

**Student:** Xiaomin Wu

**Week project:** week 39, Decision Tree & Random Forest

### Data preparation for the training:

#### Dataset creation:

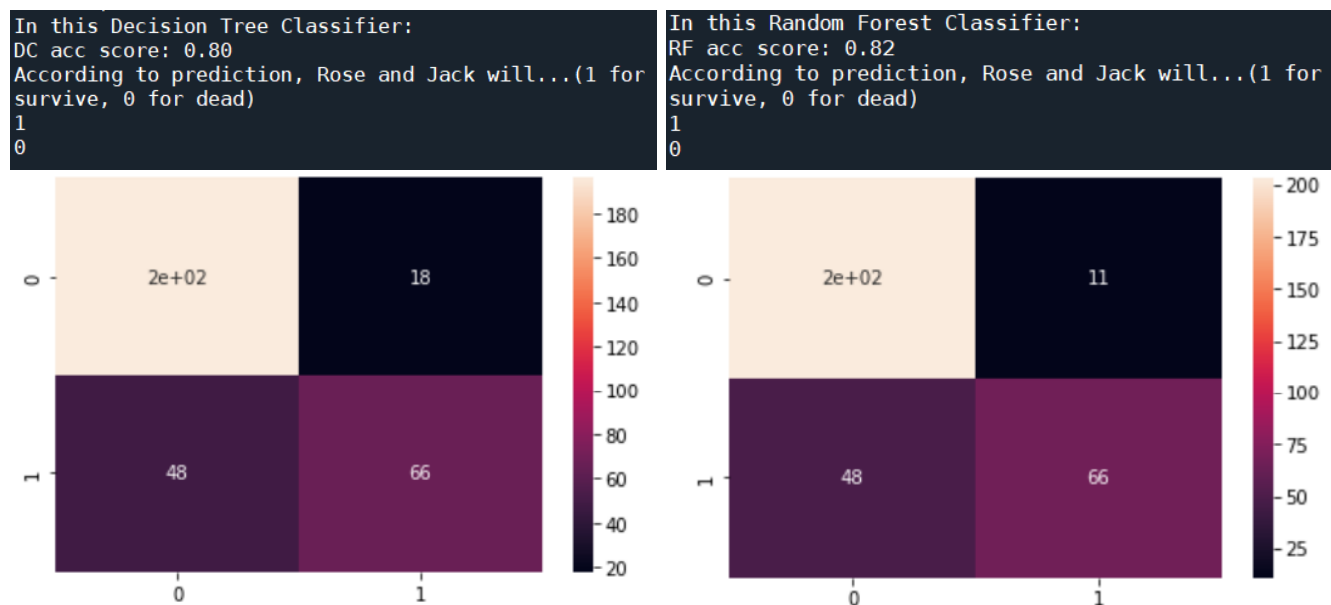
First, import dataset to pandas dataframe, then set 'Survived' as y and rest of data as X. Set test size as 25%.

There is **no scaling** and **no dummy variables** in this case, we directly use data for original format.

#### Train model and predict the survival result.:

Use different model to train, one for Decision Tree Classifier and another one for Random Forest Classifier. Set same depth for both Classifiers, to easier compare two training model. I set 5 here.

#### Relevant metrics for the case:



#### Conclusions of the results:

As we can see in result, for this dataset, Random Forest Classifier is more accurate than Decision Tree Classifier, but not a big difference.