# Machine Learning Based Real-Time Geosteering

Ngoc Tran

May 06th 2018

Data Analytics Final Project

MEWBOURNE COLLEGE OF EARTH AND ENERGY
MEWBOURNE SCHOOL OF PETROLEUM AND GEOLOGICAL ENGINEERING
The UNIVERSITY of OKLAHOMA

# Introduction & Problems

- 14000 horizontal well drilled in 2018 targeted shale plays
- However, a few shale plays have multiple target zone.
- Real-time geosteering applications: remain in-zone or steer towards a target
- Knowledge of petrophysical/ rock mechanical properties surrounding drilling bit
- LWD -> information with considerable addition costs

**Can we use Drilling measurements to interpret rock properties ?**


**Machine Learning Techniques to predict lithology by using MWD data**

# Data Acquisition

- Dataset released by Equinor and taken from Volve field in Norwegian Continental Shelf
- 12 wells (open hole and MWD logs) and 3 cored wells
  - Open Hole logs: neutron, density, GR, sonic and PE logs
  - MWD: surface RPM, WOB, Torque, ROP, Downhole Pressure and Temperature, Mud Pump Flow Rate, ECD, etc.

equinor

# Workflow

**Log Data – Sonic (DT), Gamma Ray (GR), Density (RHOB), MWD**

- Basic statistic analysis: Boxplots and Linear Regression

**Identify Lithology Clusters on DT, GR and RHOB**

- *15 wells*
- K-means and Self-Organizing Map (SOM)

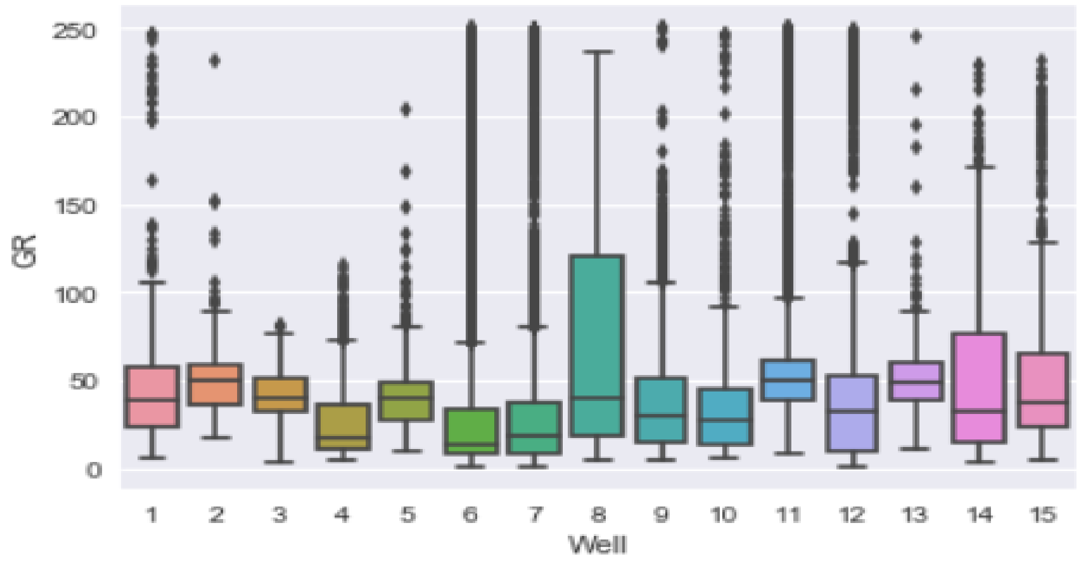
**Petrophysical Significance of Lithology clusters**

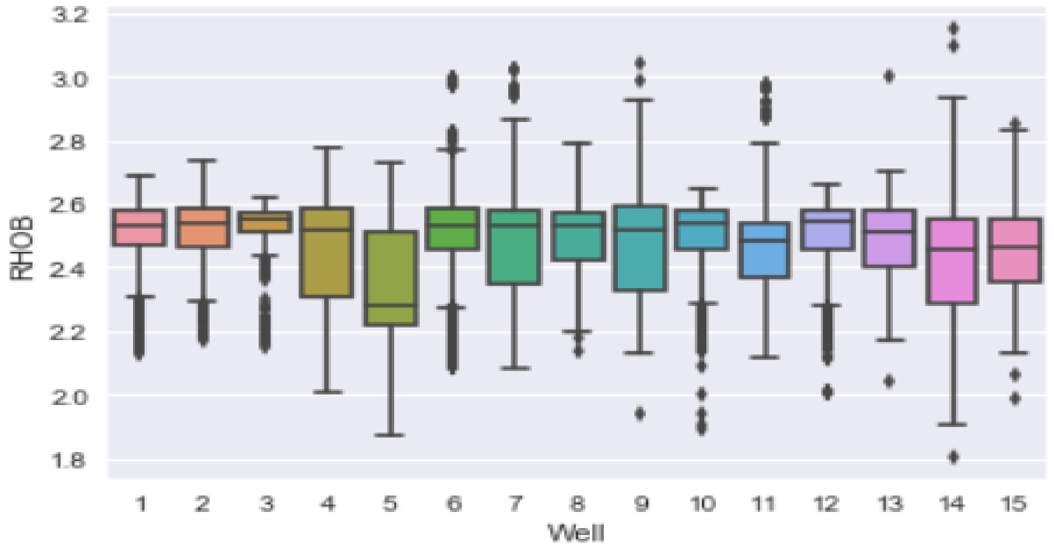
- Identify cluster properties using core data (3 wells)

**Predict Lithology clusters using MWD data**

- *Outlier removal (Z-score), Scaling and Centering on 12 wells*
- KNN, Support Vector Machine and Random Forest

