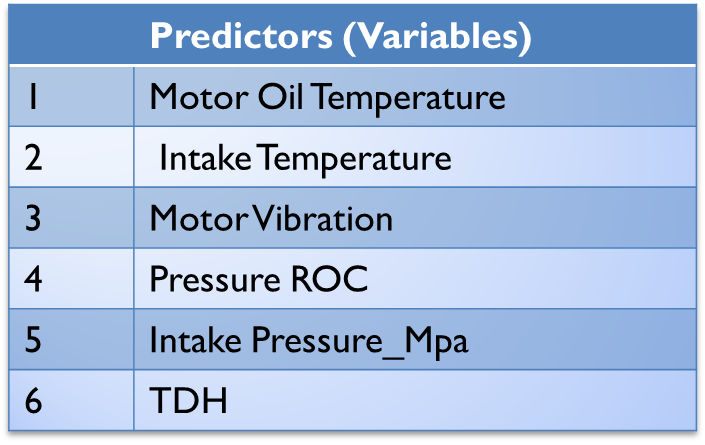
**Day 4: Exercise: Building A Decision Tree for ESP’s**

Digital Foundations, O&G University, September 2017

Goal: Build A Decision Tree to PREDICT ESP Condition

Scenario: You have data on 1000 ESP’s. (Use the data set Wells\_ESP3.csv). Some of those ESP’s have failed, while others are normal. The task is find the Combination of variables that see to cause the ESP’s to fail. We will use a Decision Tree Classifier (from sklearn) for this task.

  
Steps:

* Step 0. Read the file Wells\_ESP3.csv (Store the data in a DATAFRAME called **ESP3**)
* Step 1: What is the Target Variable? Failed ESP’s are known to depend on (See Table)
* Store the Target Variable as ‘y’ and the set of predictive variables as X.
* Step 2: Split the data into Training & Testing  
  Use 800 Rows for Training
* Question: Why do we need Training & Testing? Why not use all the available data?
* Step 2a: Verification Step  
  After splitting, Check the frequency of Normal & Failed ESP’s in Training, Test (and Total)

(How many failed and normal in both the data sets?)

* Step 3: Build A Decision Tree (using the Training Dataset) .
* Step 3a: Visualize The Decision Tree

(Use GraphViz and pydot for visualization)

* Step 4: Make Predictions using the Decision Tree Classifier
  + You will make predictions for each of the ESP’s (observations) in the Test dataset
  + 200 predictions in all
* Step 4a: Create a CONFUSION MATRIX to calculate the Accuracy of your Predictions.