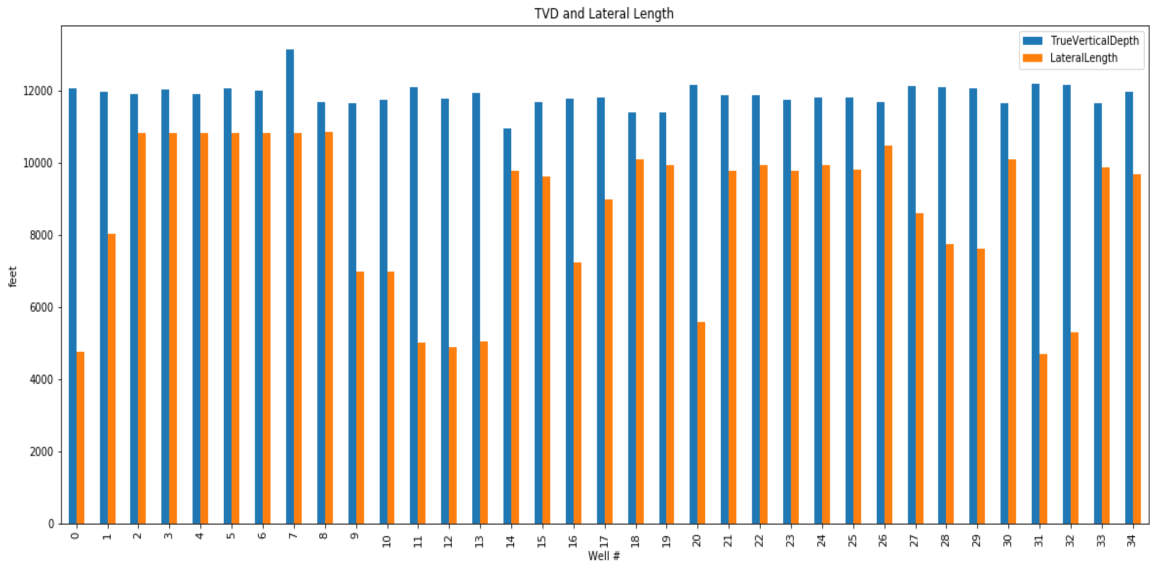
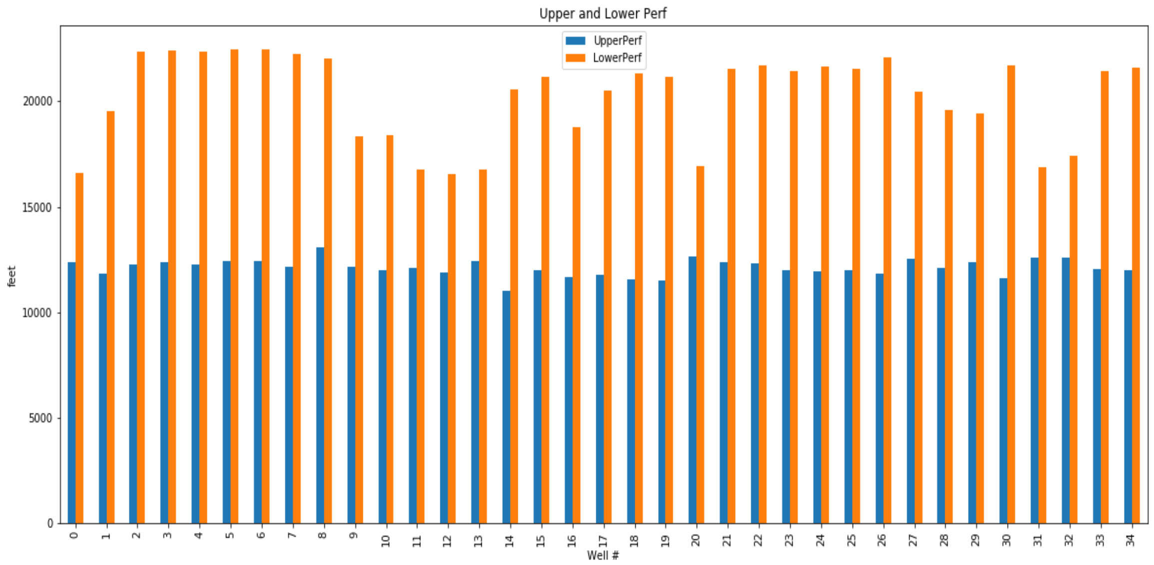
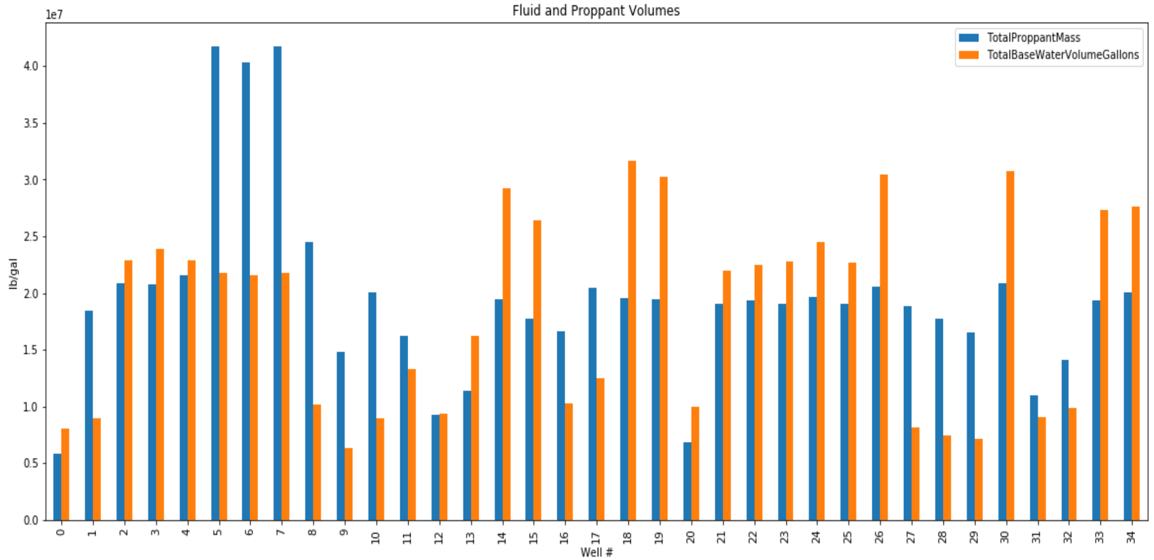
Applying Data Science to Optimize and Maximize Hydrocarbon Recovery from Oil and Gas wells in North America

