Applied Analytics on Surface Drilling and Geomechanical Data

Ngoc Tran, Deepankar Dangwal

Final Project DSA 5103

December 11th 2019

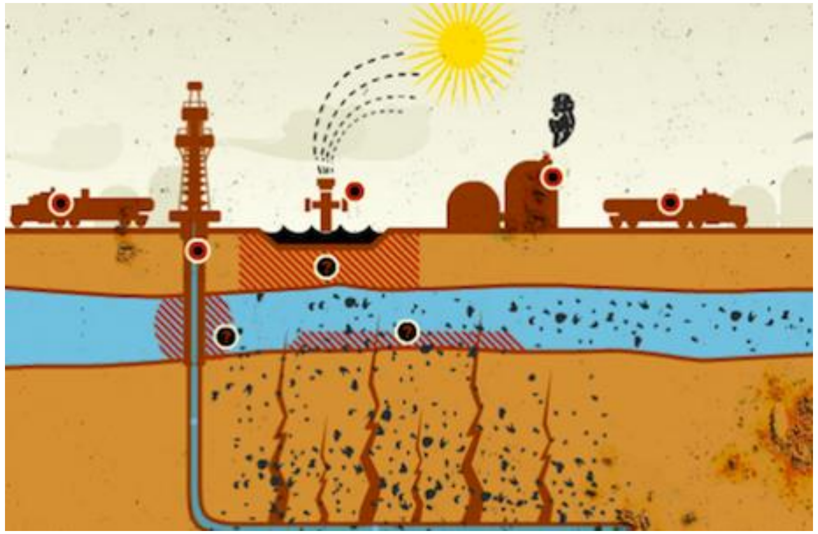
Overview

- Introduction to the Problem
- Data Description
- Data Understanding
- Modeling Approach
- Results and Discussions
- Conclusion

Introduction to the Problems

- More than 14000 horizontal wells fracked in targeted shale plays
- However, a few shale plays have multiple target zone.
- Real-time logging applications: remain in-zone or steer towards a target
- Knowledge of petrophysical/ rock geomechanical properties surrounding drilling bit
- Conventional logging (LWD) -> information with considerable addition costs

Can we use Drilling measurements (MWD data) to interpret "fracking" ability from rock?



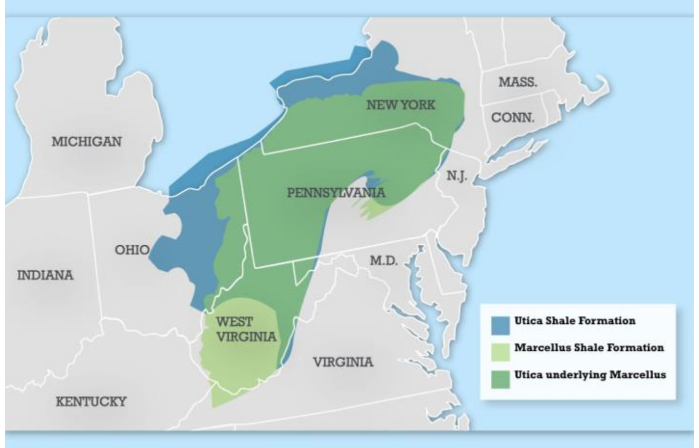
Machine Learning Techniques to predict geomechanical properties by using MWD data

Data Acquisition

Marcellus Shale Energy and Environment Laboratory

- Dataset released by Marcellus Shale Energy and Environment Laboratory and taken from Marcellus Shale Fields, West Virginia

- LWD and MWD logs from horizontal wells
  - LWD Log: Gamma Ray (GR), Young's Modulus (YME), Poisson's ratio (PR), Rock Density log (RHOB), Neutron Porosity (NPHI), Sonic slowness (DTCO)
  - MWD Log: contains 26 attributes - surface rotations per minute (RPM), weight on bit (WOB), Torque, rate of penetration (ROP), Downhole Pressure, Mud Pump Flow Rate, Mechanic Specific Energy , etc.



(Marcellus Shale Fields)

Modeling Schematic

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

- Exploratory Data Analysis (EDA)
- Missingness
- Feature Engineering – Brittleness Index variable

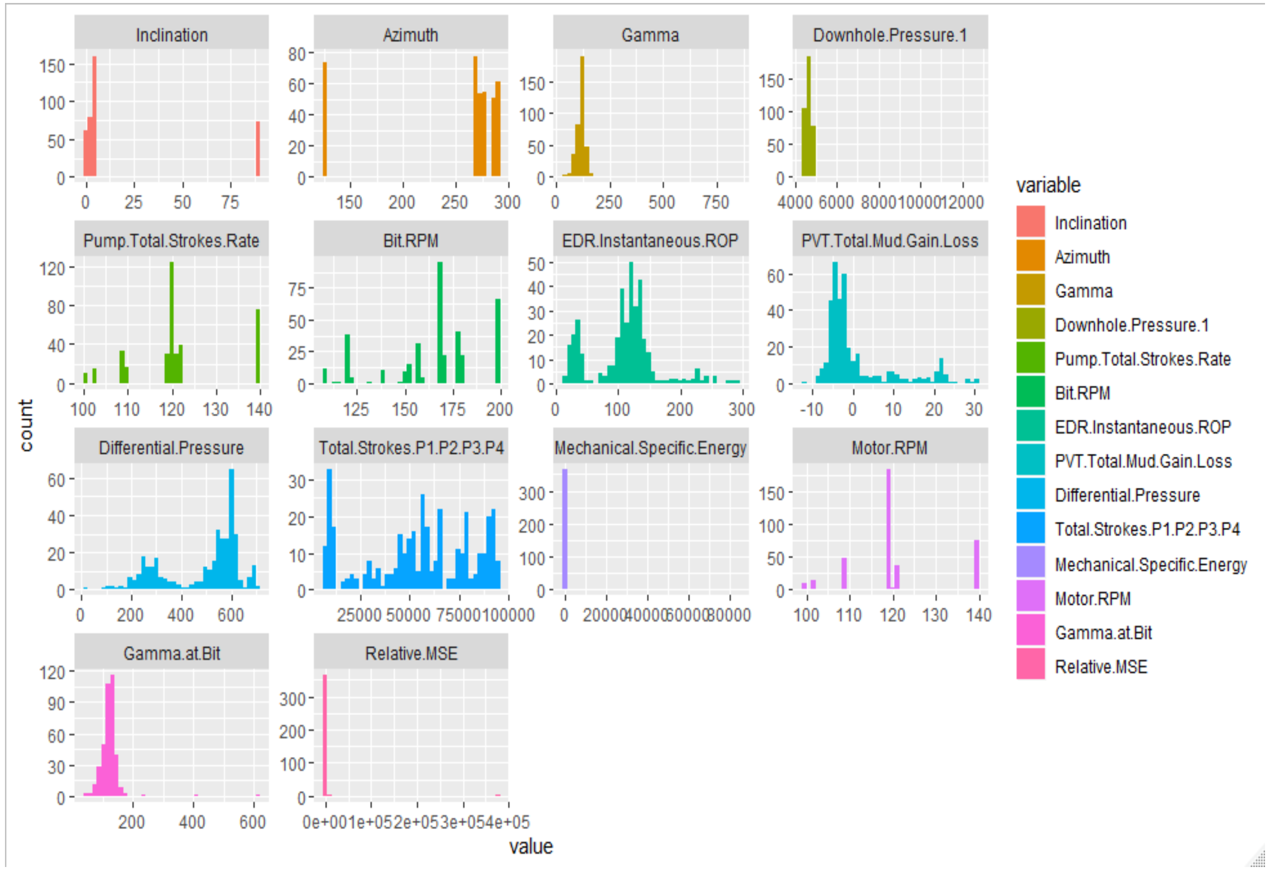
Identify Brittleness/Frackability Clusters on YME, PR, Brittleness Index