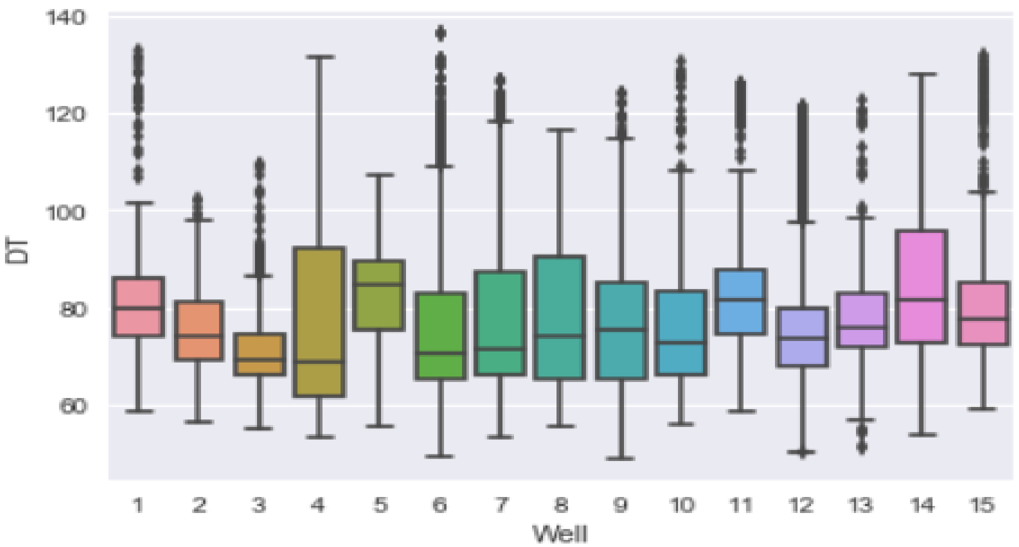
# Basic Statistics Analysis - Lithology



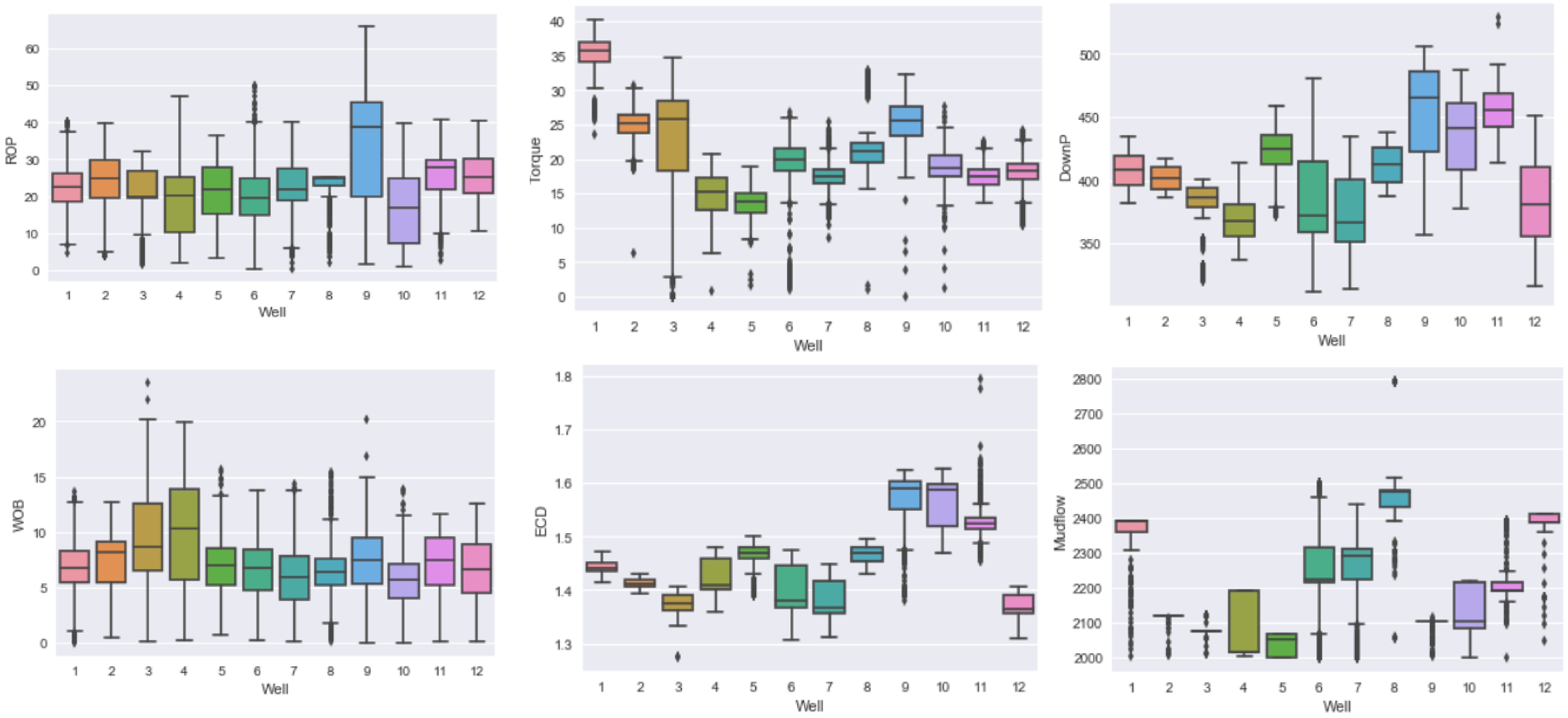




- Box plots for log variables in 15 wells:
  - GR
  - Compressional P-wave slowness (DT)
  - Bulk Density (RHOB)

