

# JADAVPUR UNIVERSITY

## MACHINE LEARNING LABORATORY

### ASSIGNMENT #2 - C) IONOSPHERE DATASET

NAME - RITIK BAID

DEPARTMENT - INFORMATION TECHNOLOGY

ROLL NO - 001811001035

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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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(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()
```

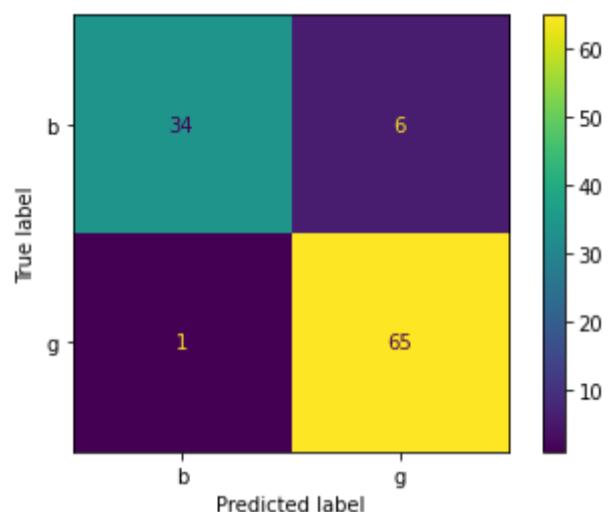
Confusion Matrix:

```
[[34  6]
 [ 1 65]]
```

```
-----
```

| Performance Evaluation |           |        |          |         |
|------------------------|-----------|--------|----------|---------|
|                        | precision | recall | f1-score | support |
| b                      | 0.97      | 0.85   | 0.91     | 40      |
| g                      | 0.92      | 0.98   | 0.95     | 66      |
| accuracy               |           |        | 0.93     | 106     |
| macro avg              | 0.94      | 0.92   | 0.93     | 106     |
| weighted avg           | 0.94      | 0.93   | 0.93     | 106     |

-----  
Accuracy:  
0.9339622641509434



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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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))

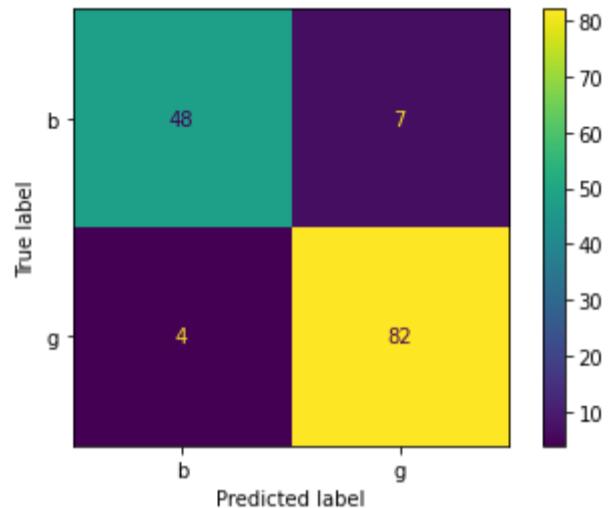
import matplotlib.pyplot as plt
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(classifier, X_test, y_test)
plt.show()
```

```

Confusion Matrix:
[[48  7]
 [ 4 82]]
-----
Performance Evaluation
      precision    recall   f1-score   support
      b          0.92     0.87     0.90      55
      g          0.92     0.95     0.94      86
      accuracy           0.92
      macro avg       0.92     0.91     0.92      141
  weighted avg       0.92     0.92     0.92      141

```

Accuracy:  
0.9219858156028369



```

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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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:

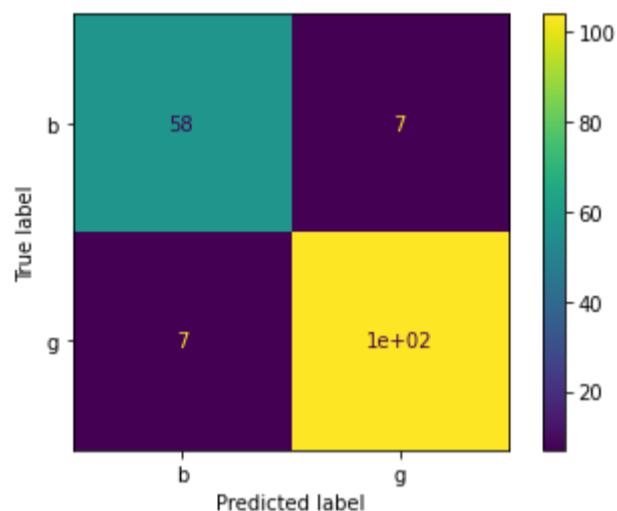
```
[[ 58  7]
 [ 7 104]]
```

-----  
Performance Evaluation

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 0.89      | 0.89   | 0.89     | 65      |
| g            | 0.94      | 0.94   | 0.94     | 111     |
| accuracy     |           |        | 0.92     | 176     |
| macro avg    | 0.91      | 0.91   | 0.91     | 176     |
| weighted avg | 0.92      | 0.92   | 0.92     | 176     |

-----  
Accuracy:

```
0.9204545454545454
```



In [ ]:

```

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

import pandas as pd
import numpy as np

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

col_name = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32', '33', '34', 'Class']

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:

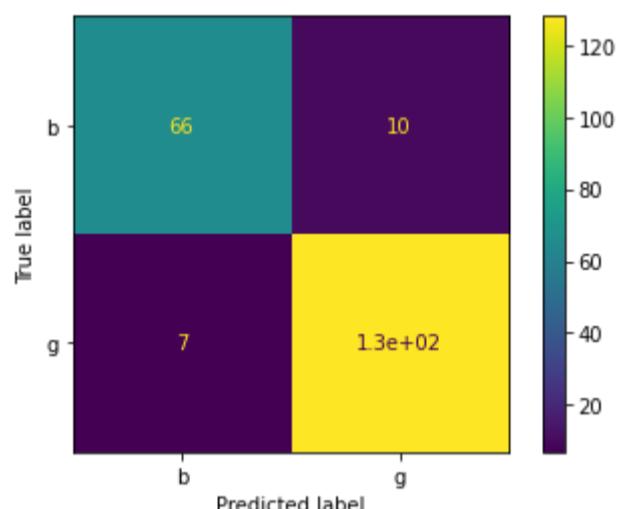
```
[[ 66 10]
 [ 7 128]]
```

Performance Evaluation

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 0.90      | 0.87   | 0.89     | 76      |
| g            | 0.93      | 0.95   | 0.94     | 135     |
| accuracy     |           |        | 0.92     | 211     |
| macro avg    | 0.92      | 0.91   | 0.91     | 211     |
| weighted avg | 0.92      | 0.92   | 0.92     | 211     |

Accuracy:

```
0.919431279620853
```



In [ ]:

```

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

import pandas as pd
import numpy as np

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

col_name = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32', '33', '34', 'Class']

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:

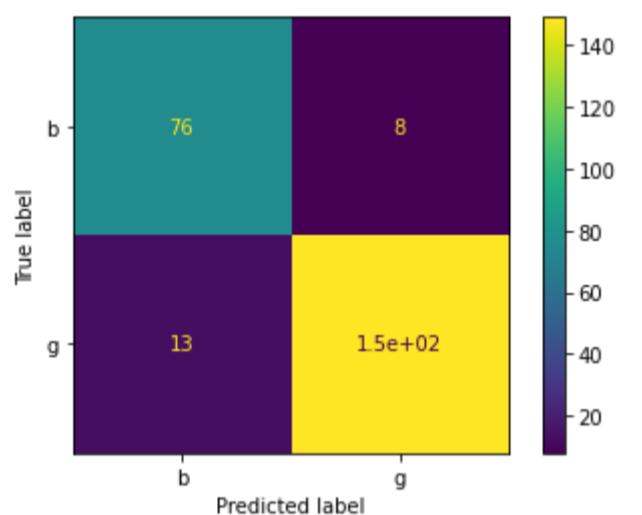
```
[[ 76   8]
 [ 13 149]]
```

Performance Evaluation

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 0.85      | 0.90   | 0.88     | 84      |
| g            | 0.95      | 0.92   | 0.93     | 162     |
| accuracy     |           |        | 0.91     | 246     |
| macro avg    | 0.90      | 0.91   | 0.91     | 246     |
| weighted avg | 0.92      | 0.91   | 0.92     | 246     |

Accuracy:

```
0.9146341463414634
```



In [ ]: #####

In [ ]: # IONOSPHERE DATASET

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

```

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
           '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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()

#####
# 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:  54.0s
[Parallel(n_jobs=-1)]: Done 158 tasks    | elapsed:  3.6min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  6.8min finished
```

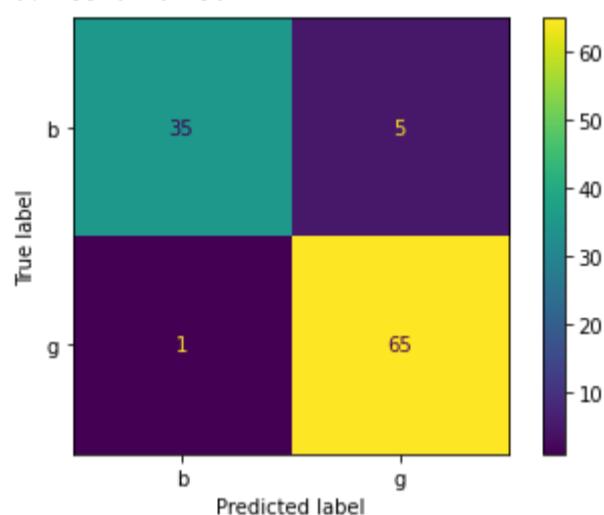
Confusion Matrix:

```
[[35  5]
 [ 1 65]]
```

| Performance Evaluation |   | precision | recall | f1-score | support |
|------------------------|---|-----------|--------|----------|---------|
| b                      | g | 0.97      | 0.88   | 0.92     | 40      |
| g                      | b | 0.93      | 0.98   | 0.96     | 66      |
|                        |   | accuracy  |        | 0.94     | 106     |
| macro avg              |   | 0.95      | 0.93   | 0.94     | 106     |
| weighted avg           |   | 0.95      | 0.94   | 0.94     | 106     |

Accuracy:

```
0.9433962264150944
```



In [ ]:

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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
           '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

df.columns = col_name

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

from sklearn.model_selection import train_test_split

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

# Feature Scaling
from sklearn.preprocessing import StandardScaler

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

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier()

#####
# Showing all the parameters

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

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

```

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

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

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

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

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

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

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

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

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

```

Parameters currently in use:

```

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

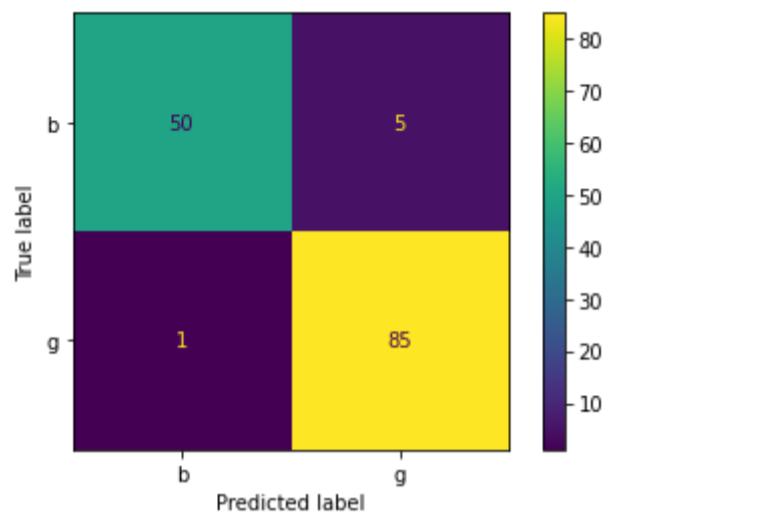
```

```

[Parallel(n_jobs=-1)]: Using backend LokyBackend with 2 concurrent workers.
[Parallel(n_jobs=-1)]: Done 37 tasks      | elapsed:  50.8s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.5min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  6.6min finished
Confusion Matrix:
[[50  5]
 [ 1 85]]
-----
```

| Performance Evaluation |           |        |          |         |
|------------------------|-----------|--------|----------|---------|
|                        | precision | recall | f1-score | support |
| b                      | 0.98      | 0.91   | 0.94     | 55      |
| g                      | 0.94      | 0.99   | 0.97     | 86      |
| accuracy               |           |        | 0.96     | 141     |
| macro avg              | 0.96      | 0.95   | 0.95     | 141     |
| weighted avg           | 0.96      | 0.96   | 0.96     | 141     |

Accuracy:  
0.9574468085106383



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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
           '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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:  49.3s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.4min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  6.4min finished

```

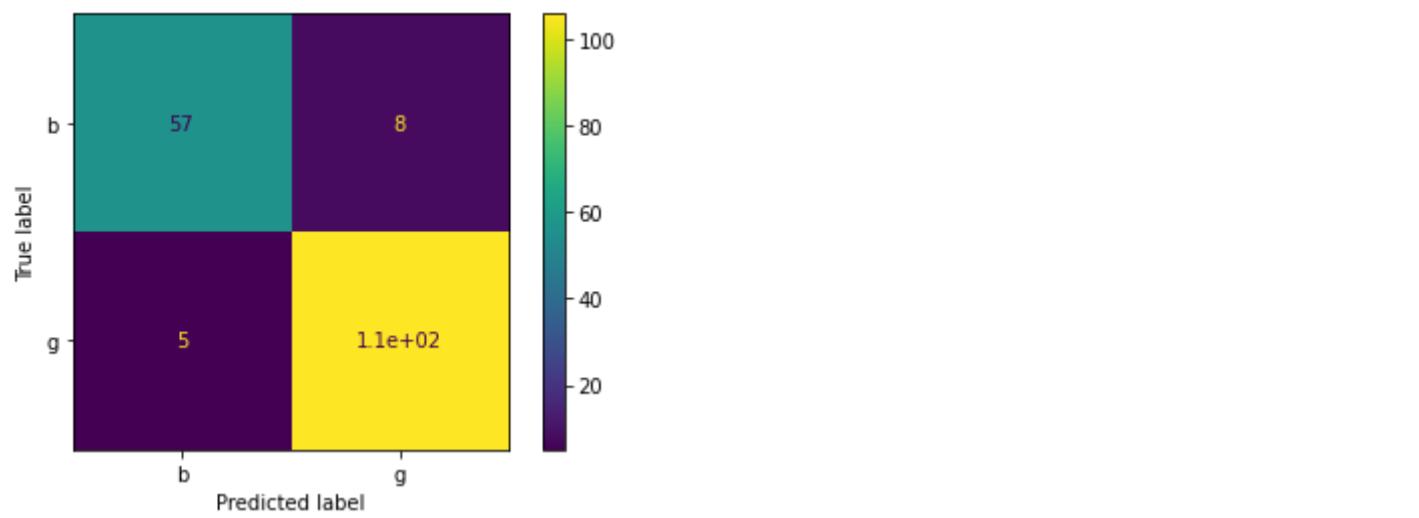
Confusion Matrix:

```

[[ 57  8]
 [ 5 106]]
-----
```

| Performance Evaluation |           |        |          |         |
|------------------------|-----------|--------|----------|---------|
|                        | precision | recall | f1-score | support |
| b                      | 0.92      | 0.88   | 0.90     | 65      |
| g                      | 0.93      | 0.95   | 0.94     | 111     |
| accuracy               |           |        | 0.93     | 176     |
| macro avg              | 0.92      | 0.92   | 0.92     | 176     |
| weighted avg           | 0.93      | 0.93   | 0.93     | 176     |

Accuracy:  
0.9261363636363636



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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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:  47.1s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.5min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  6.9min finished

```

Confusion Matrix:

```

[[ 66 10]
 [ 6 129]]
-----
```

```

Performance Evaluation
      precision    recall  f1-score   support

          b       0.92      0.87      0.89       76
          g       0.93      0.96      0.94      135

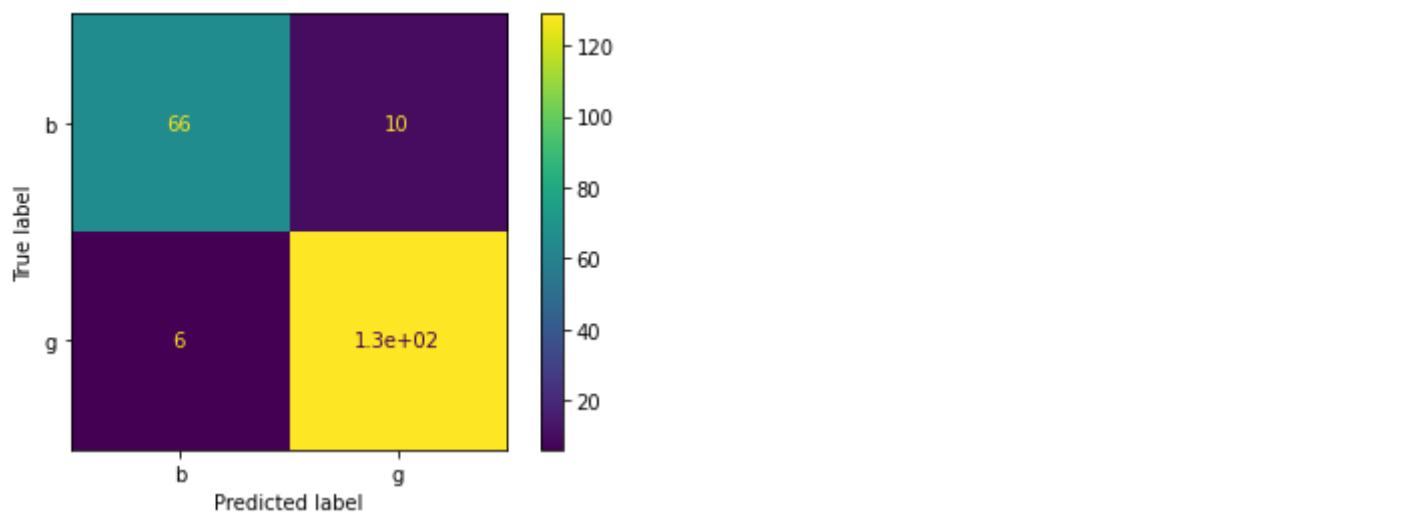
   accuracy                           0.92      211
  macro avg       0.92      0.91      0.92      211
weighted avg       0.92      0.92      0.92      211

```

```

-----
```

Accuracy:  
0.9241706161137441



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

```
import pandas as pd
import numpy as np

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

col_name = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32', '33', '34', 'Class']

df.columns = col_name

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

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

# Feature Scaling
from sklearn.preprocessing import StandardScaler

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

# Classification
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier()

#####
# Showing all the parameters

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

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

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

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

```

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

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

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

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

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

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

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

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