Data Pre-processing

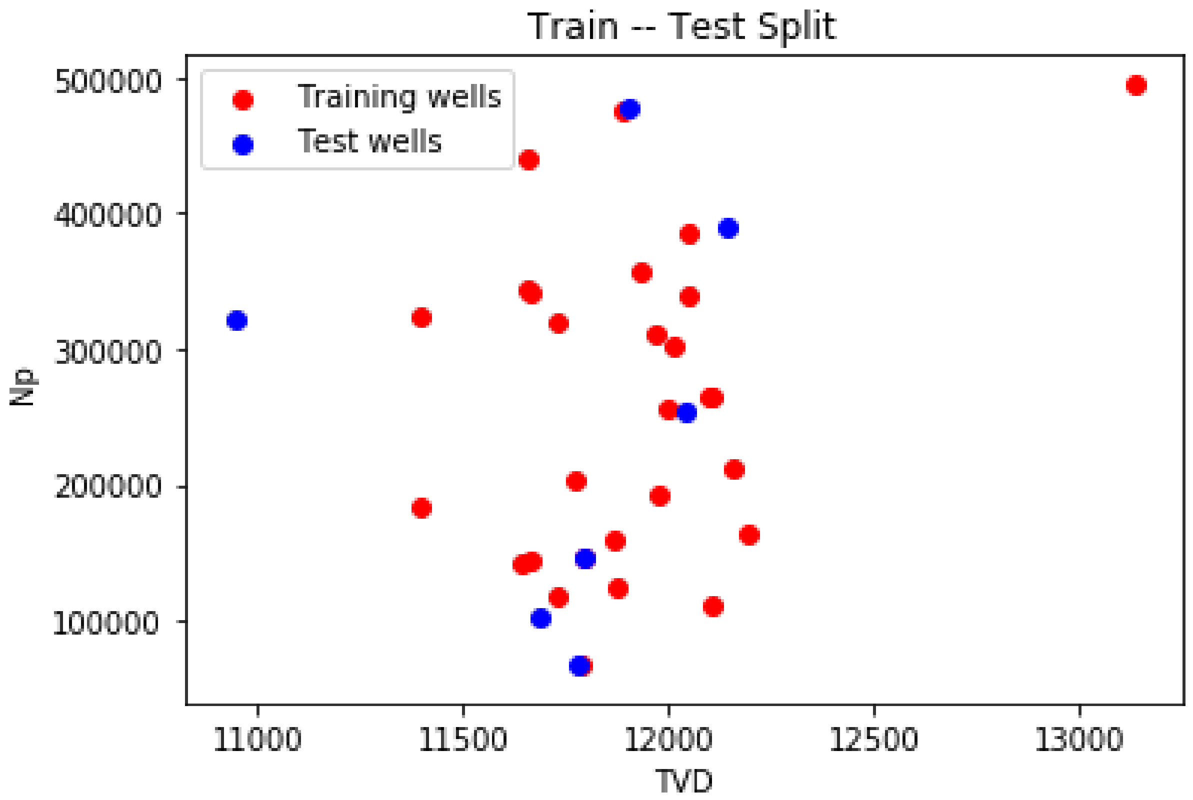
Input Drivers





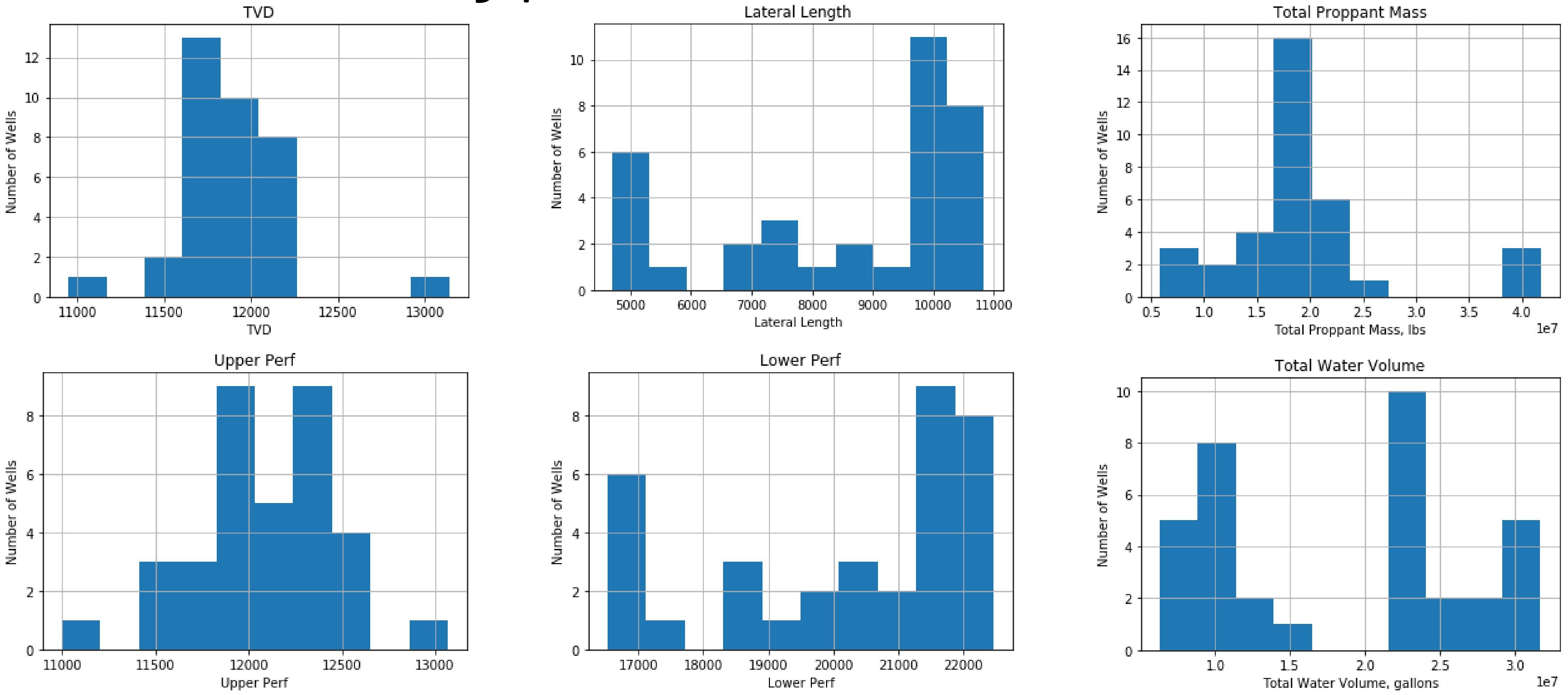


Splitting Dataset in Training and Test Sets (80-20%)

