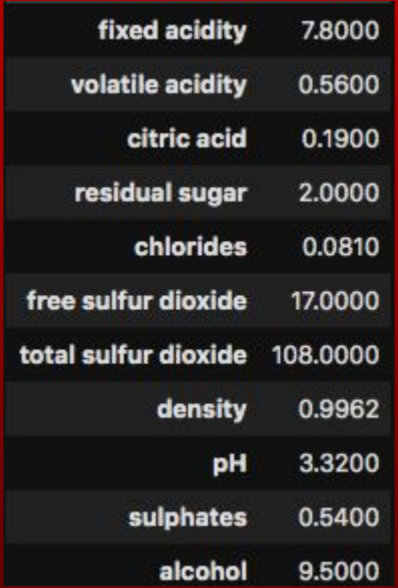
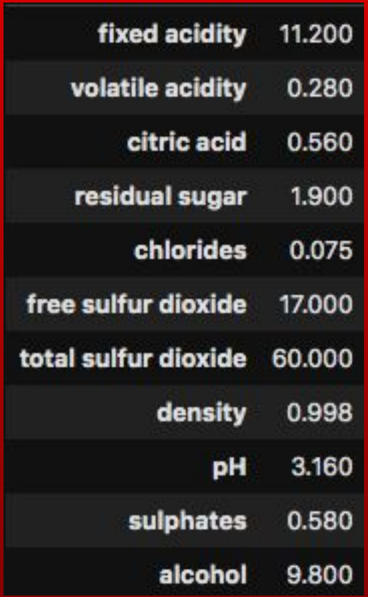
VINHO TINTO BOM

Kiryl Pashkouski
ds_042219

WINE 1



fixed acidity	7.8000
volatile acidity	0.5600
citric acid	0.1900
residual sugar	2.0000
chlorides	0.0810
free sulfur dioxide	17.0000
total sulfur dioxide	108.0000
density	0.9962
pH	3.3200
sulphates	0.5400
alcohol	9.5000



fixed acidity	11.200
volatile acidity	0.280
citric acid	0.560
residual sugar	1.900
chlorides	0.075
free sulfur dioxide	17.000
total sulfur dioxide	60.000
density	0.998
pH	3.160
sulphates	0.580
alcohol	9.800

WINE 2

WHAT IF YOU HAVE AN APP

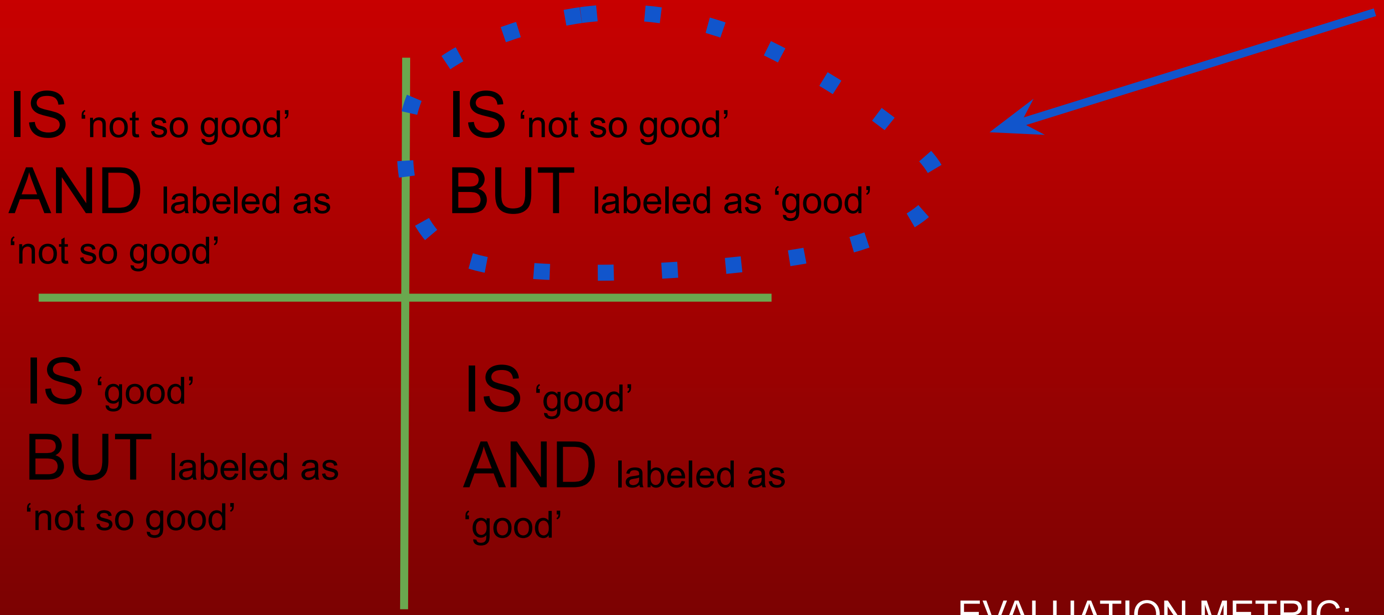
WHICH PREDICTS A QUALITY

OF A GIVEN WINE:

3 times

out of 4 attempts

!!!CORRECTLY!!!



