In [1]:
```
import pandas as pd
from sklearn.tree import  DecisionTreeClassifier
from sklearn.model_selection import train_test_split
```

In [2]:
```
test=pd.read_csv("test titanic.csv")
data=pd.read_csv("train titanic.csv")
test_ids=test["PassengerId"]
def clean(data):
    data = data.drop(["Name","PassengerId","Cabin","Ticket","Embarked","Fare"],axis=
    data["Age"].fillna(data["Age"].median(),inplace=True)
    data["SibSp"].fillna(data["SibSp"].median(),inplace=True)

    return data
data=clean(data)
test=clean(test)
```

In [3]:
```
#string to laber
from sklearn import preprocessing
le=preprocessing.LabelEncoder()
data["Sex"]=le.fit_transform(data["Sex"])
test["Sex"]=le.transform(test["Sex"])
data.head(5)

x=data.drop(columns=['Survived'])
y=data['Survived']
```

In [4]:
```
#building model using ml model
x_train,x_val,y_train,y_val=train_test_split(x,y,test_size=0.5)
len(x_train)
len(x_val)
model=DecisionTreeClassifier()
model.fit(x_train,y_train)
```

Out[4]:
```
DecisionTreeClassifier()
```

In [5]:
```
model.score(x_val,y_val)
```

Out[5]:
```
0.7937219730941704
```

In [6]:
```
predictions=model.predict(x_val)
submission_preds=model.predict(test)
```

In [7]:
```
df=pd.DataFrame({"PassengerId":test_ids.values,"Survived":submission_preds,})
```

In [8]:
```
df.to_csv("Titanicsubmission.csv",index=False)
```

In [ ]:

In [ ]: