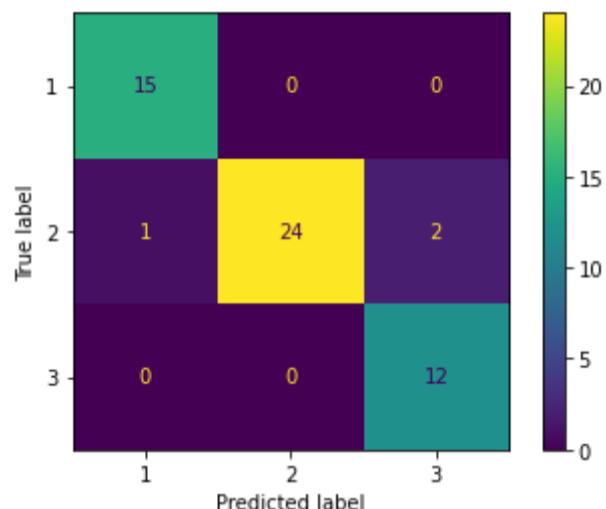
```
Confusion Matrix:
```

```
[[15  0  0]
 [ 1 24  2]
 [ 0  0 12]]
```

```
-----
```

Performance Evaluation				
	precision	recall	f1-score	support
1	0.94	1.00	0.97	15
2	1.00	0.89	0.94	27
3	0.86	1.00	0.92	12
accuracy			0.94	54
macro avg	0.93	0.96	0.94	54
weighted avg	0.95	0.94	0.94	54

Accuracy:
0.9444444444444444



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

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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 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:
[[20  0  0]
 [ 1 29  3]
 [ 0  0 19]]
-----
-----
Performance Evaluation
      precision    recall   f1-score   support
1         0.95   1.00     0.98      20
2         1.00   0.88     0.94      33
3         0.86   1.00     0.93      19
accuracy          0.9444444444444444
macro avg       0.94   0.96     0.95      72
weighted avg    0.95   0.94     0.94      72
-----
-----
Accuracy:
0.9444444444444444



```

In []:

```

# WINE DATASET
# Multi Layer Perceptron(Without Tuning)[50-50 split]

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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 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:

```

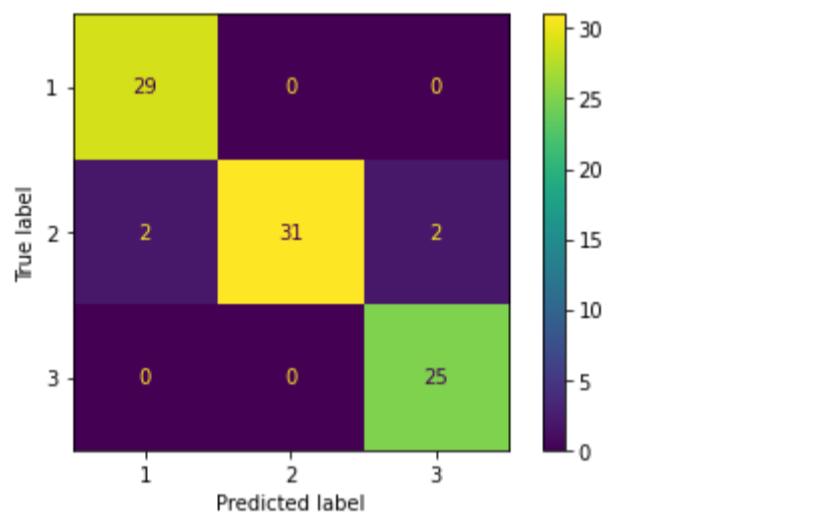
[[29  0  0]
 [ 2 31  2]
 [ 0  0 25]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
1	0.94	1.00	0.97	29
2	1.00	0.89	0.94	35
3	0.93	1.00	0.96	25
accuracy			0.96	89
macro avg	0.95	0.96	0.96	89
weighted avg	0.96	0.96	0.95	89

Accuracy:

0.9550561797752809



In []:

```

# WINE DATASET
# Multi Layer Perceptron(Without Tuning)[40-60 split]

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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 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:

```

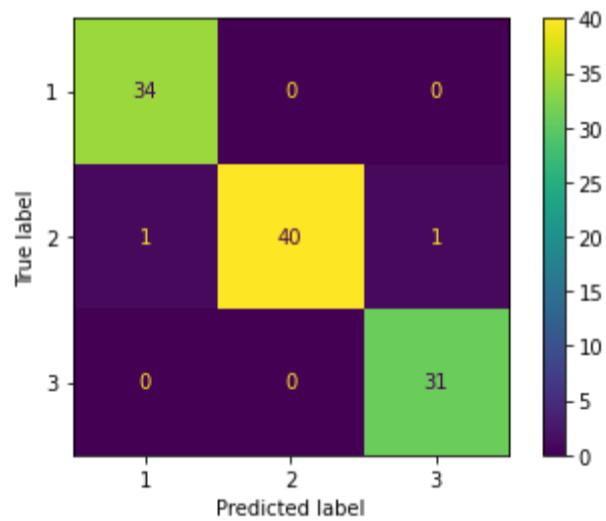
[[34  0  0]
 [ 1 40  1]
 [ 0  0 31]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
1	0.97	1.00	0.99	34
2	1.00	0.95	0.98	42
3	0.97	1.00	0.98	31
accuracy			0.98	107
macro avg	0.98	0.98	0.98	107
weighted avg	0.98	0.98	0.98	107

Accuracy:

0.9813084112149533



In []:

```

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

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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 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()

```

Confusion Matrix:

```

[[41  0  0]
 [ 1 46  2]
 [ 0  1 34]]
-----
```

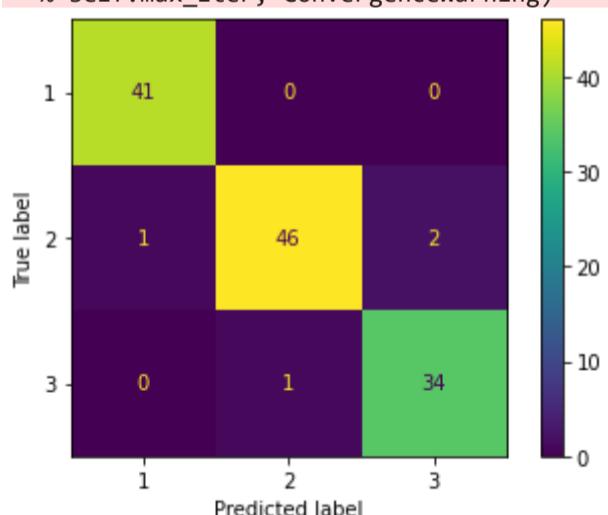
Performance Evaluation

	precision	recall	f1-score	support
1	0.98	1.00	0.99	41
2	0.98	0.94	0.96	49
3	0.94	0.97	0.96	35
accuracy			0.97	125
macro avg	0.97	0.97	0.97	125
weighted avg	0.97	0.97	0.97	125

Accuracy:

0.968

```
/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)
```



In []:

```

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

import pandas as pd
import numpy as np

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

col_name = ['Class','Alcohol','Malic acid','Ash','Alcalinity of ash','Magnesium','Total phenols','Flavanoids',
           'Nonflavanoid phenols','Proanthocyanins','Color intensity','Hue','OD280/OD315 of diluted wines','Proline']

