In [1]:
```
#importing libraries
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
symptoms=pd.read_csv("D:\Detecting parkinsons disease\cleaned-data.csv")
symptoms.head()
```

Out[2]:

	Unnamed: 0	name	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DI
0	0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.000070	0.00370	0.00554	0.011
1	1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.000080	0.00465	0.00576	0.013
2	2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.000090	0.00544	0.00576	0.016
3	3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.000090	0.00502	0.00576	0.015
4	4	phon_R01_S01_5	116.014	141.781	110.655	0.01101	0.000037	0.00593	0.00576	0.017

5 rows × 25 columns

In [3]:
```
symptoms.columns
```

Out[3]:
```
Index(['Unnamed: 0', 'name', 'MDVP:Fo(Hz)', 'MDVP:Fhi(Hz)', 'MDVP:Flo(Hz)',
       'MDVP:Jitter(%)', 'MDVP:Jitter(Abs)', 'MDVP:RAP', 'MDVP:PPQ',
       'Jitter:DDP', 'MDVP:Shimmer', 'MDVP:Shimmer(dB)', 'Shimmer:APQ3',
       'Shimmer:APQ5', 'MDVP:APQ', 'Shimmer:DDA', 'NHR', 'HNR', 'status',
       'RPDE', 'DFA', 'spread1', 'spread2', 'D2', 'PPE'],
      dtype='object')
```

In [4]:
```
#independent variables
x=symptoms.drop(['status','name'],axis=1)

#dependent variable
y=symptoms['status']
```

In [5]:
```
#splitting the data set into training and test set
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25,random_state=42)
```

In [6]:
```
model=GaussianNB()
model.fit(x_train,y_train)
```

Out[6]:
```
GaussianNB()
```

In [7]:
```
#prediction
y_pred=model.predict(x_test)
pd.DataFrame({'actual status':y_test,"predicted status:":y_pred})
```

Out[7]:

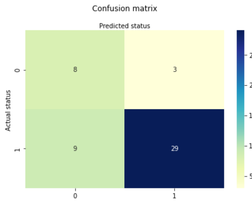
	actual status	predicted status:
138	1	1
16	1	0
155	1	1
96	1	1
68	1	1
153	1	1
55	1	1
15	1	0
112	1	1
111	1	0
184	0	0
18	1	1
82	1	1
9	1	1
164	1	1
117	1	1
69	1	1
113	1	0
192	0	1
119	1	0
123	1	1
144	1	0
66	1	1
45	0	0
158	1	1
115	1	1
67	1	1
93	1	1
30	0	0
101	1	1
118	1	0
75	1	1
24	1	0
172	0	0
127	1	1
169	0	0
19	1	1
168	0	1
73	1	0
5	1	1
135	1	1
122	1	1
167	0	0
85	1	1
56	1	1
95	1	1
35	0	0
190	0	1
42	0	0

In [8]:
```
#Model Evaluation
```

In [9]:
```
#Confusion matrix
from sklearn.metrics import confusion_matrix
conf_matrix=confusion_matrix(y_test,y_pred)
conf_matrix
```

Out[9]:
```
array([[ 8,  3],
       [ 9, 29]], dtype=int64)
```

In [10]:
```
#Heatmap for confusionmatrix
class_names=[0,1] # name  of classes
fig, ax = plt.subplots()
tick_marks = np.arange(len(class_names))
plt.xticks(tick_marks, class_names)
plt.yticks(tick_marks, class_names)
# create heatmap
sns.heatmap(pd.DataFrame(conf_matrix), annot=True, cmap="YlGnBu" ,fmt='g')
ax.xaxis.set_label_position("top")
plt.tight_layout()
plt.title('Confusion matrix', y=1.1)
plt.ylabel('Actual status')
plt.xlabel('Predicted status')
plt.savefig("HeatMap")
```



In [11]:
```
#Classification Accuracy
from sklearn import metrics
accuracy=accuracy_score(y_test,y_pred)*100
print("Accuracy_score is :",accuracy)
print("Precision:",metrics.precision_score(y_test, y_pred)*100)
print("Recall:",metrics.recall_score(y_test, y_pred)*100)
```

```
Accuracy_score is : 75.51020408163265
Precision: 90.625
Recall: 76.31578947368422
```

In [12]:
```
#LogLoss
from sklearn.metrics import log_loss
logLoss=log_loss(y_test,y_pred)
print("LogLoss: %.2f" % (logLoss))
```

```
Logloss: 8.46
```

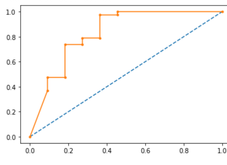
In [13]:
```
from sklearn.metrics import roc_auc_score, roc_curve

# predict probabilities
probs = model.predict_proba(x_test)
# keep probabilities for the positive outcome only
probs = probs[:, 1]

auc = roc_auc_score(y_test, probs)
print('AUC - Test Set: %.2f%%' % (auc*100))

# calculate roc curve
fpr, tpr, thresholds = roc_curve(y_test, probs)
# plot no skill
plt.plot([0, 1], [0, 1], linestyle='--')
# plot the roc curve for the model
plt.plot(fpr, tpr, marker='.')
# show the plot
plt.show()
```

```
AUC - Test Set: 83.25%
```



In [14]:
```
#F score
from sklearn.metrics import f1_score
f1 = f1_score(y_test, y_pred)
print('F1 score: %f' % f1)
```

```
F1 score: 0.828571
```

In [ ]:
```
```