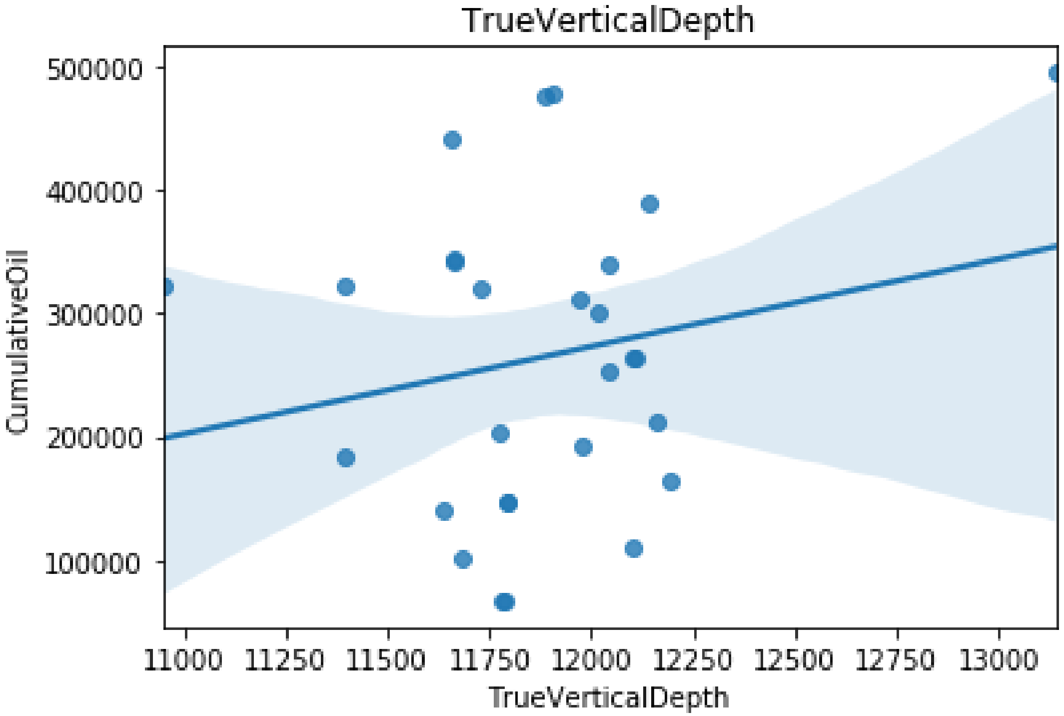


Exploratory Data Analysis

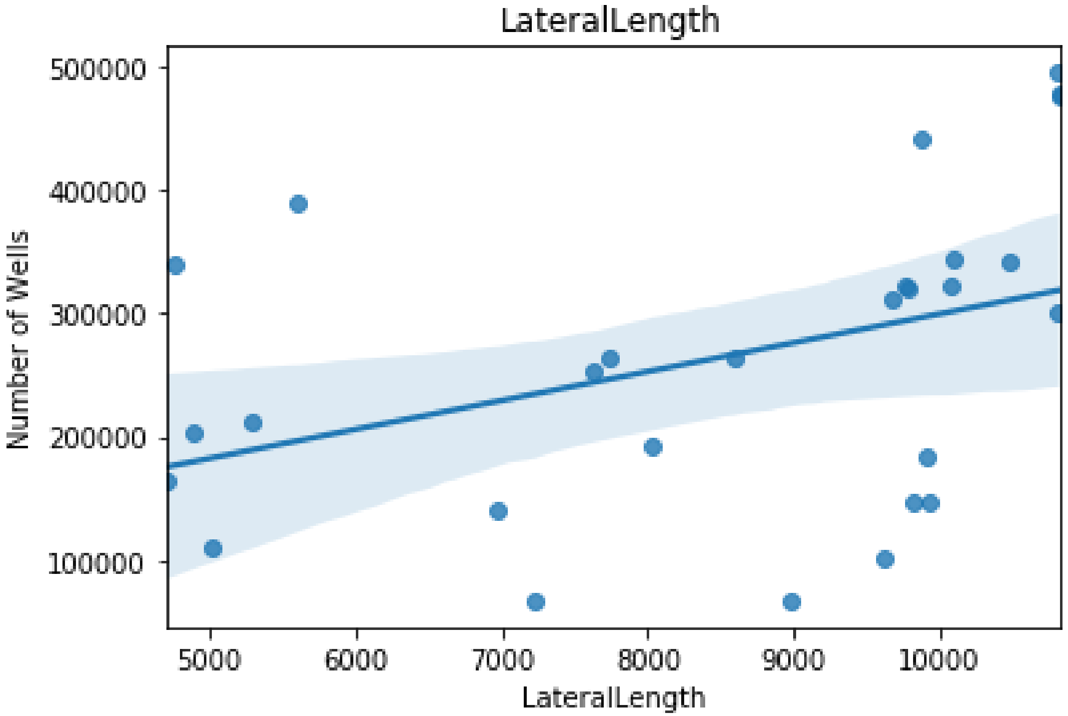
Input Features Histogram



Input Feature vs. Target -- Correlation

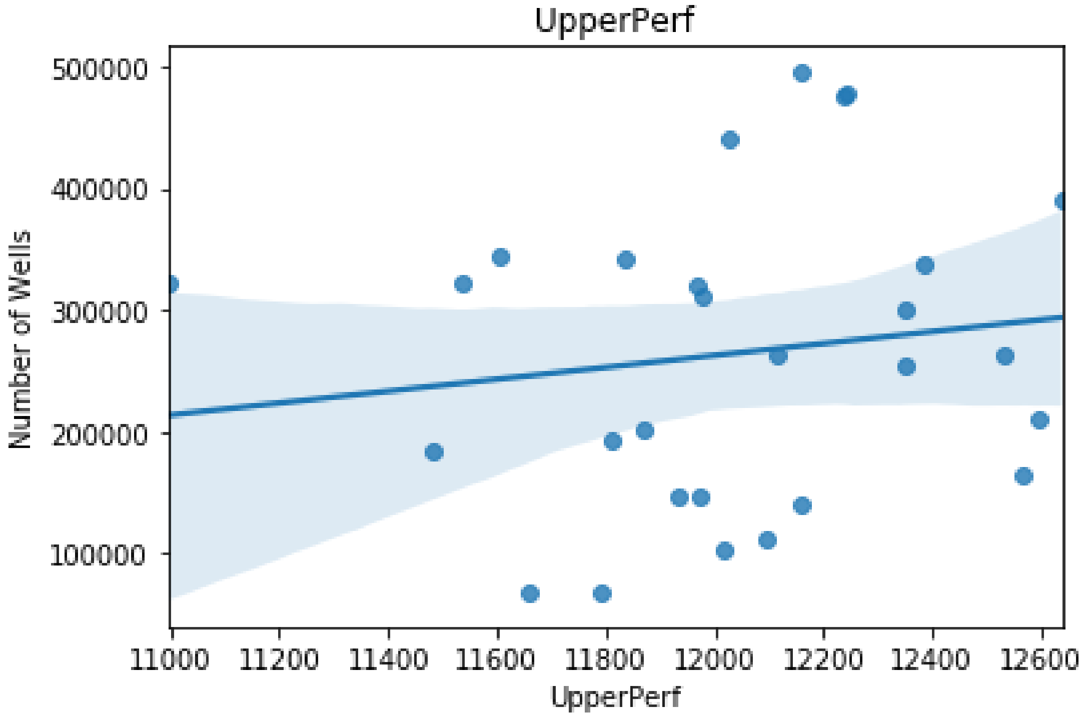


Positive Non-Linear

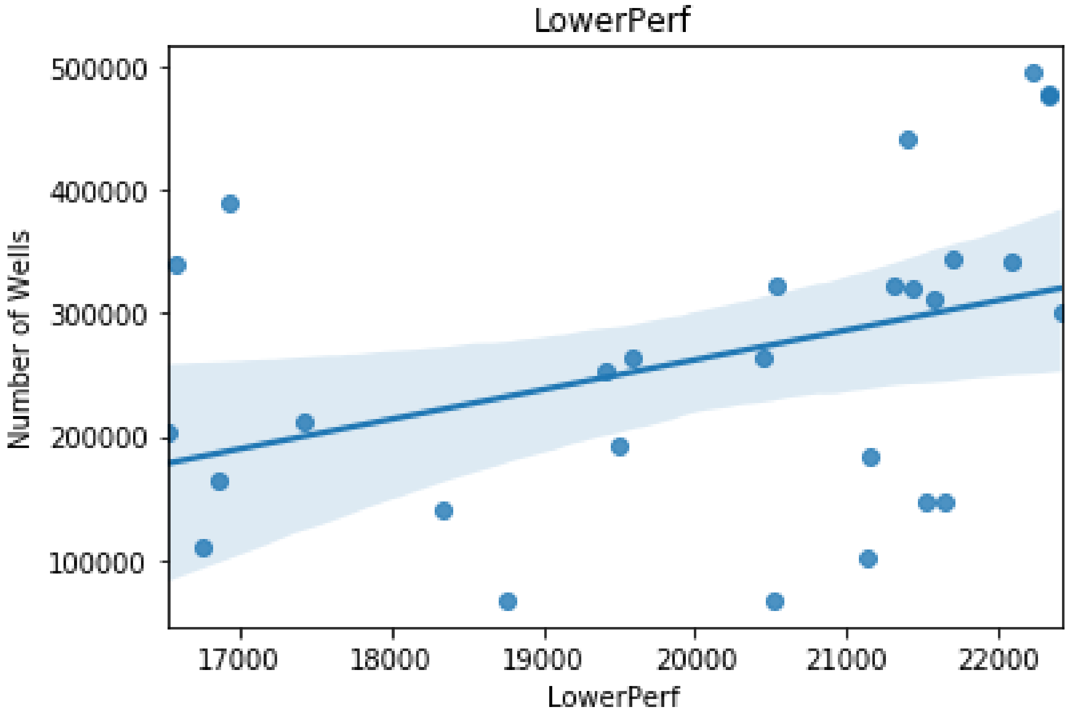


Positive Non-Linear

Input Feature vs. Target -- Correlation

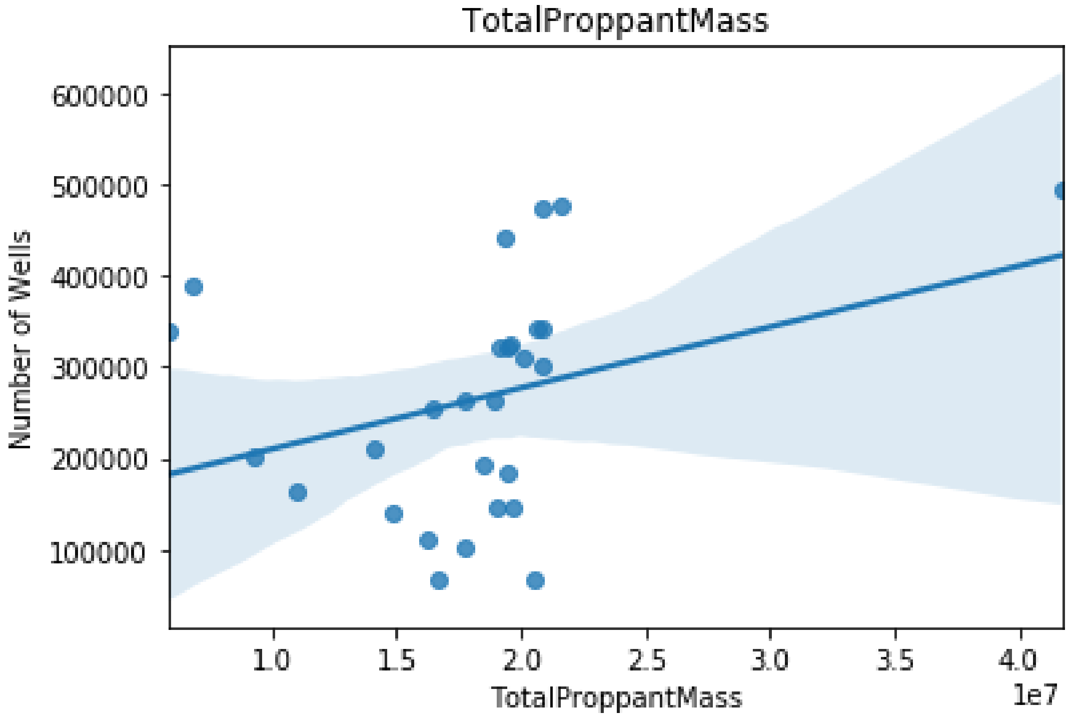


Positive Non-Linear

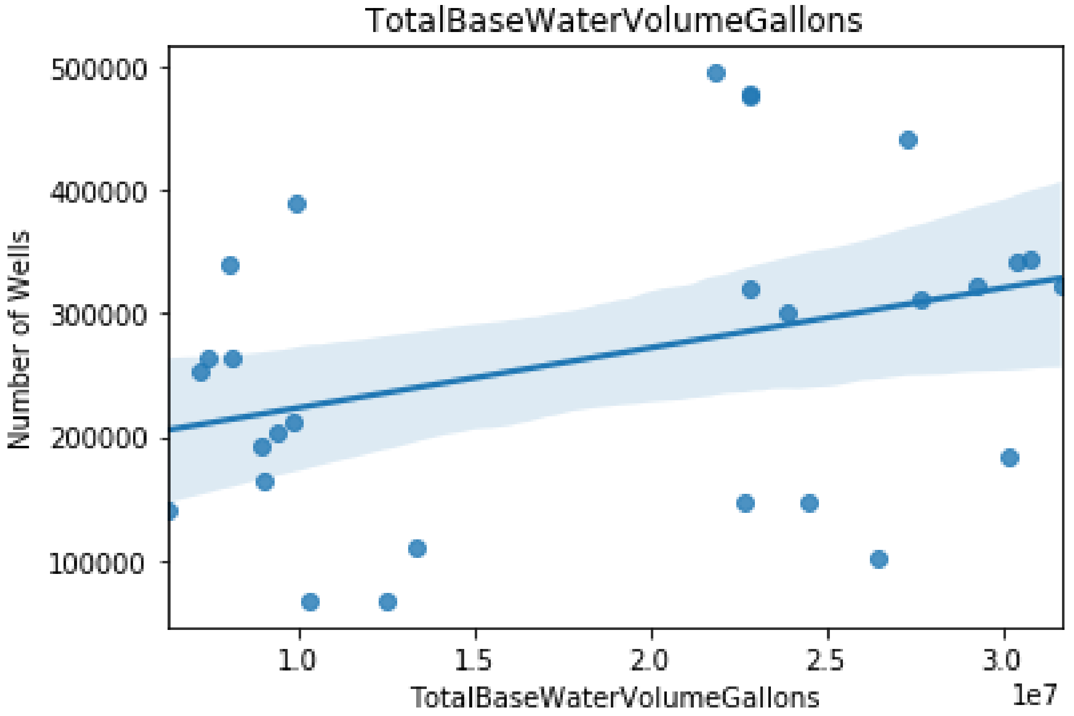


Positive Non-Linear

Input Feature vs. Target -- Correlation

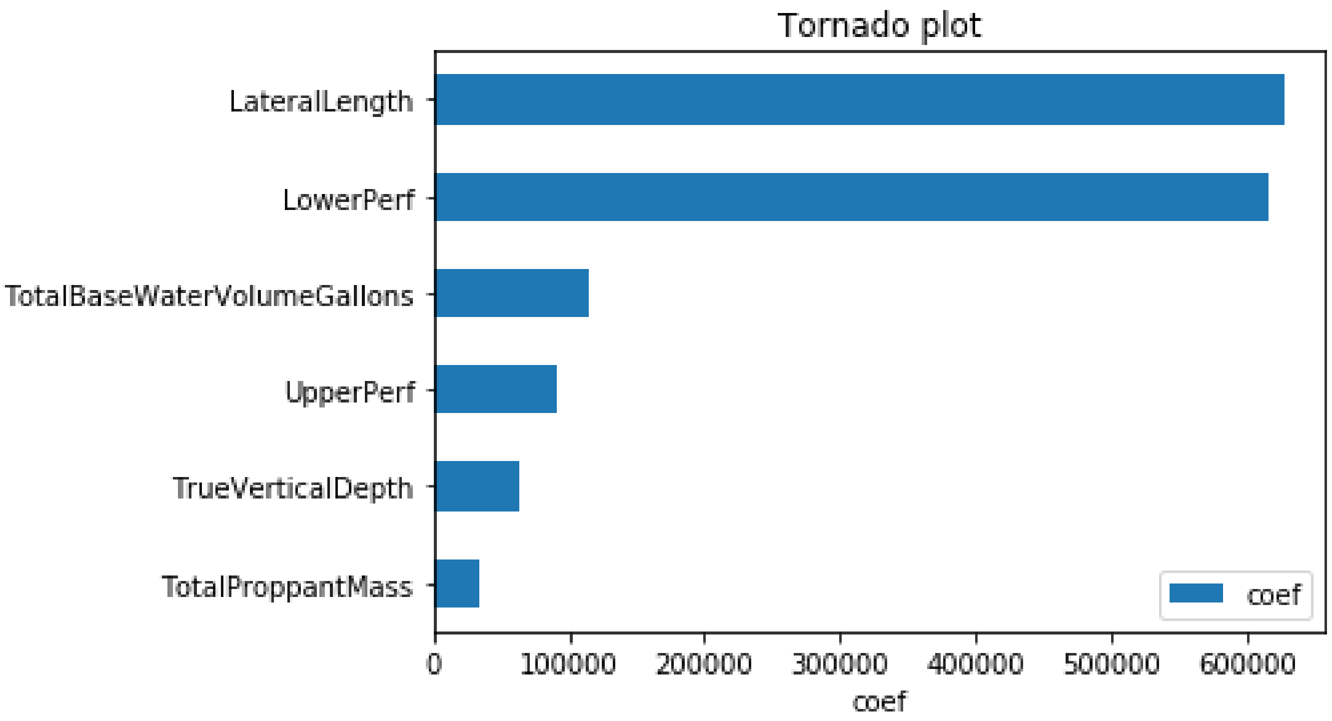


Positive Non-Linear



Positive Linear

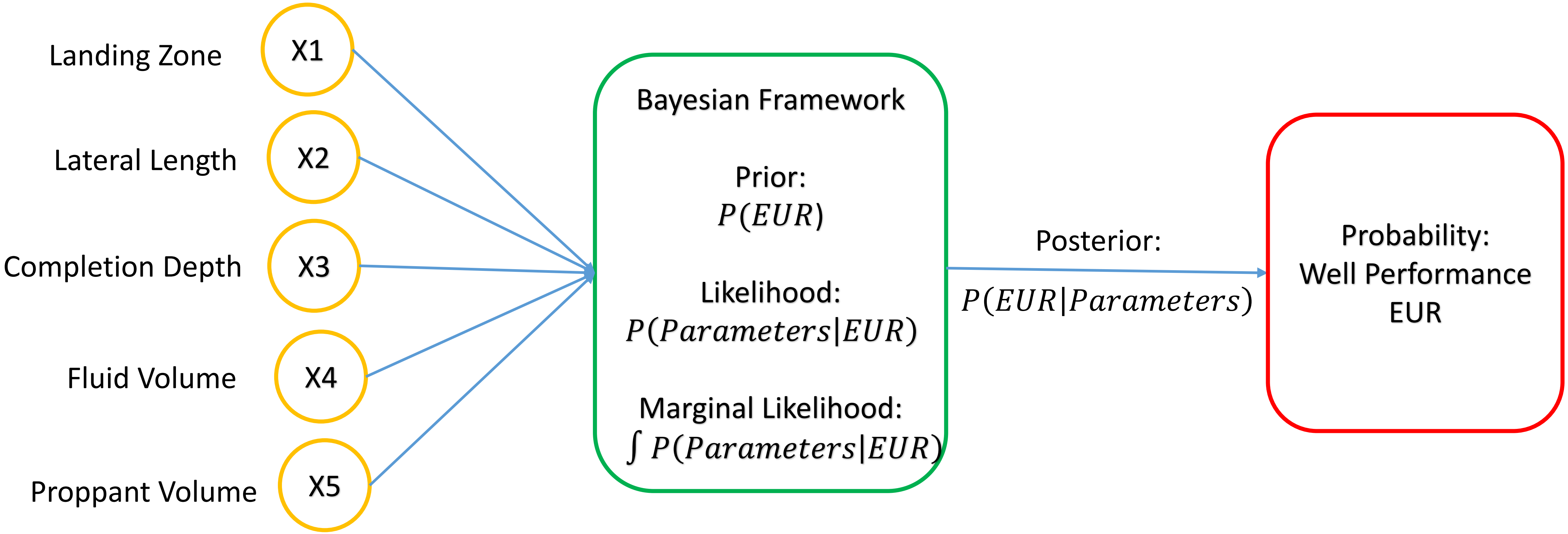
Identifying Significant Drivers – Tornado plots



- The tornado plots show the significance level of various operational parameters
- Lateral length, lower perf location and fluid volume being the most significant parameters

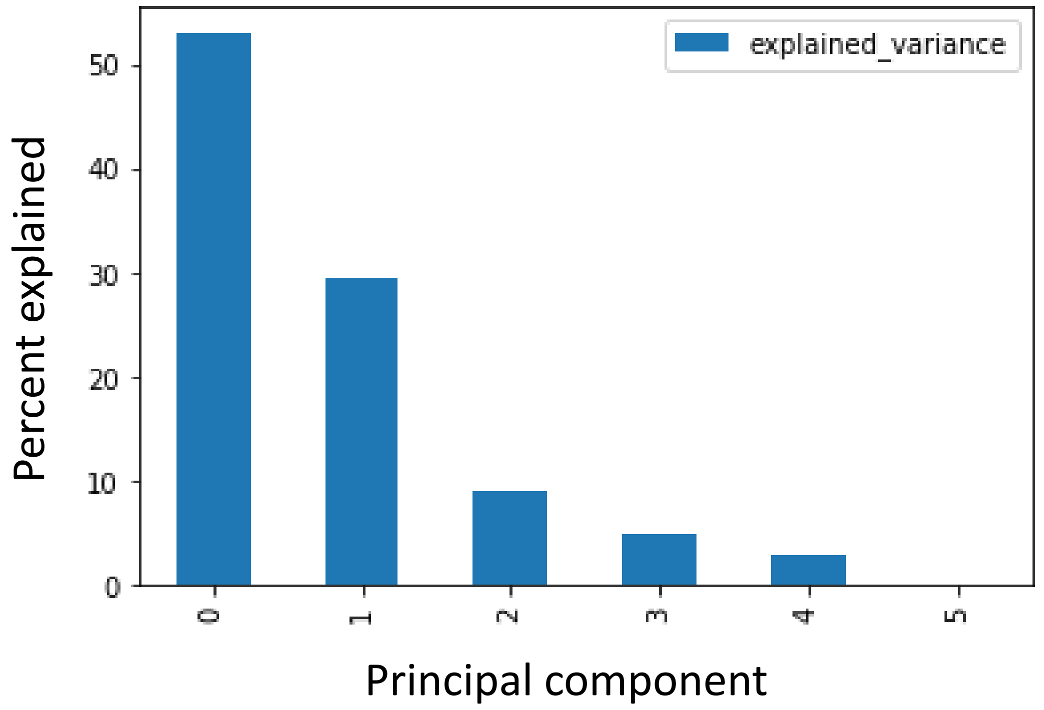
Supervised Bayesian Networks to Predict Probability of Well Performance

How does the Bayesian Network model work?



*Multiple geologic, geomechanical and operational parameters can be selected as input drivers

Dimensionality Reduction – PCA/LDA



- Principal Component Analysis (PCA) applied to identify correlations between all the input variables

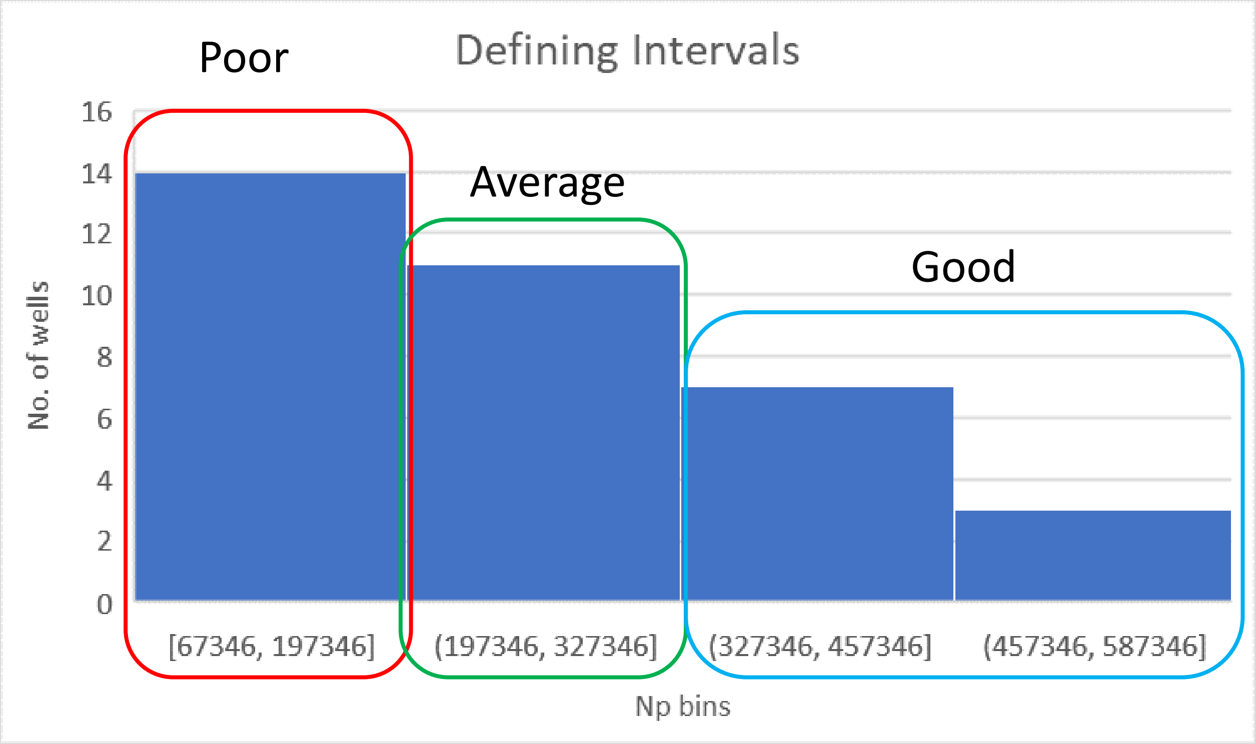
- Based on the identified pattern the number of significant dimensions was combined to 2 principal components (which are now a function of all the significant drivers)