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=0.3,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()

```

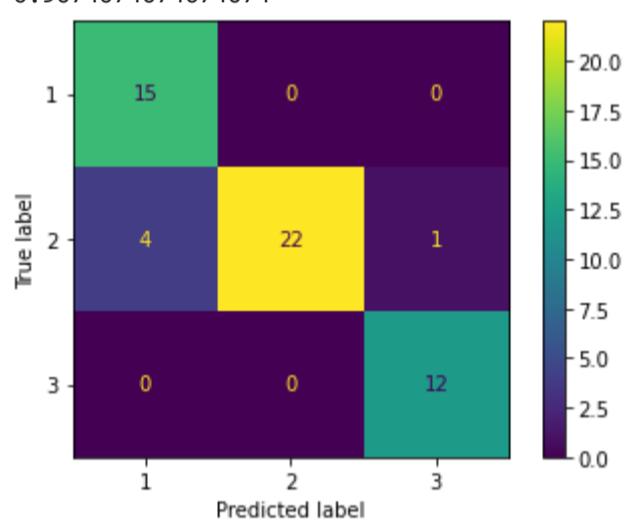
Parameters currently in use:

```
{
'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,
'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:
[[15  0  0]
 [ 4 22  1]
 [ 0  0 12]]
-----
-----
Performance Evaluation
      precision    recall   f1-score   support
1         0.79     1.00     0.88      15
2         1.00     0.81     0.90      27
3         0.92     1.00     0.96      12
accuracy                           0.91      54
macro avg       0.90     0.94     0.91      54
weighted avg    0.92     0.91     0.91      54
-----
-----
Accuracy:
0.9074074074074074

```



```

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

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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.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:

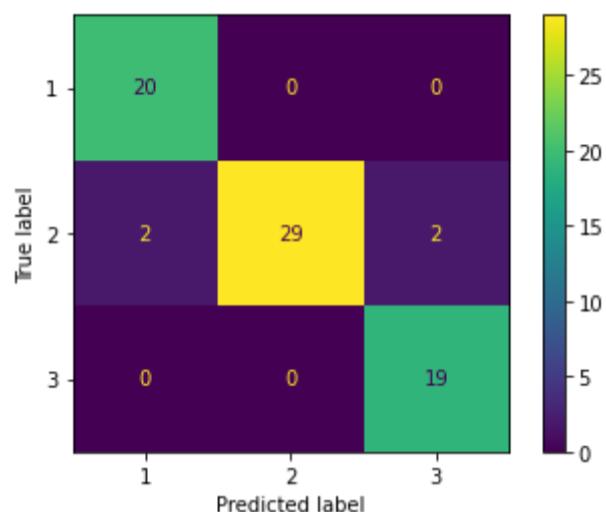
```

[[20  0  0]
 [ 2 29  2]
 [ 0  0 19]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
1	0.91	1.00	0.95	20
2	1.00	0.88	0.94	33
3	0.90	1.00	0.95	19
accuracy			0.94	72
macro avg	0.94	0.96	0.95	72
weighted avg	0.95	0.94	0.94	72

Accuracy:
0.9444444444444444



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

import pandas as pd
import numpy as np

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

col_name = ['Class','Alcohol','Malic acid','Ash','Alcalinity of ash','Magnesium','Total phenols','Flavanoids',
           'Nonflavanoid phenols','Proanthocyanins','Color intensity','Hue','OD280/OD315 of diluted wines','Proline']

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.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:

```

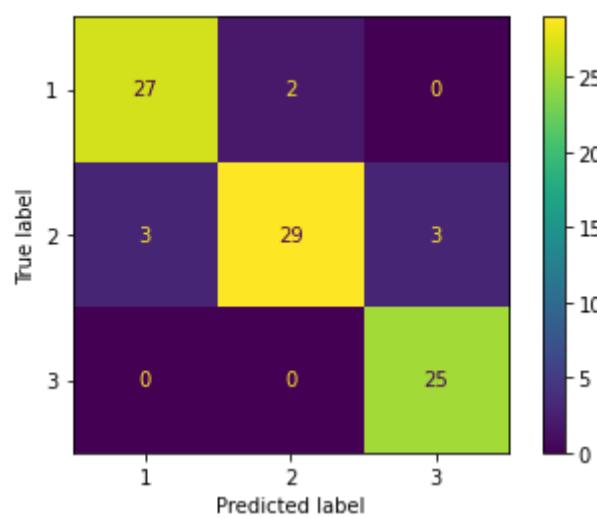
[[27  2  0]
 [ 3 29  3]
 [ 0  0 25]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
1	0.90	0.93	0.92	29
2	0.94	0.83	0.88	35
3	0.89	1.00	0.94	25
accuracy			0.91	89
macro avg	0.91	0.92	0.91	89
weighted avg	0.91	0.91	0.91	89

Accuracy:

0.9101123595505618



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

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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.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:

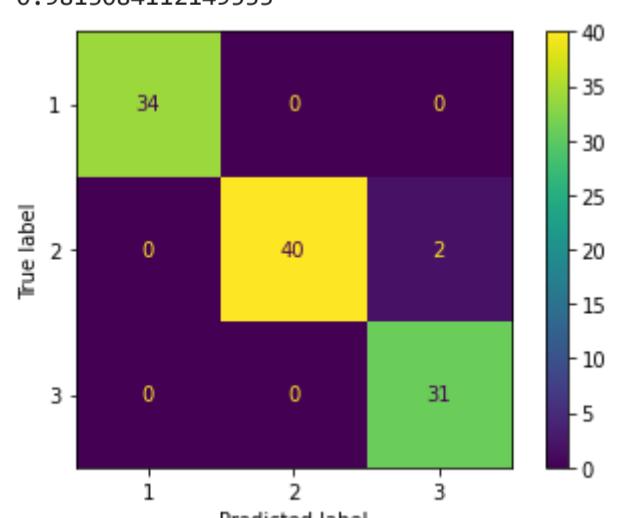
```
[[34  0  0]
 [ 0 40  2]
 [ 0  0 31]]
```

Performance Evaluation

	precision	recall	f1-score	support
1	1.00	1.00	1.00	34
2	1.00	0.95	0.98	42
3	0.94	1.00	0.97	31
accuracy			0.98	107
macro avg	0.98	0.98	0.98	107
weighted avg	0.98	0.98	0.98	107

Accuracy:

0.9813084112149533



In []:

```
# WINE DATASET
# Multi Layer Perceptron(With Tuning)[30-70 split]
```

```

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
            'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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.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()

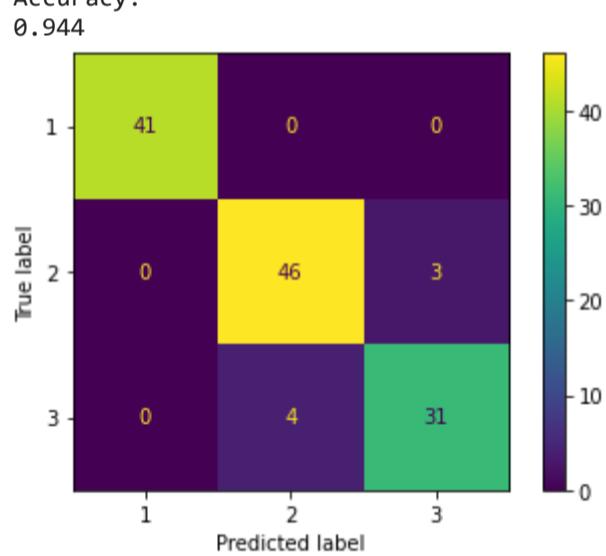
```