- Similar distribution



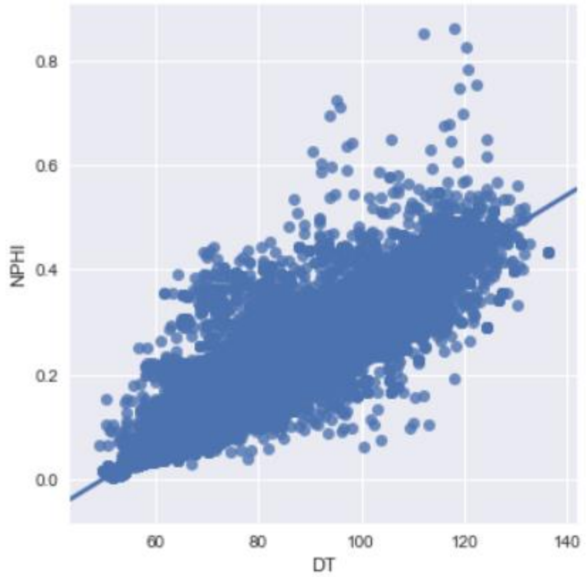
# Basic Statistics Analysis – MWD data

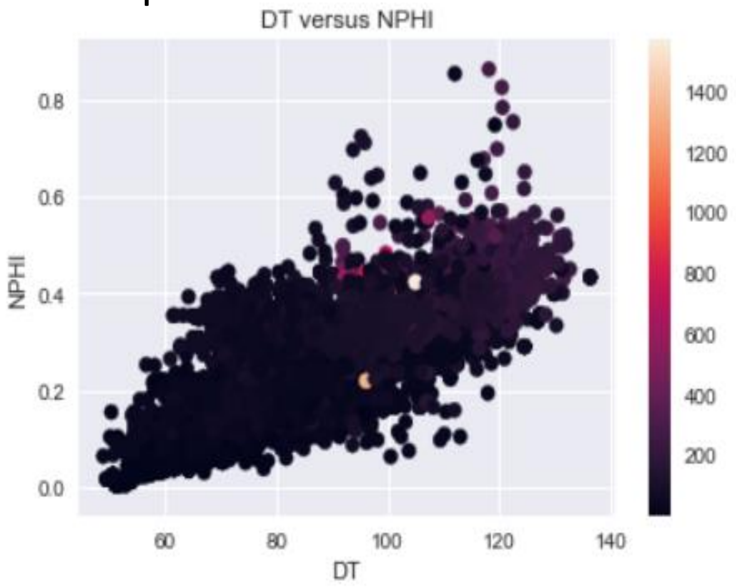



- ROP and WOB shows similar distribution
- Torque, Pressure, ECD reveals different distribution -> Different depths of zone of interest -> Depth-dependent variables