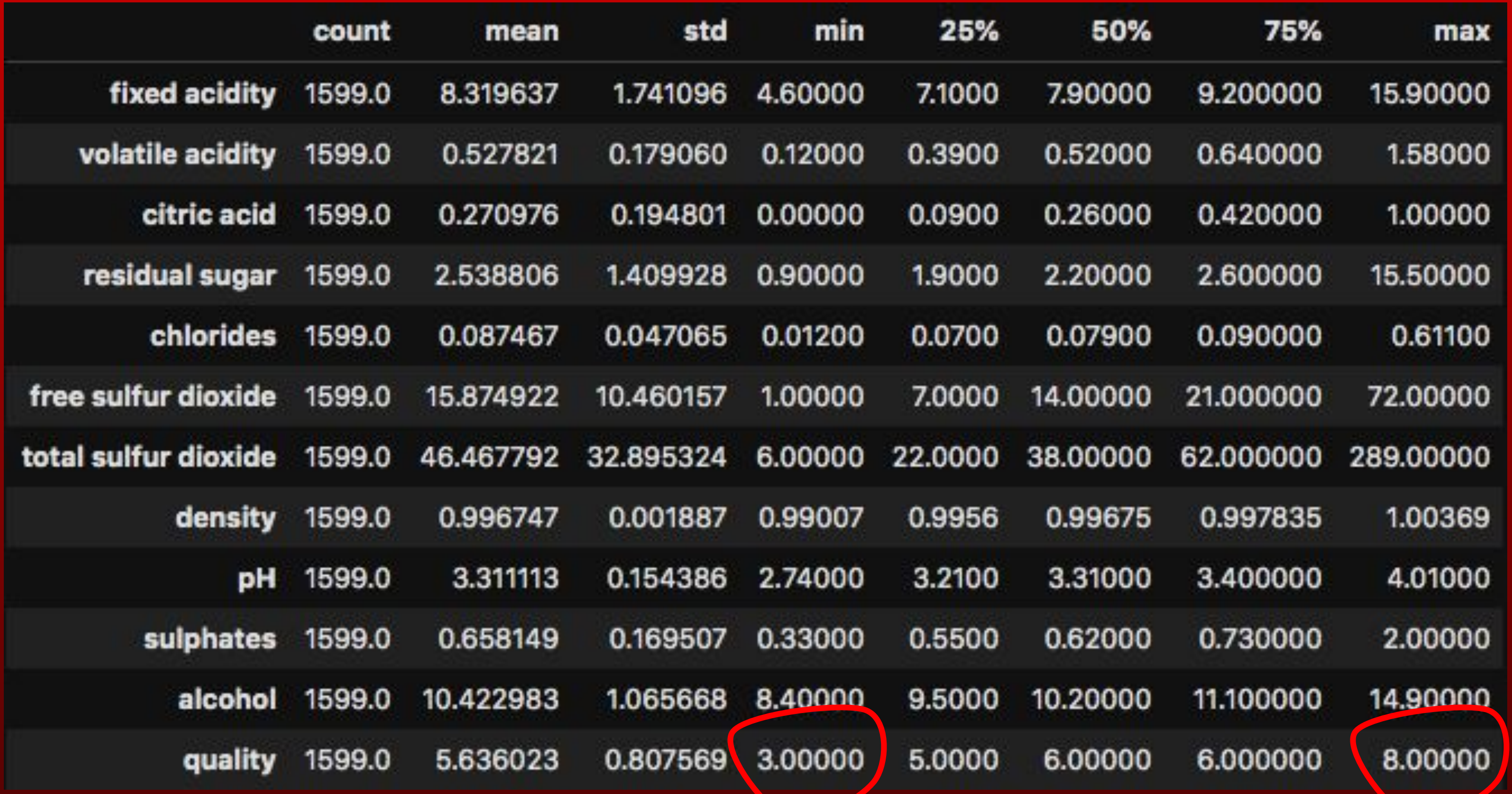
IS 'not so good' AND labeled as 'not so good'	IS 'not so good' BUT labeled as 'good'
IS 'good' BUT labeled as 'not so good'	IS 'good' AND labeled as 'good'

MINIMIZE THIS ERROR

EVALUATION METRIC:

PRECISION SCORE → MAXIMIZE

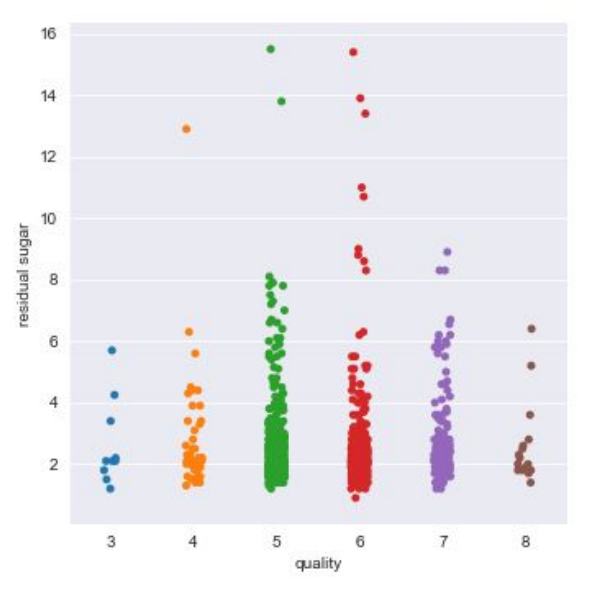
DESCRIPTION OF DATA SET

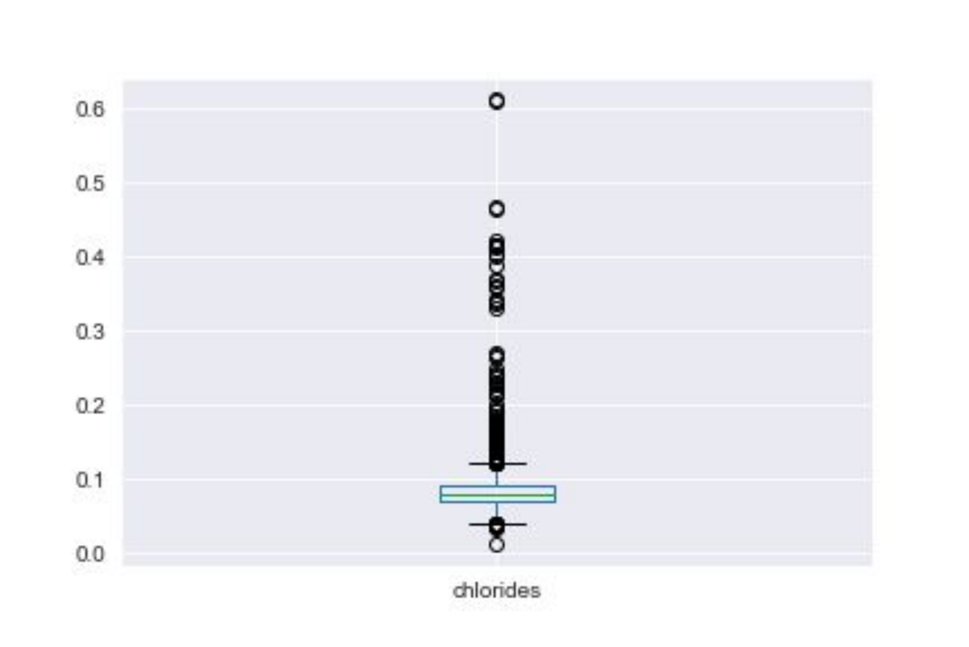


	count	mean	std	min	25%	50%	75%	max
fixed acidity	1599.0	8.319637	1.741096	4.60000	7.1000	7.90000	9.200000	15.90000
volatile acidity	1599.0	0.527821	0.179060	0.12000	0.3900	0.52000	0.640000	1.58000
citric acid	1599.0	0.270976	0.194801	0.00000	0.0900	0.26000	0.420000	1.00000
residual sugar	1599.0	2.538806	1.409928	0.90000	1.9000	2.20000	2.600000	15.50000
chlorides	1599.0	0.087467	0.047065	0.01200	0.0700	0.07900	0.090000	0.61100
free sulfur dioxide	1599.0	15.874922	10.460157	1.00000	7.0000	14.00000	21.000000	72.00000
total sulfur dioxide	1599.0	46.467792	32.895324	6.00000	22.0000	38.00000	62.000000	289.00000
density	1599.0	0.996747	0.001887	0.99007	0.9956	0.99675	0.997835	1.00369
pH	1599.0	3.311113	0.154386	2.74000	3.2100	3.31000	3.400000	4.01000
sulphates	1599.0	0.658149	0.169507	0.33000	0.5500	0.62000	0.730000	2.00000
alcohol	1599.0	10.422983	1.065668	8.40000	9.5000	10.20000	11.100000	14.90000
quality	1599.0	5.636023	0.807569	3.00000	5.0000	6.00000	6.000000	8.00000

CLEANING

- Duplicates removed
- Outliers in residual sugar, chlorides removed

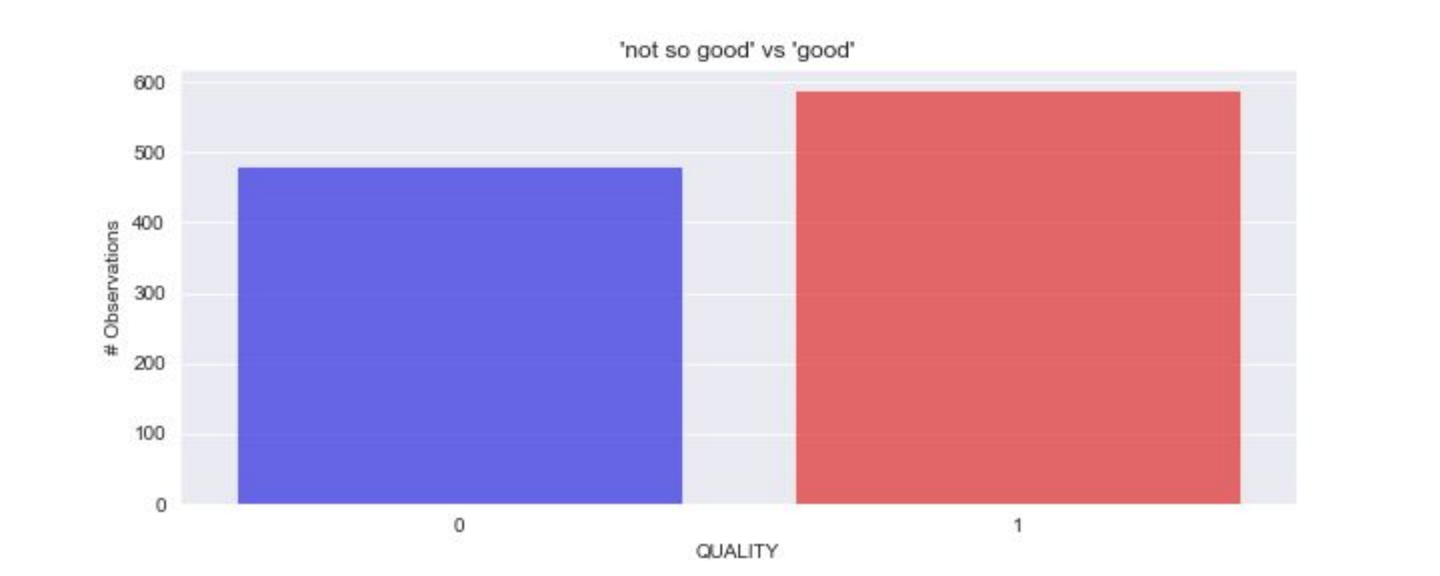




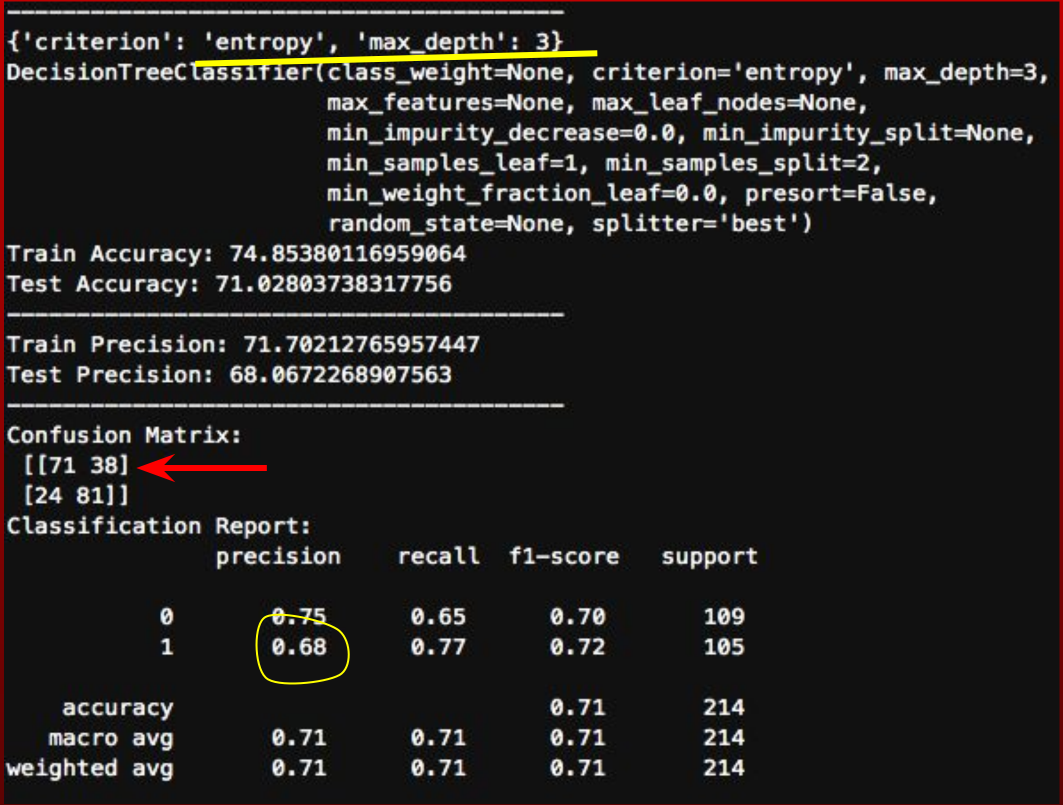
TARGET VARIABLES:

ALL OBSERVATION DIVIDED IN TWO GROUPS:

‘GOOD’ and ‘NOT SO GOOD’



BASE MODEL: DECISION TREE



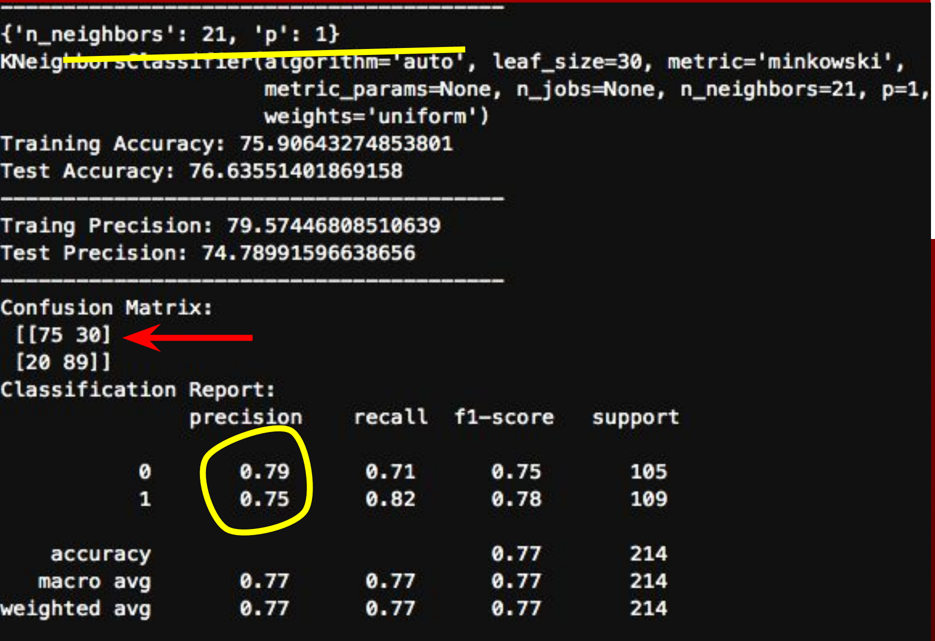
```
----------------------------------------
{'criterion': 'entropy', 'max_depth': 3}
DecisionTreeClassifier(class_weight=None, criterion='entropy', max_depth=3,
                       max_features=None, max_leaf_nodes=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, presort=False,
                       random_state=None, splitter='best')
Train Accuracy: 74.85380116959064
Test Accuracy: 71.02803738317756
----------------------------------------
Train Precision: 71.70212765957447
Test Precision: 68.0672268907563
----------------------------------------
Confusion Matrix:
 [[71 38]
 [24 81]]
Classification Report:
              precision    recall  f1-score   support

           0       0.75      0.65      0.70       109
           1       0.68      0.77      0.72       105

    accuracy                           0.71       214
   macro avg       0.71      0.71      0.71       214
weighted avg       0.71      0.71      0.71       214
```

KNN:

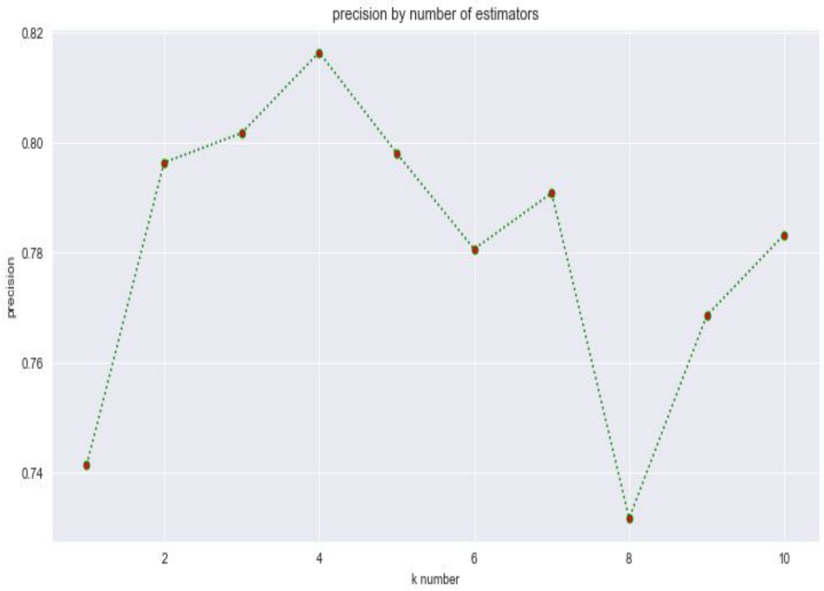
HYPERPARAMETERS TUNED:

- Number of neighbors,
- Distance

PERFORMANCE:



```
-----------------------------------------
{'n_neighbors': 21, 'p': 1}
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                     metric_params=None, n_jobs=None, n_neighbors=21, p=1,
                     weights='uniform')
Training Accuracy: 75.90643274853801
Test Accuracy: 76.63551401869158
-----------------------------------------
Traing Precision: 79.57446808510639
Test Precision: 74.78991596638656
-----------------------------------------
Confusion Matrix:
 [[75 30]
 [20 89]]
Classification Report:
              precision    recall  f1-score   support

           0       0.79      0.71      0.75       105
           1       0.75      0.82      0.78       109

    accuracy                           0.77       214
   macro avg       0.77      0.77      0.77       214
weighted avg       0.77      0.77      0.77       214
```

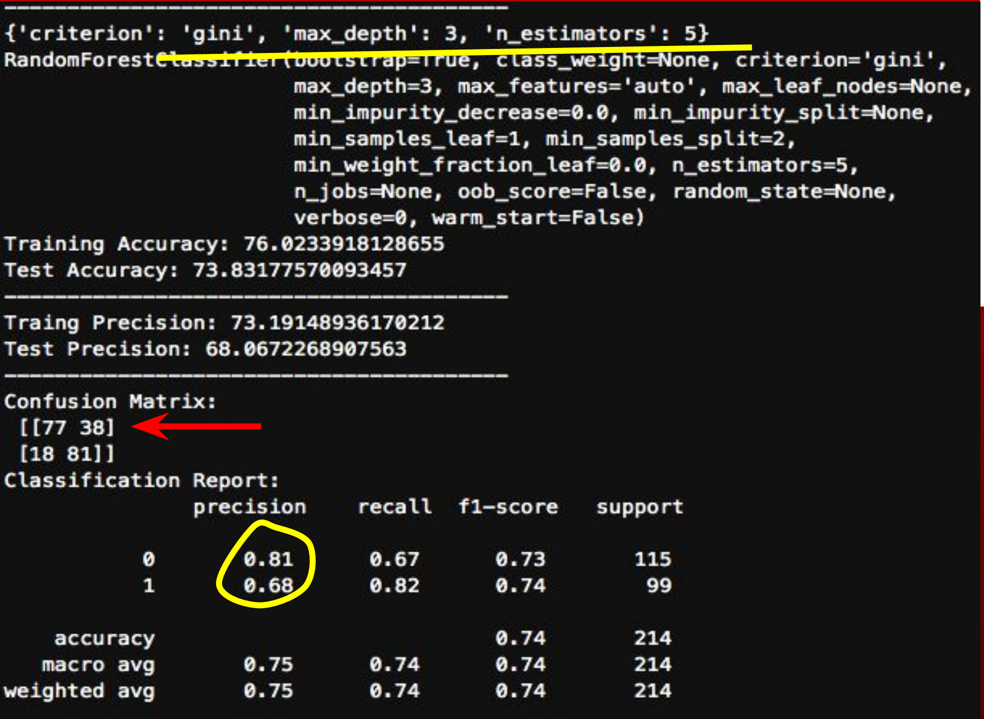


RANDOM FOREST:

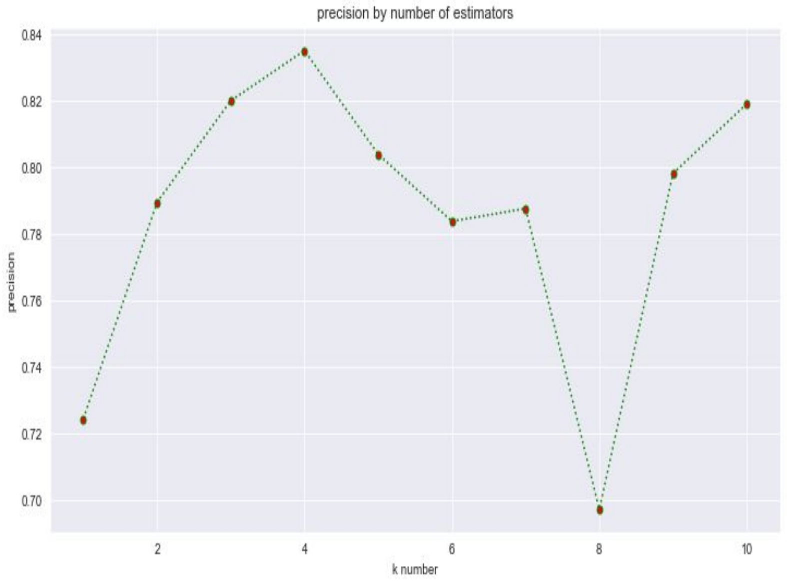
HYPERPARAMETERS:

Criterion, N_estimators, Max_Depth

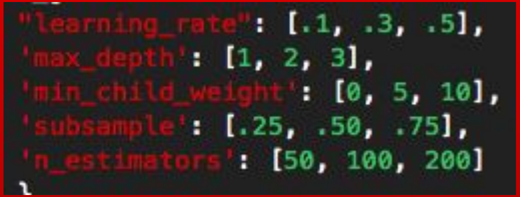
PERFORMANCE:



```
----------------------------------------
{'criterion': 'gini', 'max_depth': 3, 'n_estimators': 5}
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
                       max_depth=3, max_features='auto', max_leaf_nodes=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=5,
                       n_jobs=None, oob_score=False, random_state=None,
                       verbose=0, warm_start=False)
Training Accuracy: 76.0233918128655
Test Accuracy: 73.83177570093457
----------------------------------------
Traing Precision: 73.19148936170212
Test Precision: 68.0672268907563
----------------------------------------
Confusion Matrix:
 [[77 38]
 [18 81]]
Classification Report:
               precision    recall  f1-score   support

           0       0.81      0.67      0.73       115
           1       0.68      0.82      0.74        99

    accuracy                           0.74       214
   macro avg       0.75      0.74      0.74       214
weighted avg       0.75      0.74      0.74       214
```

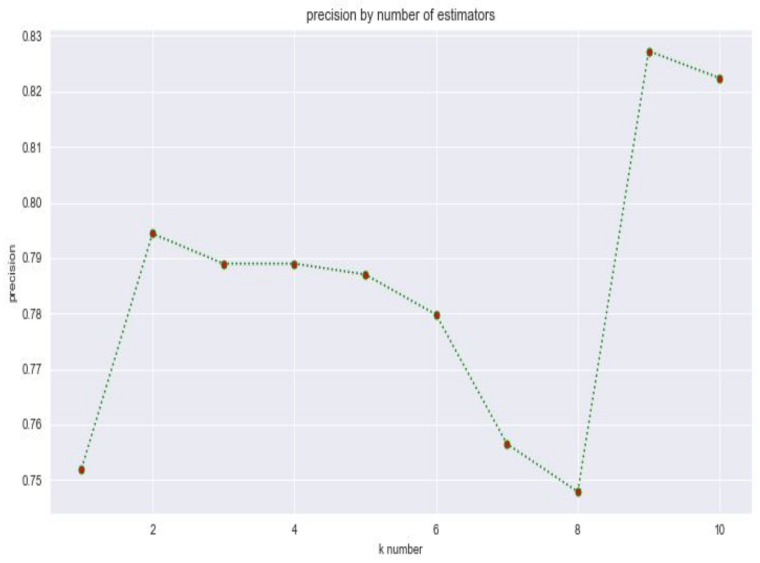


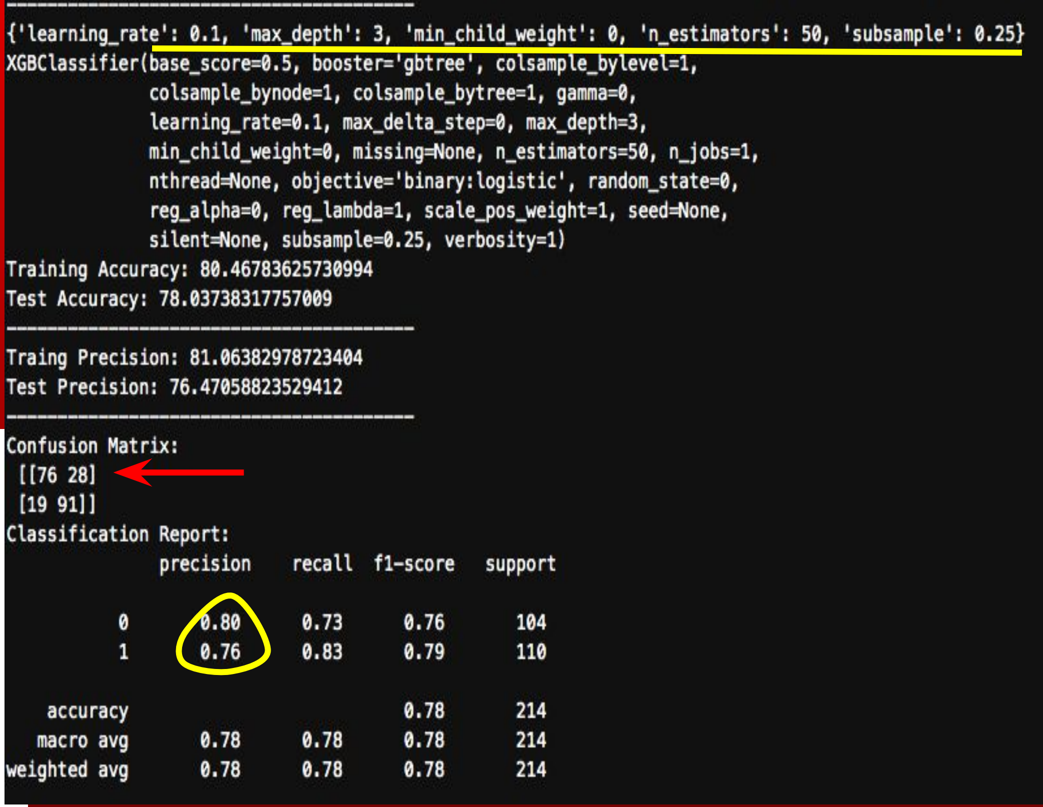
HYPERPARAMETERS:



```
"learning_rate": [.1, .3, .5],
'max_depth': [1, 2, 3],
'min_child_weight': [0, 5, 10],
'subsample': [.25, .50, .75],
'n_estimators': [50, 100, 200]
}
```

PERFORMANCE:

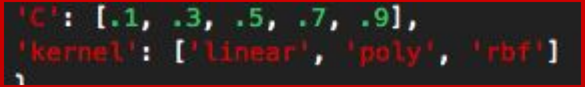




```
{'learning_rate': 0.1, 'max_depth': 3, 'min_child_weight': 0, 'n_estimators': 50, 'subsample': 0.25}
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=1, gamma=0,
              learning_rate=0.1, max_delta_step=0, max_depth=3,
              min_child_weight=0, missing=None, n_estimators=50, n_jobs=1,
              nthread=None, objective='binary:logistic', random_state=0,
              reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None,
              silent=None, subsample=0.25, verbosity=1)
Training Accuracy: 80.46783625730994
Test Accuracy: 78.03738317757009
----------------------------------------

Traing Precision: 81.06382978723404
Test Precision: 76.47058823529412
----------------------------------------
Confusion Matrix:
 [[76 28]
 [19 91]]
Classification Report:
              precision    recall  f1-score   support

           0       0.80      0.73      0.76       104
           1       0.76      0.83      0.79       110

    accuracy                           0.78       214
   macro avg       0.78      0.78      0.78       214
weighted avg       0.78      0.78      0.78       214
```

