**World Quant University**

**Professor: Tiberiu Stoica**

**Machine Learning**

Nikolas Lippmann Pareschi - [nikolaslippmann@gmail.com](mailto:nikolaslippmann@gmail.com)

**Unit 3 Project**

### Decision Trees

From the lecture notes, and other assignments, you should have the conceptual knowledge to understand decision trees, and how they are split off.  The ease of its output in making informed decisions, allows individuals to reduce complexity into binary outcomes.  We will now look at decision trees from a different angle.  Can we infer that the “depth” of the tree, meaning the layers of decisions,  adds more predictive power. In the first example in the lecture notes the added sibsp variable actually did reduce the MSE and gave us more intuition as to who would survive or die.  But does added complexity( more data, variables) help?

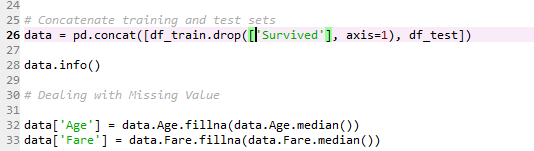
The first step is to download the data:



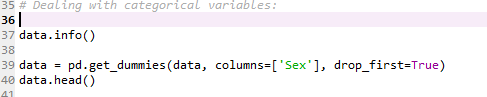
We need to define the target variable (y) in the training set:



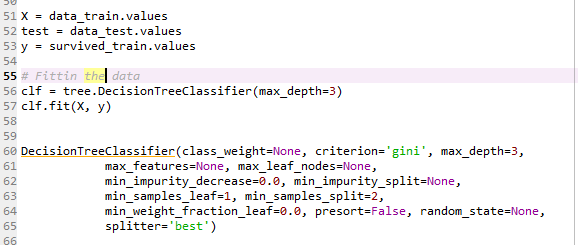
Concatenating in one dataframe and using median values to deal with the missin (NA) value in our data:



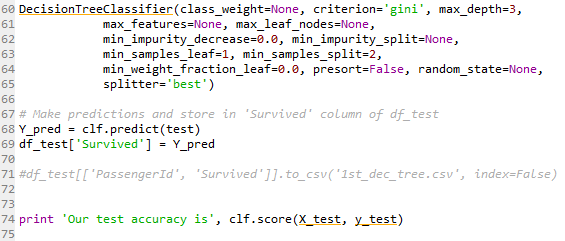
Defining categorical inputs:



Fitting the data:



Making predictions considering a tree with depth 3:





For this project I used these references:

<http://scikit-learn.org/stable/auto_examples/linear_model/plot_lasso_coordinate_descent_path.html#sphx-glr-auto-examples-linear-model-plot-lasso-coordinate-descent-path-py>

<https://www.datacamp.com/community/tutorials/kaggle-tutorial-machine-learning>

<https://blog.patricktriest.com/titanic-machine-learning-in-python/>

http://hamelg.blogspot.com/2015/11/python-for-data-analysis-part-29.html