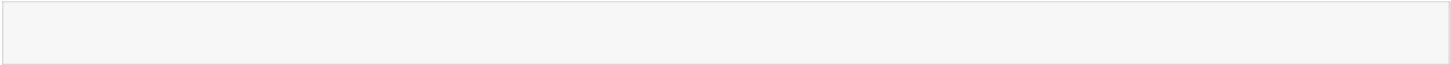
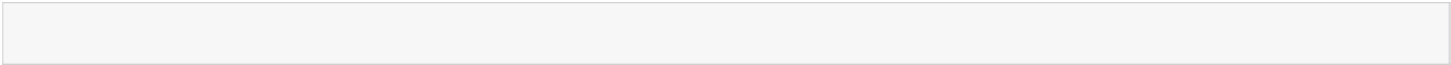
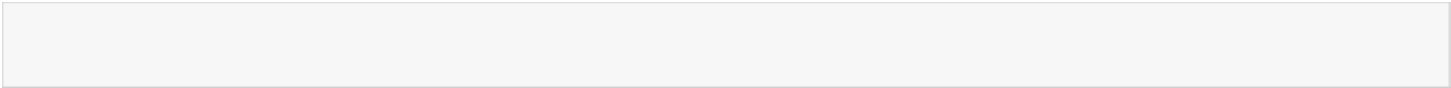
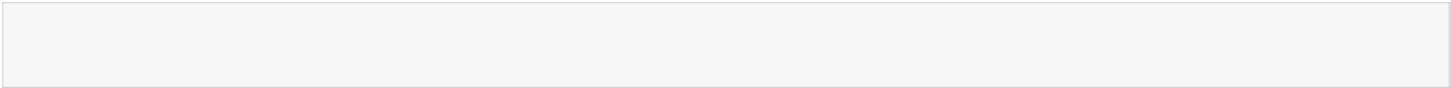
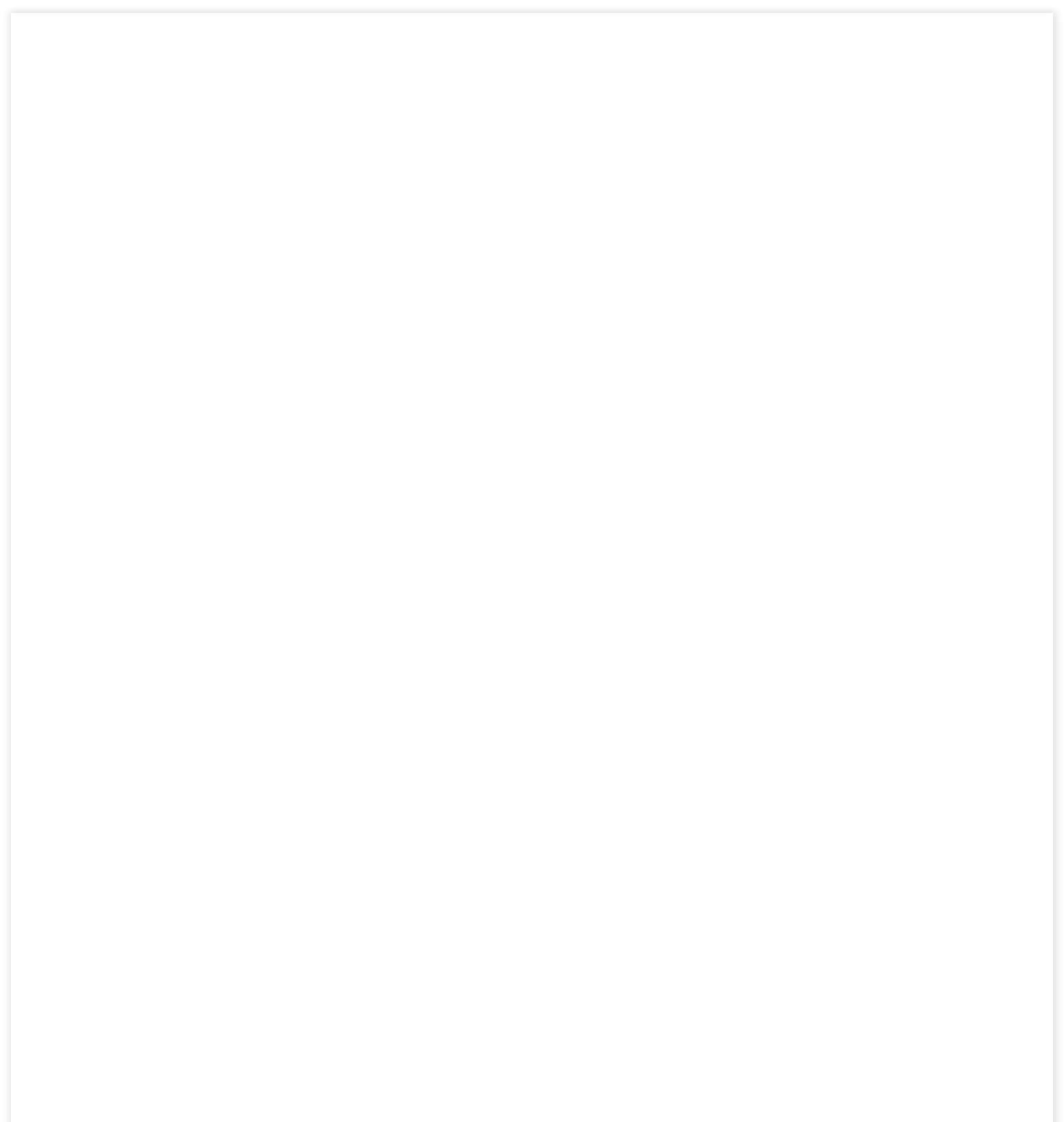
In [52]:



