THE POWER TO KNOW®

Demystifying Data Mining

Applying data mining, predictive modeling and real-time analytics in oil and gas operations

WHITE PAPER

Table of Contents

About the Author: Edward Phillips is a Solution Architect for the Global Oil and Gas Division of SAS. Phillips is a technology industry veteran who has been working with digital data since 1991. His background includes work for Claris Corporation (a subsidiary of Apple), Oracle and Internet Security Systems. Phillips came to SAS from Network General, a leading network performance and application optimization firm. He has spent the last six years helping organizations understand how to apply analytics and data mining.

Managing and Analyzing Data as a Strategic Asset

Oil producers have been trying to make sense of oilfield data since the first wireline log was employed more than 80 years ago at the Pechelbronn Oil Company oilfield in France. Much has changed since that time, and the use of instrumentation in the oil industry has exploded. The advent of advanced sensor technologies, improved control systems and real-time data acquisition has created massive stores of data to be explored, mined and exploited.

The phrase “data is as important as oil” has become common in the oil and gas industry. There are significant opportunities to extract value from these data stores with advanced data mining techniques and technologies. Mining these large reservoirs of data involves committing to key processes and technologies – and embracing new ways of thinking about problem solving. Many successful data mining projects have yielded significant results in very short time frames, and advanced analytics is a mainstay of innovators across the globe.

Analytical technologies have made it possible to understand massive amounts of data to assist in decision making across the enterprise. These same technologies – advanced data mining techniques and real-time analytical and data processing capabilities – can change the way organizations make decisions. And when improved decision making is applied in a structured manner, it can yield significant returns. When these time-tested tools and techniques are adopted by business analysts, they can enable rapid results without requiring the analysts to obtain advanced statistical training.

When time-tested tools and techniques are adopted by business analysts, they can enable rapid results without requiring the analysts to obtain advanced statistical training.

This paper defines data mining and discusses the practical application of approaches, workflows and techniques for applying data mining, predictive modeling and real-time analytics in oil and gas operations. After reviewing several use cases, the paper will explore data mining behind the scenes. This will include the role of exploratory data analysis; model development and modeling techniques; and approaches to putting models into production.

What Is Data Mining?

Data mining is the process of selecting and exploring data to discover previously unknown patterns and historical trends that can be used to develop models for predicting future outcomes. Quantitative techniques uncover patterns and relationships in data that are used to build descriptive and predictive models.

The terms data mining and predictive modeling are often used interchangeably – but they are distinct. Data mining is the process of uncovering patterns in a sample set of data and then developing models that find the same desired pattern across a much larger universe of data. Predictive modeling is the process of applying these models during the course of a business process to predict an outcome.

Data mining is the process of selecting and exploring data to discover previously unknown patterns and historical trends that can be used to develop models for predicting future outcomes.

Techniques and Methodologies Common to Data Mining

Type	Definition	Applications	Relevant SPE Papers[1]
Descriptive Models	Descriptive models classify elements into groups based on patterns, statistical relationships and the quantification of these relationships. These models are used to support the development of predictive models.	Identifying groups of statistically similar attributed wells for better decisions around treatments, stimulation, drilling approaches.	Data Mining Techniques for Optimizing Fast Track Re-engineering of Mature Fields – SPE 78333 Exploratory Data Analysis in Reservoir Characterization Projects – SPE 125368
Predictive Models	Predictive models analyze patterns and past performance in relationship to a particular desired outcome to predict the probability of that outcome.	Analyzing past performance on a particular group of wells to predict recovery, given a particular stimulation approach – or to predict the failure of a piece of equipment.	Drilling Optimization in Unconventional and Tight Gas Fields – SPE 142509 Tight Gas Well Performance Evaluation with Neural Network Analysis – SPE 135523 An Innovative Approach for the Analysis of Production History in Mature Fields – SPE 62880
Optimization Models	Optimization models use the outputs of descriptive models, predictive models and constraints to predict the results of the impact of many variables on decisions to maximize certain outcomes according to a target like cost, time or the relationship between the two.	Determining how to get the right resources to the right location at the right time in the most efficient manner – based on constraints, historical performance and the probability of similar future performance.	Attitude of Collaboration, Real-Time Decision Making in Operated Asset Management – SPE 128730 Increased Upstream Asset NPV with Forecasting, Prediction and Operational Plan Adaptation in Real Time – SPE 133450

1 These papers from the Society of Petroleum Engineers provide more examples of each data mining technique described in this table. All SPE papers are available at onepetro.org.

The scope of activities related to data mining and predictive modeling includes:

- Data preparation to merge multiple data sets, resolve missing values or outliers, and reformat data as needed.
- Exploratory data analysis to discover relationships and anomalies in the data.
- Variable transformation, enrichment and selection to better focus the modeling process.
- Model building using competitive algorithms to search for data combinations that reliably predict the outcome.
- Testing and validation of the champion model to ensure that the model generates output as expected when applied to new data.
- Putting models into production in applications and databases to optimize business processes and improve business decisions.
- Monitoring the model performance to ensure the model is predicting well and does not need to be recalibrated.

How Does Data Mining Return Value to Oil and Gas Companies?

Advanced data mining techniques can yield significant value to oil and gas companies by identifying opportunities to improve decision making. One unconventional operator identified fracture optimization opportunities in 25 percent of a population of almost 200 wells that were studied. In another case, an operator was able to understand the statistical significance of the effect of proppant on production by geography – and as a result, the operator identified opportunities to save almost US$10 million by reducing proppant volumes in approximately 200 wells.

Refiners also recognize the significant business value of data mining. For example, refiners can conduct multivariate analysis in real time to understand the effect of complex relationships on the probability of key events, improving asset uptime and reliability. In one case, a refiner needed to analyze leaks, evaporation and changes in mixture composition associated with glycol consumption. By modeling the process and associated data, the gas producer identified the problem, took corrective action, and increased product quality and production. This same technology allowed for improved maintenance for key filters – further increasing yield and reducing costs along with health, safety and environmental (HSE) impacts. In a similar fashion, real-time monitoring of complex systems allowed a major European gas producer to anticipate and prevent turbine trips, reducing unplanned downtime by 25 percent.

Refiners can conduct multivariate analysis in real time to understand the effect of complex relationships on the probability of key events.

Data mining has become a core competency of digital oilfield monitoring centers. A North Sea operator applied analytics to better anticipate maintenance requirements and improve the integrated planning system to deploy the right resources at the right time for maximum efficiency. See the Appendix for more information on the benefits of shifting from a proactive to a predictive asset maintenance model. As a result of these efforts, the operator increased production by 2 percent, lowered operating costs by 15 percent and achieved a 90 percent project completion rate (up from 30 percent).

Data mining has become a core competency of digital oilfield monitoring centers.

Behind the Scenes with Data Mining

The process of data mining includes several key activities. These include data aggregation and preparation, exploratory data analysis, modeling and deployment of models in production environments.

Data Aggregation and Preparation

Data preparation is typically 80 percent of the effort of an analytical project. One reason is that many organizations lack a single source of complete, high-quality data required for an analytical exercise. Some choose to wait for the completion of a corporate data warehouse that promises to organize, arrange, standardize and clean the data. Unfortunately, these warehouses seldom address all of the relevant data that may be critical to the analytical problem.

Preparing data for data mining should result in an analytical base table, or data set, that has variables associated with the problem being modeled. This data is prepared very differently than a warehouse for historical reporting because it is gathered from many more databases and stored in very different subsets that are tailored for the analysis at hand.

Aggregation and preparation of data for an analytical project must be conducted by members of the team who have the relevant skill set. Most data aggregation issues are not technical, but rather are related to domain knowledge and data ownership. Once the business logic is provided by the domain experts, the process of accessing, joining and reformatting data can be performed by technical staff.

Most data aggregation issues are not technical, but rather are related to domain knowledge and data ownership.

Data standardization and data quality also require collaboration between domain and technical staff. For instance, standardizing a supplier name across multiple systems is not likely to affect the result of an analytical exercise. However, decisions about how to fill missing range values in sensor tag information or how to collapse time measurement intervals should be made by the domain expert advising the analyst.

Once the workflow for aggregating and cleansing the relevant data is determined, the processes and rules related to conducting these tasks can be automated. Automation of data preparation can be done in real time when the data is created, or at intervals that are appropriate to the data's specific, time-sensitive requirements.

Exploratory Data Analysis

Exploratory data analysis (EDA) is an approach to analyzing data for the purpose of formulating hypotheses that are suitable for testing. EDA is an iterative approach that allows an analyst to scrutinize data to identify patterns that merit further statistical testing or modeling. This process of pattern discovery is referred to as exploratory rather than confirmatory, because it directs the analyst toward or away from a hypothesis through statistical testing.

Figure 1: Exploratory data analysis reveals the relative distribution of well characteristics in relation to the variable selected.

Modeling Techniques

Best practices for data mining involve utilizing multiple competing analytical techniques to determine which technique will produce the best model and therefore the best prediction. Some of these modeling techniques include decision trees, neural networks, least-angle regressions, logistic regressions, memory-based reasoning and rule induction.

A software solution with multiple modeling techniques allows an analyst to quickly and easily apply a particular modeling technique to a data set and interactively work with the parameters to try different configurations. This capability allows modelers to test many different approaches while relying on the software to pick the most accurate model, based on a set of user-selected statistical diagnostic tests. It is also important to be able to build an ensemble model by combining techniques if the combination of two models is more effective than a single model. Advanced data mining software that provides these capabilities allows an analyst to spend time developing true insights as opposed to programming analytical models.

Advanced data mining software allows an analyst to spend time developing true insights as opposed to programming analytical models.

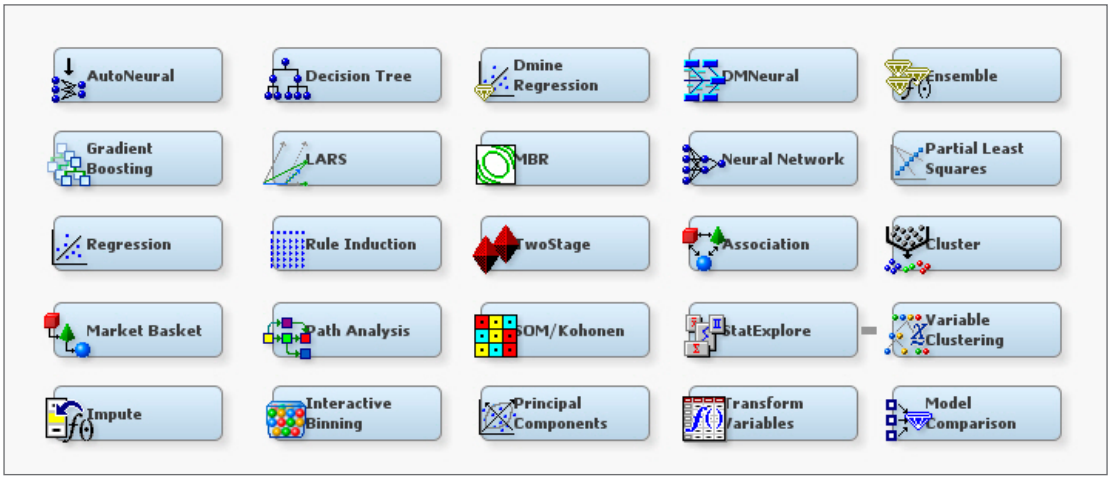

Figure 2: Examples of analytical methods that might be applied to oilfield data.

Deploying Models in a Scalable Environment

Significant business value is lost without the ability to put analytical models into production at an enterprise scale. But putting models into production involves key enabling technologies that must be flexible and robust enough to support the various technical requirements dictated by the business problem. The technical solution must support data access regardless of size and source, and must provide flexible scalability regardless of computational complexity or the time window established to return a computed value.

Data mining is most effective when deployed as part of an integrated information delivery strategy that is supported by strong business domain specialists, IT and skilled analysts. Determining who leads for each data mining exercise requires consideration of the target audience, the timing of the results, and the anticipated action as a result of the analytical insight. This multidisciplinary approach ensures that the project can have technical success and also generate business value.

Data mining is most effective when deployed as part of an integrated information delivery strategy that is supported by strong business domain specialists, IT and skilled analysts.

The technical infrastructure selected for both modeling and production must be able to support modeling at scale. When modeling large data sets, it is important to be able to process large data sets with computationally intensive tasks and to visualize large data sets that support exploratory analysis. In the case of very large data sets or complex computational processes, it may be necessary to use high-performance computing to return value within a tight time frame.

Just as business outcomes shift with the economy, data constantly changes. As a result, companies should establish models that natively adapt to changes in the data without significant human intervention. It is possible for key variables in a multivariate predictive model to shift in a statistically significant way not anticipated by the model. The ability to monitor and alert key stakeholders to ongoing model performance is the final step in deploying advanced analytical models to scale.

The ability to monitor and alert key stakeholders to ongoing model performance is the final step in deploying advanced analytical models to scale.

SAS®: A Comprehensive Approach

Data mining technologies can be utilized to exploit vast amounts of data to yield significant results in short time frames. SAS has helped many organizations apply advanced analytics to achieve significant benefits in the digital oilfield. This journey from data exploration to optimized decisions is achievable and can be deployed at scale in oilfield operations.

Learn more

For more information, see the references provided in this paper, or visit sas.com/oilgas.