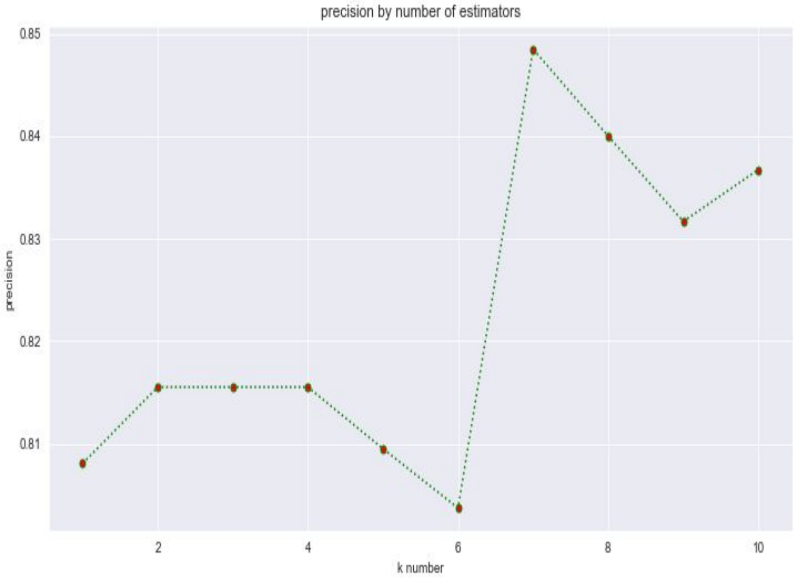
XGB

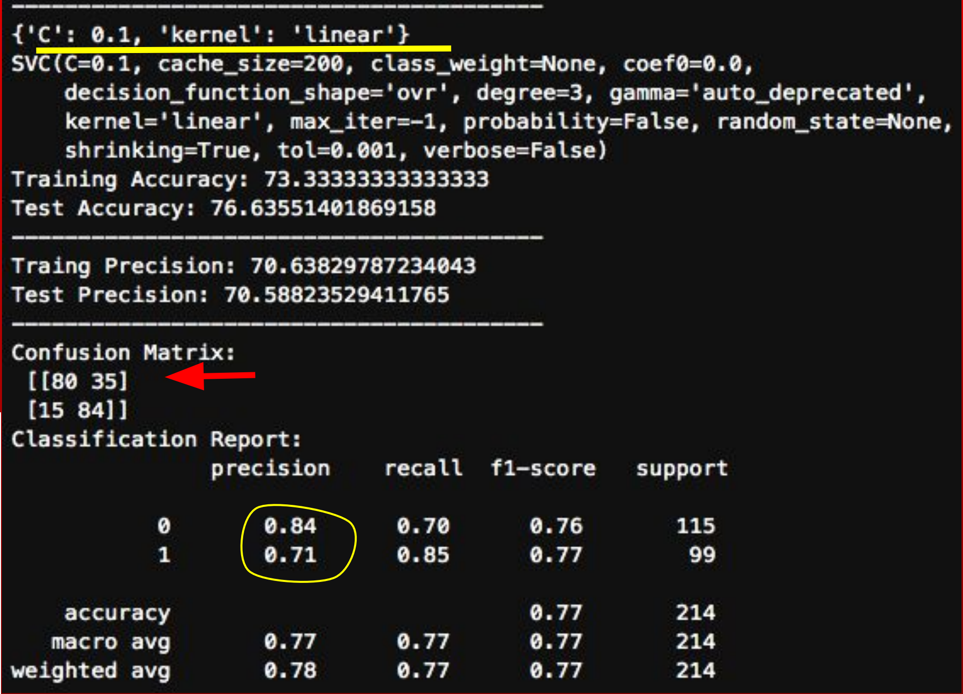
HYPERPARAMETERS:



```
'C': [.1, .3, .5, .7, .9],
'kernel': ['linear', 'poly', 'rbf']
}
```

PERFORMANCE:

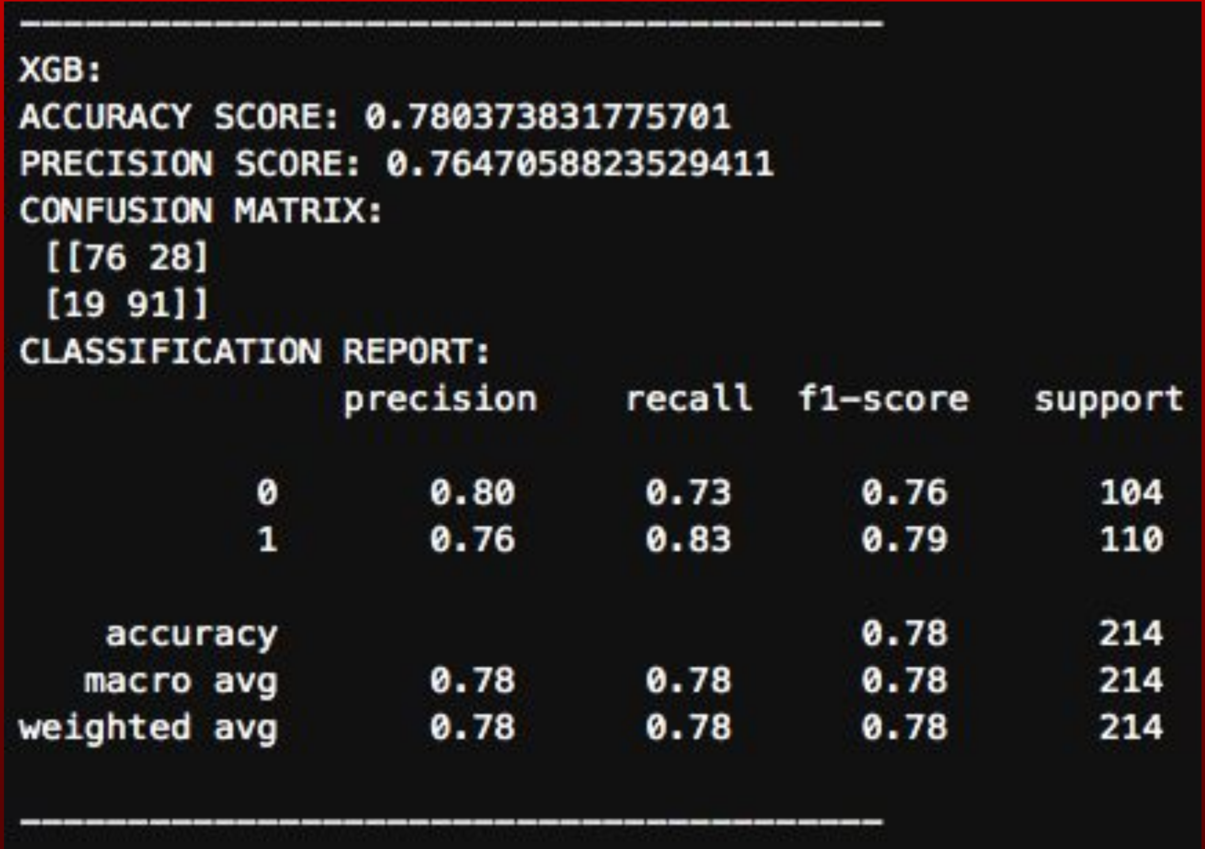




```
{'C': 0.1, 'kernel': 'linear'}
SVC(C=0.1, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
    kernel='linear', max_iter=-1, probability=False, random_state=None,
    shrinking=True, tol=0.001, verbose=False)
Training Accuracy: 73.33333333333333
Test Accuracy: 76.63551401869158
----------------------------------------
Traing Precision: 70.63829787234043
Test Precision: 70.58823529411765
----------------------------------------
Confusion Matrix:
 [[80 35]
 [15 84]]
Classification Report:
               precision    recall  f1-score   support

           0       0.84      0.70      0.76       115
           1       0.71      0.85      0.77        99

    accuracy                           0.77       214
   macro avg       0.77      0.77      0.77       214
weighted avg       0.78      0.77      0.77       214
```

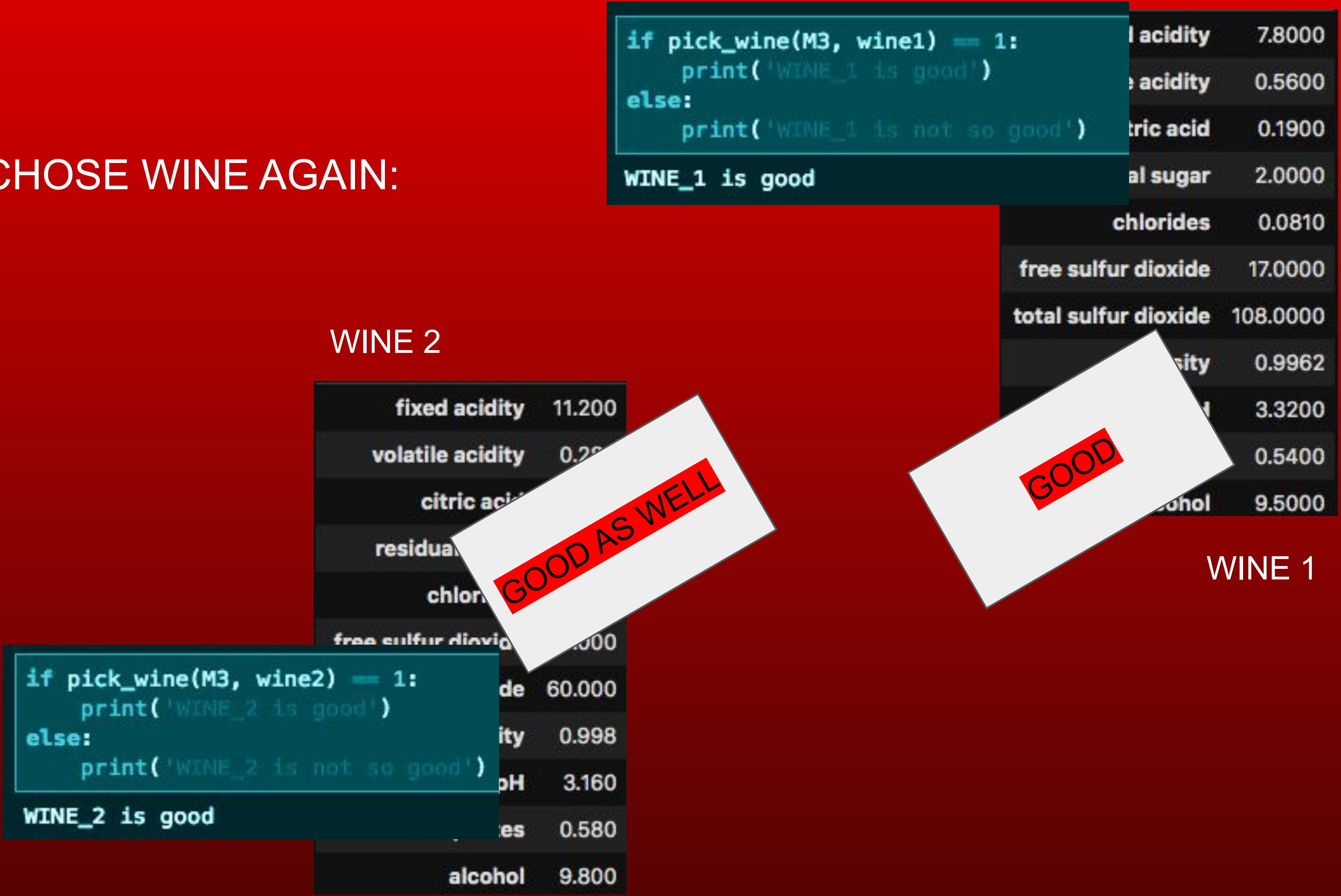
SVC

CONCLUSION



```
----------------------------------------
XGB:
ACCURACY SCORE: 0.780373831775701
PRECISION SCORE: 0.7647058823529411
CONFUSION MATRIX:
 [[76 28]
 [19 91]]
CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       0.80      0.73      0.76       104
           1       0.76      0.83      0.79       110

    accuracy                           0.78       214
   macro avg       0.78      0.78      0.78       214
weighted avg       0.78      0.78      0.78       214

----------------------------------------
```

CHOSE WINE AGAIN:



```
if pick_wine(M3, wine1) == 1:
    print('WINE_1 is good')
else:
    print('WINE_1 is not so good')
```

```
WINE_1 is good
```

WINE 1

l acidity	7.8000
e acidity	0.5600
tric acid	0.1900
al sugar	2.0000
chlorides	0.0810
free sulfur dioxide	17.0000
total sulfur dioxide	108.0000
sity	0.9962
l	3.3200
	0.5400
ohol	9.5000

GOOD

WINE 2

fixed acidity	11.200
volatile acidity	0.2
citric aci	
residua	
chlor	
free sulfur dioxid	.000
de	60.000
ity	0.998
pH	3.160
es	0.580
alcohol	9.800

GOOD AS WELL

```
if pick_wine(M3, wine2) == 1:
    print('WINE_2 is good')
else:
    print('WINE_2 is not so good')
```

```
WINE_2 is good
```
]