- The 2 PC dimensions explain more than 80% variance of the input variables

- The 2 principal components are now the new hyper-parameters combining all the significant drivers/features into 2 input features

- Linear Discriminant Analysis (LDA) gives similar results

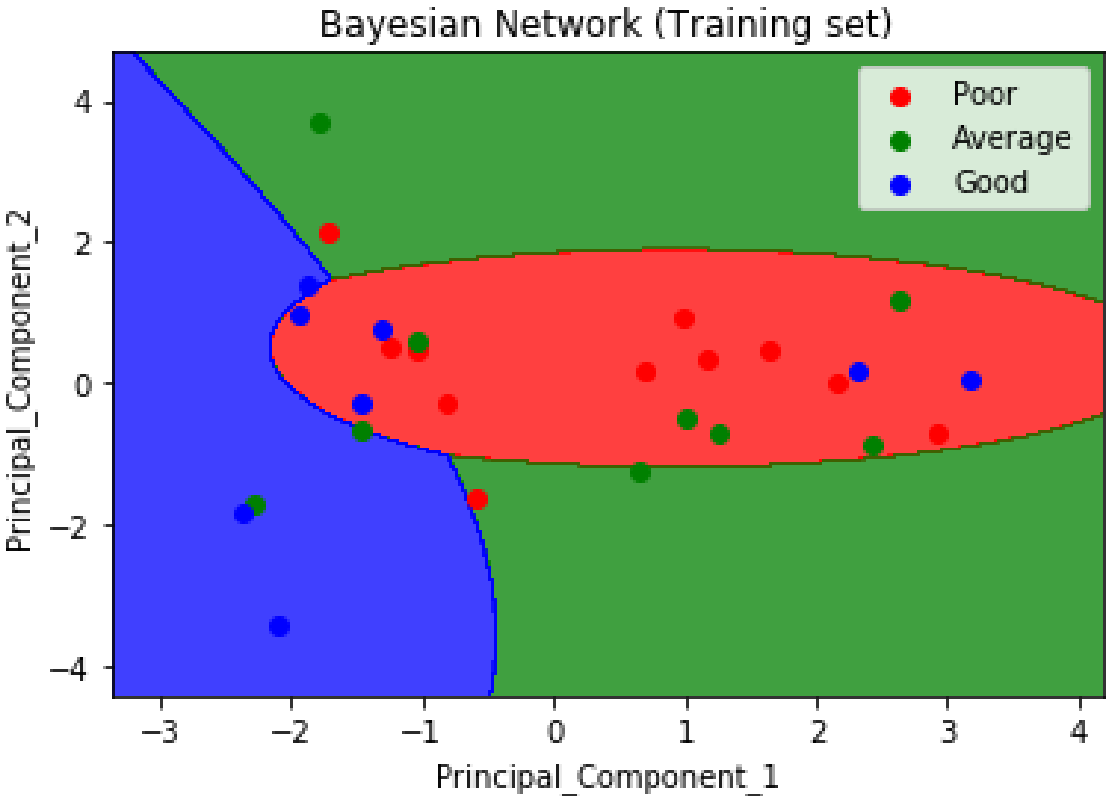
Defining Intervals to predict well probability



Defining Intervals

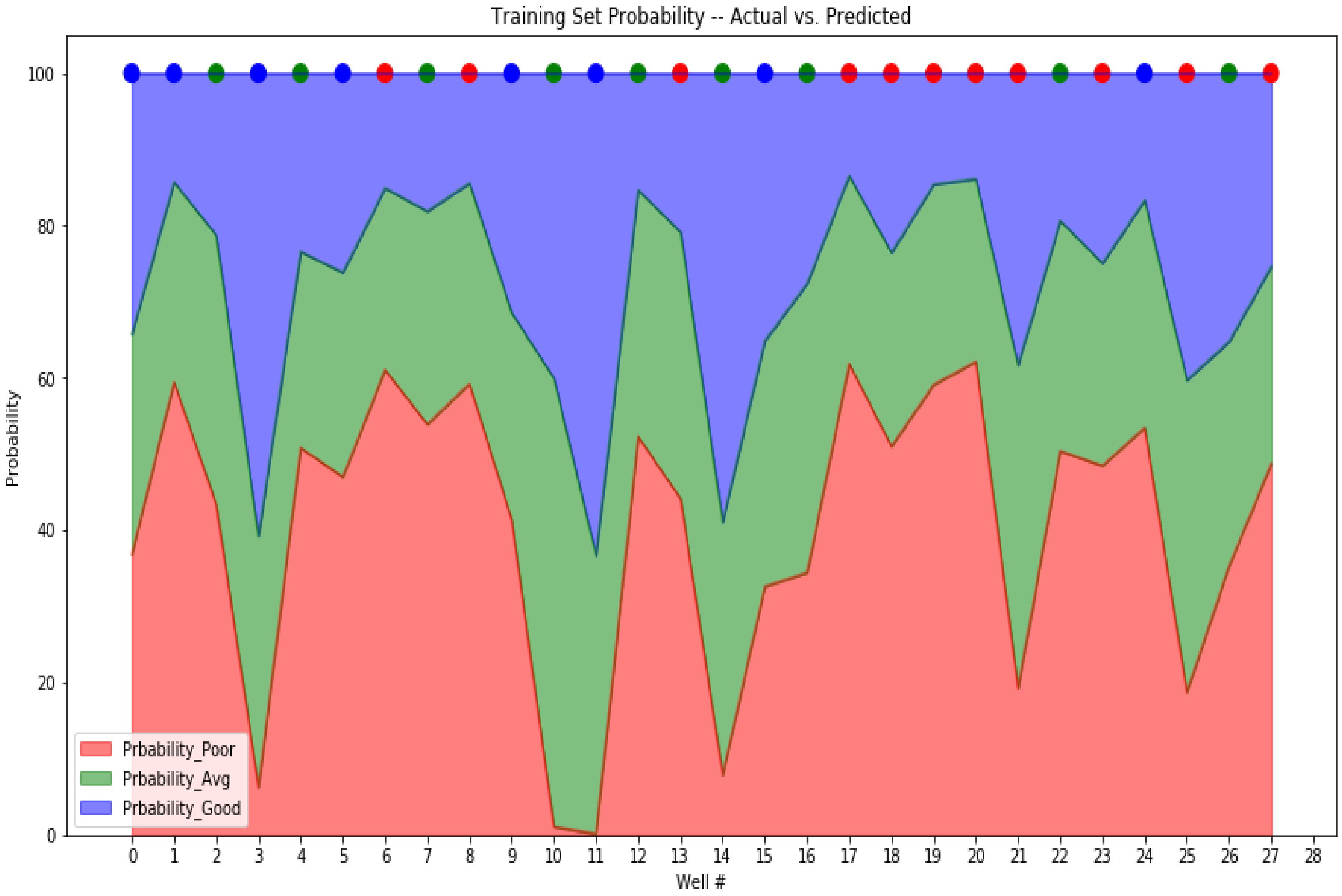
	Np Range (MSTB)
Poor	67-197
Average	197-327
Good	327+

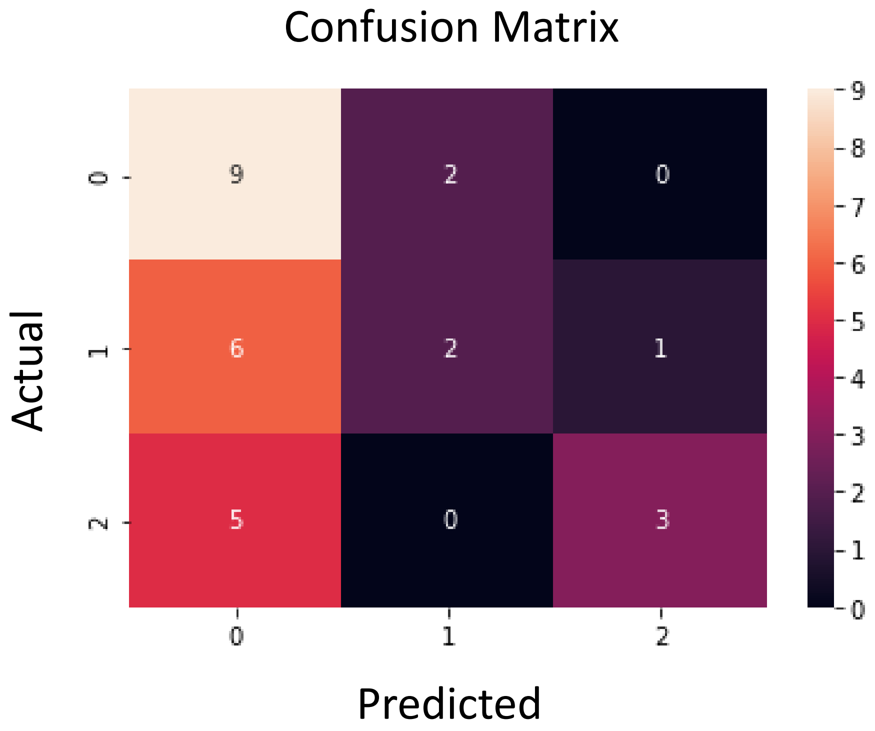
Map View: Predicted Intervals based on Probabilities vs Actual Values – Training Set



- PC1, PC2 – Scaled Principal Component Dimensions
- Dots represent actual values from wells
- Background color represents classifying boundaries and predicted ranges based on the Bayesian model

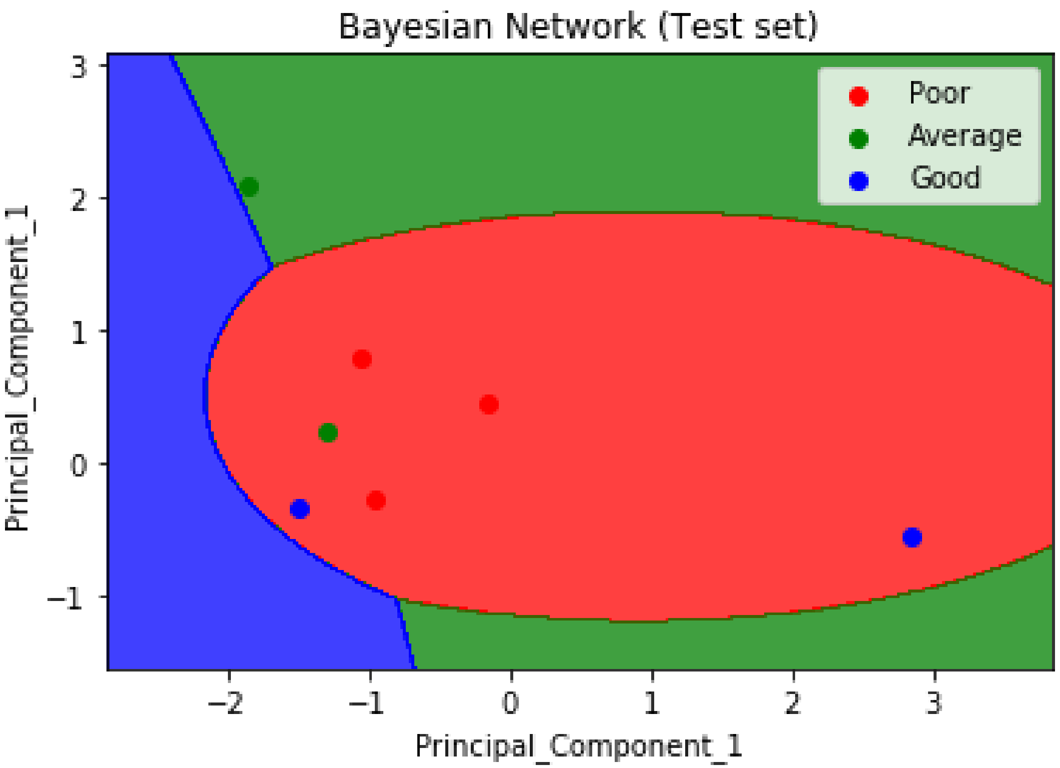
Training Set: Actual vs. Predicted Probabilities





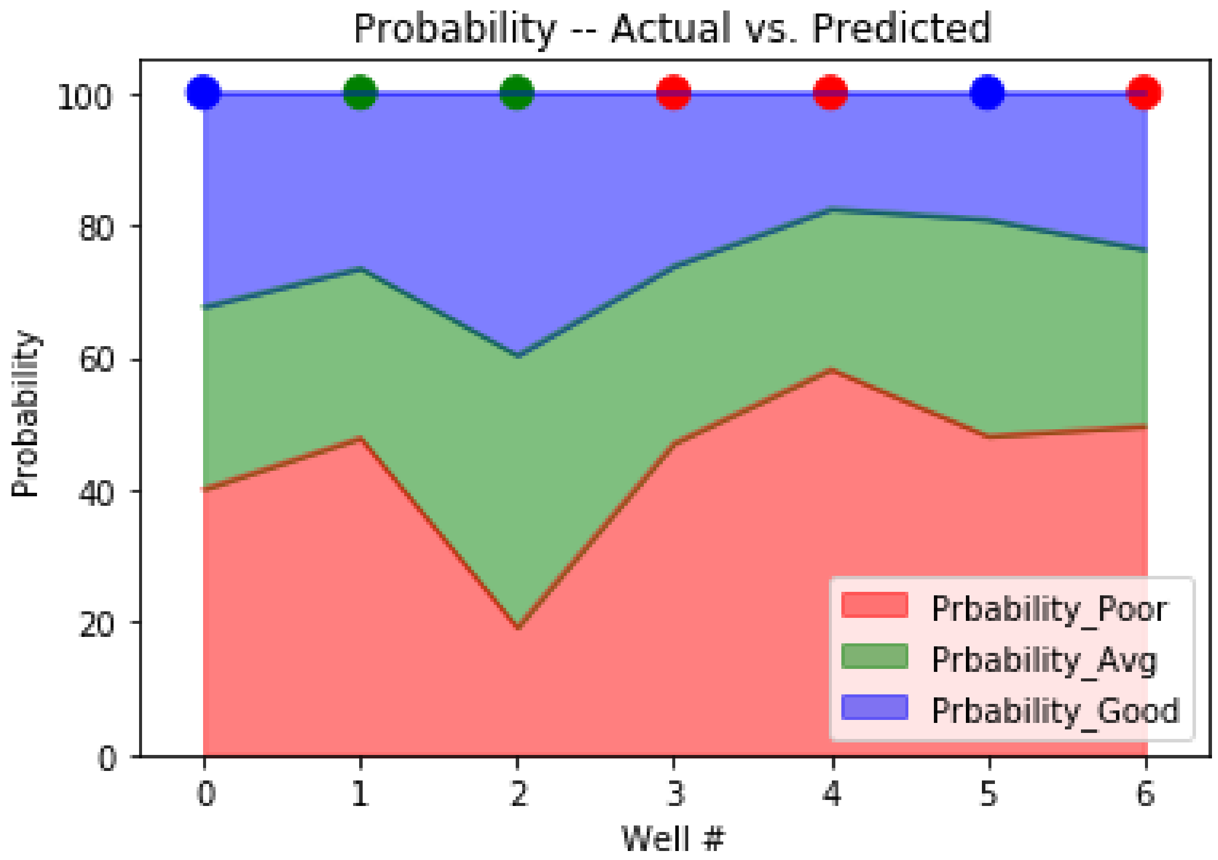
- Training Accuracy = 50%
- Dots represent actual values for wells
- Area plot shows relative predicted probabilities based on the Bayesian model

Map View: Predicted Intervals based on Probabilities vs Actual Values – Test Set

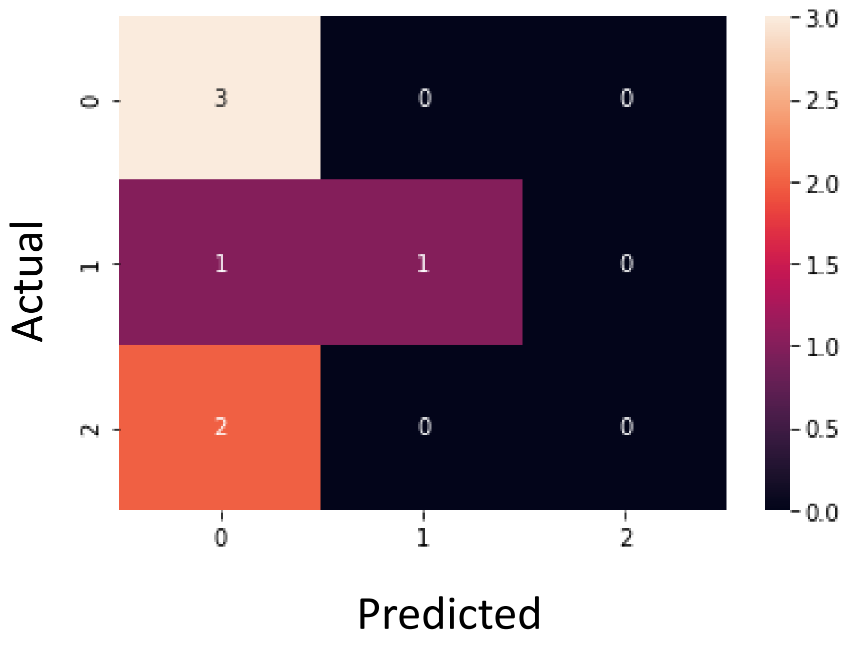


- PC1, PC2 – Scaled Principal Component Dimensions
- Dots represent actual values from wells
- Background color represents classifying boundaries and predicted ranges based on the Bayesian model

