# THE TEST\_TRAIN PROJECT:

THE CODE:

import pandas as pd

import numpy as np

from sklearn import tree

from sklearn import preprocessing

df = pd.read\_csv(r"C:\letsupgrage assignment\train.csv")

df["Age"].mean()

new\_age\_var = np.where(df["Age"].isnull(),32,df["Age"])

df["Age"] = new\_age\_var

label\_encoder = preprocessing.LabelEncoder()

encoded\_sex = label\_encoder.fit\_transform(df["Age"])

tree\_model = tree.DecisionTreeClassifier()

tree\_model.fit(X = pd.DataFrame(encoded\_sex), y = df["Survived"])

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

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

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

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

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

print("the model accuracy")

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

THE OUTPUT:

the model accuracy

0.7795275590551181

This model is made to find the accuracy of the model by using the independent variable like “sex”, “age” and the last “fair”. By using this we defined the accuracy of the model. The accuracy of the model is “0.779”. This shows the model has a good accuracy by using this independent variable .