Machine learning for better wells

System enables data conditioning and direct access to curve data.

Fred Jenson, CGG GeoSoftware

Machine learning is rapidly becoming a standard technology within the oil and gas industry. This is especially true in petrophysics, where Big Data tend to need more efficient and faster data analysis.

The term "machine learning" was coined in 1959 by Arthur Samuel and can be defined as data-driven predictions of behavior rather than rule-based algorithms. Essentially, it is a computer science that uses statistical techniques to give computer systems the ability to learn with data and without being explicitly programmed.

A simple example is to record many measurements of the time required for objects of differing attributes to fall various distances and then build a predictive model using linear regression. This predictive model would not be based upon the theory of gravity or the gravita-

tional constant. Instead, through many observations, the model would learn the underlying order in the data. Supplying more data to the model would increase the model's accuracy. Thus, machine learning models should improve and become better over time as more data become available.

Two common types of machine learning for petrophysics are multilinear regression and clustering. This article focuses on the clustering of data to determine facies, a description of distinguishing rock characteristics.

Data clustering

The two types of clustering analysis are unsupervised and supervised. Unsupervised clustering organizes data into classes that have high intra-class similarity and low inter-class similarity and no defined target attributes. Supervised learning discovers patterns in the data that relate input data to a target attribute. An expedient way
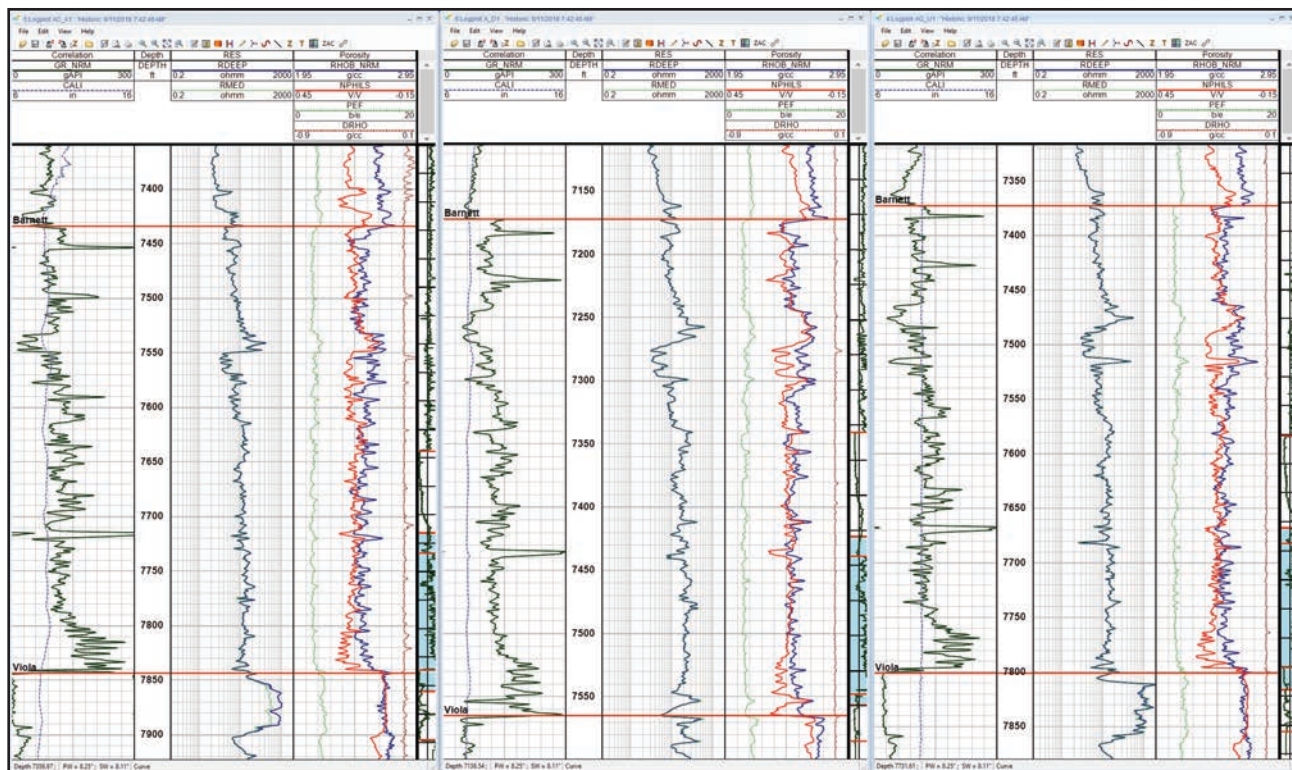

FIGURE 1. The normalized raw curve data: gamma ray, deep resistivity, density, neutron and PE curves were used in clustering analysis for the Barnett Formation. (Source: CGG)

of conducting facies classification over a large number of wells would be to perform unsupervised analysis on selected wells and then use the results as a facies target log to run supervised classification.

Unsupervised classification is far more compute- and memory-intensive than supervised classification, so it is impractical today to generate unsupervised facies on hundreds or thousands of wells.

Cluster analysis

Machine learning for data preparation requires petrophysical software, and CGG's PowerLog Ecosystem was the petrophysical platform used in this analysis. Environmentally corrected, normalized and depth-shifted input data are needed to ensure valid interpretation results. The system enables data conditioning and direct access to corrected curves and other curve data along with tops, zones, zone parameters, and other project and well data. This facilitates the machine learning process by eliminating difficult data export, editing and formatting required for machine learning when using other petrophysical software.

Python is an open-source interpreted language used for writing code to perform machine learning. Python has utilities and programs that include hundreds of scientific calculations, data analysis and visualization libraries. The interpretation example in this article uses Jupyter as the Python Interpreter for developing the clustering workflows. Jupyter is a web-based interactive application that is easy to use and has numerous features that make it ideal for modeling missing log curves and using data clustering for facies classification.

In this example it was used to perform an unsupervised classification for generating facies for a set of well logs. A "Gaussian Mixture" method was selected to generate facies in the Barnett wells used in this example. This method supports irregularly shaped clusters that are commonly observed in log curve data in unconventional reservoirs.

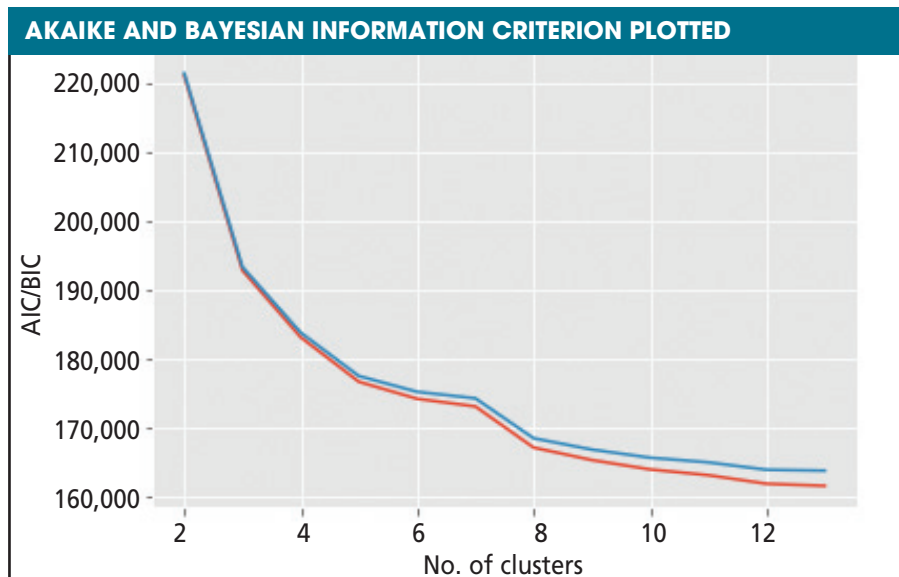**AKAIKE AND BAYESIAN INFORMATION CRITERION PLOTTED**

FIGURE 2. This visual representation shows the number of clusters (horizontal axis) versus the size of clusters (AIC/BIC vertical axis). (Source: CGG)

The potential for machine learning to improve understanding of wells, reservoirs and producing fields is virtually unlimited.

The analysis uses wells from the Barnett Formation of North Central Texas, and all the boreholes have a reliable and complete suite of well logs. A sample of the well data used in this interpretation is presented in Figure 1. A key decision to be made when clustering data is which input curves to use in the analysis. The requirement for input curves is that they exist on all wells and provide coverage over the zones of interest.

The curves used in this analysis are the bulk density log, neutron curve, gamma ray, photoelectric factor and logarithm of the resistivity. The logarithm of the resistivity is a superior input for classification as it does a better job at the lower resistivities, where variations in the measurement are more significant.

To neither overfit nor underfit and bias the model, it is important to choose the number of clusters that will properly represent the facies in the Barnett wells being evaluated. This method includes two information criteria that help optimize the number of clusters to best fit the data, as shown in Figure 2.

The inflection points on the plot indicate where the benefits of additional clusters do not add to the characterization of the data. The leveling off of AIC/BIC curves at nine clusters, as shown in this plot, corresponds to a model that adequately characterizes the
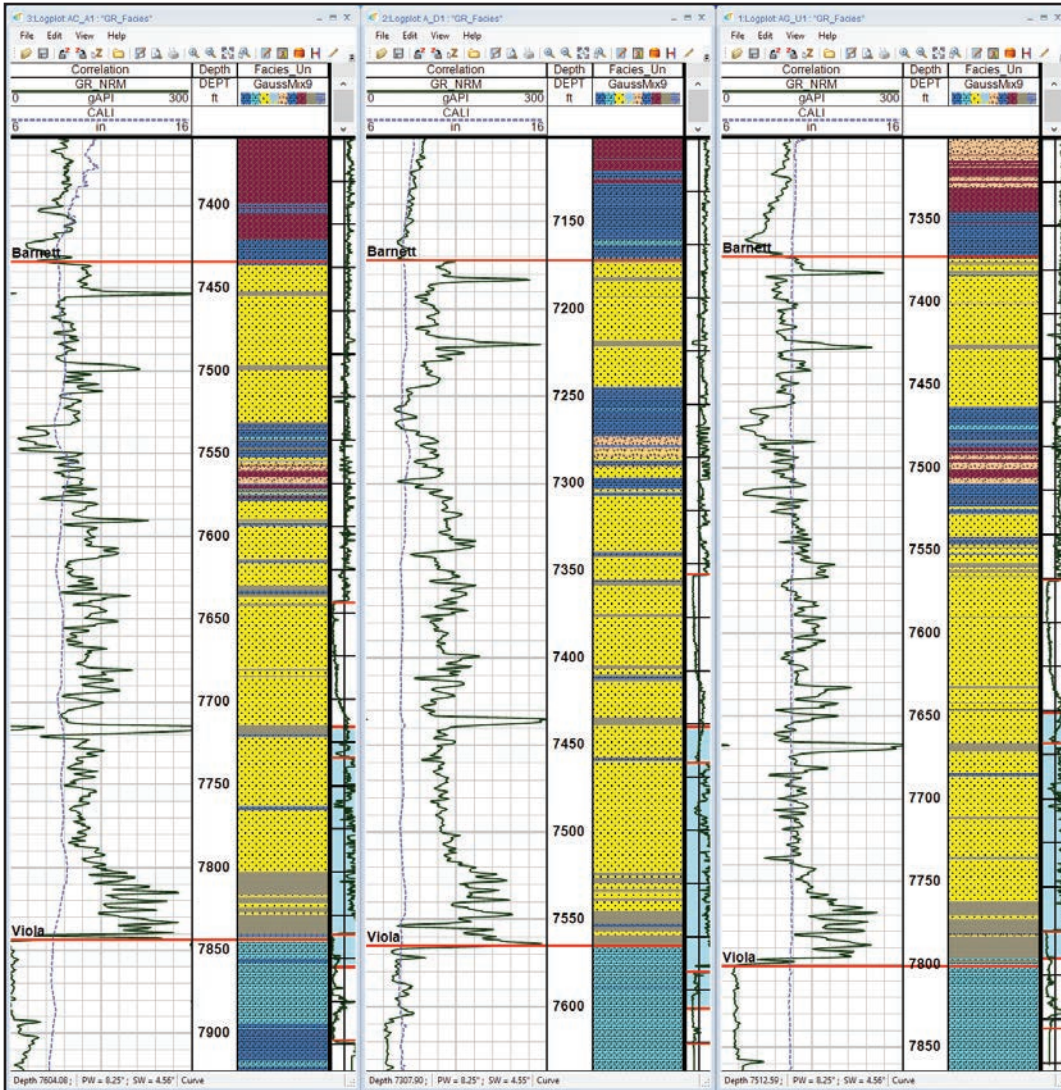
FIGURE 3. Facies determination by the Gaussian Mixture clustering analysis was used for selected wells across the Barnett Formation. (Source: CGG)

This machine learning unsupervised cluster analysis of selected wells in the Barnett and surrounding formations has been generated and evaluated, and various clustering packages were tested before selecting the Gaussian Mixture. The large number of methods available, along with the ease of use, makes this system the preferred platform for clustering analysis and other machine and deep learning workflows. Geologists can use clusters associated with specific facies to aid in correlating wells and picking tops, and geophysicists can use the facies logs to determine the accuracy of seismic inversions and to calibrate inversion results to specific facies. Quantitative seismic interpreters frequently aim to use 3-D seismic data to determine facies and calibrate the process to well log-determined facies, and engineers can use facies to aid in selecting stage intervals in multistage frac designs where limiting a stage to a single facies will maximize frac efficiency.

facies described by the Barnett well logs. The Jupyter workflow using Gaussian Mixture and generating nine clusters was run on selected wells in the project, and results are displayed in Figure 3.

Consistent results generated from well to well provide confidence in the validity of the analysis. The carbonate interval (dark blue and light blue sections) in the Barnett is easily identifiable as the Forestburg limestone, the dividing marker between upper and lower Barnett sections. The higher kerogen sections of the lower interval are correlated with grey facies and the high-quartz sections with the yellow sand facies. These facies results are consistent with prior knowledge regarding the Barnett Formation geology.

Machine learning and deep learning are technologies with multiple applications in oil and gas. Using this system for clustering petrophysical data is just one example of applying machine learning to gain better reservoir understanding. This can significantly improve completion design and help E&P companies drill more productive wells. The potential for machine learning to improve understanding of wells, reservoirs and producing fields is virtually unlimited, and to some extent, it all begins with well log data. **E&P**