Project 1:

Build Decision Tree(DV-"Survived",IDV-"Age,Gender and Fare") and Prediction

**import** **pandas** **as** **pd**

**import** **numpy** **as** **np**

**from** **sklearn** **import** preprocessing

**from** **sklearn** **import** tree

**import** **sklearn** **as** **sk**

titanic\_train=pd.read\_csv("train.csv")

new\_age\_var=np.where(titanic\_train["Age"].isnull(),28,titanic\_train["Age"])

titanic\_train["Age"]=new\_age\_var

label\_encoder=preprocessing.LabelEncoder()

encoded\_sex=label\_encoder.fit\_transform(titanic\_train["Sex"])

tree\_model=tree.DecisionTreeClassifier()

predictors=pd.DataFrame([encoded\_sex,titanic\_train["Age"],titanic\_train["Fare"]]).T

tree\_model=tree.DecisionTreeClassifier(max\_depth=8)

tree\_model.fit(X=predictors,y=titanic\_train["Survived"])

Out[83]:

DecisionTreeClassifier(ccp\_alpha=0.0, class\_weight=None, criterion='gini',

max\_depth=8, 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='deprecated',

random\_state=None, splitter='best')

with open("TrainDtree.dot",'w') as f:

f=tree.export\_graphviz(tree\_model,feature\_names=["Age","Sex","Fare"],out\_file=f);



tree\_model.score(X=predictors,y=titanic\_train["Survived"])

Out[85]: 0.9763779527559056

titanic\_test=pd.read\_csv("test.csv")

new\_age1\_var=np.where(titanic\_test["Age"].isnull(),28,titanic\_test["Age"])

titanic\_test["Age"]=new\_age1\_var

encoded\_sex\_test=label\_encoder.fit\_transform(titanic\_test["Sex"])

test\_features=pd.DataFrame([encoded\_sex\_test,titanic\_test["Age"],titanic\_test["Fare"]]).T

test\_preds=tree\_model.predict(X=test\_features)

Predicted\_Output=pd.DataFrame({"Sex ":titanic\_test["Sex "],"Survived":test\_preds})

Predicted\_Output.to\_csv("SexPredictedOutput.csv",index=False);

Predicted\_Output=pd.DataFrame({"Age":titanic\_test["Age"],"Survived":test\_preds})

Predicted\_Output.to\_csv("AgePredictedOutput.csv",index=False);

Predicted\_Output=pd.DataFrame({"Fare":titanic\_test["Fare"],"Survived":test\_preds})

Predicted\_Output.to\_csv("FarePredictedOutput.csv",index=False);