```

Parameters currently in use:

```

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

```

```

[Parallel(n_jobs=-1)]: Using backend LokyBackend with 2 concurrent workers.
[Parallel(n_jobs=-1)]: Done 37 tasks      | elapsed:  48.1s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  3.4min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  6.4min finished

```

Confusion Matrix:

```

[[ 75  9]
 [ 9 153]]
-----
```

```

Performance Evaluation
      precision    recall  f1-score   support

          b       0.89      0.89      0.89       84
          g       0.94      0.94      0.94      162

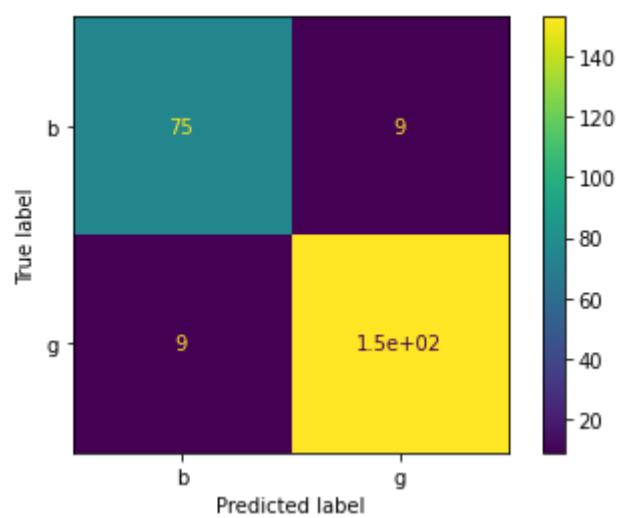
   accuracy                           0.93      246
  macro avg       0.92      0.92      0.92      246
weighted avg       0.93      0.93      0.93      246

```

```

-----
```

Accuracy:  
0.926829268292683



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

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

```
import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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 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:
```

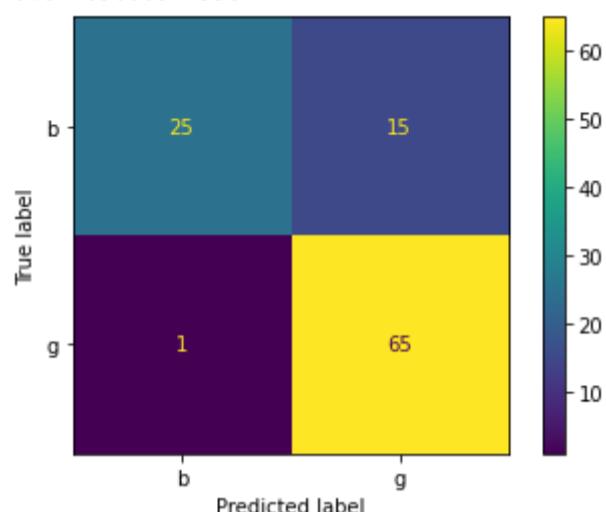
```
[[25 15]
 [ 1 65]]
```

```
-----
```

```
Performance Evaluation
```

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 0.96      | 0.62   | 0.76     | 40      |
| g            | 0.81      | 0.98   | 0.89     | 66      |
| accuracy     |           |        | 0.85     | 106     |
| macro avg    | 0.89      | 0.80   | 0.82     | 106     |
| weighted avg | 0.87      | 0.85   | 0.84     | 106     |

Accuracy:  
0.8490566037735849



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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
           '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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:
```

```
[[42 13]
 [ 1 85]]
```

```
-----
```

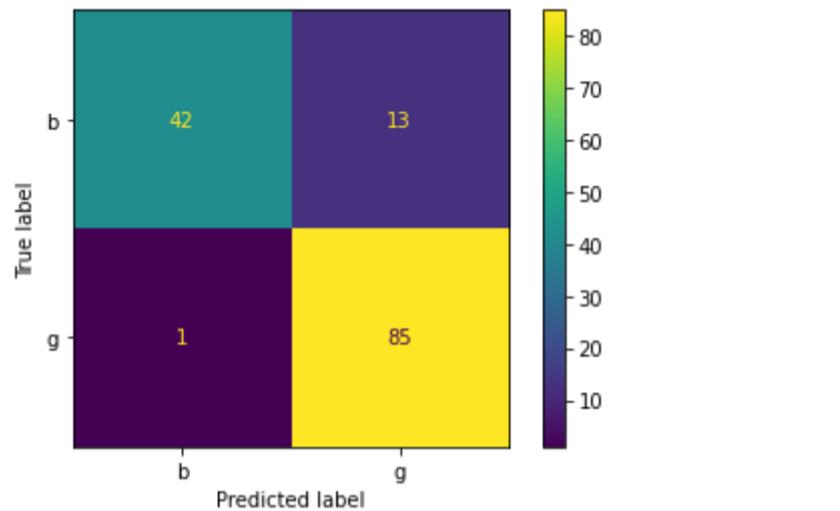
```
Performance Evaluation
```

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 0.98      | 0.76   | 0.86     | 55      |
| g            | 0.87      | 0.99   | 0.92     | 86      |
| accuracy     |           |        | 0.90     | 141     |
| macro avg    | 0.92      | 0.88   | 0.89     | 141     |
| weighted avg | 0.91      | 0.90   | 0.90     | 141     |

```
-----
```

```
Accuracy:
```

```
0.900709219858156
```



```
In [ ]:
```

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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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:
[[ 49  16]
 [  2 109]]
-----
Performance Evaluation
      precision    recall   f1-score   support
        b       0.96     0.75     0.84      65
        g       0.87     0.98     0.92     111
        accuracy           0.90      176
        macro avg       0.92     0.87     0.88      176
        weighted avg     0.90     0.90     0.89      176
-----
Accuracy:
0.8977272727272727

```

```

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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
           '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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:

```

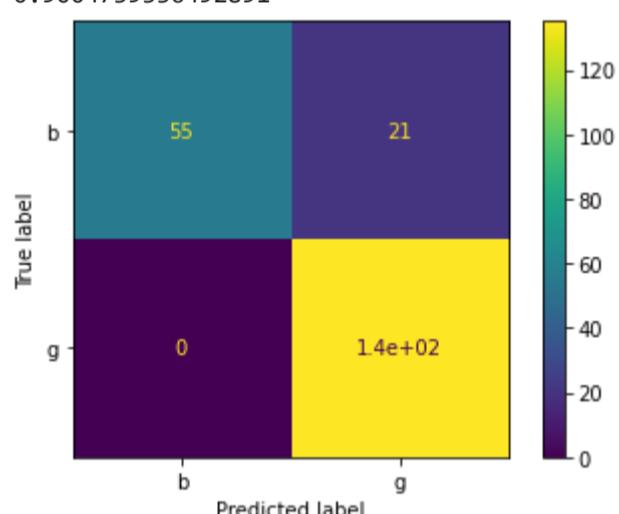
[[ 55  21]
 [  0 135]]
-----
```

Performance Evaluation

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 1.00      | 0.72   | 0.84     | 76      |
| g            | 0.87      | 1.00   | 0.93     | 135     |
| accuracy     |           |        | 0.90     | 211     |
| macro avg    | 0.93      | 0.86   | 0.88     | 211     |
| weighted avg | 0.91      | 0.90   | 0.90     | 211     |

Accuracy:

0.9004739336492891



In [ ]:

```

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

import pandas as pd
import numpy as np

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

col_name = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32', '33', '34', 'Class']

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()

```

/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:

```
[[ 63 21]
 [ 3 159]]
```

-----

-----

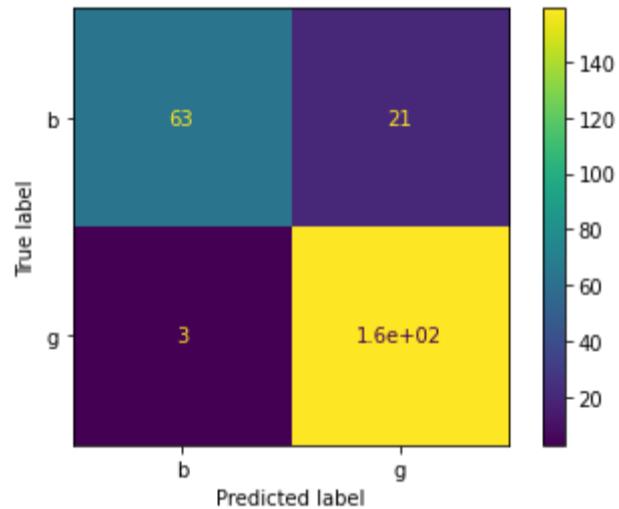
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 0.95      | 0.75   | 0.84     | 84      |
| g            | 0.88      | 0.98   | 0.93     | 162     |
| accuracy     |           |        | 0.90     | 246     |
| macro avg    | 0.92      | 0.87   | 0.88     | 246     |
| weighted avg | 0.91      | 0.90   | 0.90     | 246     |

-----

-----

Accuracy:

0.9024390243902439



In [ ]:

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

In [ ]:

```

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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
           '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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()

```

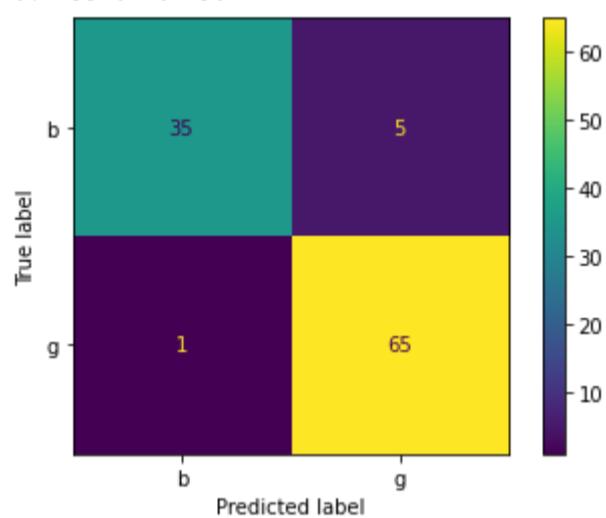
#### Parameters currently in use:

```
{'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,
 '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:
[[35  5]
 [ 1 65]]
-----
-----
Performance Evaluation
precision    recall   f1-score   support
b      0.97     0.88     0.92      40
g      0.93     0.98     0.96      66
accuracy                           0.94      106
macro avg      0.95     0.93     0.94      106
weighted avg     0.95     0.94     0.94      106
-----
-----
Accuracy:
0.9433962264150944

```



```

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

import pandas as pd
import numpy as np

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

col_name = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32', '33', '34', 'Class']

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,
'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:

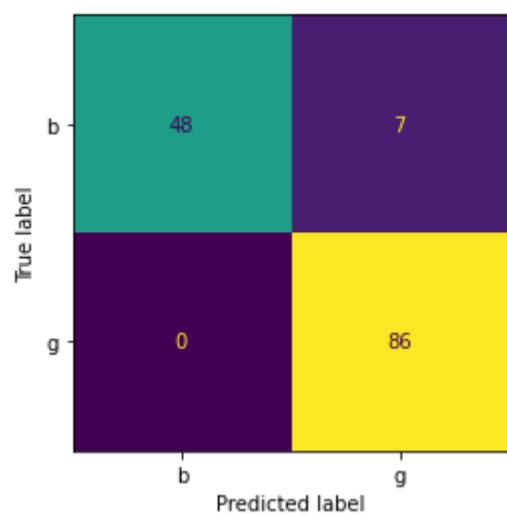
```

[[48  7]
 [ 0 86]]
-----
```

Performance Evaluation

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 1.00      | 0.87   | 0.93     | 55      |
| g            | 0.92      | 1.00   | 0.96     | 86      |
| accuracy     |           |        | 0.95     | 141     |
| macro avg    | 0.96      | 0.94   | 0.95     | 141     |
| weighted avg | 0.95      | 0.95   | 0.95     | 141     |

Accuracy:  
0.950354609929078



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

```
import pandas as pd  
import numpy as np  
  
# Dataset Preparation  
df = pd.read_csv("ionosphere.data", header=None)  
  
col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19'  
           , '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']  
  
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:

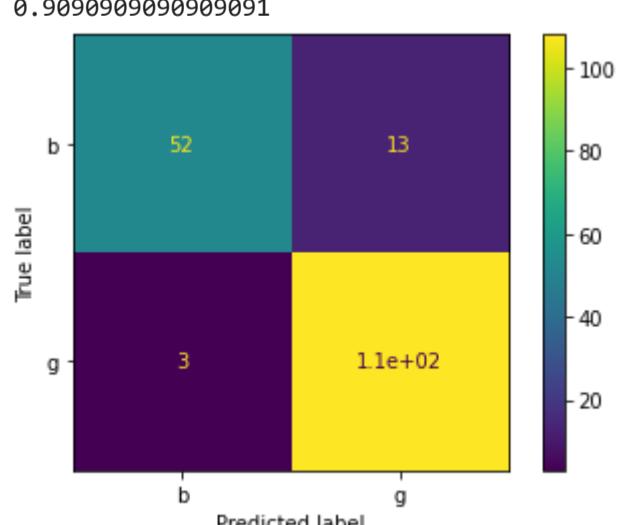
```
[[ 52 13]
 [ 3 108]]
```

```

-----
-----
Performance Evaluation
      precision    recall   f1-score   support
          b       0.95     0.80     0.87      65
          g       0.89     0.97     0.93     111

      accuracy                           0.91      176
   macro avg       0.92     0.89     0.90      176
weighted avg       0.91     0.91     0.91      176
```

Accuracy:  
0.9090909090909091



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

```

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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:

```

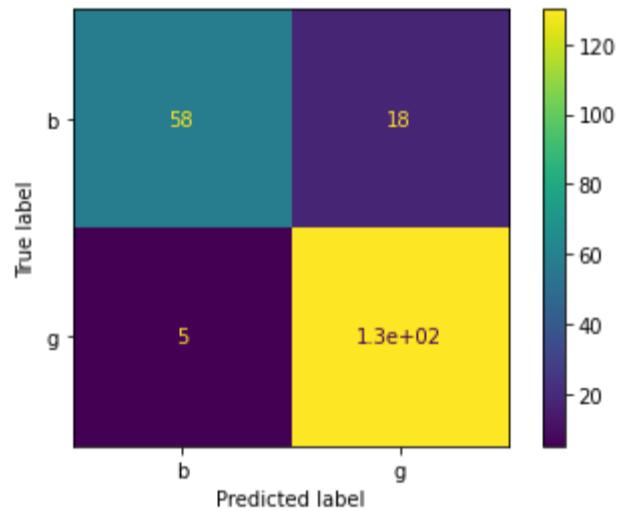
{'activation': 'relu',
 'alpha': 0.0001,
 'batch_size': 'auto',
 'beta_1': 0.9,
 'beta_2': 0.999,
 'early_stopping': False,
 'epsilon': 1e-08,
 'hidden_layer_sizes': (100,),
 'learning_rate': 'constant',
 'learning_rate_init': 0.001,
 'max_fun': 15000,
 'max_iter': 100,
 'momentum': 0.9,
 'n_iter_no_change': 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']}
Confusion Matrix:
[[ 58 18]
 [ 5 130]]
-----
```

| Performance Evaluation |   | precision    | recall       | f1-score     | support   |
|------------------------|---|--------------|--------------|--------------|-----------|
| b                      | g | 0.92<br>0.88 | 0.76<br>0.96 | 0.83<br>0.92 | 76<br>135 |
|                        |   |              |              | 0.89         | 211       |
|                        |   | macro avg    |              | 0.90         | 0.86      |
|                        |   | weighted avg |              | 0.89         | 0.89      |
|                        |   |              |              | 0.89         | 211       |

Accuracy:  
0.8909952606635071

```
/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)
```



In [ ]:

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

import pandas as pd
import numpy as np

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

col_name = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32', '33', '34', 'Class']

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:

```

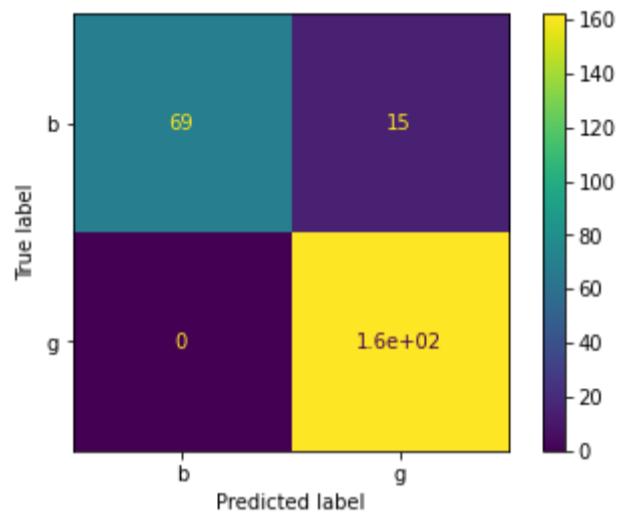
{'activation': 'relu',
 'alpha': 0.0001,
 'batch_size': 'auto',
 'beta_1': 0.9,
 'beta_2': 0.999,
 'early_stopping': False,
 'epsilon': 1e-08,
 'hidden_layer_sizes': (100,),
 'learning_rate': 'constant',
 'learning_rate_init': 0.001,
 'max_fun': 15000,
 'max_iter': 100,
 'momentum': 0.9,
 'n_iter_no_change': 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:
[[ 69 15]
 [ 0 162]]
-----
Performance Evaluation
      precision    recall   f1-score   support
      b       1.00     0.82     0.90      84
      g       0.92     1.00     0.96     162
      accuracy           0.94      246
      macro avg       0.96     0.91     0.93      246
      weighted avg     0.94     0.94     0.94      246
      
```

Accuracy:  
0.9390243902439024



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

```

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

import pandas as pd
import numpy as np

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

col_name = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32', '33', '34', 'Class']

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:

|         |
|---------|
| [34 6]  |
| [ 0 66] |

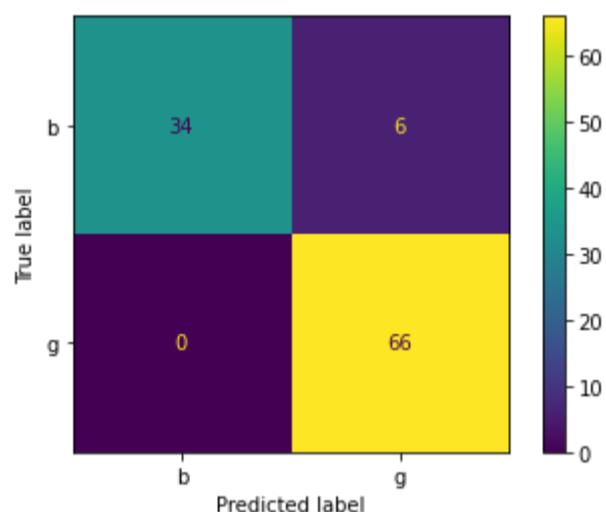
-----

Performance Evaluation

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 1.00      | 0.85   | 0.92     | 40      |
| g            | 0.92      | 1.00   | 0.96     | 66      |
| accuracy     |           |        | 0.94     | 106     |
| macro avg    | 0.96      | 0.93   | 0.94     | 106     |
| weighted avg | 0.95      | 0.94   | 0.94     | 106     |

-----

Accuracy:  
0.9433962264150944



In [ ]:

```

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

import pandas as pd
import numpy as np

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

col_name = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
           '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32', '33', '34', 'Class']

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:

```

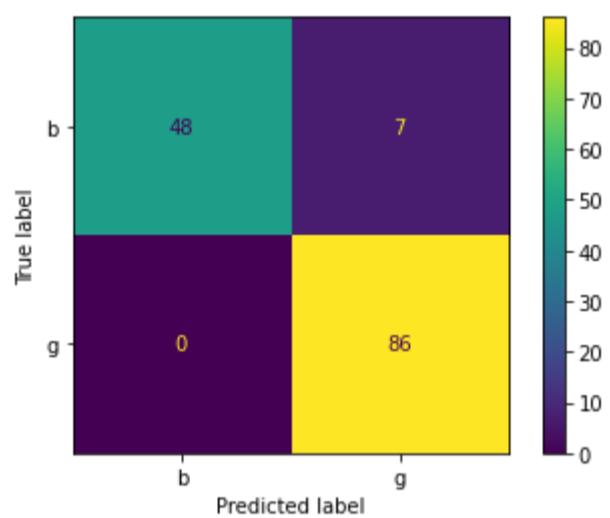
[[48  7]
 [ 0 86]]
-----
```

Performance Evaluation

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 1.00      | 0.87   | 0.93     | 55      |
| g            | 0.92      | 1.00   | 0.96     | 86      |
| accuracy     |           |        | 0.95     | 141     |
| macro avg    | 0.96      | 0.94   | 0.95     | 141     |
| weighted avg | 0.95      | 0.95   | 0.95     | 141     |

Accuracy:

```
0.950354609929078
```



In [ ]:

```

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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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:

|           |
|-----------|
| [[ 55 10] |
| [ 2 109]] |

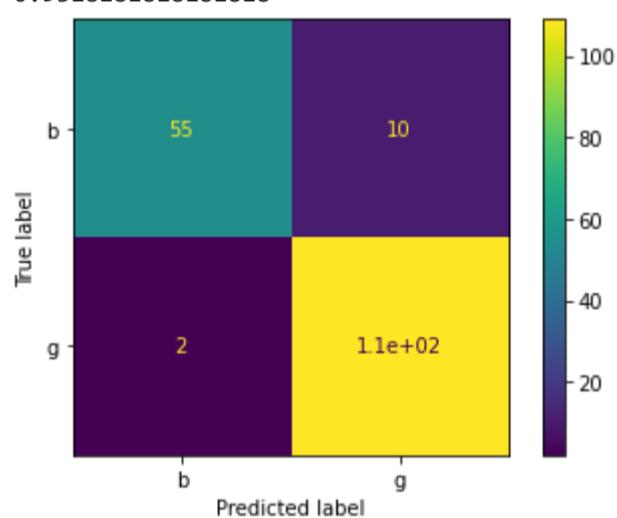
-----

Performance Evaluation

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 0.96      | 0.85   | 0.90     | 65      |
| g            | 0.92      | 0.98   | 0.95     | 111     |
| accuracy     |           |        | 0.93     | 176     |
| macro avg    | 0.94      | 0.91   | 0.92     | 176     |
| weighted avg | 0.93      | 0.93   | 0.93     | 176     |

-----

Accuracy:  
0.9318181818181818



In [ ]:

```

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

import pandas as pd
import numpy as np

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

col_name = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
           '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32', '33', '34', 'Class']

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:

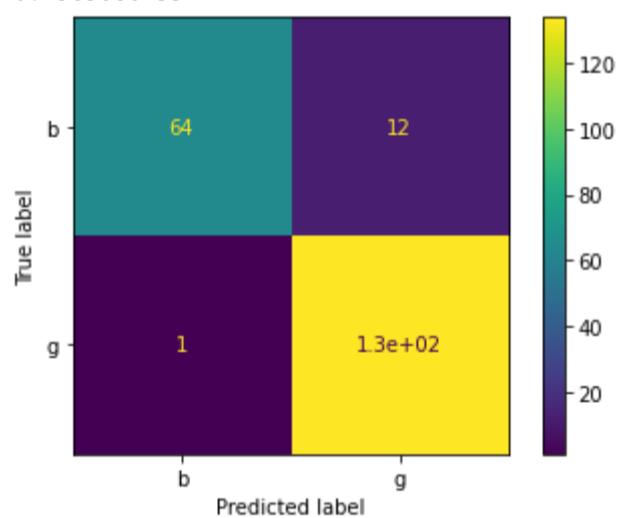
```
[[ 64 12]
 [ 1 134]]
```

Performance Evaluation

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 0.98      | 0.84   | 0.91     | 76      |
| g            | 0.92      | 0.99   | 0.95     | 135     |
| accuracy     |           |        | 0.94     | 211     |
| macro avg    | 0.95      | 0.92   | 0.93     | 211     |
| weighted avg | 0.94      | 0.94   | 0.94     | 211     |

Accuracy:

0.9383886255924171



In [ ]:

```

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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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:

```

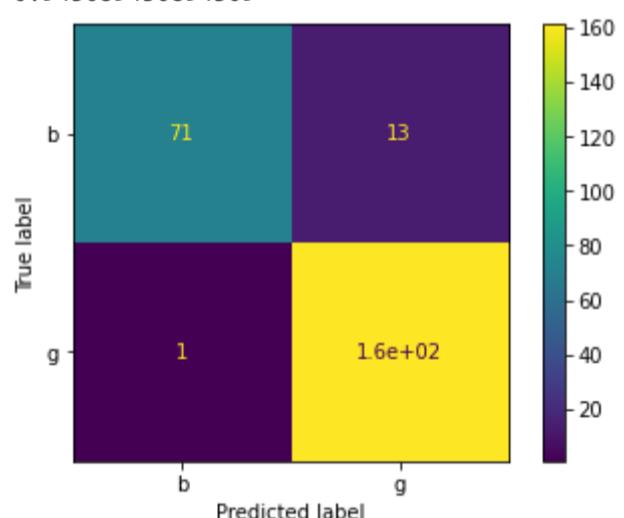
[[ 71 13]
 [ 1 161]]
-----
```

Performance Evaluation

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 0.99      | 0.85   | 0.91     | 84      |
| g            | 0.93      | 0.99   | 0.96     | 162     |
| accuracy     |           |        | 0.94     | 246     |
| macro avg    | 0.96      | 0.92   | 0.93     | 246     |
| weighted avg | 0.95      | 0.94   | 0.94     | 246     |

Accuracy:

0.943089430894309



In [ ]: #####

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

```

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
           '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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']}

```









### [CV] ..... Confusion Matrix:

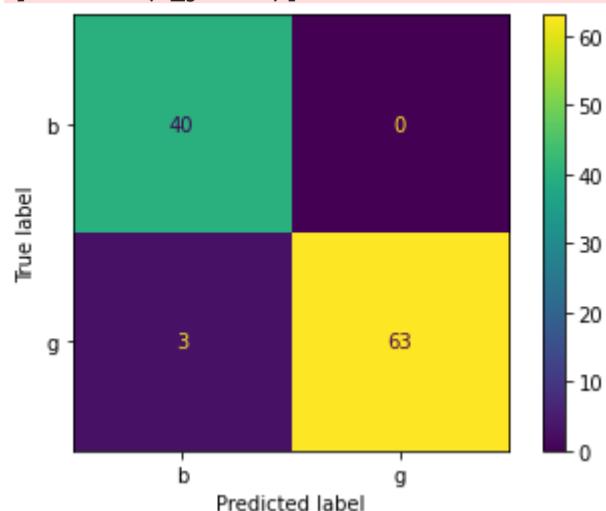
confusion  
[[40 0]  
 [ 3 63]]

Performance Evaluation

| Performance Evaluation |           |        |          |         |
|------------------------|-----------|--------|----------|---------|
|                        | precision | recall | f1-score | support |
| b                      | 0.93      | 1.00   | 0.96     | 40      |
| g                      | 1.00      | 0.95   | 0.98     | 66      |
| accuracy               |           |        | 0.97     | 106     |
| macro avg              | 0.97      | 0.98   | 0.97     | 106     |
| weighted avg           | 0.97      | 0.97   | 0.97     | 106     |

```
Accuracy:  
0.9716981132075472
```

```
[Parallel(n_jobs=1)]: Done 240 out of 240 | elapsed: 1.3s finished
```



```
In [ ]:
```

```
# IONOSPHERE DATASET  
# SVM(With Tuning)[60-40 split]  
  
import pandas as pd  
import numpy as np  
  
# Dataset Preparation  
df = pd.read_csv("ionosphere.data", header=None)  
  
col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',  
           '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']  
  
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:









```
[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=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:
```

```
[[55  0]  
 [ 6 80]]
```