Test Set: Actual vs. Predicted Probabilities



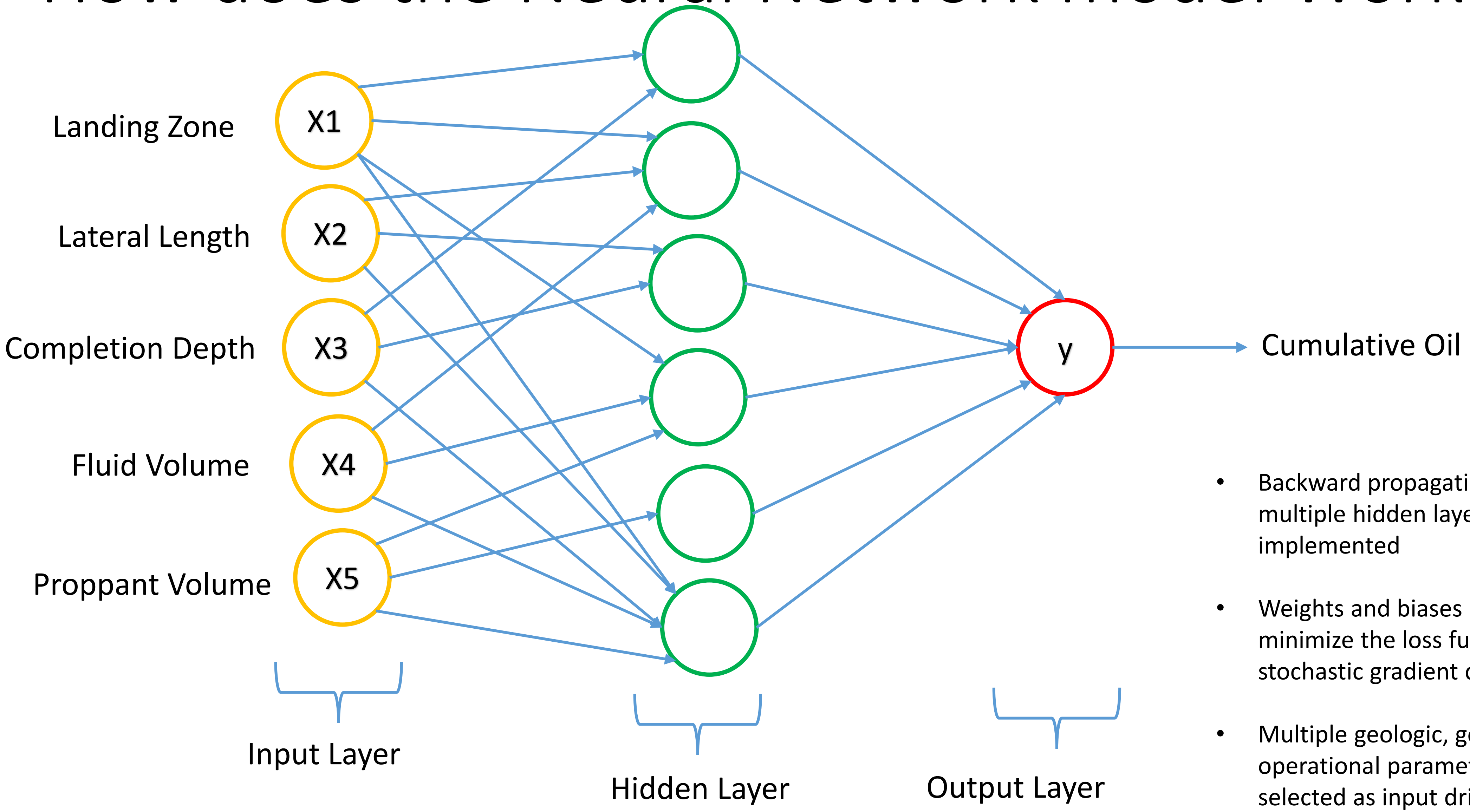
Confusion Matrix



- Testing Accuracy = 57%
- Dots represent actual values for wells
- Area plot shows relative predicted probabilities based on the Bayesian model

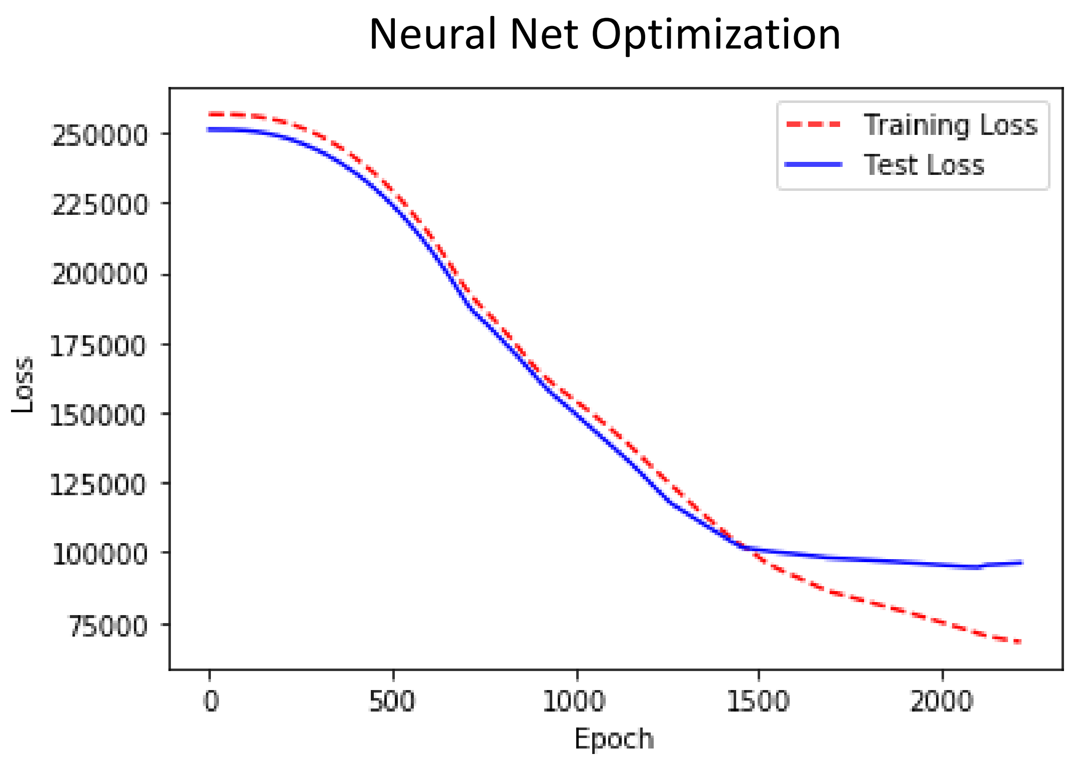
Artificial Neural Network (ANN) to Predict Well Performance

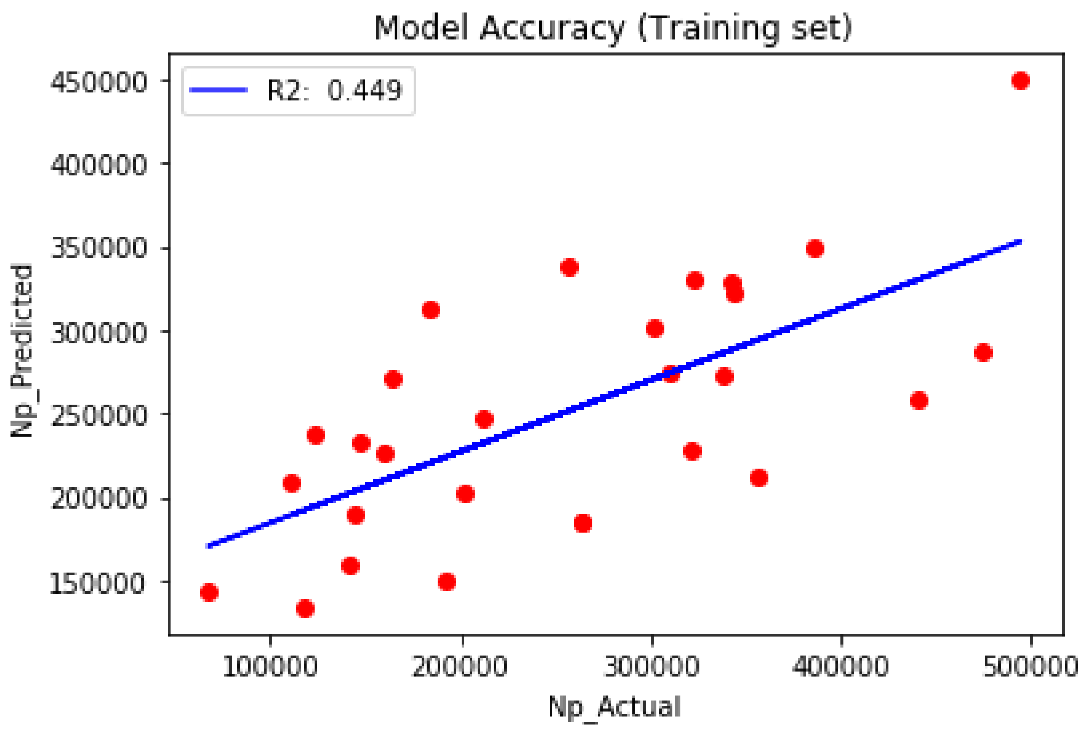
How does the Neural Network model Work?



