Давиташвили Шако ИУ5И-65Б Вариант 18

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

from sklearn.metrics import roc_curve, auc, f1_score
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.svm import SVC
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder

df = pd.read_csv('FIFA 2018 Statistics.csv')

df.head()
```

```
         Date          Team      Opponent  Goal Scored  Ball
Possession %  \
0  14-06-2018        Russia  Saudi Arabia            5
40
1  14-06-2018  Saudi Arabia        Russia            0
60
2  15-06-2018         Egypt       Uruguay            0
43
3  15-06-2018       Uruguay         Egypt            1
57
4  15-06-2018       Morocco          Iran            0
64

   Attempts  On-Target  Off-Target  Blocked  Corners  ...  Yellow Card
\
0        13          7           3        3        6  ...            0

1         6          0           3        3        2  ...            0

2         8          3           3        2        0  ...            2

3        14          4           6        4        5  ...            0

4        13          3           6        4        5  ...            1


   Yellow & Red  Red  Man of the Match  1st Goal        Round  PSO  \
0             0    0               Yes      12.0  Group Stage   No
1             0    0                No       NaN  Group Stage   No
2             0    0                No       NaN  Group Stage   No
3             0    0               Yes      89.0  Group Stage   No
4             0    0                No       NaN  Group Stage   No
```

```
   Goals in PSO  Own goals  Own goal Time
0             0        NaN            NaN
1             0        NaN            NaN
2             0        NaN            NaN
3             0        NaN            NaN
4             0        1.0           90.0

[5 rows x 27 columns]
```

```
df.shape
```

```
(128, 27)
```

```
df.isna().sum()
```

```
Date                        0
Team                        0
Opponent                    0
Goal Scored                 0
Ball Possession %           0
Attempts                    0
On-Target                   0
Off-Target                  0
Blocked                     0
Corners                     0
Offsides                    0
Free Kicks                  0
Saves                       0
Pass Accuracy %             0
Passes                      0
Distance Covered (Kms)      0
Fouls Committed             0
Yellow Card                 0
Yellow & Red                0
Red                         0
Man of the Match            0
1st Goal                   34
Round                       0
PSO                         0
Goals in PSO                0
Own goals                 116
Own goal Time             116
dtype: int64
```

```
df['Own goals'] = df['Own goals'].fillna(0.0)
df['1st Goal'] = df['1st Goal'].fillna(0.0)
df['Own goal Time'] = df['Own goal Time'].fillna(0.0)
```

```
df.isna().sum()
```

```
Date                      0
Team                      0
```

```
Opponent                  0
Goal Scored               0
Ball Possession %         0
Attempts                  0
On-Target                 0
Off-Target                0
Blocked                   0
Corners                   0
Offsides                  0
Free Kicks                0
Saves                     0
Pass Accuracy %           0
Passes                    0
Distance Covered (Kms)    0
Fouls Committed           0
Yellow Card               0
Yellow & Red              0
Red                       0
Man of the Match          0
1st Goal                  0
Round                     0
PSO                       0
Goals in PSO              0
Own goals                 0
Own goal Time             0
dtype: int64
```

```
le = LabelEncoder()
df['enc_team'] = le.fit_transform(df[['Team']])
df['enc_opponent'] = le.fit_transform(df[['Opponent']])
```

```
d:\Projects\ML\RK2\.venv\Lib\site-packages\sklearn\preprocessing\
_label.py:116: DataConversionWarning: A column-vector y was passed
when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
d:\Projects\ML\RK2\.venv\Lib\site-packages\sklearn\preprocessing\
_label.py:116: DataConversionWarning: A column-vector y was passed
when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

```
dum = pd.get_dummies(df[['Round']])

for col in dum:
    df[col] = dum[col]

df.head()
```

```
         Date          Team      Opponent  Goal Scored  Ball
Possession %  \
```

```
0  14-06-2018        Russia  Saudi Arabia            5
40
1  14-06-2018  Saudi Arabia        Russia            0
60
2  15-06-2018         Egypt       Uruguay            0
43
3  15-06-2018       Uruguay         Egypt            1
57
4  15-06-2018       Morocco          Iran            0
64

   Attempts  On-Target  Off-Target  Blocked  Corners  ...  Own
goals  \
0        13          7           3        3        6  ...        0.0

1         6          0           3        3        2  ...        0.0

2         8          3           3        2        0  ...        0.0

3        14          4           6        4        5  ...        0.0

4        13          3           6        4        5  ...        1.0


   Own goal Time  enc_team  enc_opponent  Round_3rd Place  Round_Final
\
0            0.0        23            24            False        False

1            0.0        24            23            False        False

2            0.0         8            31            False        False

3            0.0        31             8            False        False

4           90.0        17            13            False        False


   Round_Group Stage  Round_Quarter Finals  Round_Round of 16  \
0               True                 False              False
1               True                 False              False
2               True                 False              False
3               True                 False              False
4               True                 False              False

   Round_Semi- Finals
0               False
1               False
2               False
3               False
```

```
4               False

[5 rows x 35 columns]
```

```
df['Man of the Match'] = df['Man of the Match'].apply(lambda x: True
if x == 'Yes' else False)
```

```
df.head()
```

```
         Date          Team      Opponent  Goal Scored  Ball
Possession %  \
0  14-06-2018        Russia  Saudi Arabia            5
40
1  14-06-2018  Saudi Arabia        Russia            0
60
2  15-06-2018         Egypt       Uruguay            0
43
3  15-06-2018       Uruguay         Egypt            1
57
4  15-06-2018       Morocco          Iran            0
64

   Attempts  On-Target  Off-Target  Blocked  Corners  ...  Own
goals  \
0        13          7           3        3        6  ...        0.0

1         6          0           3        3        2  ...        0.0

2         8          3           3        2        0  ...        0.0

3        14          4           6        4        5  ...        0.0

4        13          3           6        4        5  ...        1.0


   Own goal Time  enc_team  enc_opponent  Round_3rd Place  Round_Final
\
0            0.0        23            24            False        False

1            0.0        24            23            False        False

2            0.0         8            31            False        False

3            0.0        31             8            False        False

4           90.0        17            13            False        False


   Round_Group Stage  Round_Quarter Finals  Round_Round of 16  \
```

```
0                True                  False              False
1                True                  False              False
2                True                  False              False
3                True                  False              False
4                True                  False              False

   Round_Semi- Finals
0                False
1                False
2                False
3                False
4                False

[5 rows x 35 columns]
```

```
df = df.drop(columns=['Date', 'Team', 'Opponent', 'Round', 'PSO'])
```

```
X_train, X_test, y_train, y_test =
train_test_split(df.drop(columns=['Man of the Match']), df[['Man of
the Match']], test_size=0.2)
```

```
svc = SVC()
svc.fit(X_train, y_train)
```

```
d:\Projects\ML\RK2\.venv\Lib\site-packages\sklearn\utils\
validation.py:1143: DataConversionWarning: A column-vector y was
passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

SVC()
```

```
svc_predicted = svc.predict(X_test)
```

```
grd = GradientBoostingClassifier()
grd.fit(X_train, y_train)
```

```
d:\Projects\ML\RK2\.venv\Lib\site-packages\sklearn\ensemble\
_gb.py:437: DataConversionWarning: A column-vector y was passed when a
1d array was expected. Please change the shape of y to (n_samples, ),
for example using ravel().
  y = column_or_1d(y, warn=True)

GradientBoostingClassifier()
```

```
grd_predicted = grd.predict(X_test)
```