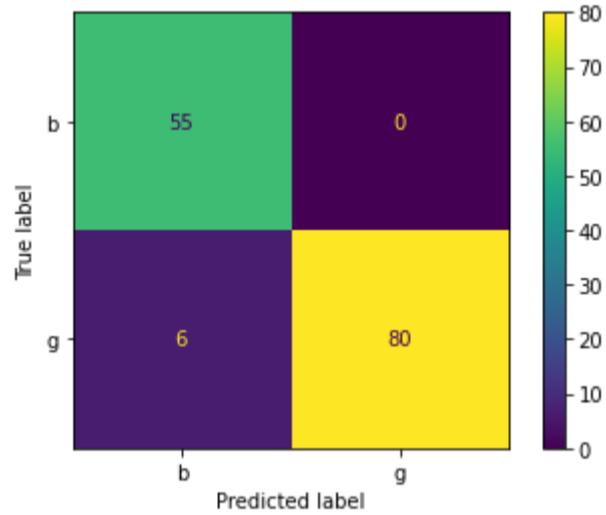
---

| Performance Evaluation |   | precision | recall | f1-score | support |
|------------------------|---|-----------|--------|----------|---------|
|                        | b | 0.90      | 1.00   | 0.95     | 55      |
|                        | g | 1.00      | 0.93   | 0.96     | 86      |
| accuracy               |   |           |        | 0.96     | 141     |
| macro avg              |   | 0.95      | 0.97   | 0.96     | 141     |
| weighted avg           |   | 0.96      | 0.96   | 0.96     | 141     |

---

Accuracy:  
0.9574468085106383

```
[Parallel(n_jobs=1)]: Done 240 out of 240 | elapsed: 1.1s finished
```



In [ ]:

```
# IONOSPHERE DATASET  

# SVM(With Tuning)[50-50 split]

import pandas as pd  

import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19'  

           , '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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:

```

{'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=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

```

```
[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
```

```
[Parallel(n_jobs=1)]: Done    1 out of    1 | elapsed:    0.0s remaining  
[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] 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=sigmoid .....  
[CV] ..... C=0.1, gamma=0.001, kernel=sigmoid, total= 0.0s
```







### Confusion Matrix:

```
[[ 63  2]
 [ 9 102]]
```

Performance Evaluation

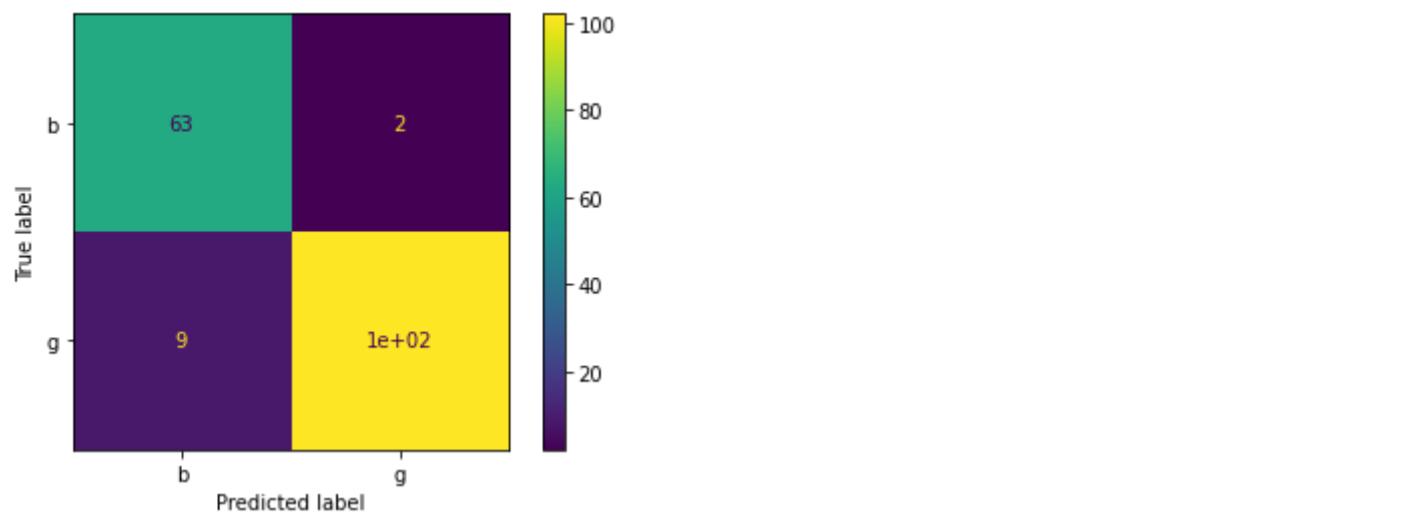
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 0.88      | 0.97   | 0.92     | 65      |
| g            | 0.98      | 0.92   | 0.95     | 111     |
| accuracy     |           |        | 0.94     | 176     |
| macro avg    | 0.93      | 0.94   | 0.93     | 176     |
| weighted avg | 0.94      | 0.94   | 0.94     | 176     |

#### Accuracy:

Accuracy  
93.75

[Page 17]

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



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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
           '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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:







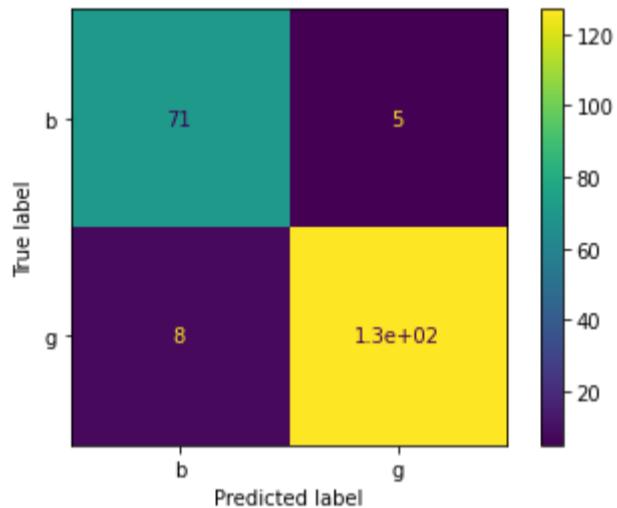


```
[CV] ..... C=100, gamma=0.001, kernel=rbf, total= 0.0s
[CV] C=100, gamma=0.001, kernel=poly .....
[CV] ..... C=100, gamma=0.001, kernel=poly, total= 0.0s
[CV] C=100, gamma=0.001, kernel=poly .....
[CV] ..... C=100, gamma=0.001, kernel=poly, total= 0.0s
[CV] C=100, gamma=0.001, kernel=poly .....
[CV] ..... 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:
[[ 71  5]
 [ 8 127]]
```

| Performance Evaluation |   | precision | recall | f1-score | support |
|------------------------|---|-----------|--------|----------|---------|
|                        | b | 0.90      | 0.93   | 0.92     | 76      |
|                        | g | 0.96      | 0.94   | 0.95     | 135     |
| accuracy               |   |           |        | 0.94     | 211     |
| macro avg              |   | 0.93      | 0.94   | 0.93     | 211     |
| weighted avg           |   | 0.94      | 0.94   | 0.94     | 211     |

Accuracy:  
0.9383886255924171

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



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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
           '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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:









### Confusion Matrix:

```
[[ 78   6]
 [ 14 148]]
```

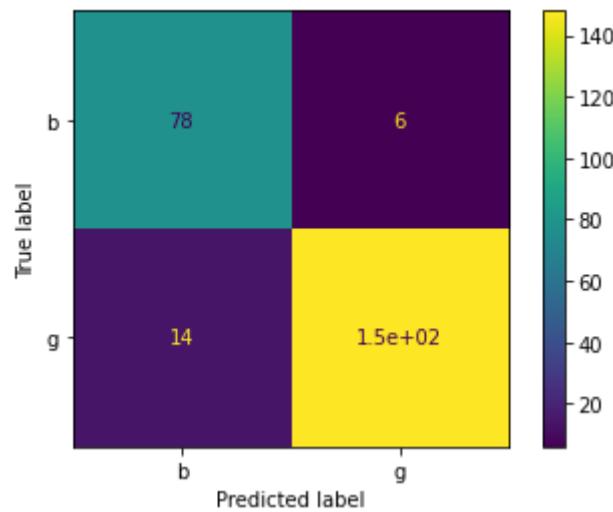
Performance Evaluation

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 0.85      | 0.93   | 0.89     | 84      |
| g            | 0.96      | 0.91   | 0.94     | 162     |
| accuracy     |           |        | 0.92     | 246     |
| macro avg    | 0.90      | 0.92   | 0.91     | 246     |
| weighted avg | 0.92      | 0.92   | 0.92     | 246     |

Accuracy:

Accuracy:

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



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

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

```
import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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=34)
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 15 important parameters

pca = PCA(n_components=15)
pca.fit(X_train)
X_train = pca.transform(X_train)
X_test = pca.transform(X_test)