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:
[[41  0  0]
 [ 0 46  3]
 [ 0  4 31]]
-----
-----
Performance Evaluation
      precision    recall   f1-score   support
      1         1.00     1.00     1.00      41
      2         0.92     0.94     0.93      49
      3         0.91     0.89     0.90      35
      accuracy                           0.944      125
      macro avg       0.94     0.94     0.94      125
  weighted avg       0.94     0.94     0.94      125
  -----
  Accuracy: 0.944

```



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

```

In [ ]:
# WINE DATASET
# SVM(Without Tuning)[70-30 split]

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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=0.3, 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:

```

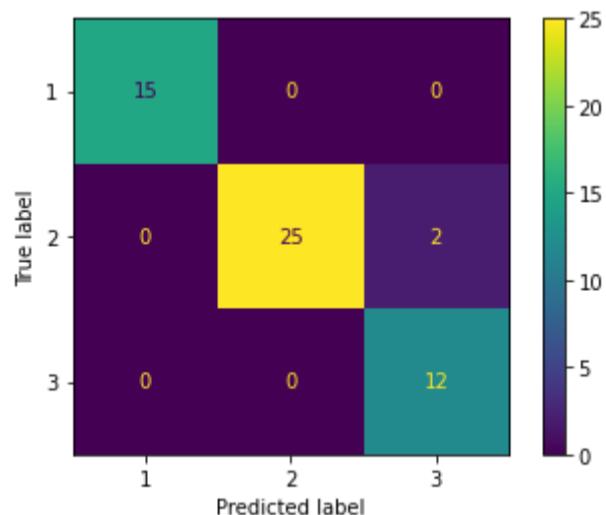
[[15  0  0]
 [ 0 25  2]
 [ 0  0 12]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
1	1.00	1.00	1.00	15
2	1.00	0.93	0.96	27
3	0.86	1.00	0.92	12
accuracy			0.96	54
macro avg	0.95	0.98	0.96	54
weighted avg	0.97	0.96	0.96	54

Accuracy:

0.9629629629629629



In []:

```

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

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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.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:

```

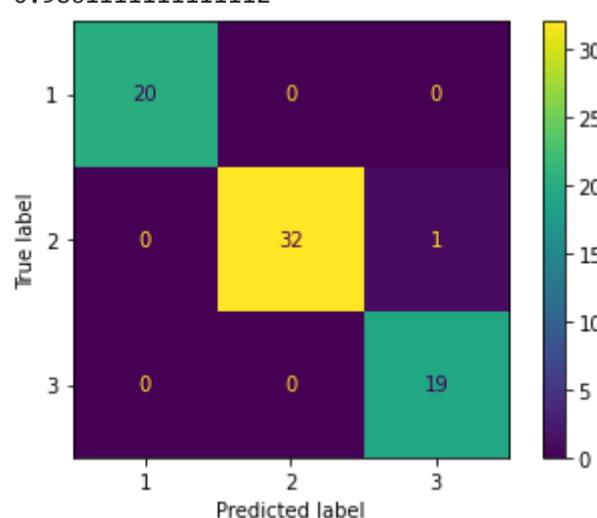
[[20  0  0]
 [ 0 32  1]
 [ 0  0 19]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
1	1.00	1.00	1.00	20
2	1.00	0.97	0.98	33
3	0.95	1.00	0.97	19
accuracy			0.99	72
macro avg	0.98	0.99	0.99	72
weighted avg	0.99	0.99	0.99	72

Accuracy:

0.986111111111112



In []:

```
# WINE DATASET
# SVM(Without Tuning)[50-50 split]
```

```
import pandas as pd
import numpy as np
```

```

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
            'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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.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:

```

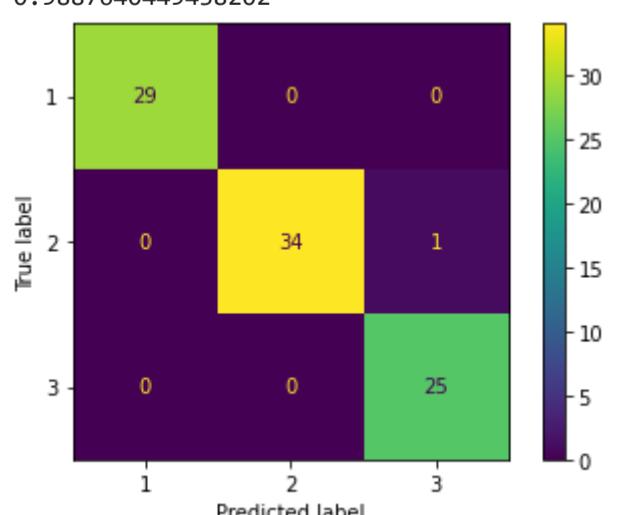
[[29  0  0]
 [ 0 34  1]
 [ 0  0 25]]
-----
```

Performance Evaluation

	precision	recall	f1-score	support
1	1.00	1.00	1.00	29
2	1.00	0.97	0.99	35
3	0.96	1.00	0.98	25
accuracy			0.99	89
macro avg	0.99	0.99	0.99	89
weighted avg	0.99	0.99	0.99	89

Accuracy:

0.9887640449438202



```
In [ ]: # WINE DATASET
# SVM(Without Tuning)[40-60 split]

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
            'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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.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:

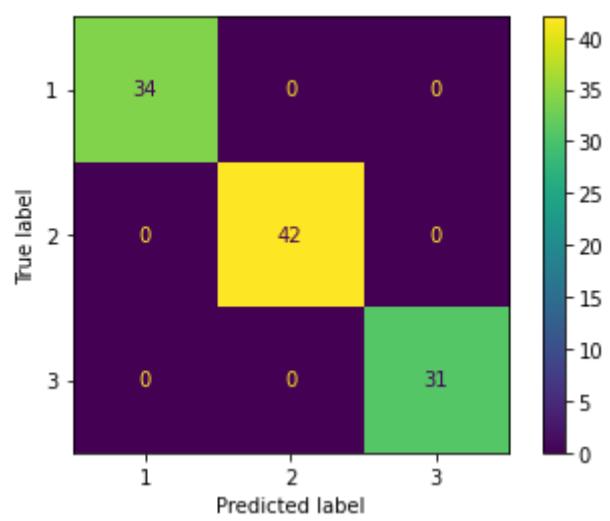
```
[[34  0  0]
 [ 0 42  0]
 [ 0  0 31]]
```

Performance Evaluation

	precision	recall	f1-score	support
1	1.00	1.00	1.00	34
2	1.00	1.00	1.00	42
3	1.00	1.00	1.00	31
accuracy			1.00	107
macro avg	1.00	1.00	1.00	107
weighted avg	1.00	1.00	1.00	107

Accuracy:

1.0



