# Prediction of Downhole Pressure Using Machine Learning

### Project 3: Report Suparna Bhattacharjee, Sruthi Karicheri, Jennie Ran

# Problem Statement

To determine a model that can accurately predict the downhole pressure for a oil well given data from the Volve oil field from the Southern part of the Norwegian sea. Analysis will provide a deeper understanding of the relationships between six variables. (choke size, well head pressure, temperature, oil volume, gas volume, and water volume)

# Data-sets

Original file was in the format of a CSV. Data was sourced from Equinor’s website. Contained variables for five wells from 2008-2014.

# <https://www.equinor.com/en/how-and-why/digitalisation-in-our-dna.html>

# Data Wrangling

Pairwise plots provided insight on the correlation between variables. Per discussion with the SME, some variables were dropped based on those correlations and the usability of the measure to predict downhole pressure. The resulting cleaned data had the choke size, well head pressure, temperature, oil volume, gas volume, and water volume measures for five wells.

# Exploratory Data Analysis

Three Models:

1. Multi-linear
2. Random Forest
3. Neural Network

All models’ MSE (mean squared error) and R-squared measures were evaluate on train and test data. Residual plots were generated to visualize these results. Tuning was done on each model to generate higher accuracy.

# Modeling

Three Models:

1. Multi-linear
   * To explore the linear correlation between variables with down hole pressure.
2. Random Forest
3. Neural Network
   * The use of deep learning to run liner modeling on the data to find the best model for predictions on test data.

### Train-Test Split

<what split was used and how it was acheived>

### Multi-Linear Regression

#### Model Selection

<tuning details>

#### Model Scores

<relevant scores for the model>

### Random Forest

#### Model Selection

<tuning details>

#### Model Scores

<relevant scores for the model>

### Neural Network

#### Model Selection

Deep learning with the use of neural networks. Keras sequential model used as data was relatively simple, numerical only, and from one input source. Tuning with done the hyperas. Created dropout layers to increase accuracy.

#### Model Scores

Model Score: 0.886

Test

MSE: 0.1392

R-squared: 0.8608

Train

MSE: 0.14

R-squared: 0.8577

# Conclusion

<inferences from analysis and modeling>

### Further Exploration

<scope for further work>

# References

Keras Metric Tuning

<https://github.com/keras-team/keras/issues/7947>

Journal pf Petroleum Exploration and Production Technology

<https://link.springer.com/article/10.1007/s13202-019-0728-4>