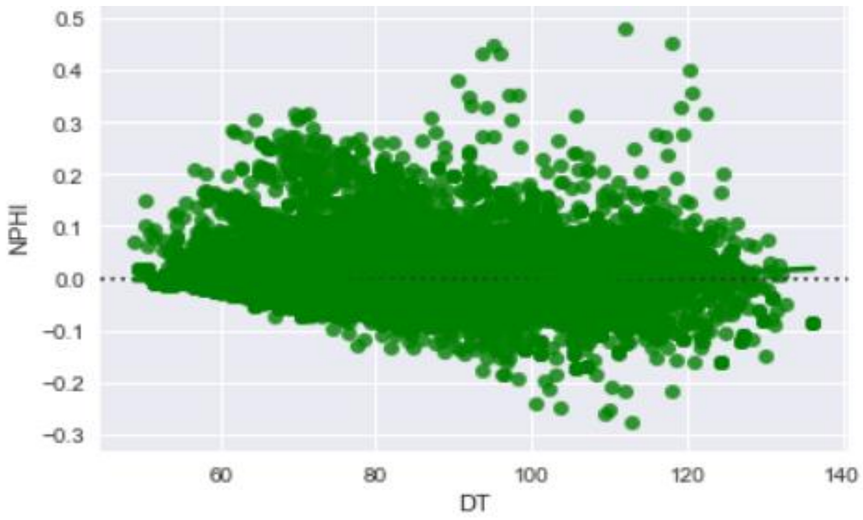
# Linear Regression: DT vs. NPHI

## Linear regression plots




## Residual plot



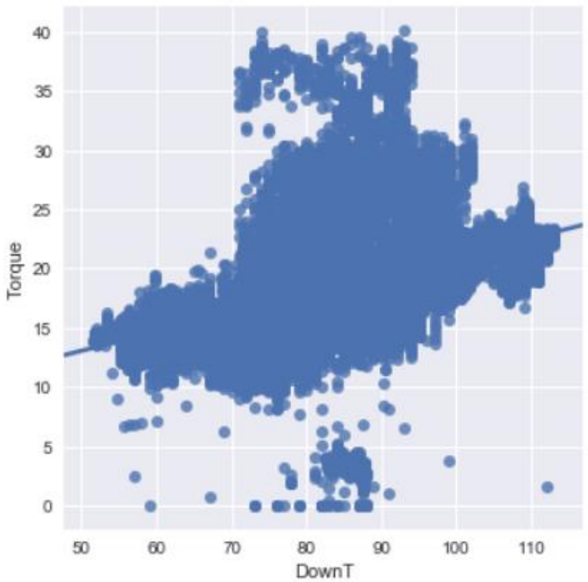

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   NPHI   R-squared:                       0.851
Model:                            OLS   Adj. R-squared:                  0.851
Method:                 Least Squares   F-statistic:                 3.192e+05
Date:                Sun, 05 May 2019   Prob (F-statistic):               0.00
Time:                        18:49:57   Log-Likelihood:             1.0526e+05
No. Observations:               55945   AIC:                        -2.105e+05
Df Residuals:                   55943   BIC:                        -2.105e+05
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.2998      0.001   -357.486      0.000      -0.301      -0.298
DT             0.0060   1.07e-05    564.976      0.000       0.006       0.006