- K-means and SOM
- Identify Brittleness/Frackability significance

Predict Brittleness/Frackability clusters using MWD

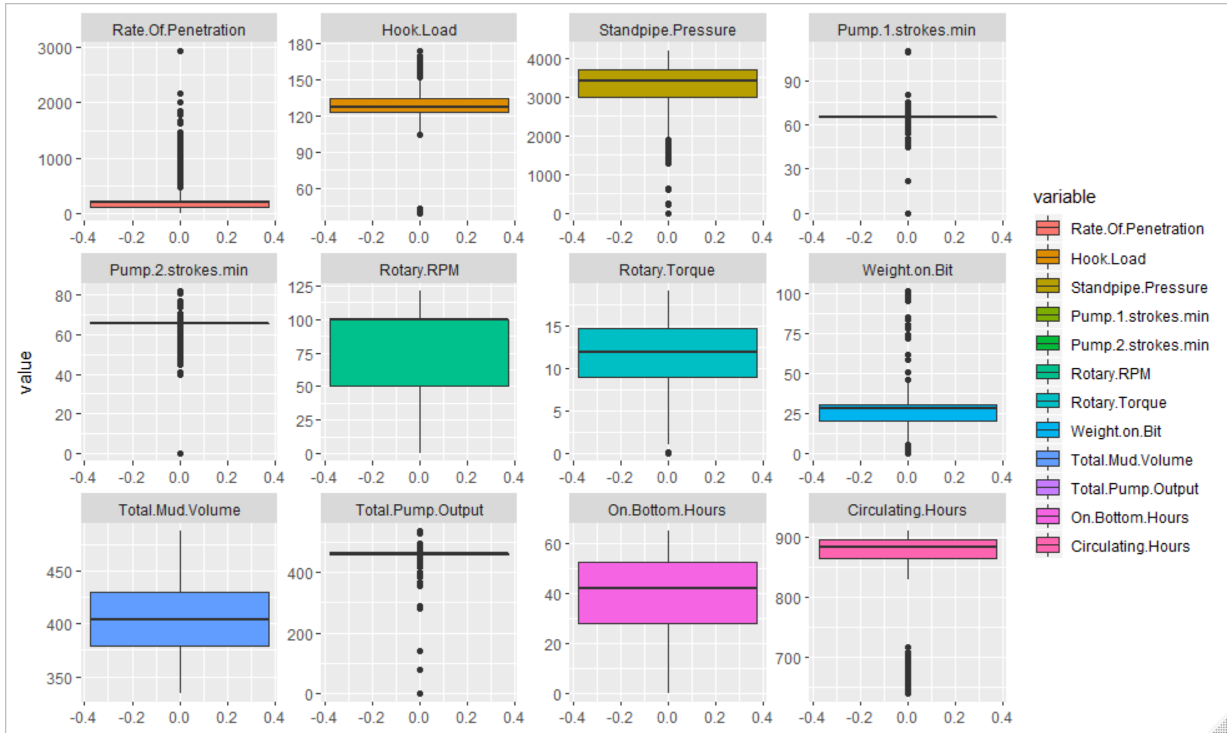
- Sample split data into Train/ Test sets (65-35%), Centering and Scaling
- K Nearest Neighbors , Gradient Boosted Random Forest, Multi Layer Perceptron

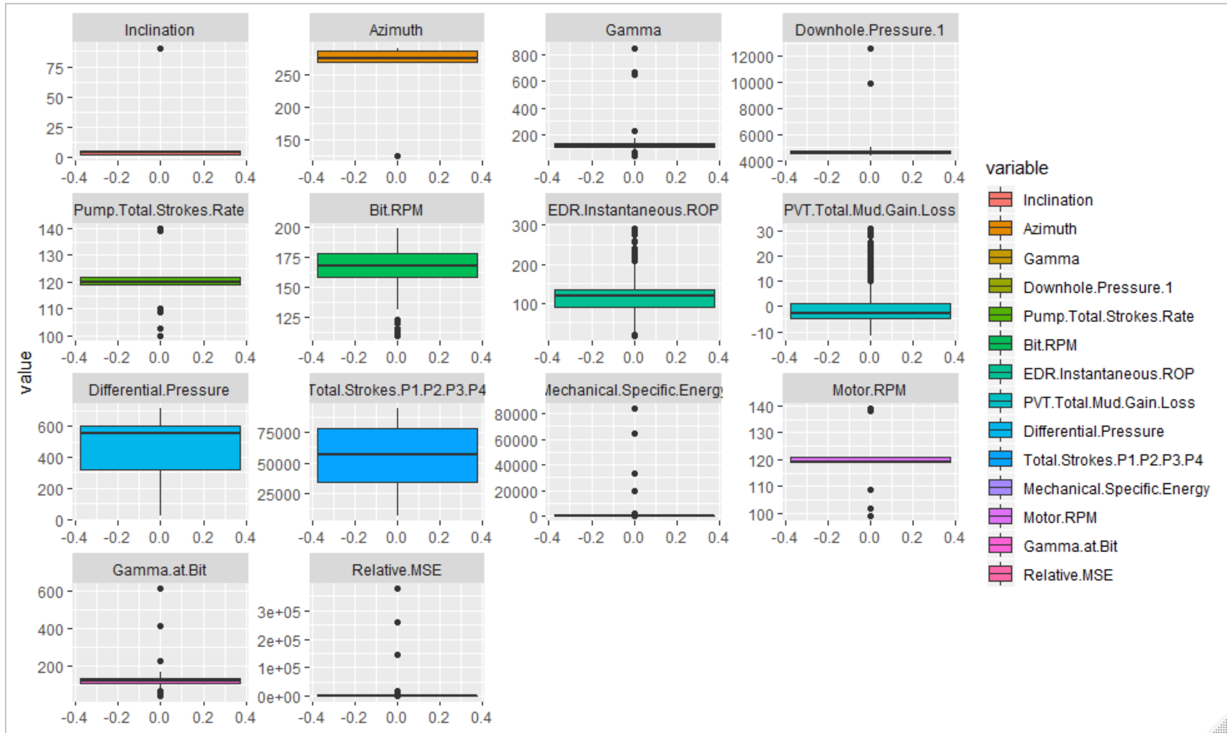
EDA- Data Visualization for Drilling Data (Histogram)



- Total mud Volume is normally distributed
- Rate of Penetration, Hook Load, Standpipe Pressure, Rotary Torque, Weight on Bit, On Bottom Hours, and PVT Total Mud Gain/Loss, Gamma at Bit show the highly left/ right skewness
- District distributions of Rotary RPM, Pump Total Strokes rate, Motor RPM

EDA- Data Visualization for Drilling Data (Boxplots)