References

- Data Mining 101: Applying Business Analytics Webinar Series:
  - On Demand webinar: sas.com/reg/web/corp/907013
  - Conclusions paper: sas.com/reg/wp/corp/28224
- SAS-authored papers for the Society of Petroleum Engineers: sas.com/industry/oilgas/spe-resources.html

Appendix

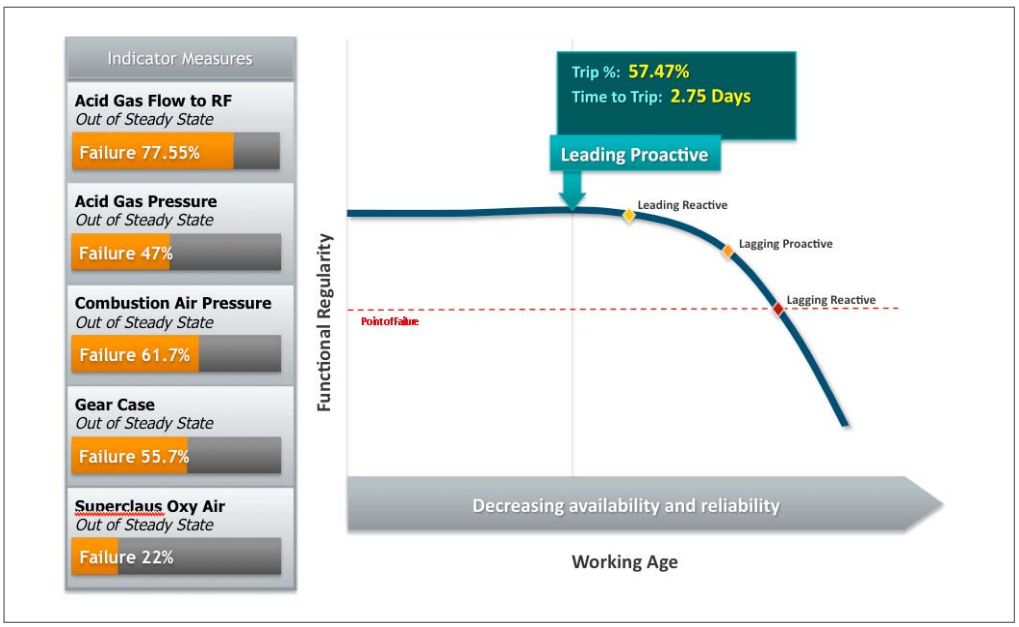

Figure 3: More proactive examination of indicators lengthens time for mitigation.

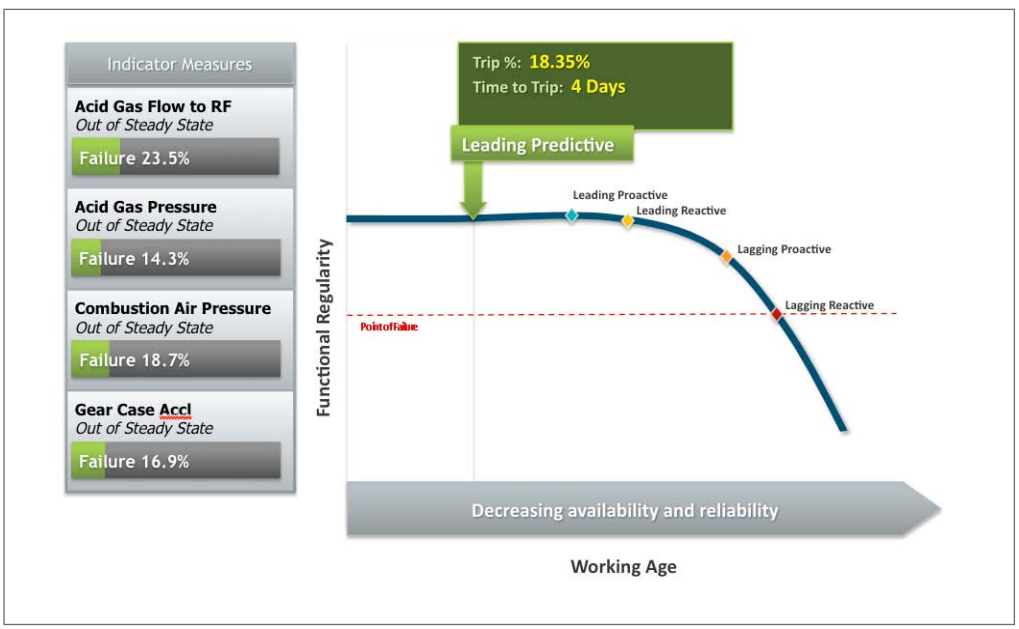

Figure 4: Predictive analysis optimizes asset maintenance.

About SAS

SAS is the leader in business analytics software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions, SAS helps customers at more than 55,000 sites improve performance and deliver value by making better decisions faster. Since 1976, SAS has been giving customers around the world THE POWER TO KNOW®. For more information on SAS® Business Analytics software and services, visit **sas.com**.

SAS THE POWER TO KNOW®

SAS Institute Inc. World Headquarters +1 919 677 8000

To contact your local SAS office, please visit: **www.sas.com/offices**