# Classification
from sklearn.ensemble import RandomForestClassifier

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

y_pred = classifier.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

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

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

```

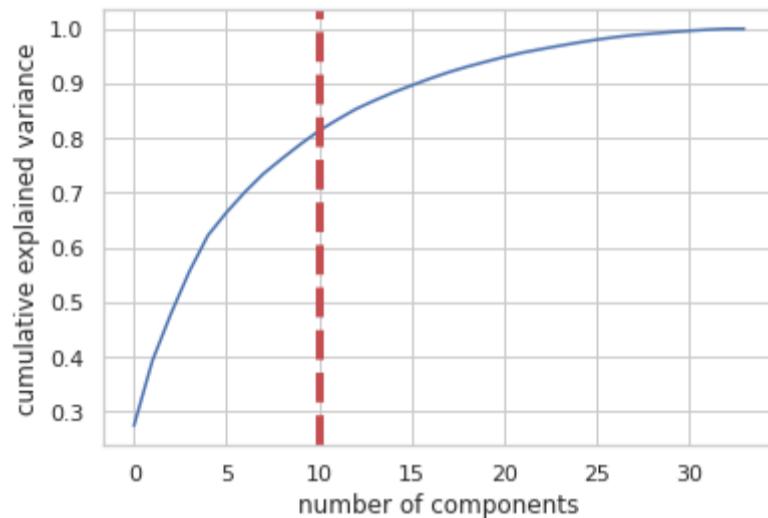
print("Performance Evaluation")
print(classification_report(y_test, y_pred))

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

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

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

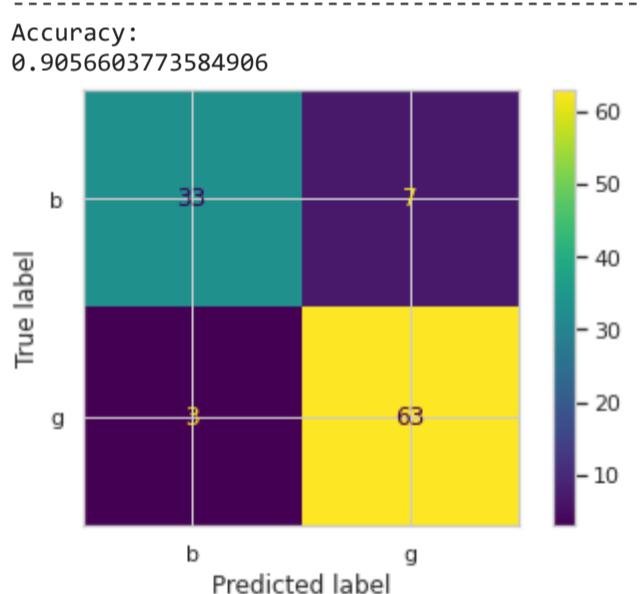
```



```

None
Confusion Matrix:
[[33  7]
 [ 3 63]]
-----
-----
Performance Evaluation
      precision    recall   f1-score   support
        b       0.92     0.82     0.87      40
        g       0.90     0.95     0.93      66
        accuracy           0.91      106
        macro avg       0.91     0.89     0.90      106
        weighted avg     0.91     0.91     0.90      106

```



```

In [ ]:
# IONOSPHERE DATASET
# Random Forest Classifier[With Tuning][70-30 split]

import pandas as pd
import numpy as np

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

col_name = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
            '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32', '33', '34', 'Class']

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=34)
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 15 important parameters

pca = PCA(n_components=15)
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")

```

```

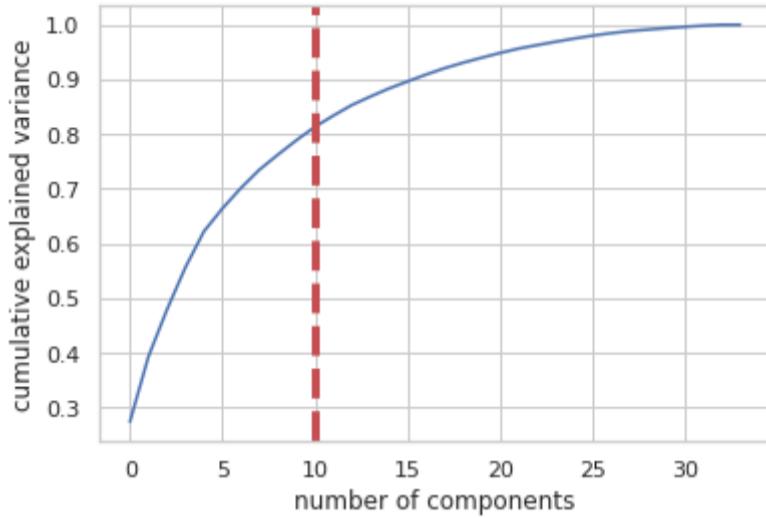
print(classification_report(y_test, y_pred))

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:   37.6s
[Parallel(n_jobs=-1)]: Done 158 tasks      | elapsed:  2.5min
[Parallel(n_jobs=-1)]: Done 300 out of 300 | elapsed:  4.8min finished

```

Confusion Matrix:

```

[[34  6]
 [ 6 60]]
-----
```

```

Performance Evaluation
      precision    recall  f1-score   support

        b       0.85     0.85     0.85      40
        g       0.91     0.91     0.91      66

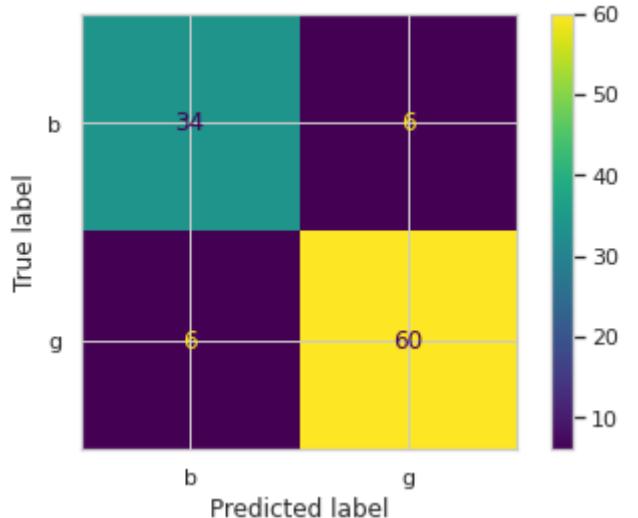
   accuracy                           0.89      106
  macro avg       0.88     0.88     0.88      106
weighted avg       0.89     0.89     0.89      106

```

```

-----
```

Accuracy:  
0.8867924528301887



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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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=34)
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=15)
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))
```

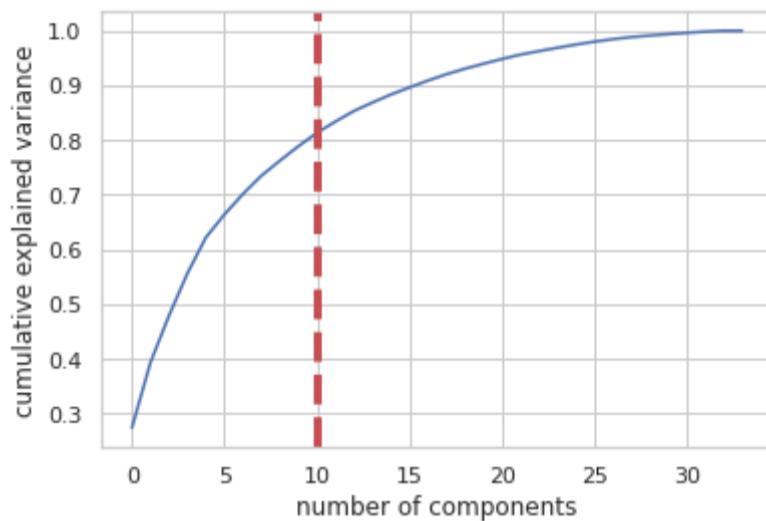
```

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()

```

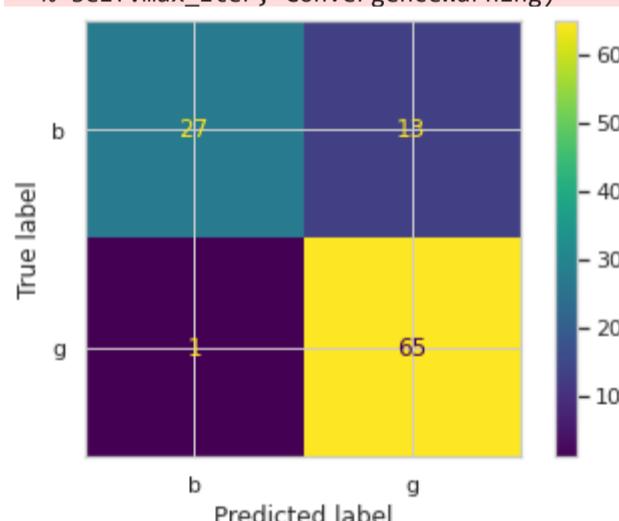


None  
Confusion Matrix:  
[[27 13]  
 [ 1 65]]

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 0.96      | 0.68   | 0.79     | 40      |
| g            | 0.83      | 0.98   | 0.90     | 66      |
| accuracy     |           |        | 0.87     | 106     |
| macro avg    | 0.90      | 0.83   | 0.85     | 106     |
| weighted avg | 0.88      | 0.87   | 0.86     | 106     |

Accuracy:  
0.8679245283018868

/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 [ ]:
# IONOSPHERE DATASET
# Multi Layer Perceptron(With Tuning)[70-30 split]

import pandas as pd
import numpy as np

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

col_name = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
           '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '30', '31', '32', '33', '34', 'Class']

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=34)
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=15)
pca.fit(X_train)
X_train = pca.transform(X_train)
X_test = pca.transform(X_test)

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

#####
# Showing all the parameters

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

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

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

#####

from sklearn.model_selection import GridSearchCV

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

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

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

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

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

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

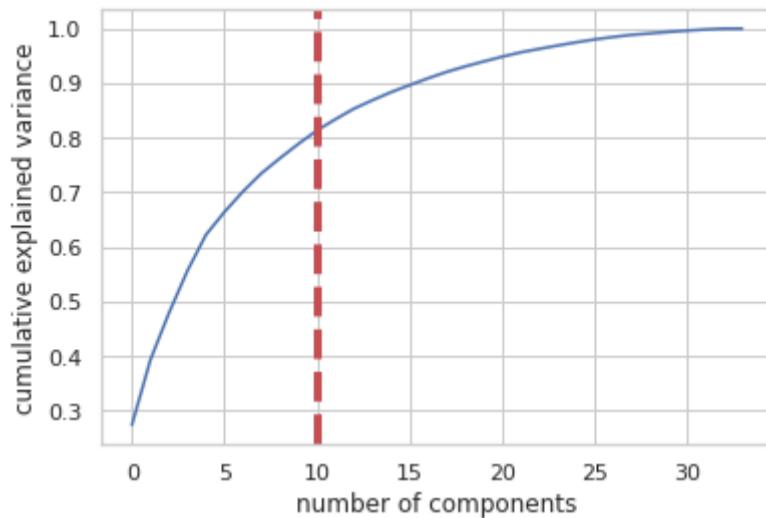
print("-----")
print("-----")

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

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

