/\*-------------------------Title-------------------------\*/

Project Title -> Titanic Survival Using Decision Tree Algorithm & Random Forest Algorithm

Subject : Data Science and Analytics

Submitted To : Prof, Abdul Wahid Memon (CSE Department,QUEST NAWABSHAH)

/\*-------------------Group Information-------------------\*/

Group Members :

1) Muzamil Arain (F-18CS 07)

2) Hina Kareem (F-18CS 14)

3) Tahseen Abbas (F-18CS 52)

<----------------------------------------------------------->

Dataset Link :https://www.kaggle.com/c/titanic

Steps:

A) Data Preprocessing

1) Imported Train Data

1.1) Removed Unwanted Columns

1.2) Splitted Data in 90-10 test ratio [Train 1.1] [Test 1.1]

1.3) Cleaned the data -> Age column with null values

1.4) Encoded the Sex column -> 1=Male and 0=Female

2) Imported Given Test Data Set [Test 1.2]

1.1) Removed Unwanted Columns

1.2) Cleaned the data -> Age and Fare columns with null values

1.4) Encoded the Sex column -> 1=Male and 0=Female

B) Training and Testing Process

1) Decision Tree

1.1) Trained Model on [Train 1.1] set

1.2) Got 97.6% training accuracy

1.3) Evaluated our model on [Test 1.1] set and got 85.5% accuracy

1.4) Again Evaluated our model on [Test 1.2] set and got 80.1% accuracy

2) Random Forest Algorithm

i) n\_estimators = 10

1.1) Trained Model on [Train 1.1] set with default estimators i.e. 10

1.2) Got 97.6% training accuracy

1.3) Evaluated our model on [Test 1.1] set and got 90% accuracy

1.4) Again Evaluated our model on [Test 1.2] set and got 86.12% accuracy

ii) n\_estimators = 20

1.1) Trained Model on [Train 1.1] set and changed n\_estimators to 20

1.2) Got 97.3% training accuracy

1.3) Evaluated our model on [Test 1.1] set and got 88.8% accuracy

1.4) Again Evaluated our model on [Test 1.2] set and got 84.4% accuracy

iii) n\_estimators = 30

1.1) Trained Model on [Train 1.1] set and changed n\_estimators to 30

1.2) Got 97.3% training accuracy

1.3) Evaluated our model on [Test 1.1] set and got 91.1% accuracy

1.4) Again Evaluated our model on [Test 1.2] set and got 84.92% accuracy

Result : It was observed that Random Forest with default estimators was performing pretty well as compared to the Single Decision Tree.