```
In [ ]:
# WINE DATASET
# SVM(Without Tuning)[30-70 split]

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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.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:

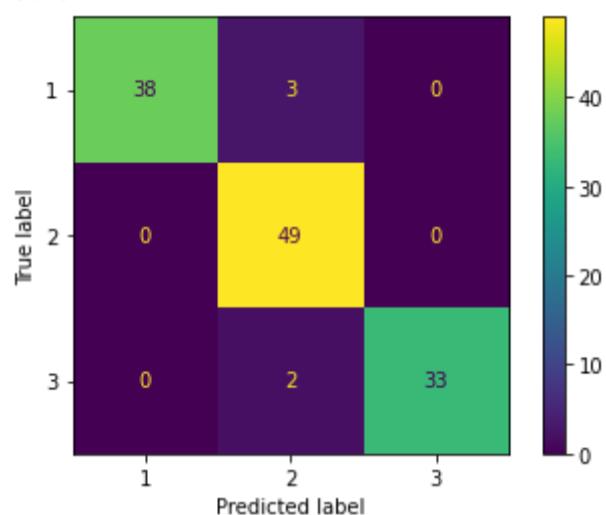
	1	2	3
1	38	3	0
2	0	49	0
3	0	2	33

Performance Evaluation

	precision	recall	f1-score	support
1	1.00	0.93	0.96	41
2	0.91	1.00	0.95	49
3	1.00	0.94	0.97	35

accuracy		0.96		125
macro avg	0.97	0.96	0.96	125
weighted avg	0.96	0.96	0.96	125

Accuracy:
0.96



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

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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=0.3, 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 .....
[CV] ..... C=0.1, gamma=1, kernel=rbf, total= 0.0s
[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 .....
[CV] ..... C=0.1, gamma=1, kernel=rbf, total= 0.0s
[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 .....
[CV] ..... C=0.1, gamma=1, kernel=rbf, total= 0.0s
[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 .....
[CV] ..... C=0.1, gamma=1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=rbf .....
[CV] ..... C=0.1, gamma=0.1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=rbf .....
[CV] ..... C=0.1, gamma=0.1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=rbf .....
[CV] ..... C=0.1, gamma=0.1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=rbf .....
[CV] ..... C=0.1, gamma=0.1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=rbf .....
[CV] ..... C=0.1, gamma=0.1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=rbf .....
[CV] ..... C=0.1, gamma=0.1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=poly .....
[CV] ..... C=0.1, gamma=0.1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=poly .....
[CV] ..... C=0.1, gamma=0.1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=poly .....
[CV] ..... C=0.1, gamma=0.1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=poly .....
[CV] ..... C=0.1, gamma=0.1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=poly .....
[CV] ..... C=0.1, gamma=0.1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=poly .....
[CV] ..... C=0.1, gamma=0.1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=poly .....
[CV] ..... C=0.1, gamma=0.1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=poly .....
[CV] ..... C=0.1, gamma=0.1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=poly .....
[CV] ..... C=0.1, gamma=0.1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=poly .....
[CV] ..... C=0.1, gamma=0.1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=0.1, kernel=sigmoid, total= 0.0s
```


Confusion Matrix:
[[15 0 0]

$$\begin{bmatrix} 15 & 0 & 0 \\ 0 & 26 & 1 \\ 0 & 0 & 12 \end{bmatrix}$$

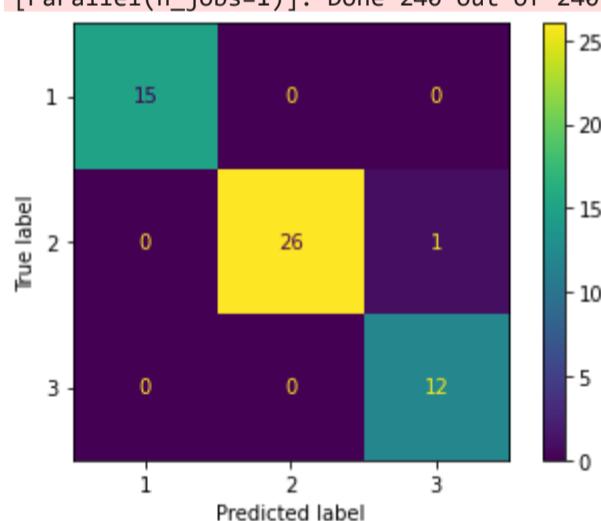
Performance Evaluation

	precision	recall	f1-score	support
1	1.00	1.00	1.00	15
2	1.00	0.96	0.98	27
3	0.92	1.00	0.96	12
accuracy			0.98	54
macro avg	0.97	0.99	0.98	54
weighted avg	0.98	0.98	0.98	54

Accuracy:

Accuracy:
0.9814814814814815

[Parallel(n_jobs=1)]: Done 240 out of 240 | elapsed: 0.9s finished



In []:

```
# WINE DATASET  
# SVM(with Tuning)[60-40 split]
```

```
import pandas as pd
import numpy as np

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

col_name = ['Class','Alcohol','Malic acid','Ash','Alcalinity of ash','Magnesium','Total phenols','Flavanoids',
            'Nonflavanoid phenols','Proanthocyanins','Color intensity','Hue','OD280/OD315 of diluted wines','Proline']

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.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 .....
[CV] ..... C=0.1, gamma=1, kernel=rbf, total= 0.0s
[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 .....
[CV] ..... C=0.1, gamma=1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=1, kernel=rbf .....

```

```
[CV] C=0.1, gamma=0.001, kernel=rbf .....  
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.  
[Parallel(n_jobs=1)]: Done 1 out of 1 | elapsed: 0.0s remaining: 0.0s  
[CV] ..... C=0.1, gamma=0.001, kernel=rbf, total= 0.0s  
[CV] C=0.1, gamma=0.001, kernel=rbf .....  
[CV] ..... C=0.1, gamma=0.001, kernel=rbf, total= 0.0s  
[CV] C=0.1, gamma=0.001, kernel=rbf .....  
[CV] ..... C=0.1, gamma=0.001, kernel=rbf, total= 0.0s  
[CV] C=0.1, gamma=0.001, kernel=rbf .....  
[CV] ..... C=0.1, gamma=0.001, kernel=rbf, total= 0.0s  
[CV] C=0.1, gamma=0.001, kernel=poly .....  
[CV] ..... C=0.1, gamma=0.001, kernel=poly, total= 0.0s  
[CV] C=0.1, gamma=0.001, kernel=poly .....  
[CV] ..... C=0.1, gamma=0.001, kernel=poly, total= 0.0s  
[CV] C=0.1, gamma=0.001, kernel=poly .....  
[CV] ..... C=0.1, gamma=0.001, kernel=poly, total= 0.0s
```


[cv] Confusion Matrix:

```
[[20  0  0]
 [ 0 32  1]
 [ 0  0 19]]
```

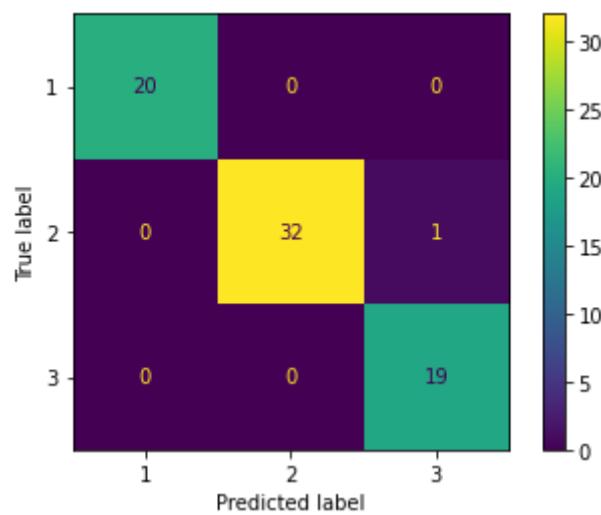
Performance Evaluation

Performance Evaluation		precision	recall	f1-score	support
1	1.00	1.00	1.00	1.00	20
2	1.00	0.97	0.98	0.98	33
3	0.95	1.00	0.97	0.97	19
accuracy				0.99	72
macro avg		0.98	0.99	0.99	72
weighted avg		0.99	0.99	0.99	72

Assumptions

Accuracy:
0.986111111111112

[Parallel(n_jobs=1)]: Done 340 out of 340 | elapsed: 0.9s finished



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

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
            'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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.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:

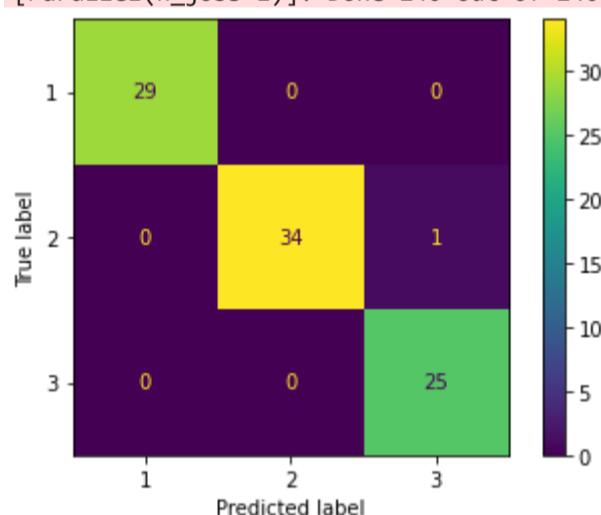

```

[CV] ..... 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=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:
[[29  0  0]
 [ 0 34  1]
 [ 0  0 25]]
-----
```

Performance Evaluation		precision	recall	f1-score	support
1	1.00	1.00	1.00	29	
2	1.00	0.97	0.99	35	
3	0.96	1.00	0.98	25	
accuracy				0.99	89
macro avg	0.99	0.99	0.99	89	
weighted avg	0.99	0.99	0.99	89	

Accuracy:
0.9887640449438202

[Parallel(n_jobs=1)]: Done 240 out of 240 | elapsed: 0.9s finished



```

In [ ]:
# WINE DATASET
# SVM(With Tuning)[40-60 split]

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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.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()

```

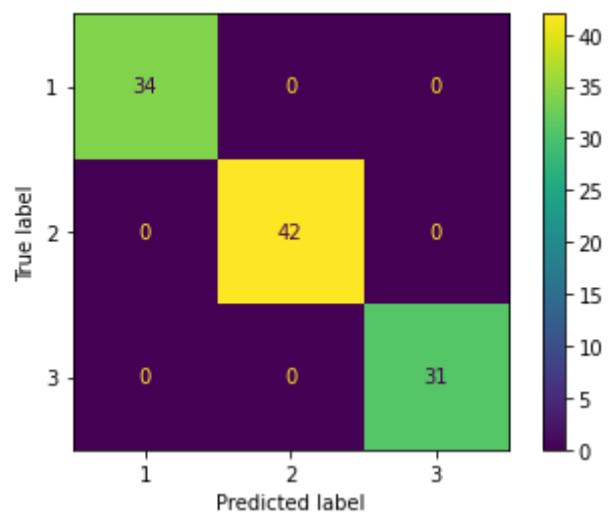
Parameters currently in use:

```

{'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 .....
[CV] ..... C=0.1, gamma=1, kernel=rbf, total= 0.0s
[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 .....
[CV] ..... C=0.1, gamma=1, kernel=rbf, total= 0.0s
[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=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s

```


Accuracy: 1.0 [Parallel](n_jobs=1): Done 240 out of 240 | elapsed: 0.9s finished



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

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
            'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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.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] ..... 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:
```

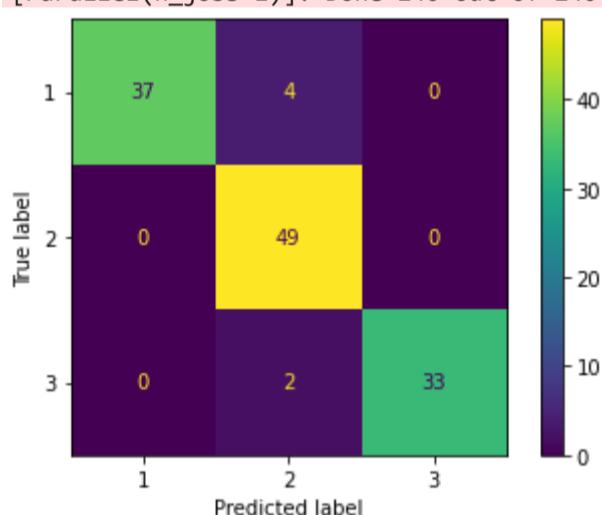
```
[[37  4  0]
 [ 0 49  0]
 [ 0  2 33]]
```

Performance Evaluation					
	precision	recall	f1-score	support	
1	1.00	0.90	0.95	41	
2	0.89	1.00	0.94	49	
3	1.00	0.94	0.97	35	
accuracy			0.95	125	
macro avg	0.96	0.95	0.95	125	
weighted avg	0.96	0.95	0.95	125	

Accuracy:

0.952

[Parallel(n_jobs=1)]: Done 240 out of 240 | elapsed: 0.9s finished



In []:

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

In []:

```
# WINE DATASET
# Random Forest Classifier(Without Tuning)[70-30 split]

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
           'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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=0.3, 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=13)
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()
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")
```

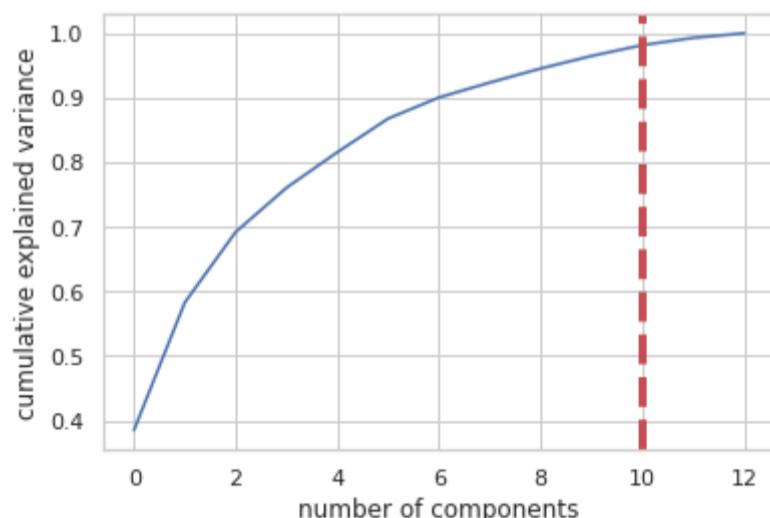
```
print(classification_report(y_test, y_pred))
```

```
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:
```

```
[[15  0  0]
 [ 2 23  2]
 [ 0  0 12]]
```

```
-----
```

