

Training\_model1 to training\_model3 are the models I trained for finding the best one to this project.

Model2 and Model3 are relatively good.

The final prediction is done by training\_model3\_randomforest. The file is *data\_test\_model3.ipynb*

#### General Procedure:

- Split the categorical and numerical data
  - o Convert the categorical data to one-hot data frame
  - o Extract the label from numerical data
- Normalization
- Cross validation preparation: 20% validate + 80% train
- Instantiate an instance of the model and then train it