==============================================================================
Omnibus:                    25724.339   Durbin-Watson:                   0.262
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           363668.706
Skew:                           1.846   Prob(JB):                         0.00
Kurtosis:                      14.932   Cond. No.                         424.
==============================================================================
```

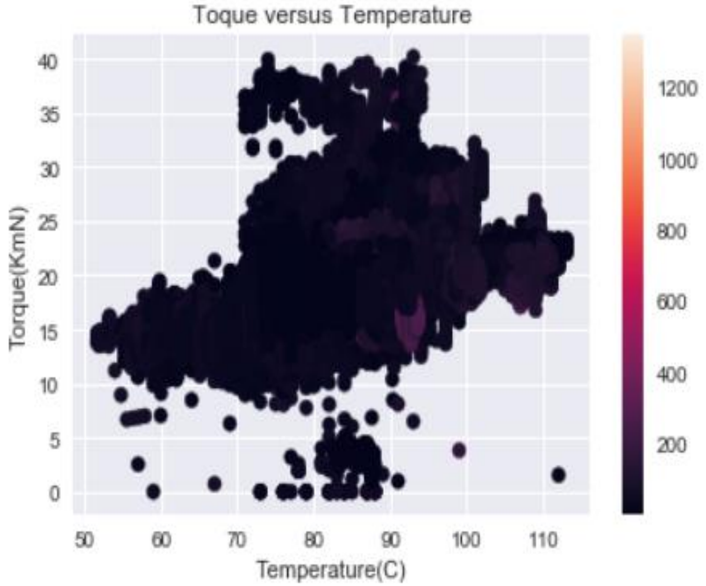
ANOVA results

| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| DT | 1.0 | 433.896557 | 433.896557 | 319197.624797 | 0.0 |
| Residual | 55943.0 | 76.045287 | 0.001359 | NaN | NaN |

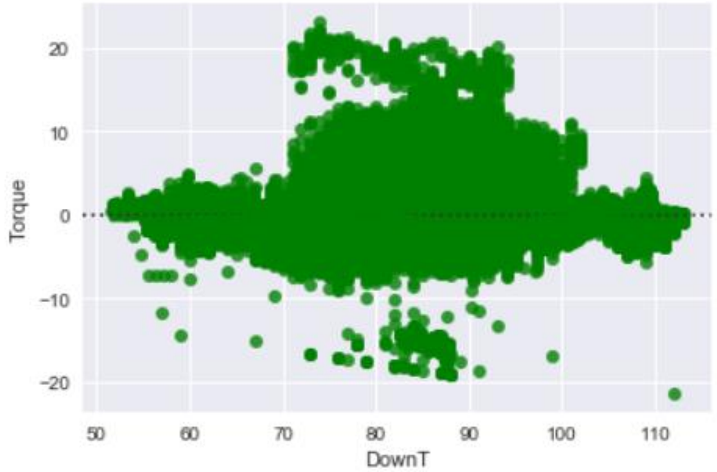
# Linear Regression on Torque vs. Temperature

Linear regression plots




Residual plot

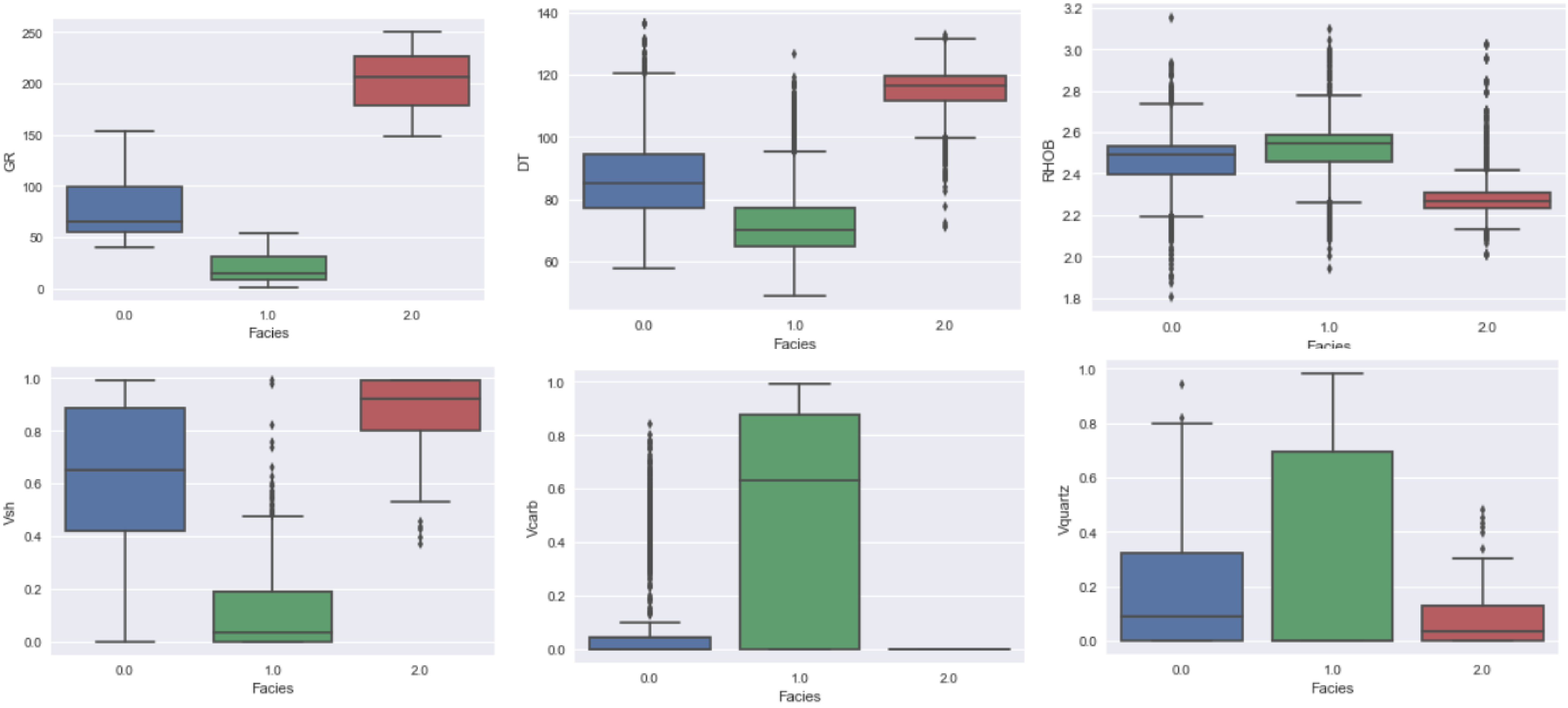



