**Titanic Dataset Decision Tree**

**Importing the packages**

import pandas as pd

import numpy as np

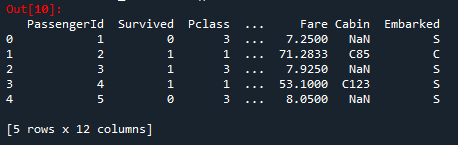
from sklearn import tree

from sklearn import preprocessing

### Loading the dataset

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

titanic\_train.head()

****

### Converting String to Int

l\_enc = preprocessing.LabelEncoder()

titanic\_train['Sex'] = l\_enc.fit\_transform(titanic\_train['Sex'])

### Decision Tree

tree\_model = tree.DecisionTreeClassifier(max\_depth=3)

predictors = pd.DataFrame([titanic\_train['Sex'],titanic\_train['Age'],titanic\_train['Fare']]).T

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

Out[16]:

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

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='deprecated',

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

with open('Dtree2.dot','w') as f:

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

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

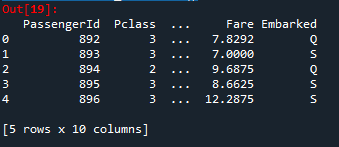
Out[18]: 0.796400449943757

### Model is accurate 79%

### Prediction

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

titanic\_test.head()



titanic\_test.isna().sum()

Out[20]:

PassengerId 0

Pclass 0

Name 0

Sex 0

Age 0

SibSp 0

Parch 0

Ticket 0

Fare 0

Embarked 0

dtype: int64

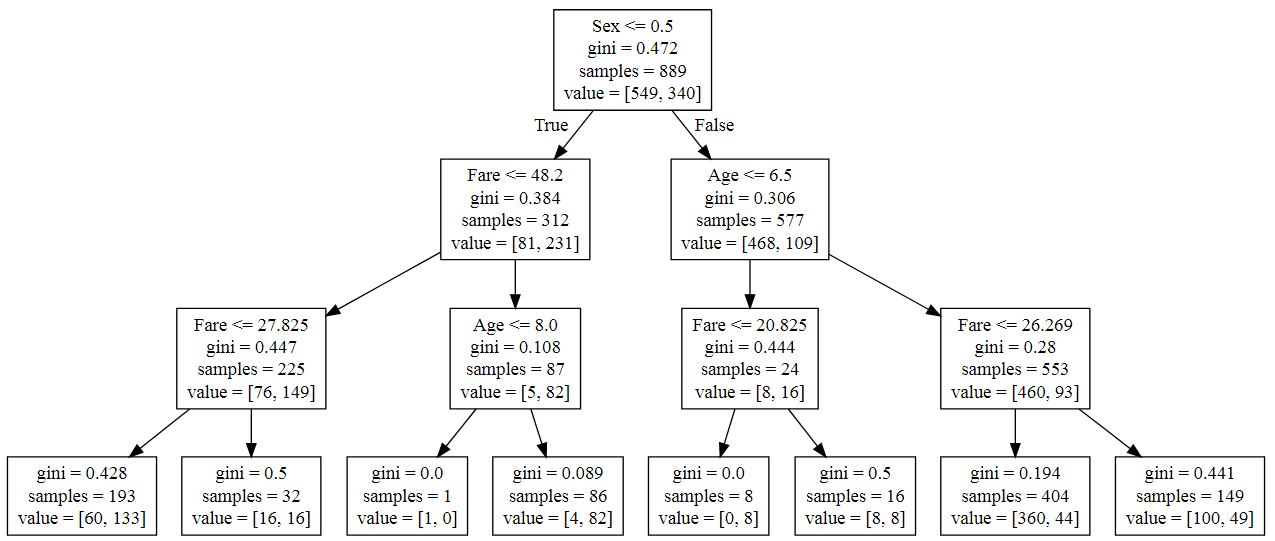
titanic\_test['Sex'] = l\_enc.fit\_transform(titanic\_test['Sex'])

test\_features = pd.DataFrame([titanic\_test['Sex'],titanic\_test['Age'],titanic\_test['Fare']]).T

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

predicted\_op = pd.DataFrame({'PassengerId':titanic\_test['PassengerId'],'Survived':test\_predict})

predicted\_op.to\_csv('output.csv',index=False)



**Rules corresponding to given tree:**

1. **From total 889 samples 549 not survived and 340 survived.**
2. **Total 312 female, 81 female not survived and 231 female survived, and From 577 Male, 468 not survived and 109 male survived.**
3. **For female, if fair<=27.825 from total 225 samples 76 not survived and 149 Female survived.**
4. **For female, if age<=8.0 from total 87 samples 5 not survived and 82 Female survived.**
5. **For female, age<=8.0 , in one set only 1 not survived and 0 survived.**
6. **For female, Age>8.0 , from 86 samples 4 not survived and 86 survived.**
7. **For female, If fair>27.825 from 32 samples 16 not survived and 16 survived.**
8. **For male,if fair<=20.825 from total 24 samples 8 not survived and 16 survived.**
9. **For male,if fair<=20.825,total 8 Male survived.**
10. **For male,if fair>20.825,total 8 not survived and 8 Male survived.**