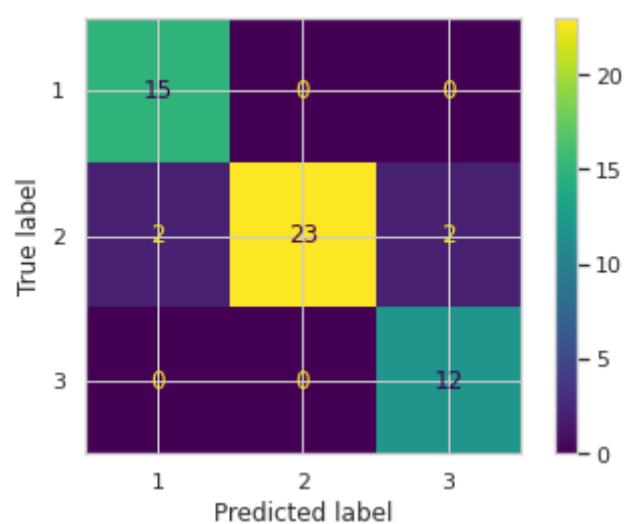
```
Performance Evaluation
```

	precision	recall	f1-score	support
1	0.88	1.00	0.94	15
2	1.00	0.85	0.92	27
3	0.86	1.00	0.92	12
accuracy			0.93	54
macro avg	0.91	0.95	0.93	54
weighted avg	0.94	0.93	0.93	54

```
-----
```

```
-----
```

Accuracy:
0.9259259259259259



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

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
            'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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)

# 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=13)
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("-----")

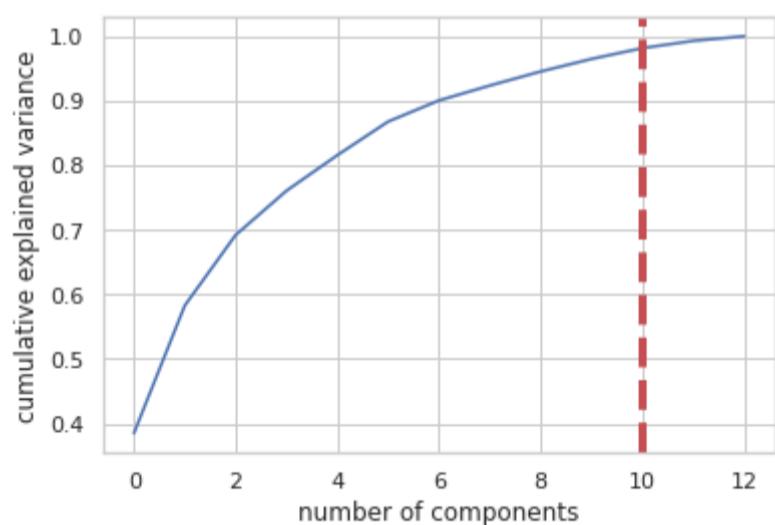
print("Performance Evaluation")
print(classification_report(y_test, y_pred))

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:

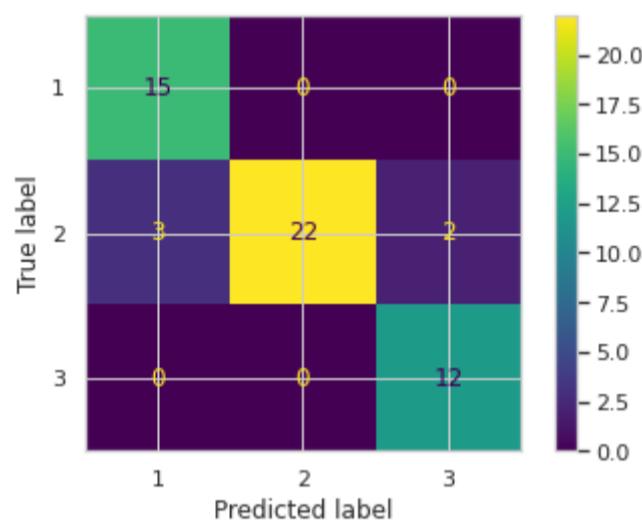
```
{
'bootstrap': True,
'ccp_alpha': 0.0,
'class_weight': None,
'criterion': 'gini',
'max_depth': None,
'max_features': 'auto',
'max_leaf_nodes': None,
'max_samples': None,
'min_impurity_decrease': 0.0,
'min_impurity_split': None,
'min_samples_leaf': 1,
'min_samples_split': 2,
'min_weight_fraction_leaf': 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:  53.5s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.6min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  6.8min finished
Confusion Matrix:
[[15  0  0]
 [ 3 22  2]
 [ 0  0 12]]
-----
```

Performance Evaluation		precision	recall	f1-score	support
	1	0.83	1.00	0.91	15
	2	1.00	0.81	0.90	27
	3	0.86	1.00	0.92	12
accuracy				0.91	54
macro avg		0.90	0.94	0.91	54
weighted avg		0.92	0.91	0.91	54

Accuracy:
0.9074074074074074



```

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

import pandas as pd
import numpy as np

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

col_name = ['Class','Alcohol','Malic acid','Ash','Alcalinity of ash','Magnesium','Total phenols','Flavanoids',
           'Nonflavanoid phenols','Proanthocyanins','Color intensity','Hue','OD280/OD315 of diluted wines','Proline']

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)

# 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=13)
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.axline(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("-----")

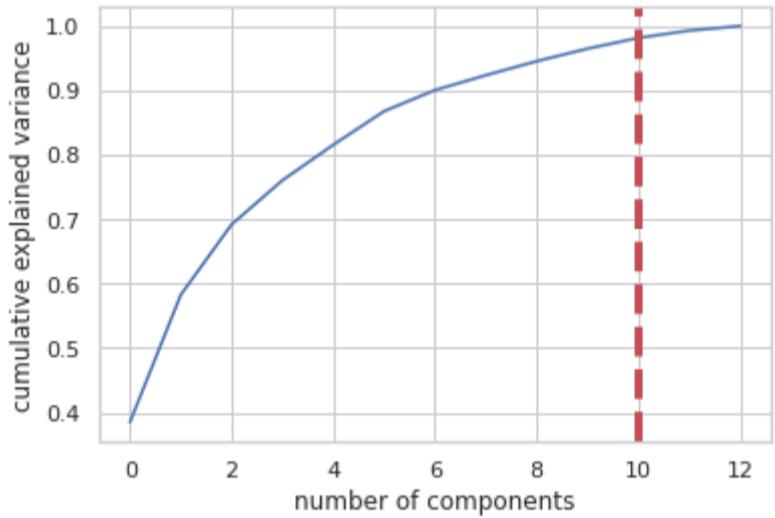
print("Performance Evaluation")
print(classification_report(y_test, y_pred))

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

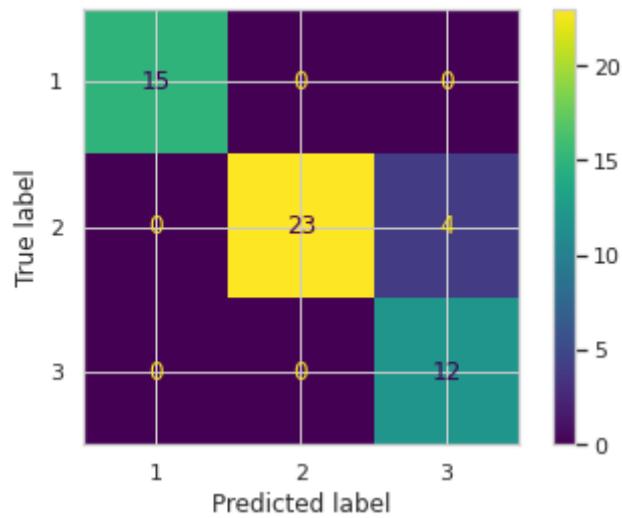
```
/usr/local/lib/python3.7/dist-packages/sklearn/neural_network/_multilayer_perceptron.py:571: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (200) reached and the optimization hasn't converged yet.
  % self.max_iter, ConvergenceWarning)
```

Confusion Matrix:

```
[[15  0  0]
 [ 0 23  4]
 [ 0  0 12]]
```

	precision	recall	f1-score	support
1	1.00	1.00	1.00	15
2	1.00	0.85	0.92	27
3	0.75	1.00	0.86	12
accuracy			0.93	54
macro avg	0.92	0.95	0.93	54
weighted avg	0.94	0.93	0.93	54

Accuracy:
0.9259259259259259



In []:

```
# WINE DATASET
# Multi Layer Perceptron(With Tuning)[70-30 split]

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
            'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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=0.3, 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=13)
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, 0.001, 0.01, 0.1],
    'learning_rate': ['constant', 'linear'],
    'learning_rate_init': [0.001, 0.01, 0.1, 1],
    'power_t': [0.5, 0.9],
    'warm_start': [False, True]
}
```

```

        '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("-----")

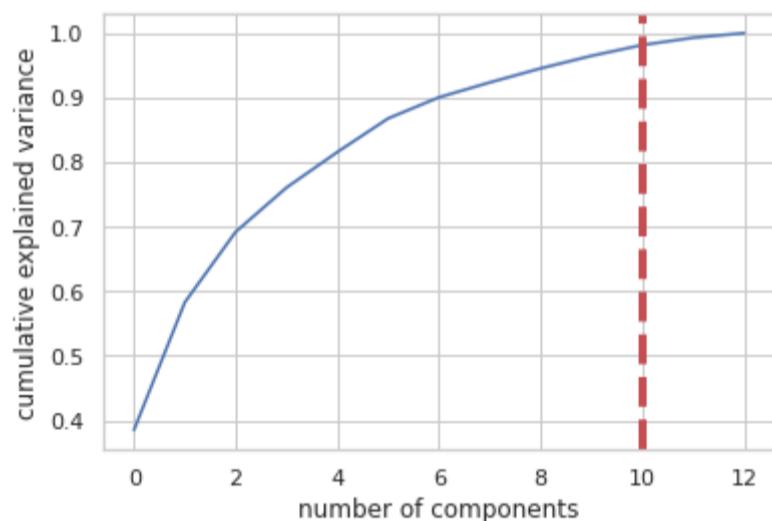
print("Performance Evaluation")
print(classification_report(y_test, y_pred))

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:

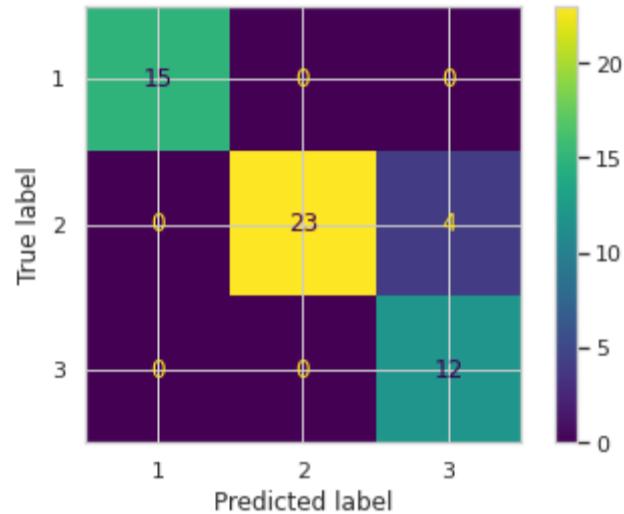
```
[[15  0  0]
 [ 0 23  4]
 [ 0  0 12]]
```

Performance Evaluation

	precision	recall	f1-score	support
1	1.00	1.00	1.00	15
2	1.00	0.85	0.92	27
3	0.75	1.00	0.86	12
accuracy			0.93	54
macro avg	0.92	0.95	0.93	54
weighted avg	0.94	0.93	0.93	54

Accuracy:

```
0.9259259259259259
```



In []:

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

import pandas as pd
import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
            'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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=0.3, 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=13)
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("-----")

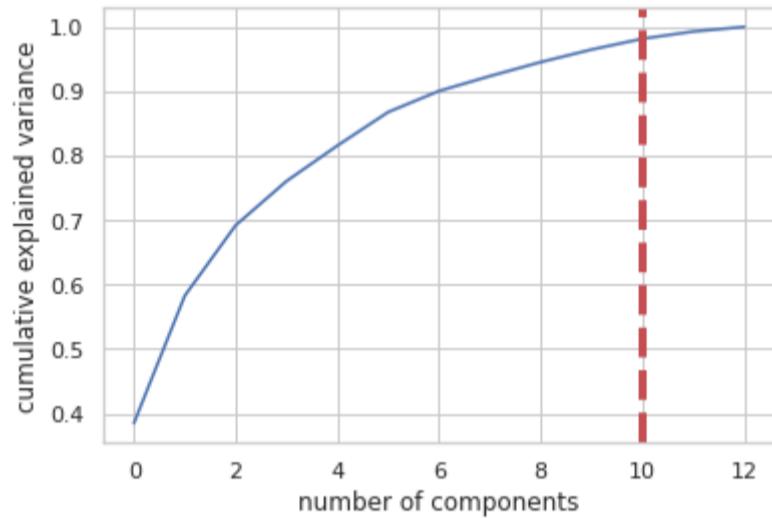
print("Performance Evaluation")
print(classification_report(y_test, y_pred))

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()