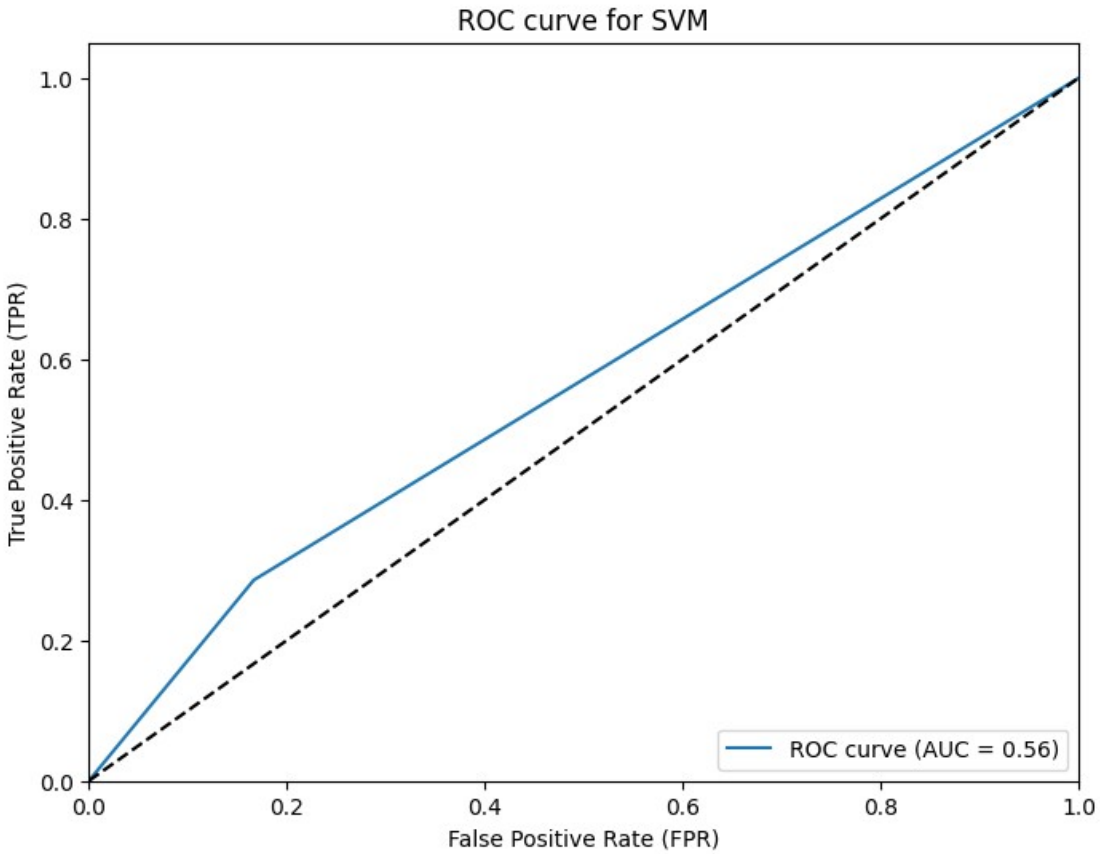
```
print('SVM')
f1_score(y_true=y_test, y_pred=svc_predicted)
```

```
SVM

0.4
```

```
print('Gradient Boosting')
f1_score(y_true=y_test, y_pred=grd_predicted)
```

```
Gradient Boosting

0.7200000000000001
```

```
def draw_roc(y_true, y_score, name):
    fpr, tpr, thresholds = roc_curve(y_true=y_true, y_score=y_score)

    roc_auc = auc(fpr, tpr)

    plt.figure(figsize=(8, 6))
    plt.plot(fpr, tpr, label='ROC curve (AUC = %0.2f)' % roc_auc)
    plt.plot([0, 1], [0, 1], 'k--')
    plt.xlim([0.0, 1.0])
    plt.ylim([0.0, 1.05])
    plt.xlabel('False Positive Rate (FPR)')
    plt.ylabel('True Positive Rate (TPR)')
    plt.title(f'ROC curve for {name}')
    plt.legend(loc="lower right")
    plt.show()

draw_roc(y_true=y_test, y_score=svc_predicted, name='SVM')
```



```
draw_roc(y_true=y_test, y_score=grd_predicted, name='Gradient
Boosting')
```