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

**from sklearn import** tree,preprocessing

In [53]:

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

In [54]:

titanic\_train.isna().sum() Out[54]:

|  |  |
| --- | --- |
| PassengerId | 0 |
| Survived | 0 |
| Pclass | 0 |
| Name | 0 |
| Sex | 0 |
| Age | 0 |
| SibSp | 0 |
| Parch | 0 |
| Ticket | 0 |
| Fare | 0 |
| Cabin | 687 |
| Embarked | 0 |
| dtype: int64 |  |

In [55]:

label\_encoder=preprocessing.LabelEncoder()

In [56]:

titanic\_train['Sex']=label\_encoder.fit\_transform(titanic\_train['Sex']) titanic\_train['Embarked']=label\_encoder.fit\_transform(titanic\_train['Embarked']) tree\_model=tree.DecisionTreeClassifier()

In [57]:

predictors=pd.DataFrame([titanic\_train["Sex"],titanic\_train["Age"],titanic\_train['Fare']]

).T

In [58]:

tree\_model=tree.DecisionTreeClassifier(max\_depth=6)

In [59]:

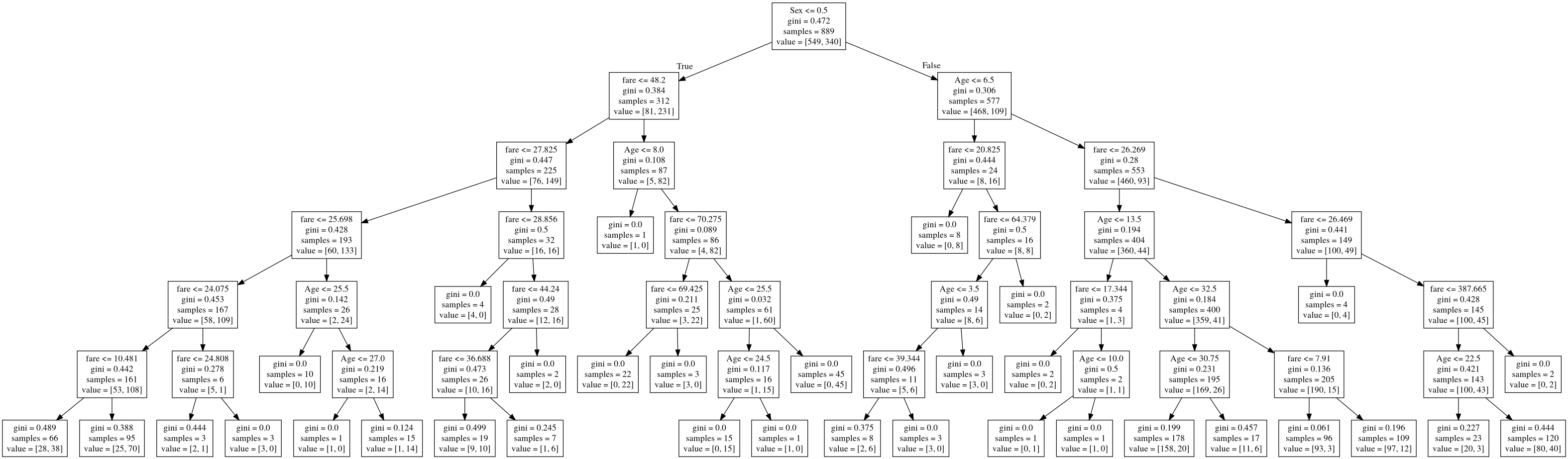
tree\_model.fit(X=predictors, y=titanic\_train['Survived']) Out[59]:

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

max\_depth=6, 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')

In [60]:

**with** open('train\_tree.dot','w')**as** f: f=tree.export\_graphviz(tree\_model,feature\_names=['Sex','Age','fare'],out\_file=f)



**RULES:**

• If sex <= 0.5 and fare<=10481 then survived .