```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Torque   R-squared:                       0.207
Model:                            OLS   Adj. R-squared:                  0.207
Method:                 Least Squares   F-statistic:                 1.355e+04
Date:                Sun, 05 May 2019   Prob (F-statistic):               0.00
Time:                        18:57:50   Log-Likelihood:            -1.3964e+05
No. Observations:               51860   AIC:                         2.793e+05
Df Residuals:                   51858   BIC:                         2.793e+05
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      5.1219      0.118     43.408      0.000       4.891       5.353
DownT          0.1588      0.001    116.425      0.000       0.156       0.161
==============================================================================
Omnibus:                    12838.246   Durbin-Watson:                   0.085
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           178226.590
Skew:                           0.808   Prob(JB):                         0.00
Kurtosis:                      11.937   Cond. No.                         651.
==============================================================================
```

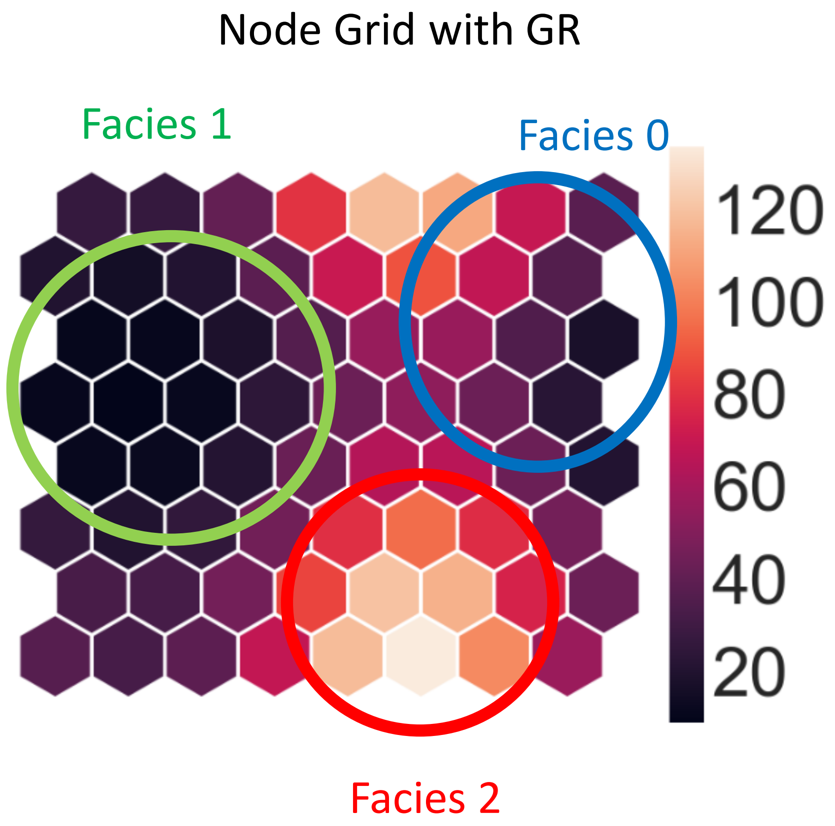
ANOVA results

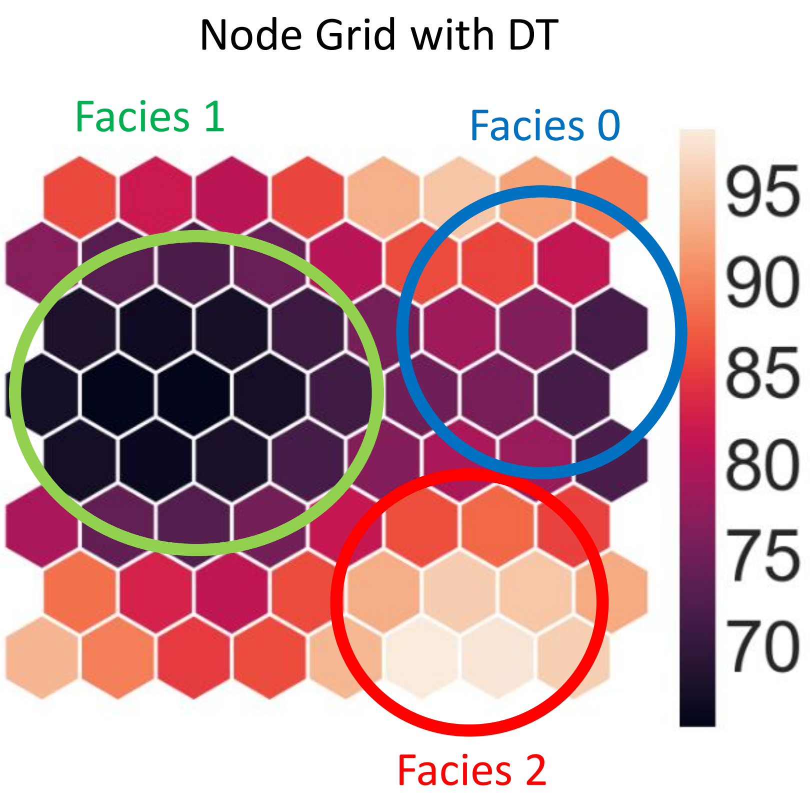
| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| DownT | 1.0 | 173133.536519 | 173133.536519 | 13554.864581 | 0.0 |
| Residual | 51858.0 | 662371.717758 | 12.772797 | NaN | NaN |

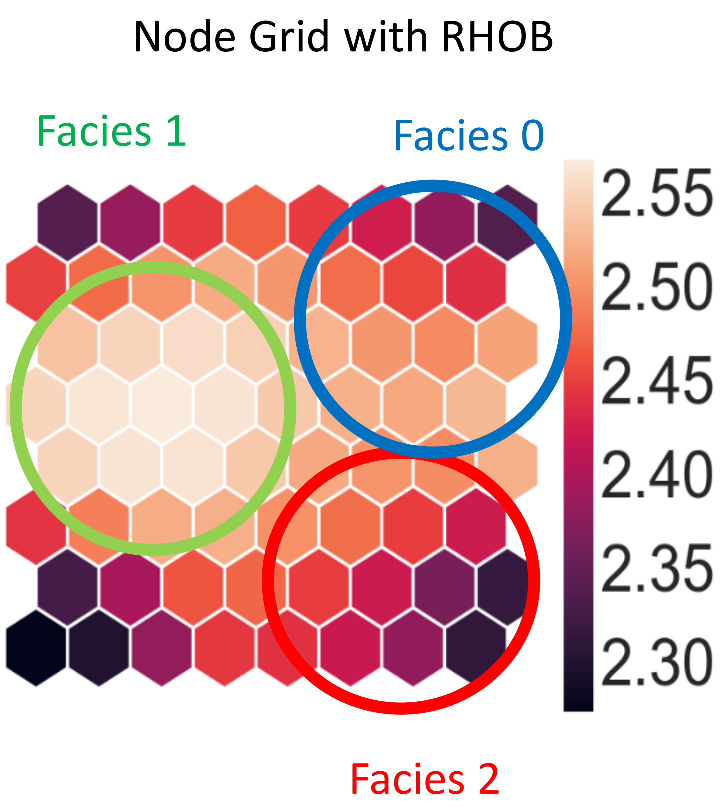
# K-Means Clustering on Log Data



# Self Organizing Map (SOM) for 3 clusters



# KNeighbors Classification

- This technique predicts facies cluster based on MWD
  - Features: MWD Data
  - Target: Facies
  - Number of Samples: 51877 & Number of features: 6
- Using train_test_split package given by Python
  - Train set: 80% Dataset & Test set: 20 % Dataset
- Hyperparameter Optimization by using GridSearch and Cross Validation KFold

| | param_n_neighbors | param_p | mean_test_score |
|---|---|---|---|
| **0** | 3 | 1 | 0.939182 |
| **1** | 3 | 1.5 | 0.935977 |
| **5** | 5 | 1 | 0.935327 |
| **2** | 3 | 2 | 0.933688 |
| **6** | 5 | 1.5 | 0.931496 |

