In [1]:
```
import pandas as pd
```

In [2]:
```
import numpy as np
```

In [3]:
```
from sklearn import tree
```

In [4]:
```
from sklearn import preprocessing
```

In [20]:
```
titanic_train=pd.read_csv(r'C:\Users\siyad\AppData\Local\Temp\Temp1_EDA-20200804T090436Z-001.zip\EDA\titanic_eda\train.csv')
```

In [21]:
```
titanic_train.head()
```

Out[21]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [22]:
```
titanic_train['Age'].mean()
```

Out[22]:
```
29.69911764705882
```

In [32]:
```
titanic_train['Sex'].value_counts()
```

Out[32]:
```
male      577
female    314
Name: Sex, dtype: int64
```

In [23]:
```
titanic_train.isnull().sum()
```

Out[23]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [24]:
```
new_age=np.where(titanic_train['Age'].isnull(),29,titanic_train['Age'])
```

In [25]:
```
titanic_train['Age']=new_age
```

In [37]:
```
titanic_train['Pclass'].value_counts()
```

Out[37]:
```
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

In [26]:
```
label_encoder=preprocessing.LabelEncoder()
```

In [107]:
```
enc_sex=label_encoder.fit_transform(titanic_train['Sex'])
```

In [28]:
```
#since categorical var
tree_model=tree.DecisionTreeClassifier()
```

In [29]:
```
tree_model.fit(X=pd.DataFrame(enc_sex),y=titanic_train['Survived'])
```

Out[29]:
```
DecisionTreeClassifier()
```

In [30]:
```
with open('Dtree1.dot','w') as f:
    f=tree.export_graphviz(tree_model,feature_names=['Sex'],out_file=f)
```

In [139]:
```
pred1=pd.DataFrame([enc_sex,titanic_train['Pclass']]).T
```

In [140]:
```
tree_model.fit(X=pred1,y=titanic_train['Survived'])
```

Out[140]:
```
DecisionTreeClassifier()
```

In [141]:
```
with open('Dtree2.dot','w') as f:
    f=tree.export_graphviz(tree_model,feature_names=['Sex','Pclass'],out_file=f)
```

In [142]:
```
tree_model=tree.DecisionTreeClassifier()
```

In [147]:
```
pred1=pd.DataFrame([enc_sex,titanic_train['Pclass'],titanic_train['Age'],titanic_train['Fare'],titanic_train['Embarked'],
                    titanic_train['SibSp']]).T
```

In [148]:
```
tree_model.fit(X=pred1,y=titanic_train['Survived'])
```

Out[148]:
```
DecisionTreeClassifier()
```

In [149]:
```
with open('Dtree5.dot','w') as f:
    f=tree.export_graphviz(tree_model,feature_names=['Sex','Pclass','Age','Fare','Embarked','SibSp'],out_file=f)
```

In [150]:
```
tree_model.score(X=pred1,y=titanic_train['Survived'])
```

Out[150]:
```
0.9752530933633295
```

In [53]:
```
titanic_test=pd.read_csv(r'C:\Users\siyad\AppData\Local\Temp\Temp1_EDA-20200804T090436Z-001.zip\EDA\titanic_eda\test.csv')
```

In [80]:
```
titanic_test.isnull().sum()
```

Out[80]:
```
PassengerId      0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          326
Embarked         0
dtype: int64
```

In [79]:
```
titanic_test.dropna(subset=['Fare'], how='all', inplace=True)
```

In [81]:
```
titanic_test['Age'].mean()
```

Out[81]:
```
30.14388489208633
```

In [82]:
```
new_age=np.where(titanic_test['Age'].isnull(),30,titanic_test['Age'])
```

In [83]:
```
titanic_test['Age']=new_age
```

In [204]:
```
enc_sex_test=label_encoder.fit_transform(titanic_test['Sex'])
```

In [85]:
```
test_features=pd.DataFrame([enc_sex_test,titanic_test['Pclass'],titanic_test['Age'],titanic_test['Fare']]).T
```

In [86]:
```
test_pred=tree_model.predict(X=test_features)
```

In [87]:
```
predicted_Output=pd.DataFrame({'PassengerID':titanic_test['PassengerId'],'Survived':test_pred})
```

In [88]:
```
predicted_Output.to_csv('Output_Titanic.csv',index=False)
```

In [89]:
```
from sklearn.ensemble import RandomForestClassifier
```

In [90]:
```
titanic_train.columns
```

Out[90]:
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [91]:
```
label_encoder=preprocessing.LabelEncoder()
```

In [92]:
```
titanic_train['Sex']=label_encoder.fit_transform(titanic_train['Sex'])
```

In [95]:
```
titanic_train.dropna(subset=['Embarked'], how='all', inplace=True)
```

In [96]:
```
titanic_train['Embarked']=label_encoder.fit_transform(titanic_train['Embarked'])
```

In [97]:
```
#oob=out of back score
rf_model=RandomForestClassifier(n_estimators=1000,max_features=2,oob_score=True)
```

In [98]:
```
features=['Sex','Pclass','SibSp','Embarked','Age','Fare']
```

In [166]:
```
features2=['Sex','Age','Fare']
```

In [155]:
```
rf_model.fit(X=titanic_train[features2],y=titanic_train['Survived'])
```

Out[155]:
```
RandomForestClassifier(max_features=2, n_estimators=1000, oob_score=True)
```

In [160]:
```
print('OOB Accuracy: ')
print(rf_model.oob_score_)
print(rf_model.score(X=titanic_train[features2],y=titanic_train['Survived']))
```

