

Project 1 RED WINE

```
import pandas as pd

from sklearn.preprocessing import StandardScaler

from sklearn.model_selection import train_test_split


# Load the dataset

data =
pd.read_csv('https://github.com/FlipRoboTechnologies/MLDatasets/blob/main/Red%20Wine/winequality-red.csv?raw=true')


# Select the input variables and the target variable

X = data.drop('quality', axis=1)
y = data['quality']


# Scale the input variables

scaler = StandardScaler()

X = scaler.fit_transform(X)


# Split the data into a training set and a test set

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

from sklearn.tree import DecisionTreeClassifier


# Create a decision tree classifier

clf = DecisionTreeClassifier(random_state=42)


# Train the classifier on the training set

clf.fit(X_train, y_train)

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score


# Predict the quality of the wines in the test set

y_pred = clf.predict(X_test)
```

```
# Calculate the accuracy, precision, recall, and F1 score

accuracy = accuracy_score(y_test, y_pred)

precision = precision_score(y_test, y_pred, average='weighted')

recall = recall_score(y_test, y_pred, average='weighted')

f1 = f1_score(y_test, y_pred, average='weighted')


print(f'Accuracy: {accuracy:.3f}')
print(f'Precision: {precision:.3f}')
print(f'Recall: {recall:.3f}')
print(f'F1 score: {f1:.3f}')


# Calculate the area under the ROC curve

roc_auc = roc_auc_score(y_test, clf.predict_proba(X_test)[:, 1], multi_class='ovr')

print(f'Area under the ROC curve: {roc_auc:.3f}')

Accuracy: 0.627

Precision: 0.613

Recall: 0.619

F1 score: 0.609

Area under the ROC curve: 0.868
```