import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn import linear\_model

from sklearn.metrics import accuracy\_score, confusion\_matrix

data = pd.read\_csv('quality.csv')

print(data.info())

label\_dict = {'B':0,

'G':1}

data.label =[label\_dict[item] for item in data.label]

X = data.drop(columns = ['label'])

y = data.label

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X,y,test\_size=.3, random\_state=15)

from sklearn.linear\_model import LogisticRegression

logreg = LogisticRegression()

logreg.fit(X\_train,y\_train)

y\_pred = logreg.predict(X\_test)

from sklearn import metrics

print('\n\nAccuracy score:')

print(accuracy\_score(y\_test,y\_pred))

print('\n\nConfusion matrix:')

print(confusion\_matrix(y\_test,y\_pred))

import matplotlib.pyplot as plt

y\_pred\_proba = logreg.predict\_proba(X\_test)[::,1]

fpr,tpr,\_ = metrics.roc\_curve(y\_test,y\_pred\_proba)

auc = metrics.roc\_auc\_score(y\_test,y\_pred\_proba)

print('\n\nauc score:')

print(auc)

plt.plot(fpr,tpr,label='data1,auc='+str(auc))

plt.legend(loc=4)

plt.show()

predict\_for = pd.DataFrame({

"num\_words": [18],

"num\_characters": [95],

"num\_misspelled": [2],

"bin\_end\_qmark": [0],

"num\_interrogative": [2],

"bin\_start\_small": [0],

"num\_sentences": [2],

"num\_punctuations": [3]

})

prediction\_result = logreg.predict(predict\_for)

print(prediction\_result)

predict\_for2 = pd.DataFrame({

"num\_words": [18],

"num\_characters": [95],

"num\_misspelled": [2],

"bin\_end\_qmark": [0],

"num\_interrogative": [2],

"bin\_start\_small": [0],

"num\_sentences": [2],

"num\_punctuations": [3]

})

prediction\_result = logreg.predict(predict\_for2)

print(prediction\_result)