```
OOB Accuracy:
0.8245219347581553
0.9741282339707537
```

In [99]:
```
rf_model.fit(X=titanic_train[features],y=titanic_train['Survived'])
```

Out[99]:
```
RandomForestClassifier(max_features=2, n_estimators=1000, oob_score=True)
```

In [101]:
```
print('OOB Accuracy: ')
print(rf_model.oob_score_)
```

```
OOB Accuracy:
0.8188976377952756
```

In [102]:
```
for feature,imp in zip(features,rf_model.feature_importances_):
    print(feature,imp)
```

```
Sex 0.26950058485898
Pclass 0.08840980448743559
SibSp 0.05122537303099227
Embarked 0.03193991905610311
Age 0.2718989080803905
Fare 0.28702541048609853
```

In [188]:
```
tree_model=tree.DecisionTreeClassifier()
```

In [189]:
```
pred=pd.DataFrame([enc_sex,titanic_train['Age'],titanic_train['Fare']]).T
```

In [190]:
```
titanic_train.isnull().sum()
```

Out[190]:
```
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 [191]:
```
tree_model.fit(X=pred,y=titanic_train['Survived'])
```

Out[191]:
```
DecisionTreeClassifier()
```

In [192]:
```
with open('Dtree4.dot','w') as f:
    f=tree.export_graphviz(tree_model,feature_names=['Sex','Age','Fare'],out_file=f)
```

In [193]:
```
tree_model.score(X=pred,y=titanic_train['Survived'])
```

Out[193]:
```
0.9741282339707537
```

In [194]:
```
titanic_test=pd.read_csv(r'C:\Users\siyad\AppData\Local\Temp\Temp1_EDA-20200804T090436Z-001.zip\EDA\titanic_eda\test.csv')
```

In [195]:
```
new_age=np.where(titanic_test['Age'].isnull(),30,titanic_test['Age'])
```

In [196]:
```
titanic_test['Age']=new_age
```

In [197]:
```
enc_sex_test=label_encoder.fit_transform(titanic_test['Sex'])
```

In [205]:
```
test_features=pd.DataFrame([enc_sex_test,titanic_test['Age'],titanic_test['Fare']]).T
```

In [206]:
```
titanic_test.isnull().sum()
```

Out[206]:
```
PassengerId      0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          326
Embarked         0
dtype: int64
```

In [207]:
```
test_pred=tree_model.predict(X=test_features)
```

In [201]:
```
titanic_test.dropna(subset=['Fare'], how='all', inplace=True)
```

In [208]:
```
predicted_Output2=pd.DataFrame({'PassengerID':titanic_test['PassengerId'],'Survived':test_pred})
```

In [210]:
```
predicted_Output2.to_csv('Output2_Titanic.csv',index=False)
```

In [ ]:
```

```