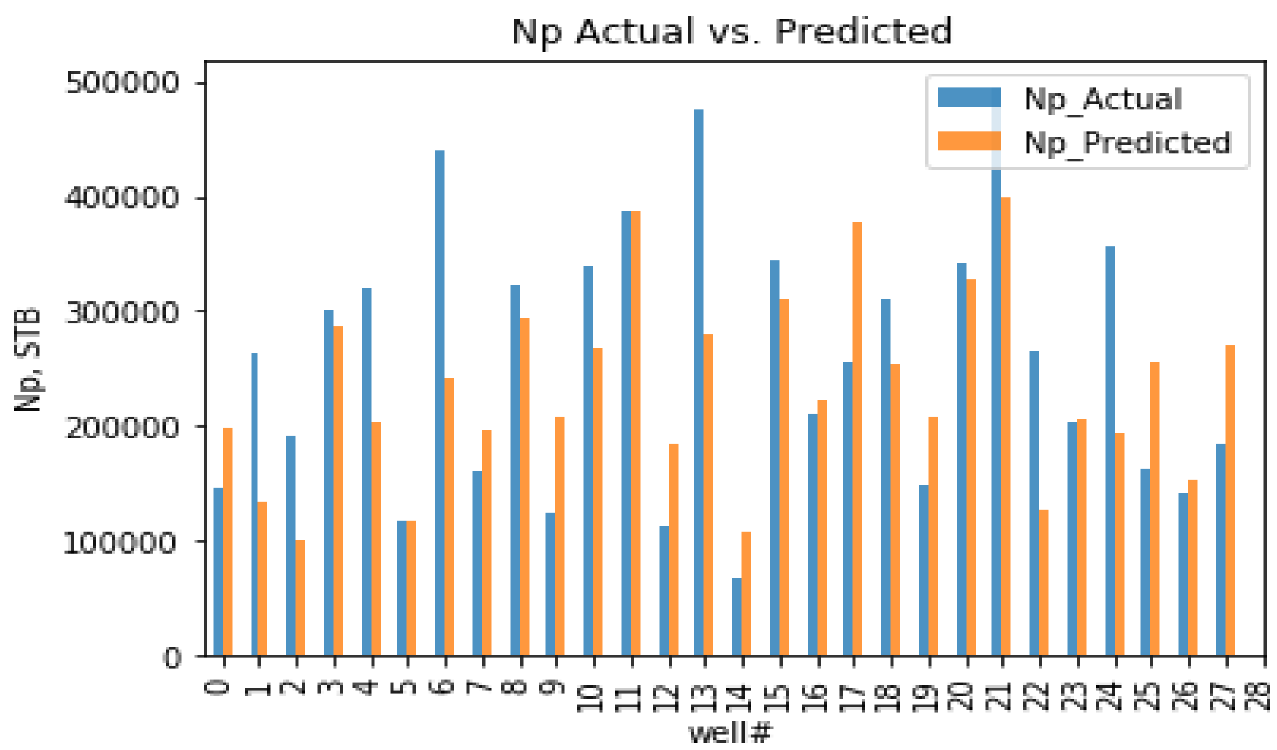
- Backward propagation model with multiple hidden layers can be implemented
- Weights and biases are optimized to minimize the loss function using a stochastic gradient descent
- Multiple geologic, geomechanical and operational parameters can be selected as input drivers

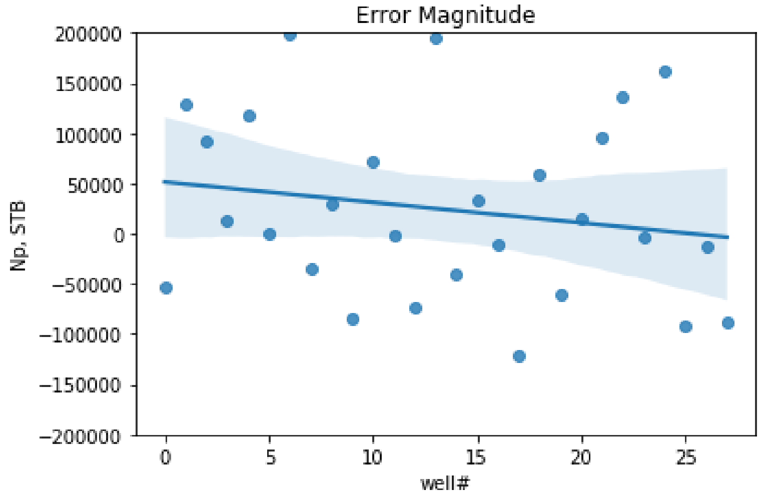
Neural Net Training and Model Accuracy – 28 wells

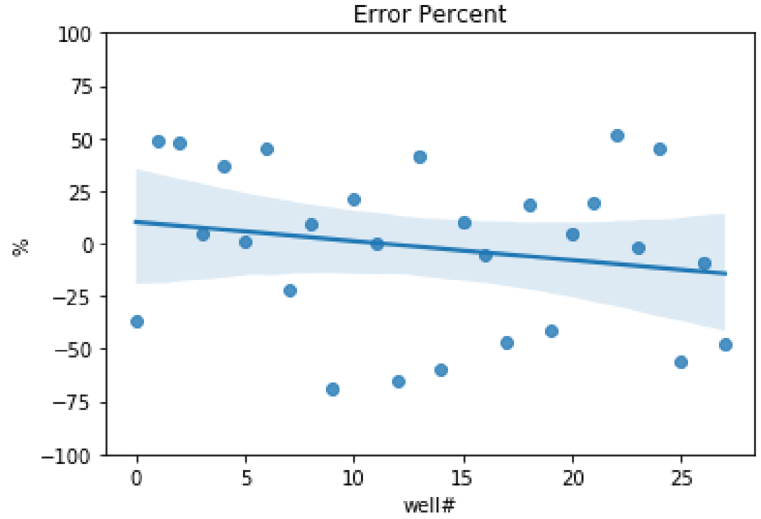




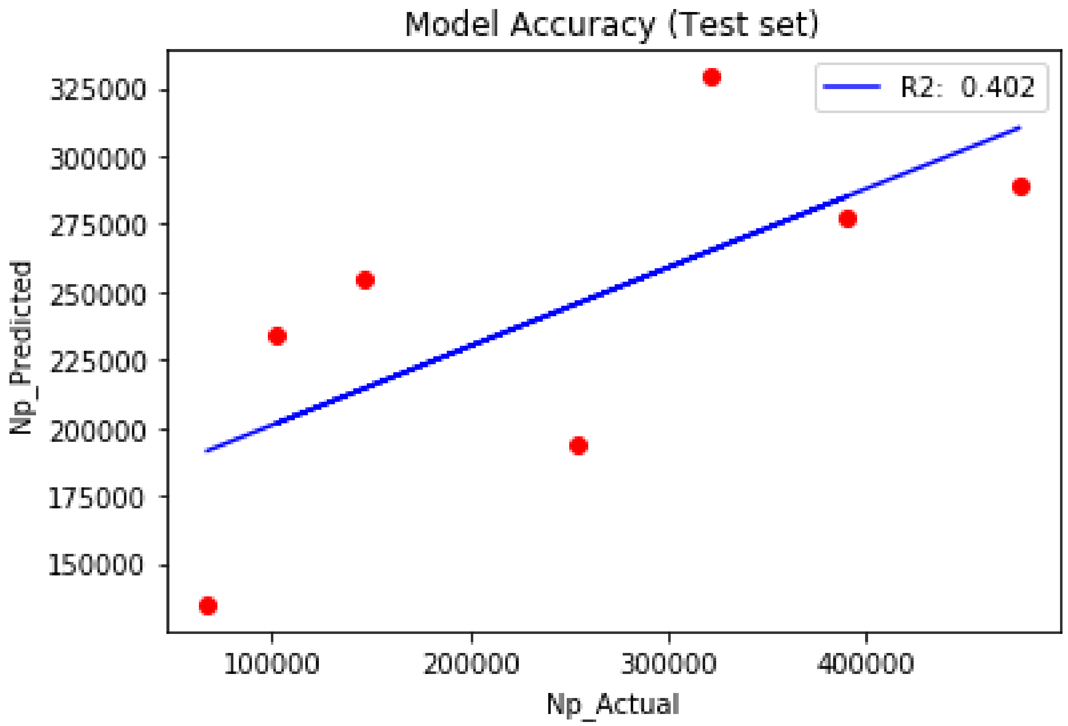
Training Set: Actual vs. Predicted Error

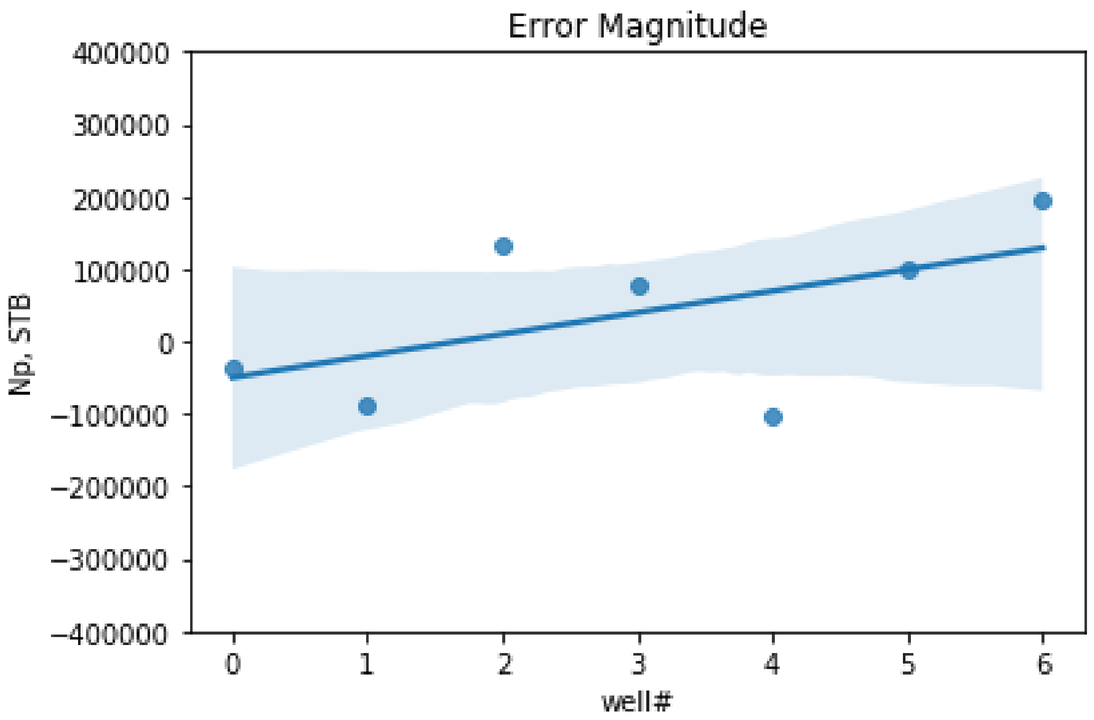






Neural Net Testing Accuracy – 7 wells





Test Set: Actual vs. Predicted Error