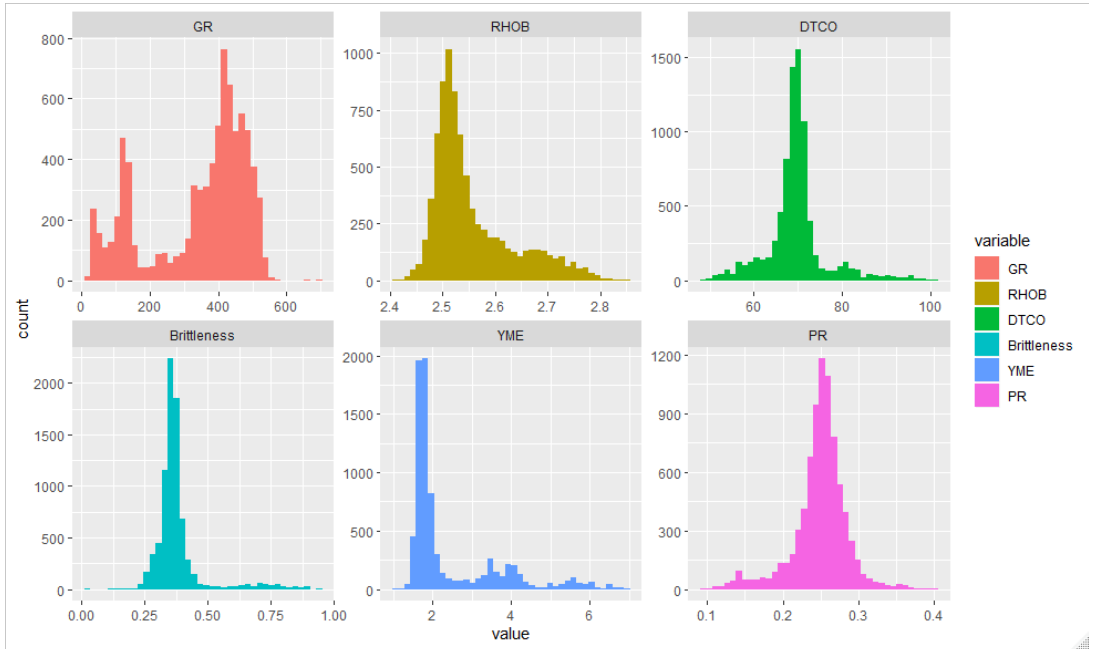




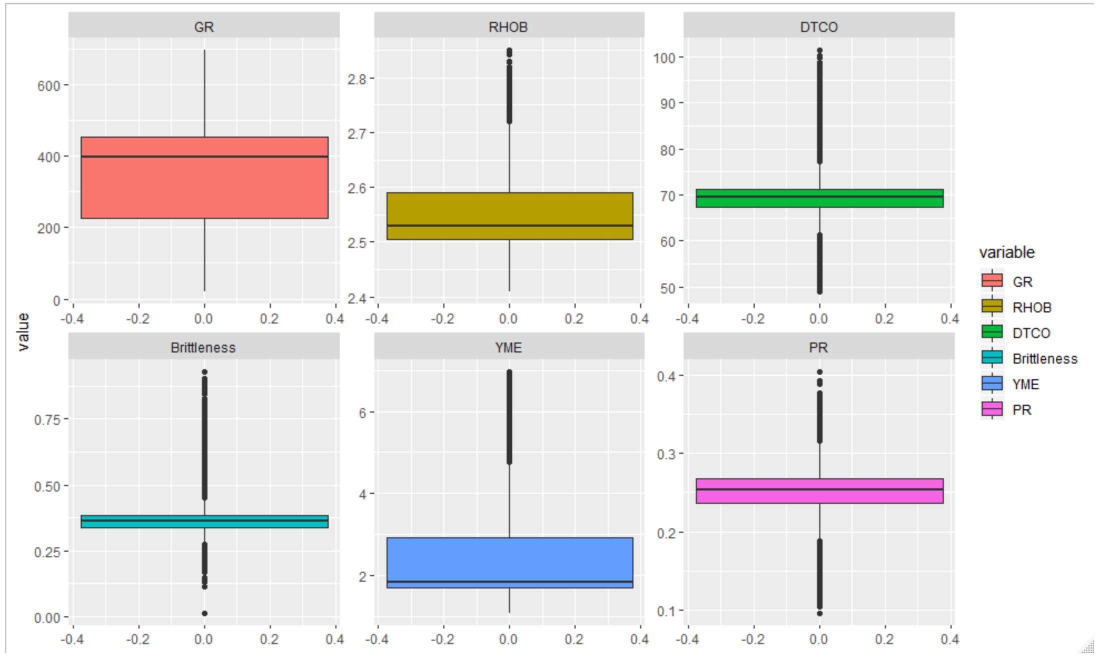
- Boxplots reveal max, min and quartiles
- It also shows outliers

Visualization for Geomechanical Parameters

Histograms for Geomechanical Data

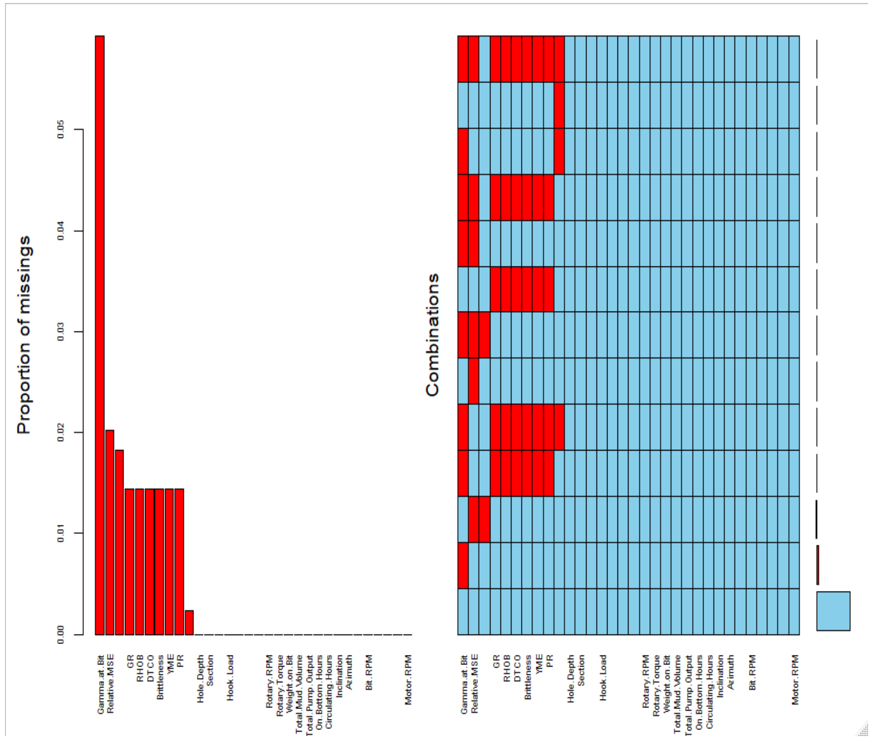


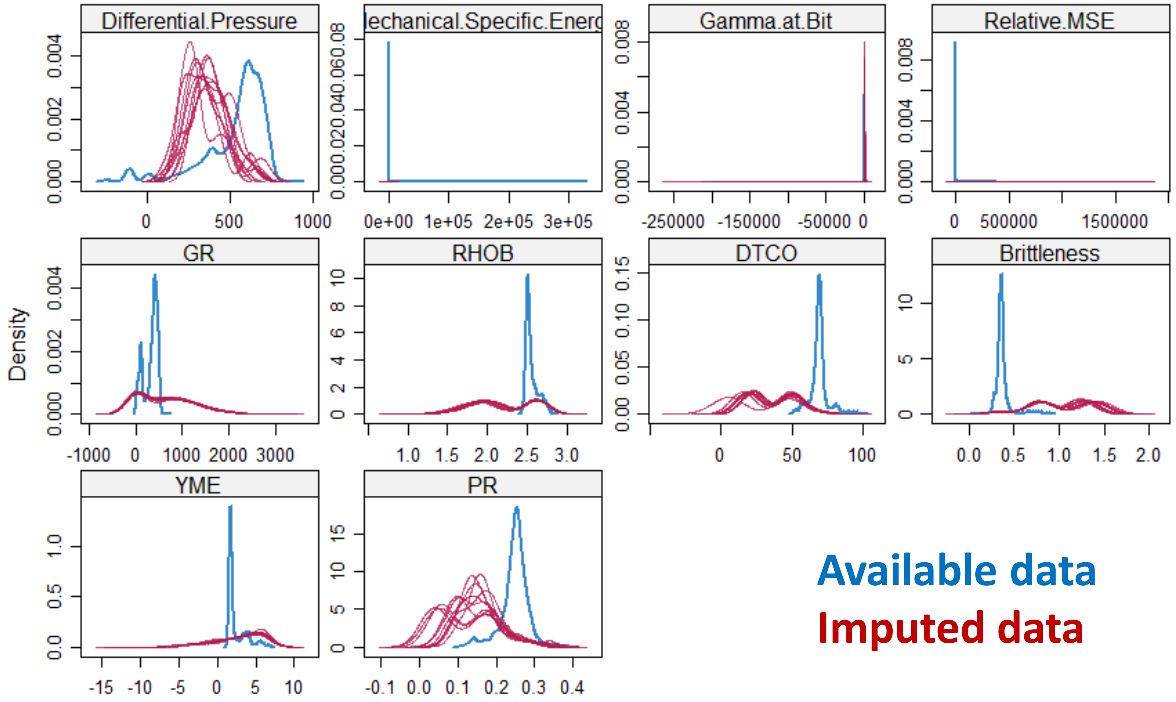
Boxplots for Geomechanical Data



- DTCO, Brittleness and Poisson's ratio are normally distributed
- RHOB and YME, show the highly left/ right skewness
- Outliers on the boxplots for these variables suggest the classification technique to group these points

Missingness Discussion



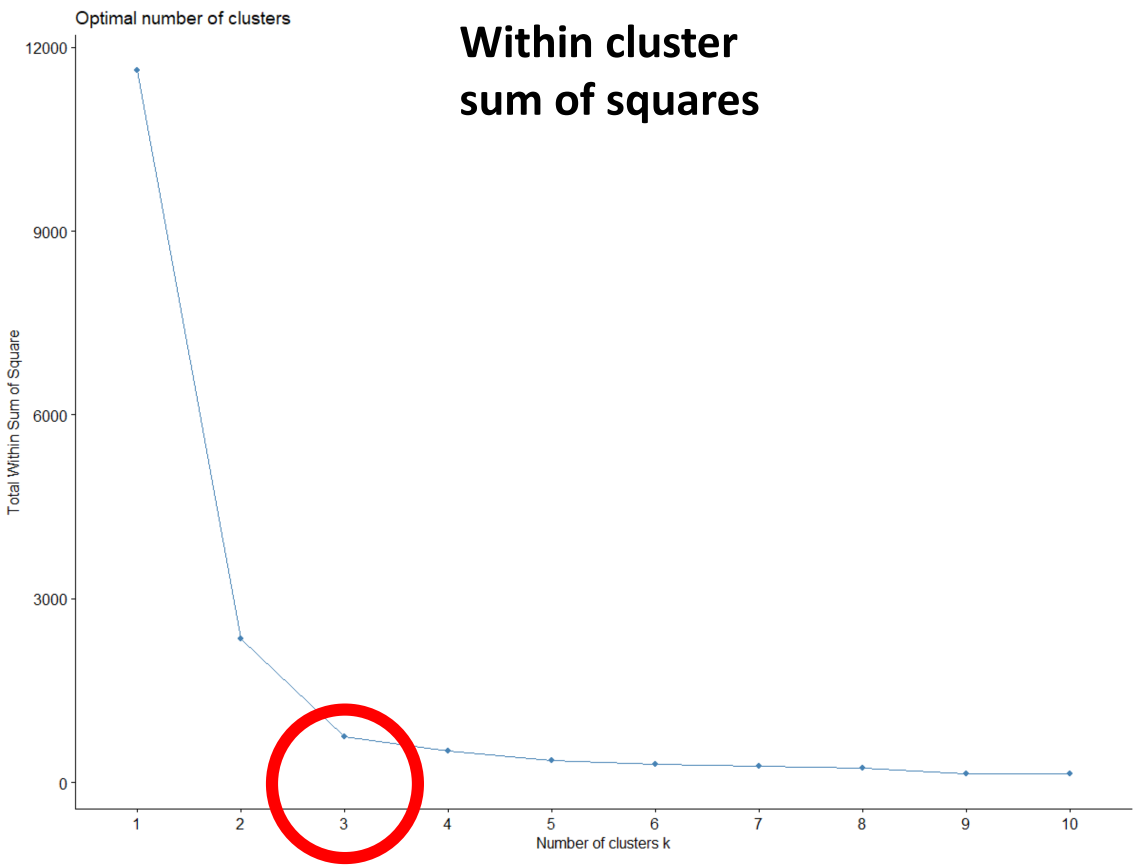


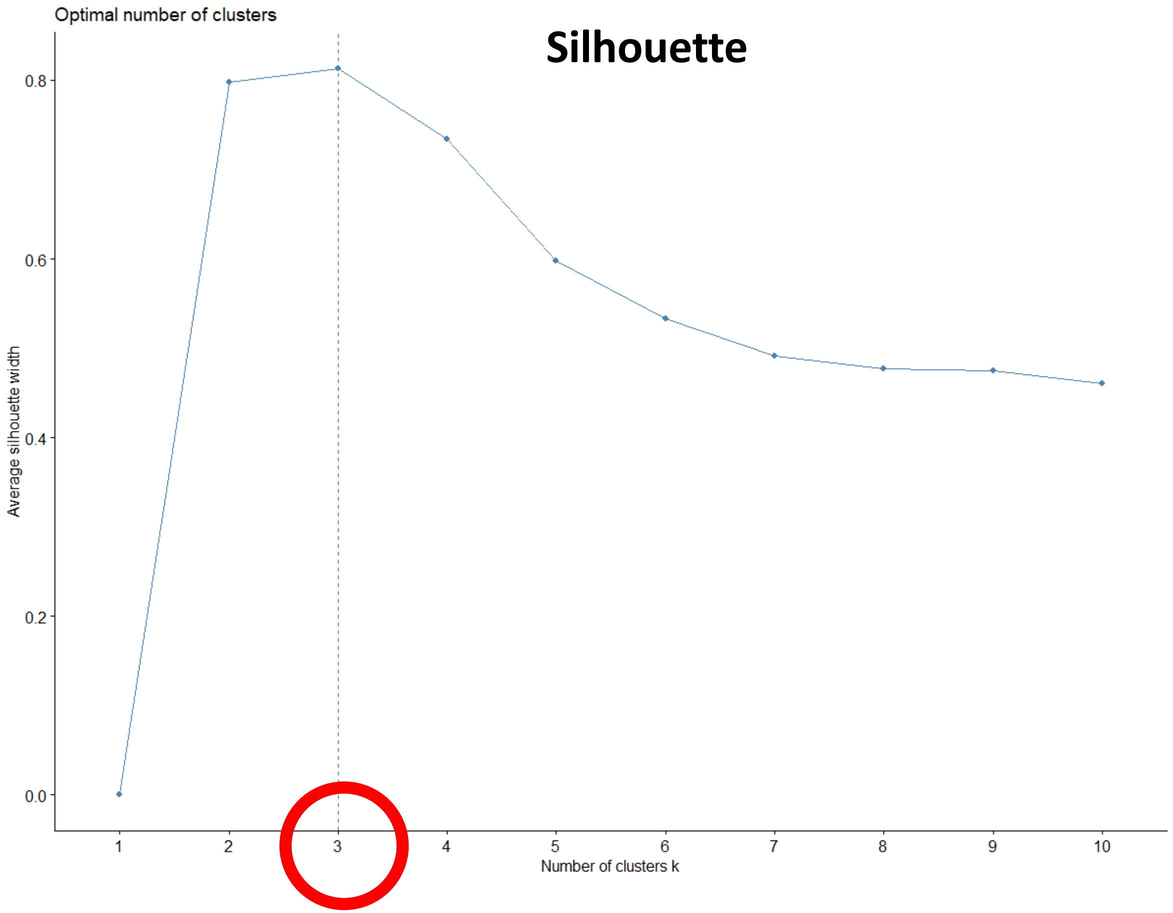
Available data
Imputed data

- GR has the most missing variables
- Probability distributions of imputed data (norm method) does not match distribution for the available data.
- We proceed by removing all the incomplete cases

Removing all incomplete cases

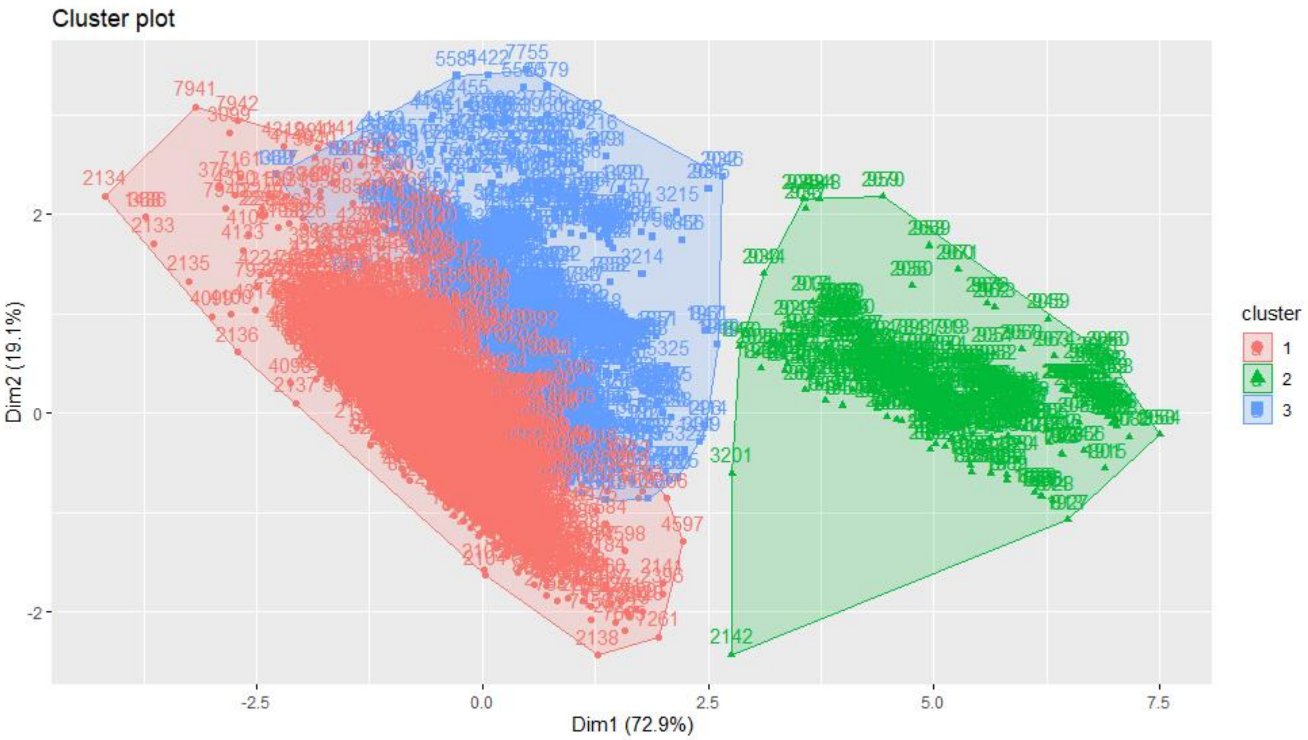
K-Means Optimal Numbers of Clusters

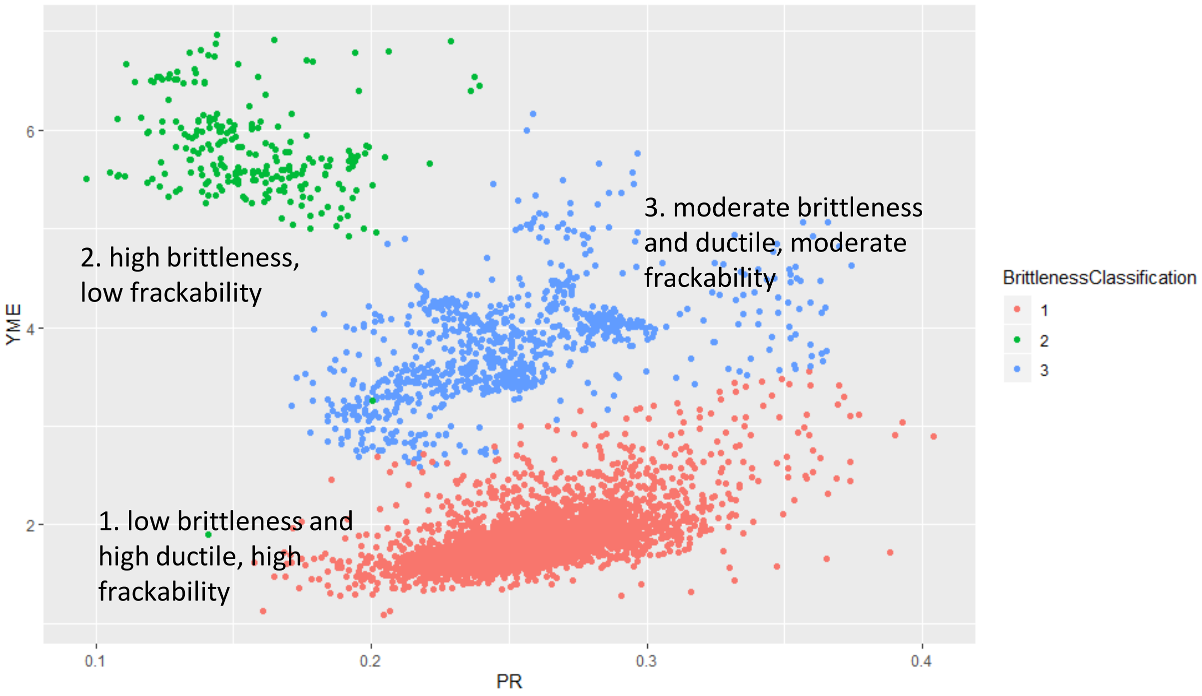




The optimal number of clusters should be 3

K-Means Clustering Results



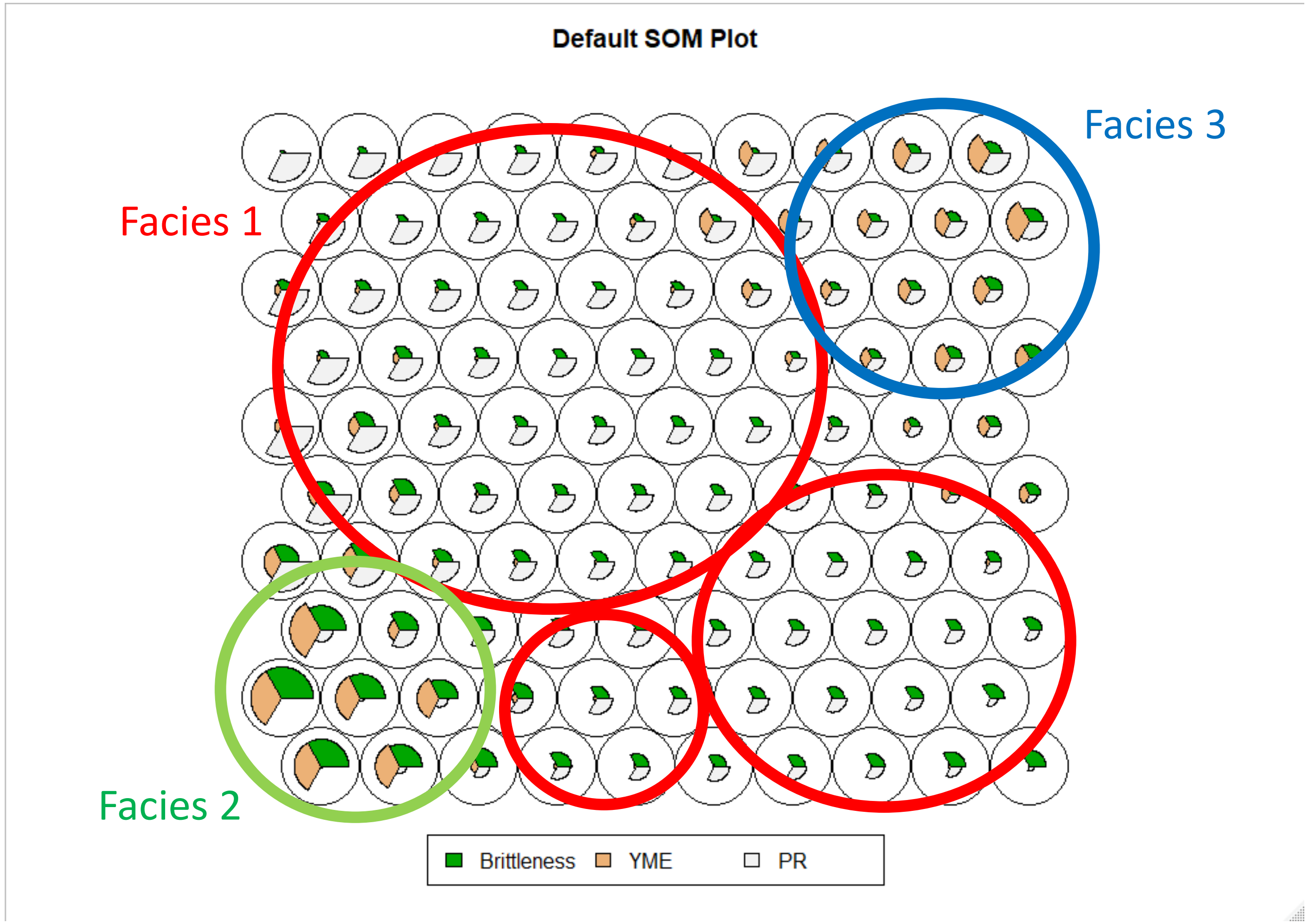


The predicted clusters plotted on principal component 1,2 plane

The predicted clusters plotted on Young's modulus, Poisson's ratio plane.

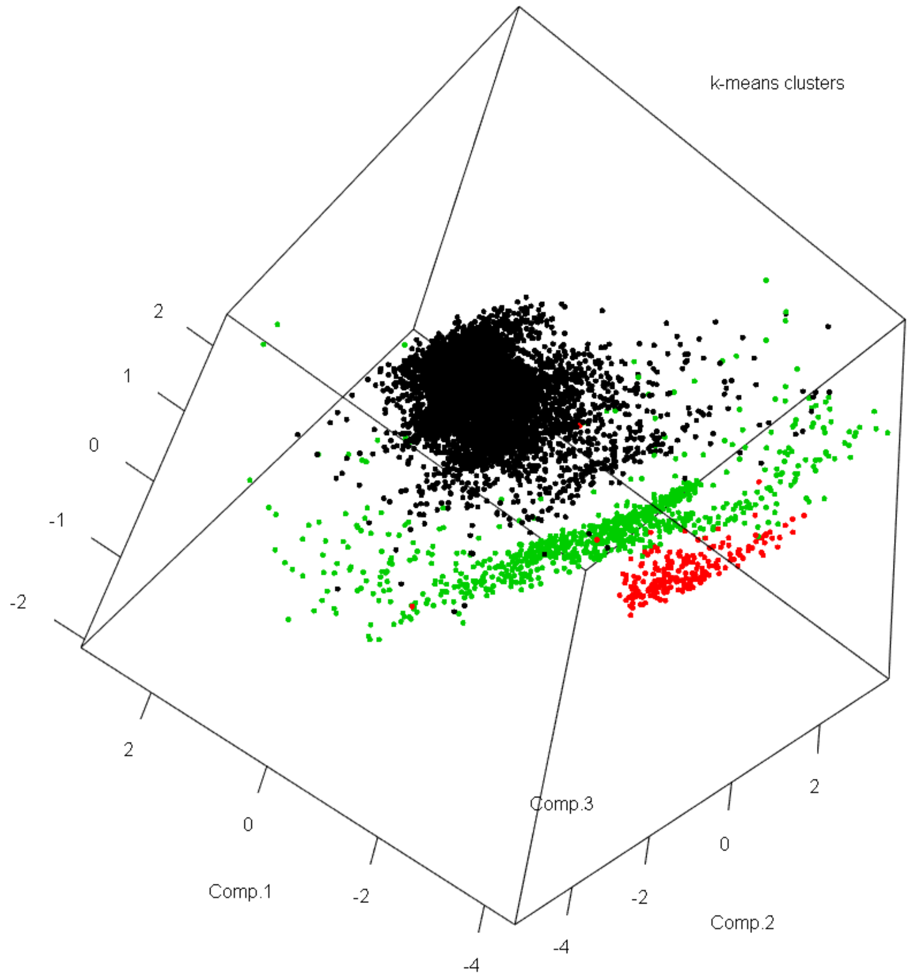
Self-Organizing Map SOM

- The left corner:
  - high brittleness and higher young's modulus
- The middle section
  - lower brittleness/ higher ductile with lower young's modulus and higher poisson's ratio
- The right corner
  - moderate brittleness and ductile, thus moderate frackability.
- Overall, SOM validates 3 clusters predicted from K-Means

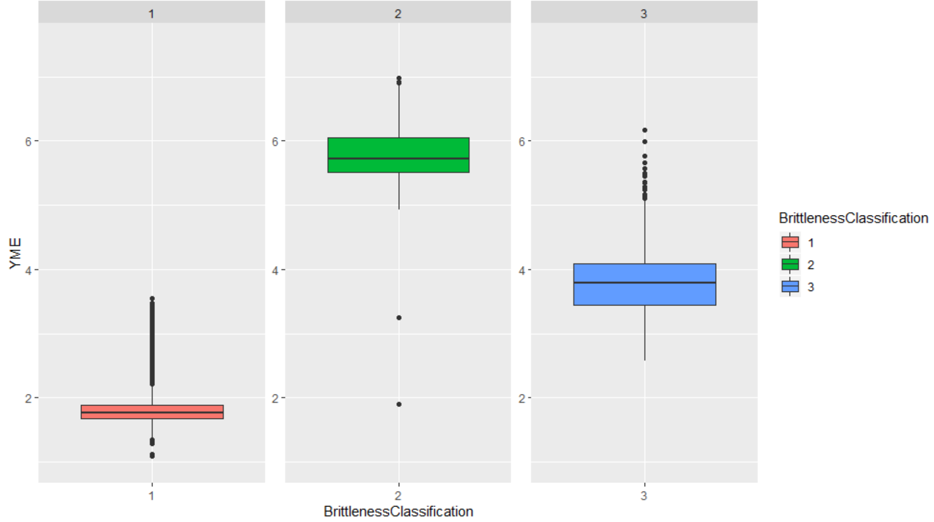


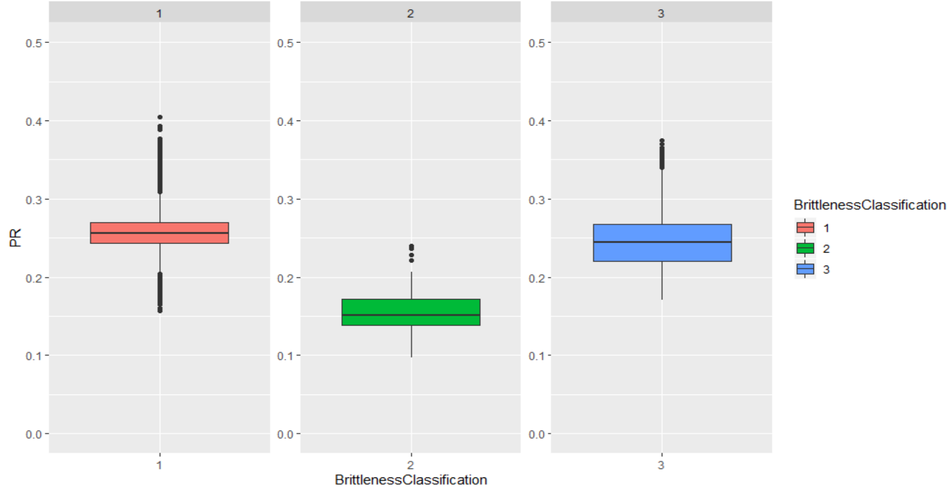
Petrophysical and Geomechanical Significance from 3 clusters

3 Principle Components of Gamma Ray, Density of Rocks and Sonic Slowness



Boxplots of Young's Modulus (YME) and Poisson ratio (PR) for 3 clusters.



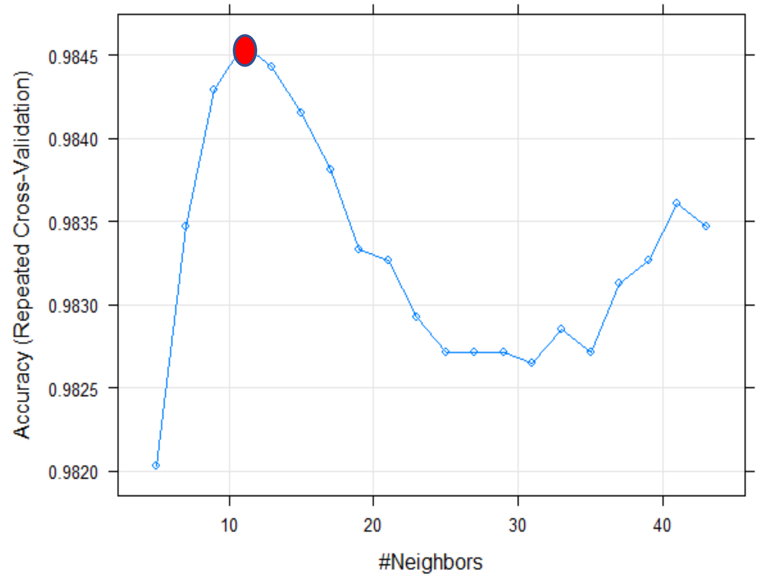


Classifications with Geomechanical Properties

Facies	Properties
1	Low brittleness, high ductile, thus high frackability
2	High brittleness, thus low frackability
3	Moderate brittleness and ductile, thus moderate frackability

Supervised KNN

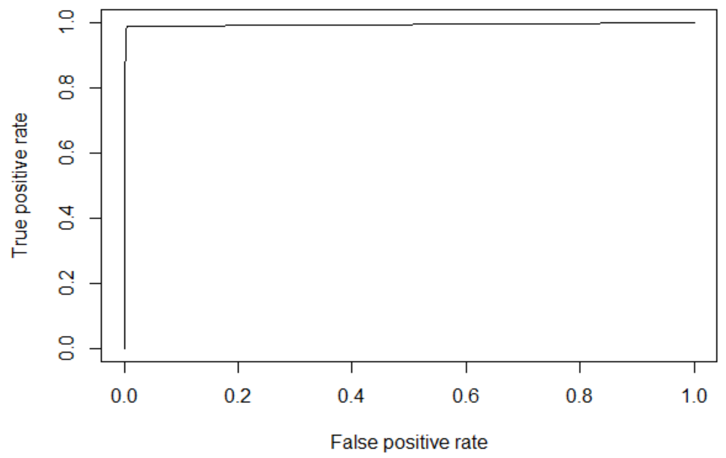
- This technique predicts facies cluster based on MWD
  - Features: MWD Data
  - Target: Geomechanical Facies
- Using random split
  - Train set: 65% Dataset & Test set: 35 % Dataset



- Hyperparameter Optimization by determining optimal numbers of neighbors -> n=11

```
Confusion Matrix and Statistics

pred_knn    1    2    3
       1 2037    2   37
       2    0   88    6
       3    6    0  442

Overall Statistics

               Accuracy : 0.9805
                 95% CI : (0.9745, 0.9855)
    No Information Rate : 0.7804
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9441

 Mcnemar's Test P-Value : 1.165e-06

Statistics by Class:

                     Class: 1 Class: 2 Class: 3
Sensitivity            0.9971  0.97778   0.9113
Specificity            0.9322  0.99763   0.9972
Pos Pred Value         0.9812  0.93617   0.9866
Neg Pred Value         0.9889  0.99921   0.9802
Prevalence             0.7804  0.03438   0.1853
Detection Rate         0.7781  0.03361   0.1688
Detection Prevalence   0.7930  0.03591   0.1711
Balanced Accuracy      0.9646  0.98770   0.9543
```

- Accuracy: 0.9805
- Kappa: 0.9441



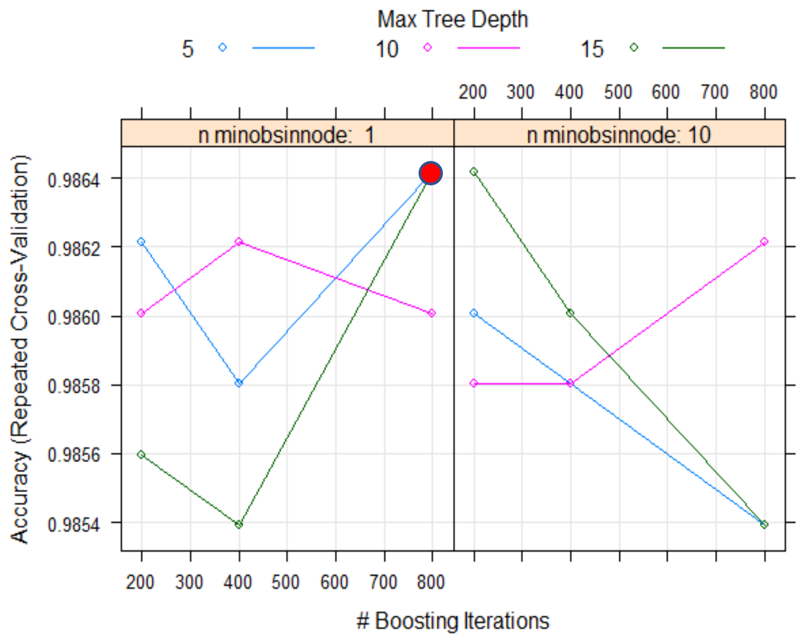
ROC Curve

Supervised Techniques- Gradient Boosted Random Forest

Features: MWD Data

Target: Geomechanical Facies

- Hyperparameters include:
  - Number of trees (Boosting iterations) => Number of trees = 800
  - Interaction depth (Max tree depth) => Max tree depth = 5
  - Minimum number of nodes in last level of a tree (n.minobsinmode) => n.minobsinmode =1




```
Confusion Matrix and Statistics

pred_gbm    1    2    3
       1 2037    2   32
       2    0   87    1
       3    6    1  452

Overall Statistics

               Accuracy : 0.984
                 95% CI : (0.9784, 0.9884)
    No Information Rate : 0.7804
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.954

 Mcnemar's Test P-Value : 0.0001877

Statistics by Class:

                     Class: 1 Class: 2 Class: 3
Sensitivity            0.9971  0.96667   0.9320
Specificity            0.9409  0.99960   0.9967
Pos Pred Value         0.9836  0.98864   0.9847
Neg Pred Value         0.9890  0.99881   0.9847
Prevalence             0.7804  0.03438   0.1853
Detection Rate         0.7781  0.03323   0.1727
Detection Prevalence   0.7911  0.03361   0.1753
Balanced Accuracy      0.9690  0.98314   0.9643
```

- Accuracy: 0.984
- Kappa: 0.9454

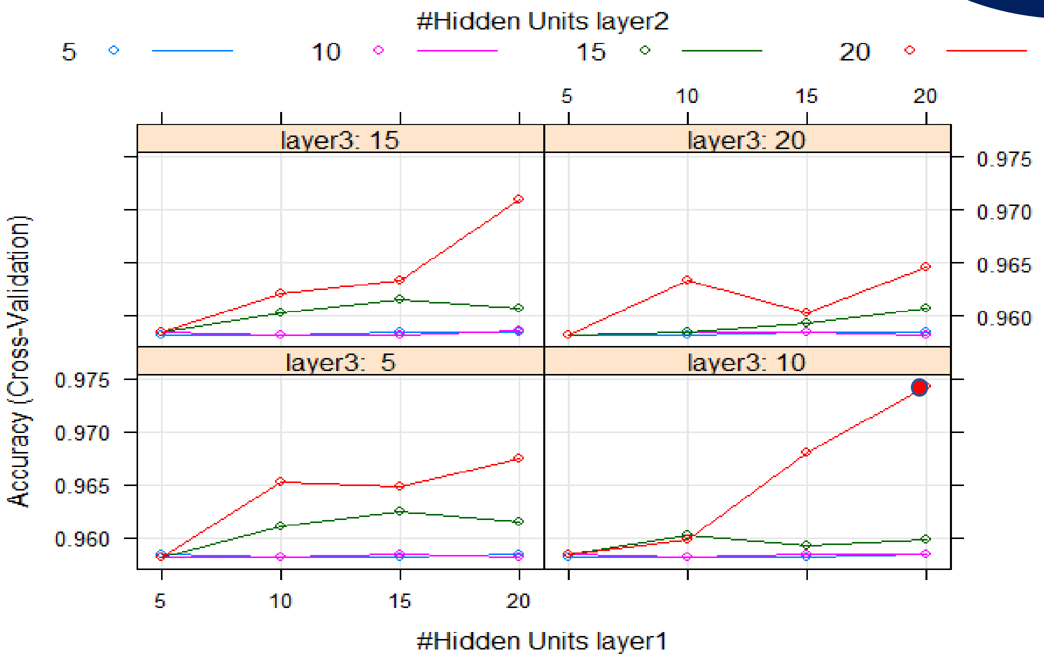
Supervised Techniques- Neural Networks multi-layer perceptron

Features: MWD Data

Target: Geomechanical Facies

- Hyper-parameters include:
  - Number of nodes in layer 1 = 20
  - Number of nodes in layer 2 = 20
  - Number of nodes in layer 3 = 20

```
Confusion Matrix and Statistics


pred_mlp    1    2    3
       1 2037    1   37
       2    3   89    1
       3    3    0  447

Overall Statistics

               Accuracy : 0.9828
                 95% CI : (0.9771, 0.9874)
    No Information Rate : 0.7804
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9507

 Mcnemar's Test P-Value : 8.923e-07

Statistics by Class:

                     Class: 1 Class: 2 Class: 3
Sensitivity            0.9971  0.98889   0.9216
Specificity            0.9339  0.99842   0.9986
Pos Pred Value         0.9817  0.95699   0.9933
Neg Pred Value         0.9890  0.99960   0.9825
Prevalence             0.7804  0.03438   0.1853
Detection Rate         0.7781  0.03400   0.1707
Detection Prevalence   0.7926  0.03552   0.1719
Balanced Accuracy      0.9655  0.99365   0.9601
```



- Accuracy = 0.9828
- Kappa = 0.9507
- This technique predicts most numbers of Facies #3 (moderately good zone for fracking)

3 Supervised Techniques Sum Up

FACIES/ CLUSTERS KNN		Predicted		
		1	2	3
Actual	1	2037	2	37
	2	0	88	6
	3	6	0	442

FACIES/ CLUSTERS GBM Random Forest		Predicted		
		1	2	3
Actual	1	2037	2	32
	2	0	87	1
	3	6	1	452

FACIES/CLUSTERS MLP Neural Networks		Predicted		
		1	2	3
Actual	1	2037	1	37
	2	3	**89**	1
	3	3	0	**447**

- 3 methods give consistent accuracy of 0.98 and kappa of 0.95
- 3 methods predict similar numbers of Facies 1 with 2037 datapoints
- MLP Neural Networks predicts most numbers of Facies 2 and 3 with 89 and 447 datapoints
- Need more data obtained from same/ different basin to generate full-field scale model

Conclusions

- Feature engineering, Exploratory Data Analysis and Missingness Handling, offer insights to the combined set of more than 263000 records
- The unsupervised and supervised techniques demonstrate the great promise.
  - Unsupervised methods allow the predictions of 3 facies
  - KNN gives 98% accuracy in cluster prediction on test dataset
  - Gradient Boosting gives 98% accuracy in cluster prediction on test dataset
  - Neural Networks gives 98% accuracy in cluster prediction on test dataset, forecasts the most numbers of moderate-good facies for fracking
- The application to drilling automation and advisory systems to prevent out-of-zone drilling, minimize rig-time and equipment and allow cost saving

THANK YOU

References

Marcellus Shale Energy and Environment Laboratory. http://mseel.org/