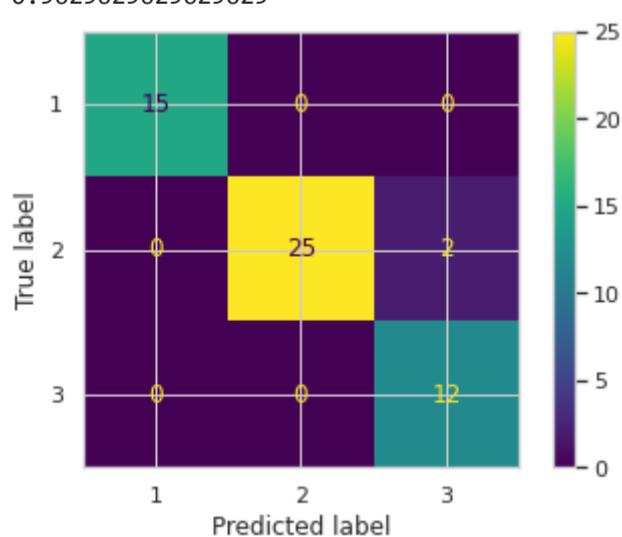
```



None
Confusion Matrix:
[[15 0 0]
 [0 25 2]
 [0 0 12]]

Performance Evaluation
precision recall f1-score support
1 1.00 1.00 1.00 15
2 1.00 0.93 0.96 27
3 0.86 1.00 0.92 12
accuracy 0.96 0.96 0.96 54
macro avg 0.95 0.98 0.96 54
weighted avg 0.97 0.96 0.96 54

Accuracy:
0.9629629629629629



In []:

```

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

import pandas as pd

```

```

import numpy as np

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

col_name = ['Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids',
            'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline']

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=0.3, 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=13)
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("-----")

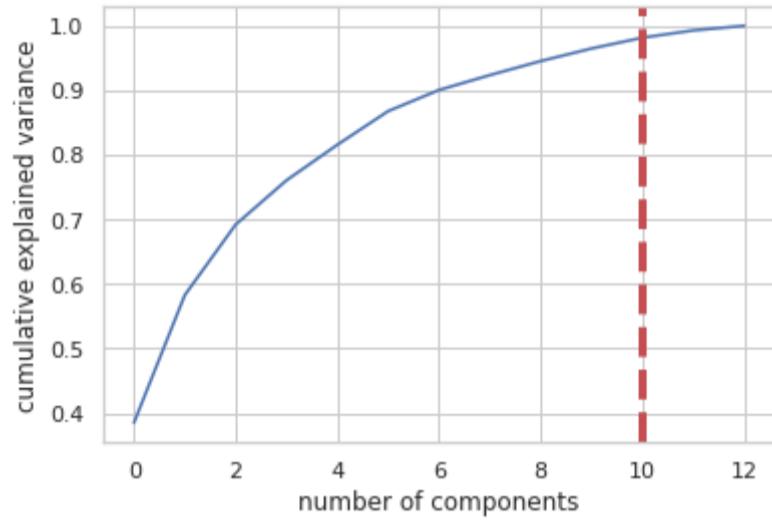
```

```
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:

```
{'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 .....
[CV] ..... C=0.1, gamma=1, kernel=rbf, total= 0.0s
[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 .....
[CV] ..... C=0.1, gamma=1, kernel=rbf, total= 0.0s
[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 .....
[CV] ..... C=0.1, gamma=1, kernel=rbf, total= 0.0s
[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=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=poly .....
[CV] ..... C=0.1, gamma=1, kernel=poly, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=1, kernel=sigmoid .....
[CV] ..... C=0.1, gamma=1, kernel=sigmoid, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=rbf .....
[CV] ..... C=0.1, gamma=0.1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=rbf .....
[CV] ..... C=0.1, gamma=0.1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=rbf .....
[CV] ..... C=0.1, gamma=0.1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=rbf .....
[CV] ..... C=0.1, gamma=0.1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=rbf .....
[CV] ..... C=0.1, gamma=0.1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=rbf .....
[CV] ..... C=0.1, gamma=0.1, kernel=rbf, total= 0.0s
[CV] C=0.1, gamma=0.1, kernel=poly .....
[CV] ..... C=0.1, gamma=0.1, kernel=poly, total= 0.0s
```

```
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.  
[Parallel(n_jobs=1)]: Done 1 out of 1 | elapsed: 0.0s remaining: 0.0s
```



```
[CV] C=100, gamma=0.01, kernel=sigmoid .....  
[CV] ..... C=100, gamma=0.01, kernel=sigmoid, total= 0.0s  
[CV] C=100, gamma=0.01, kernel=sigmoid .....  
[CV] ..... C=100, gamma=0.01, kernel=sigmoid, total= 0.0s  
[CV] C=100, gamma=0.01, kernel=sigmoid .....  
[CV] ..... C=100, gamma=0.01, kernel=sigmoid, total= 0.0s  
[CV] C=100, gamma=0.01, kernel=sigmoid .....  
[CV] ..... C=100, gamma=0.01, kernel=sigmoid, total= 0.0s  
[CV] C=100, gamma=0.01, kernel=sigmoid .....  
[CV] ..... C=100, gamma=0.01, kernel=sigmoid, total= 0.0s  
[CV] C=100, gamma=0.001, kernel=rbf .....  
[CV] ..... C=100, gamma=0.001, kernel=rbf, total= 0.0s  
[CV] C=100, gamma=0.001, kernel=rbf .....  
[CV] ..... C=100, gamma=0.001, kernel=rbf, total= 0.0s  
[CV] C=100, gamma=0.001, kernel=rbf .....  
[CV] ..... C=100, gamma=0.001, kernel=rbf, total= 0.0s  
[CV] C=100, gamma=0.001, kernel=rbf .....  
[CV] ..... C=100, gamma=0.001, kernel=rbf, total= 0.0s  
[CV] C=100, gamma=0.001, kernel=rbf .....  
[CV] ..... C=100, gamma=0.001, kernel=rbf, total= 0.0s  
[CV] C=100, gamma=0.001, kernel=rbf .....  
[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=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:  
[[15  0  0]  
 [ 0 25  2]  
 [ 0  0 12]]
```

[[15 0 0]

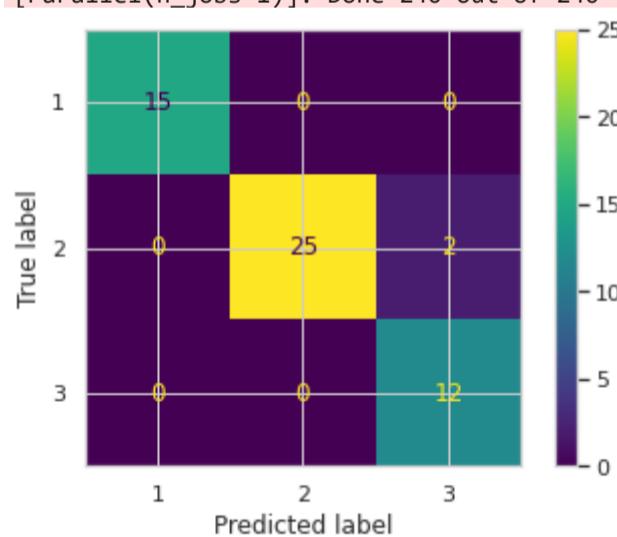
```
[[-15  0  0]
 [ 0 25  2]
 [ 0  0 12]]
```

Performance Evaluation				
	precision	recall	f1-score	support
1	1.00	1.00	1.00	15
2	1.00	0.93	0.96	27
3	0.86	1.00	0.92	12
accuracy			0.96	54
macro avg	0.95	0.98	0.96	54
weighted avg	0.97	0.96	0.96	54

Accuracy:

Accuracy:
0.9629629629629

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



In []: