**Team 5: U.S. Crude Oil and Natural Gas Price and Rotary Rigs in Operation Analysis**

UH SPE Machine Learning Bootcamp First Project: Linear Problem and Linear Classification

Dung Nguyen, Xu Yang, Charles White, Jose Benavides

Abstract

Rotary rig is an equipment commonly used for drilling purpose in most wells. Accordingly, how many rotary rigs in operation can reflect how much oil a company is getting, which in turn affects the crude oil price. The information on this relationship, if obtained, can be combined with the cost and revenue analysis from the company’s sales department to provide an insight on how to increase the its profits.

In this report, two datasets will be obtained from <https://www.eia.gov/> using the site’s provided API and the dataset’s IDs. The two specific datasets are: “U.S. Crude Oil and Natural Gas Rotary Rigs in Operation” and “Spot Price of WIT.” First, The Spot Price and Rotary Rigs Count datasets will be synchronized by shifting the time series to make them match. Next, they will be plotted respectively on the vertical and horizontal axis, the visual allowing a clue to their relationship. Finally, the data on the two variables will be evaluated side by side, using the techniques that the graph implies as being the most fitting.

Most likely, due to the nature of the assignment, the techniques to be used will be: Linear modelling (regression) and either Ridge or Lasso. Pearson correlation, regression’s iteration times and final residues analysis along with other statistics evaluation techniques are to be utilized to validate the outcome of the analysis as well.