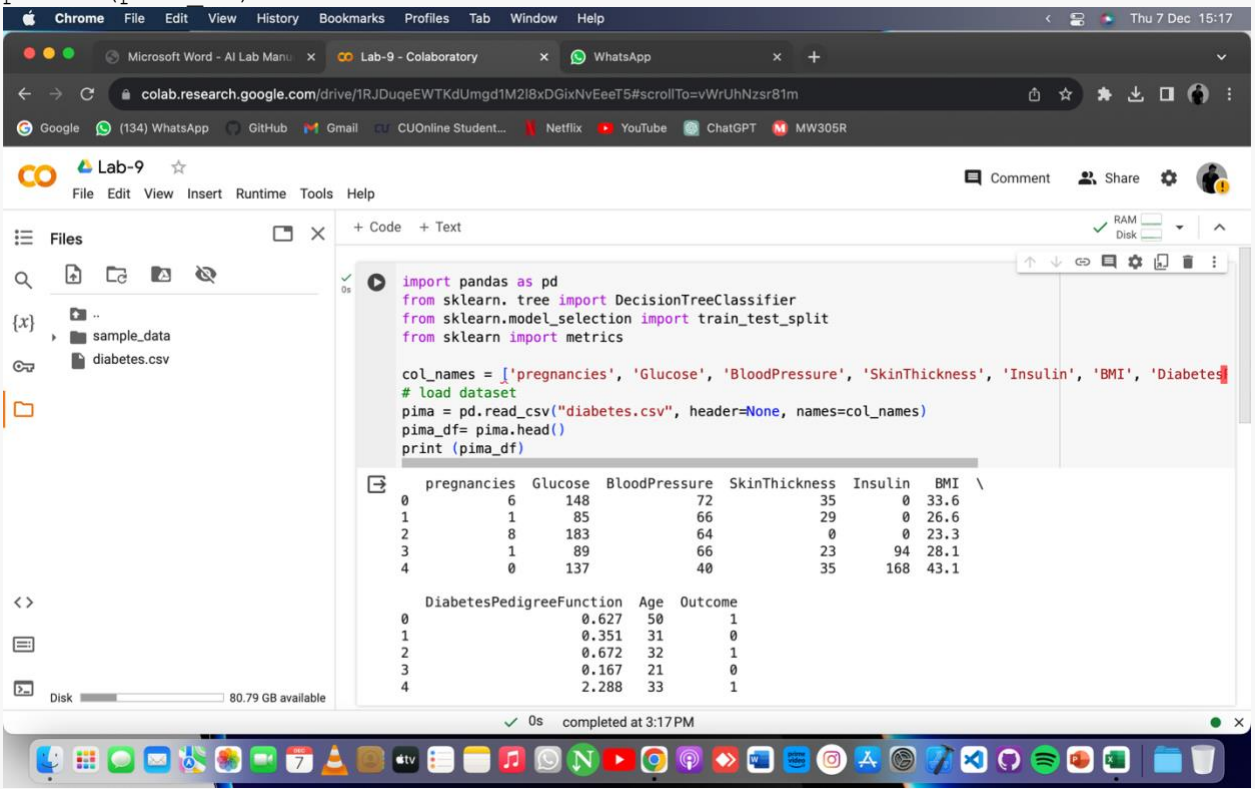
LAB 9

Importing Required Libraries:

```
import pandas as pd
from sklearn. tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
```

Loading Dataset:

```
col_names = ['pregnant', 'glucose', 'bp', 'skin', 'insulin', 'bmi',
'pedigree', 'age', 'label']
# load dataset
pima = pd.read_csv("diabetes.csv", header=None, names=col_names)
pima_df= pima.head()
print (pima_df)
```



Feature Selection:

```
#Feature Selection
```

```
feature_cols = ['pregnant', 'glucose', 'bp', 'skin', 'insulin', 'bmi',
'pedigree']
X = pima[feature_cols] # Features
y = pima.label # Target variable
```

Splitting Data:

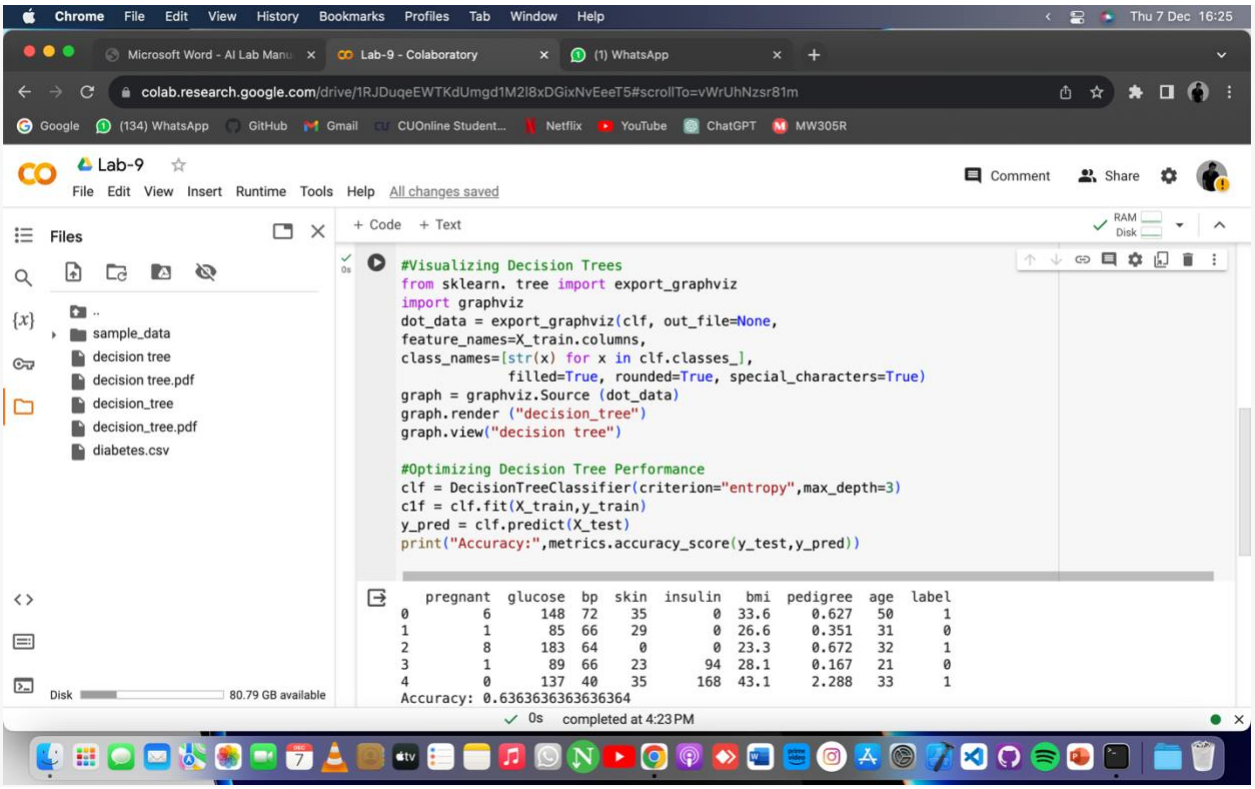
```
#splitting data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=1)
```

Building Decision Tree:

```
#building dt
clf=DecisionTreeClassifier()
clf=clf.fit(X_train,y_train)
y_pred=clf.predict(X_test)
```

Evaluating The Model:

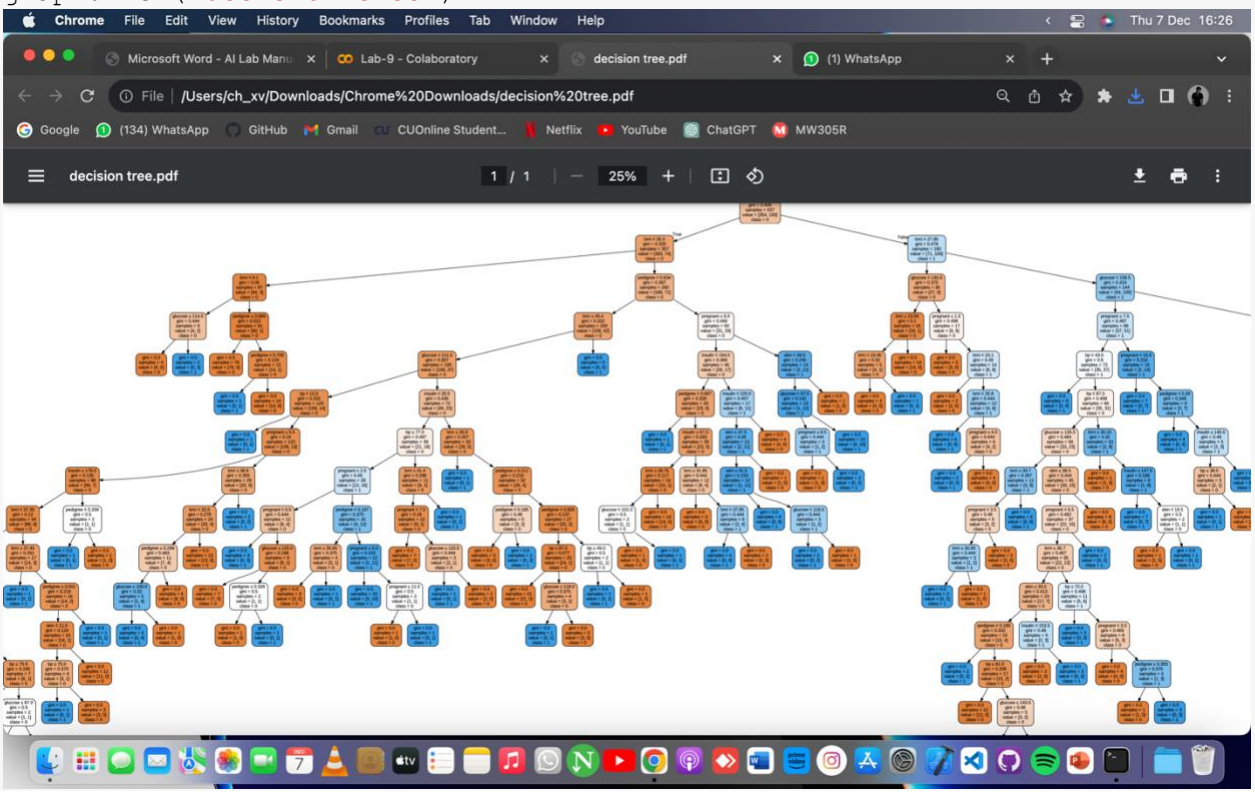
```
#evaluating
print("Accuracy:",metrics.accuracy_score(y_test,y_pred))
```



Visualizing:

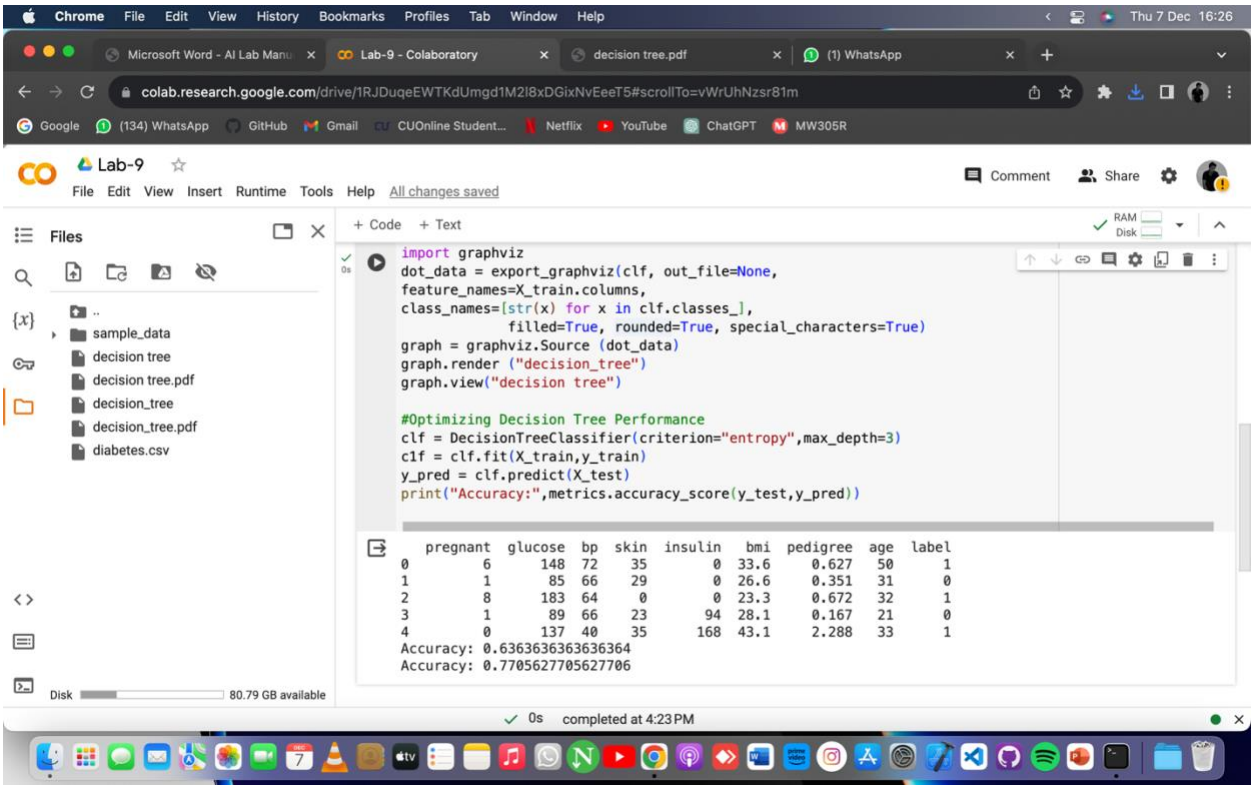
```
#Visualizing Decision Trees
from sklearn. tree import export_graphviz
import graphviz
dot_data = export_graphviz(clf, out_file=None,
feature_names=X_train.columns,
class_names=[str(x) for x in clf.classes_],
             filled=True, rounded=True, special_characters=True)
graph = graphviz.Source (dot_data)
graph.render ("decision_tree")
```

```
graph.view("decision tree")
```



Optimizing:

```
#Optimizing Decision Tree Performance
clf = DecisionTreeClassifier(criterion="entropy",max_depth=3)
c1f = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test,y_pred))
```


Lab-9
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
sample_data
decision tree
decision tree.pdf
decision_tree
decision_tree.pdf
diabetes.csv
Disk 80.79 GB available
import graphviz
dot_data = export_graphviz(clf, out_file=None,
feature_names=X_train.columns,
class_names=[str(x) for x in clf.classes_],
filled=True, rounded=True, special_characters=True)
graph = graphviz.Source (dot_data)
graph.render ("decision_tree")
graph.view("decision tree")
#Optimizing Decision Tree Performance
clf = DecisionTreeClassifier(criterion="entropy",max_depth=3)
c1f = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test,y_pred))
pregnant glucose bp skin insulin bmi pedigree age label
0 6 148 72 35 0 33.6 0.627 50 1
1 1 85 66 29 0 26.6 0.351 31 0
2 8 183 64 0 0 23.3 0.672 32 1
3 1 89 66 23 94 28.1 0.167 21 0
4 0 137 40 35 168 43.1 2.288 33 1
Accuracy: 0.6363636363636364
Accuracy: 0.7705627705627706
0s completed at 4:23 PM