```
memorization performance:  0.9738319558564854
generalization performance:  0.9457401696222051
```

# Support Vector Machine (SVM) classification

- This technique predicts facies cluster based on MWD
  - Features: MWD Data
  - Target: Facies
- Using train_test_split package given by Python
  - Train set: 80% Dataset & Test set: 20 % Dataset
  - Dataset removed outlier (Z-score) and scaled
- SVM Model Parameter: C= 10, kernel= Radial Basis Function (rbf)
- Confusion Matrix

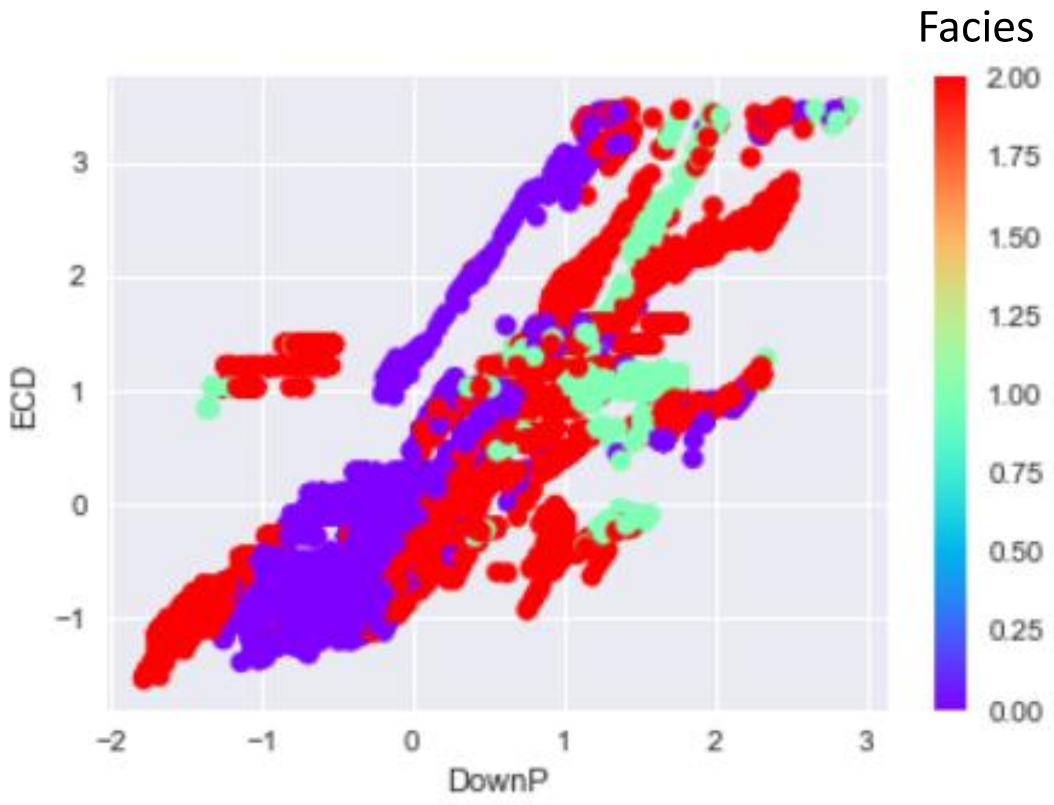
```
[[4675    3  307]
 [   6  656  100]
 [ 134   65 3309]]
```

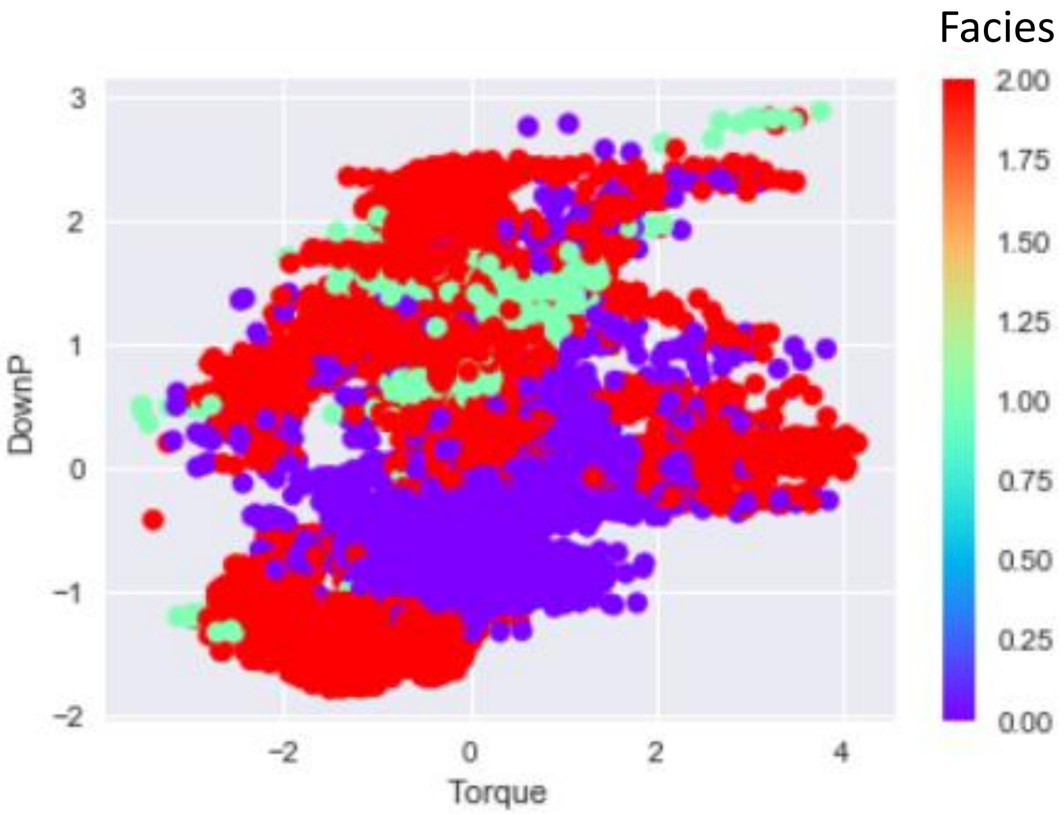
- Scores

```
             precision    recall  f1-score   support

          0       0.97      0.94      0.95      4985
          1       0.91      0.86      0.88       762
          2       0.89      0.94      0.92      3508

avg / total       0.94      0.93      0.93      9255
```

```
generalization performance:  0.93354943273906
memorization performance:  0.9317592392478928
```

# Support Vector Machine (SVM) classification

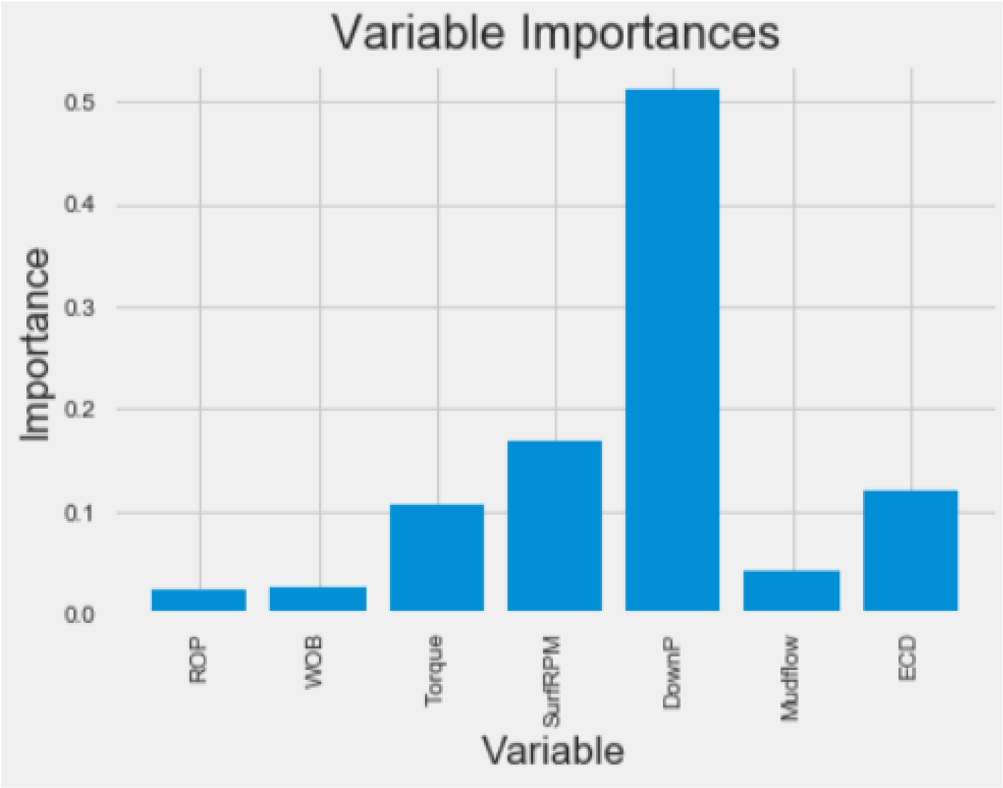


# Random Forest

- Using train_test_split package given by Python
  - Train set: 80% Dataset & Test set: 20 % Dataset
  - Dataset outlier removed (Z-score) and scaled
- This method also allows to evaluate the variable importance
- Initial tuning parameter: n_estimators= 200, max_depth =5, max_features=4
- As for initial tuning parameter
  - Memorization performance: 0.882
  - Generalization performance: 0.884

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.94 | 0.91 | 0.93 | 4985 |
| 1 | 0.93 | 0.54 | 0.69 | 762 |
| 2 | 0.81 | 0.92 | 0.86 | 3508 |
| avg / total | 0.89 | 0.88 | 0.88 | 9255 |

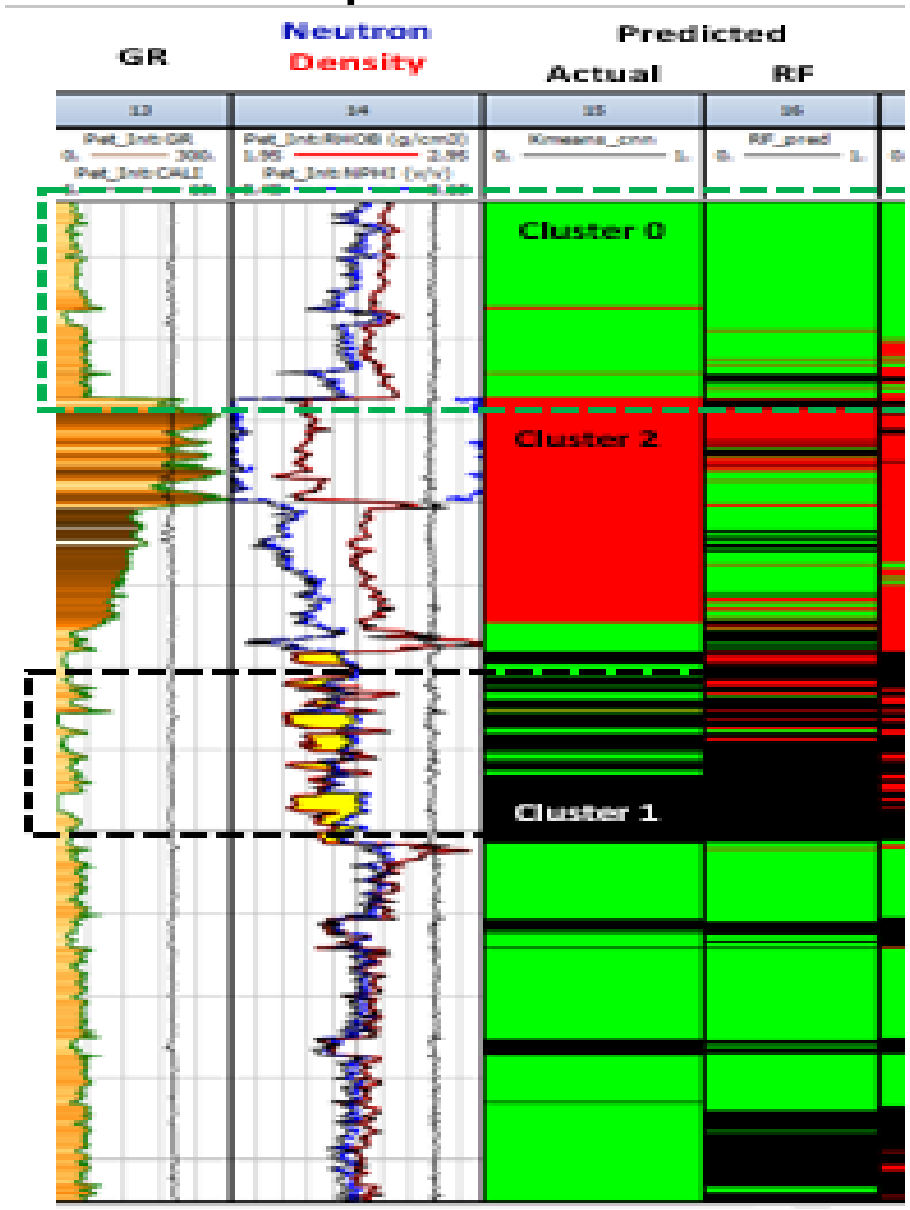


# Random Forest Hyperparameter Optimization and Validation on Log

- Hyperparameter optimization with GridSearch and KFold cross validation

| | param_max_depth | param_min_samples_split | mean_test_score | std_test_score |
|---|---|---|---|---|
| **16** | 8 | 10 | 0.929760 | 0.000478 |
| **15** | 8 | 5 | 0.928707 | 0.000932 |
| **17** | 8 | 15 | 0.928652 | 0.001016 |
| **12** | 7 | 5 | 0.917711 | 0.000801 |
| **13** | 7 | 10 | 0.917603 | 0.001510 |

```
memorization performance:  0.9375945537065052
generalization performance:  0.9349540788762831

             precision    recall  f1-score   support

          0       0.98      0.94      0.95      4985
          1       0.92      0.86      0.89       762
          2       0.89      0.95      0.92      3508

avg / total       0.94      0.93      0.94      9255
```



# Conclusions

- The unsupervised and supervised  techniques shows the great promise
- Unsupervised methods allow the predictions of 3 facies
- Supervised methods
  - KNN gives 94% accuracy in cluster prediction on test dataset
  - SVM gives 94% accuracy in cluster prediction on test dataset
  - Random Forest gives 93% accuracy in cluster prediction on test dataset
- Application to drilling automation and advisory systems
  - Prevent out-of-zone drilling
  - Minimizes rig-time and equipment
  - Cost saving

# References


- The dataset for the current study has been taken from the Volve field on the Norwegian Continental Shelf (NCS) released by Equinor ASA.  Weblink: https://data.equinor.com/authenticate. [70,000 samples. 21 features]

# Workflow

**Log Data – Sonic (DTCO), Gamma Ray (GR), Young Modulus (YME), Poisson ratio (PR), MWD**

- Exploratory Data Analysis
- Missingness
- Feature Engineering

**Identify Brittleness/Frackability Clusters on DT, GR and RHOB**

- K-means and SOM

**Petrophysical Significance of Lithology clusters**

- Identify cluster properties using core data (3 wells)

**Predict Lithology clusters using MWD data**

- *Outlier removal (), Scaling and Centering on 12 wells*
- KNN, Support Vector*Z-score* Machine and Random Forest