• If sex>0.5 and age>6.5 and fare<=387.665 then not survied

* if sex>0.5 and age<=22.5 and fare<=387.665 then survived
* if sex>0.5 and age>6.5 and fare<=26.469 then not survived
* if sex>0.5 and age<=10.0 and fare<=26.269 then not survived

In [61]:



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

Out[61]:

0.8256467941507312

***PREDICTION***

In [62]:



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

In [63]:



titanic\_test.head(10)

Out[63]:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PassengerId** | | **Pclass** | **Name** | **Sex** | **Age** | **SibSp** | **Parch** | **Ticket** | **Fare** | **Embarked** |
| **0** 892 | | 3 | Kelly, Mr. James | male | 34.5 | 0 | 0 | 330911 | 7.8292 | Q |
| **1** 893 | | 3 | Wilkes, Mrs. James (Ellen Needs) | female | 47.0 | 1 | 0 | 363272 | 7.0000 | S |
| **2** 894 | | 2 | Myles, Mr. Thomas Francis | male | 62.0 | 0 | 0 | 240276 | 9.6875 | Q |
| **3** 895 | | 3 | Wirz, Mr. Albert | male | 27.0 | 0 | 0 | 315154 | 8.6625 | S |
| **4** 896 | | 3 | Hirvonen, Mrs. Alexander (Helga E | female | 22.0 | 1 | 1 | 3101298 | 12.2875 | S |
|  |  |  | Lindqvist) |  |  |  |  |  |  |  |
| **5** | 897 | 3 | Svensson, Mr. Johan Cervin | male | 14.0 | 0 | 0 | 7538 | 9.2250 | S |
| **6** | 898 | 3 | Connolly, Miss. Kate | female | 30.0 | 0 | 0 | 330972 | 7.6292 | Q |
| **7** | 899 | 2 | Caldwell, Mr. Albert Francis | male | 26.0 | 1 | 1 | 248738 | 29.0000 | S |

**9** 901 3 Davies, Mr. John Samuel male 21.0 2 0 A/4

C

2657 7.2292

0

0

female 18.0

Abrahim, Mrs. Joseph (Sophie Halaut

Easu)

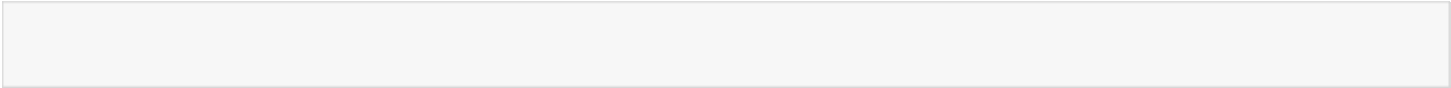
3

900

**8**

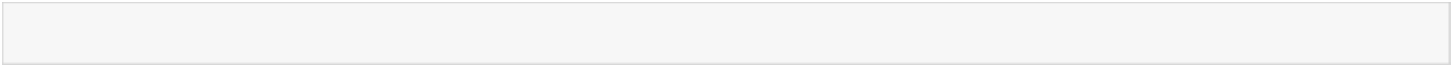
48871

In [64]:



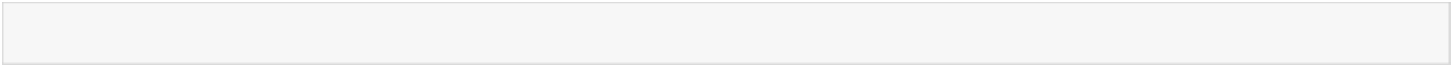
df = pd.DataFrame(titanic\_test) df.Age = df.Age.astype(int) titanic\_test['Age'] = df.Age

In [65]:



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

In [66]:



test\_features=pd.DataFrame([titanic\_test["Sex"],titanic\_test["Age"],titanic\_test['Fare']

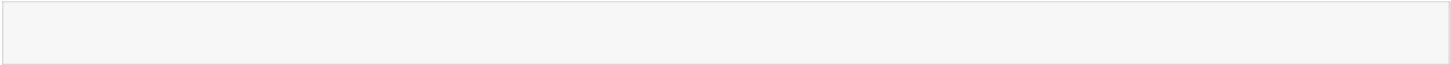
]).T

In [67]:



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

In [68]:



predicted\_output = pd.DataFrame({'PassengerId':titanic\_test['PassengerId'],'Survived':tes t\_preds})

In [69]:

24.1500 S





predicted\_output.to\_csv('Output.csv',index=**False**)