```

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



None

Parameters currently in use:

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

Confusion Matrix:

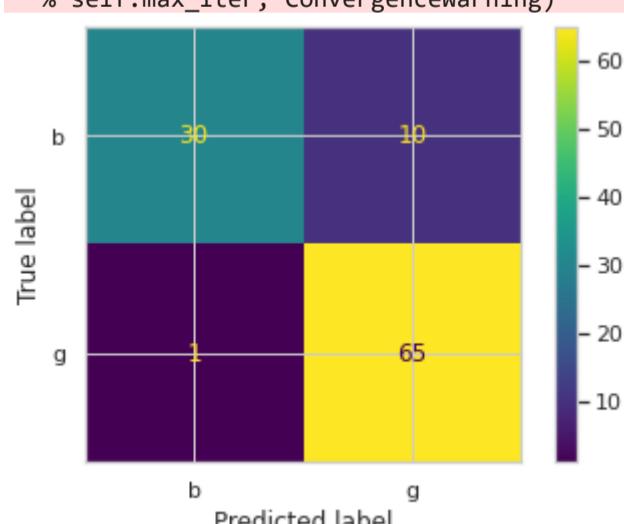
```
[[30 10]
 [ 1 65]]
```

```
Performance Evaluation
      precision    recall   f1-score   support
        b       0.97     0.75     0.85      40
        g       0.87     0.98     0.92      66
    accuracy          0.90
  macro avg       0.92     0.87     0.88      106
weighted avg       0.90     0.90     0.89      106
```

Accuracy:

```
0.8962264150943396
```

```
/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)
```



In [ ]:

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

```
import pandas as pd
```

```

import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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=34)
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=15)
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("-----")

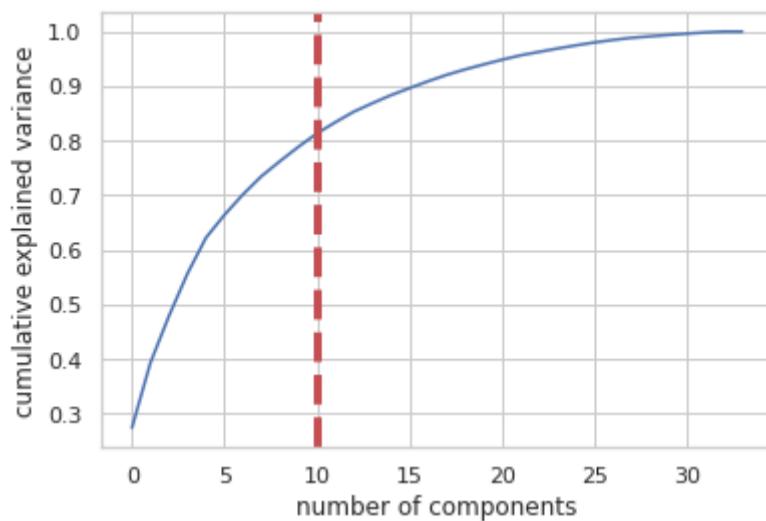
print("Performance Evaluation")
print(classification_report(y_test, y_pred))

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()

```

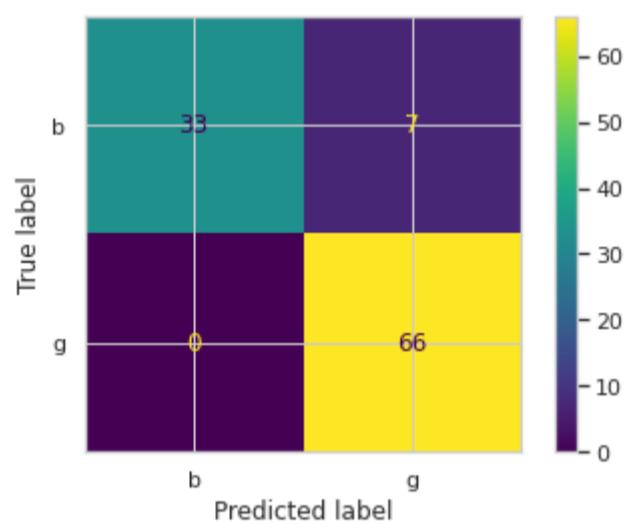


None  
Confusion Matrix:  
[[33 7]  
 [ 0 66]]

Performance Evaluation

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| b            | 1.00      | 0.82   | 0.90     | 40      |
| g            | 0.90      | 1.00   | 0.95     | 66      |
| accuracy     |           |        | 0.93     | 106     |
| macro avg    | 0.95      | 0.91   | 0.93     | 106     |
| weighted avg | 0.94      | 0.93   | 0.93     | 106     |

Accuracy:  
0.9339622641509434



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

import pandas as pd
import numpy as np

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

col_name = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19',
            '20','21','22','23','24','25','26','27','28','29','30','31','32','33','34','Class']

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=34)
```

```

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=15)
pca.fit(X_train)
X_train = pca.transform(X_train)
X_test = pca.transform(X_test)

# Classification
from sklearn.svm import SVC

classifier = SVC()

#####
# Showing all the parameters

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

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

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

#####

from sklearn.model_selection import GridSearchCV

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

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

y_pred = rf_random.predict(X_test)

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

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

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

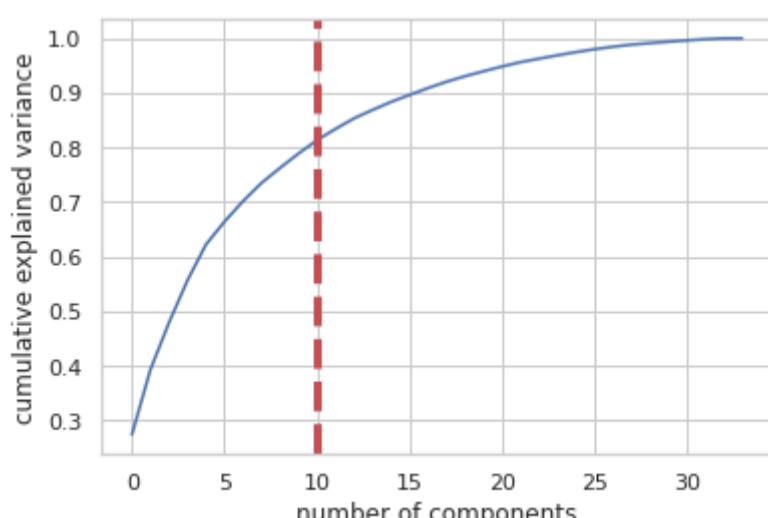
print("Performance Evaluation")
print(classification_report(y_test, y_pred))

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

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

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

```



None

Parameters currently in use:







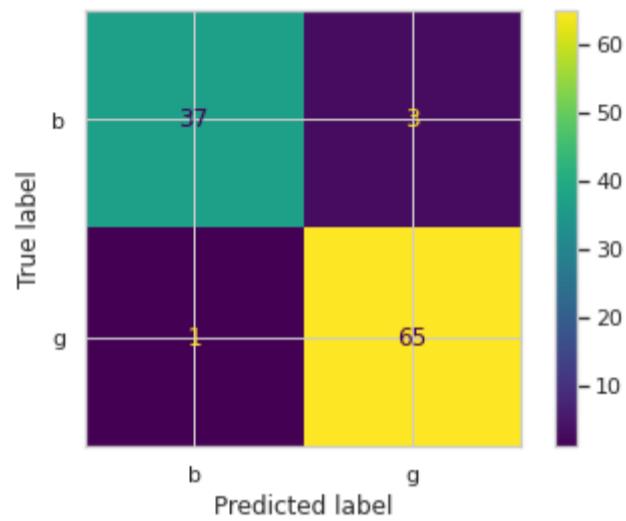


```
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
[CV] C=100, gamma=0.001, kernel=sigmoid .....
[CV] ..... C=100, gamma=0.001, kernel=sigmoid, total= 0.0s
Confusion Matrix:
[[37  3]
 [ 1 65]]
```

```
-----  
Performance Evaluation  
precision    recall   f1-score   support  
  
      b       0.97     0.93     0.95      40  
      g       0.96     0.98     0.97      66  
  
accuracy                           0.96      106  
macro avg       0.96     0.95     0.96      106  
weighted avg     0.96     0.96     0.96      106
```

```
-----  
Accuracy:  
0.9622641509433962
```

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



```
In [ ]:
```