URTeC: 2668073

The Rise of the Machines, Analytics, and the Digital Oilfield: Artificial Intelligence in the Age of Machine Learning and Cognitive Analytics

Kathy Ball, Devon Energy; Tristan Arbus, Devon Energy; Uchenna Odi, Devon Energy; Jessamyn Sneed, Devon Energy

This paper was prepared for presentation at the Unconventional Resources Technology Conference held in Austin, Texas, USA, 24-26 July 2017.

Abstract

In the next ten years, instantaneous value from digital oilfield systems will dramatically alter the oil and gas landscape as cost and operational efficiencies are attained through the reliance on artificial intelligence. The digital oilfield will radically change how oilfield workers, machines, and the holistic enterprise operate to achieve results and compete in the new digital world. The new digital oilfield will be a disruptive technology that creates new value streams for exploration and production ranging from automated decisions and reactions in real time to massively improved operational efficiencies, connected infrastructure platforms, and much better interaction between machines and humans. The days of collecting and storing large volumes of data for later analysis will become a distant memory. The digital oilfield will change expectations for all aspects of our industry ranging from how fast decisions are made to detecting patterns the human eye cannot see in order to take advantage of the insights quicker. Data silos will be reduced and information shared across all areas of the oil company. At a simple level, artificial intelligence will be used to increase the accuracy of predictions to near-cognitive robotic comprehension in machine learning.

To quote Erik Brynjolfsson, director of the MIT Initiative on the Digital Economy, "Humans must adapt to collaborate with machines, and when that collaboration happens, the end result is stronger." This session will outline three ways advanced analytics and artificial intelligence can help bridge the gap between the digital oilfield and analytics: (1) predictive analytics from a case study in predictive asset failures, (2) machine learning in completions, and (3) text analytics in drilling. The session will outline the people, process and technologies needed to enable this infrastructure. Additionally, key pitfalls to avoid regarding systems, silos, and the human barriers to understanding artificial intelligence and getting past the hype associated with new technologies will be discussed.

Introduction

Historically, exploration and production companies have benefited from high oil prices. The industry has been able to hide operational inefficiencies due to large budgets permitted by high oil prices. Financial risk was often ignored because the expectation of the next return on investment was enough to justify the risks. Recent changes in the economics of the oil and gas industry have caused exploration and production companies to rethink their budgets and to institute cost-cutting measures to mitigate the loss of revenue. Outdated technologies and practices are under continual scrutiny to ensure that budgets are met as margins tighten. This has resulted in oil and gas companies instituting data analytics as a means of improving cost savings. (Mitchell, 2016)

To prove this theory, Accenture and Microsoft partnered with Oklahoma-based Penn Energy to survey 229 oil and gas professionals, including engineers, managers, geoscientists and other staff, about their views towards data and analytics (Accenture and Microsoft, 2015). Their results indicated that investments in the cloud, collaboration, infrastructure, and mobility will shift to big data, analytics, the Industrial Internet of Things (IIOT), and automation in the next three to five years. This survey also found that the professionals believed machine learning to have the highest growth potential.

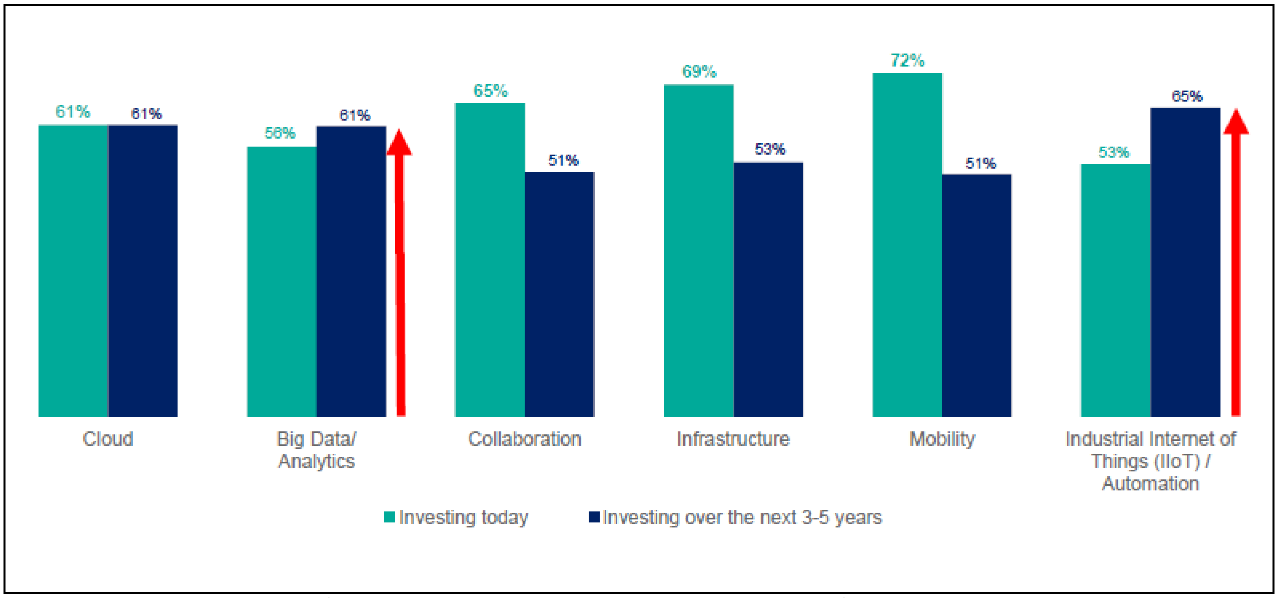

Figure 1: Investment breakdown for oil and gas in the next 3-5 years (Accenture and Microsoft, 2015)

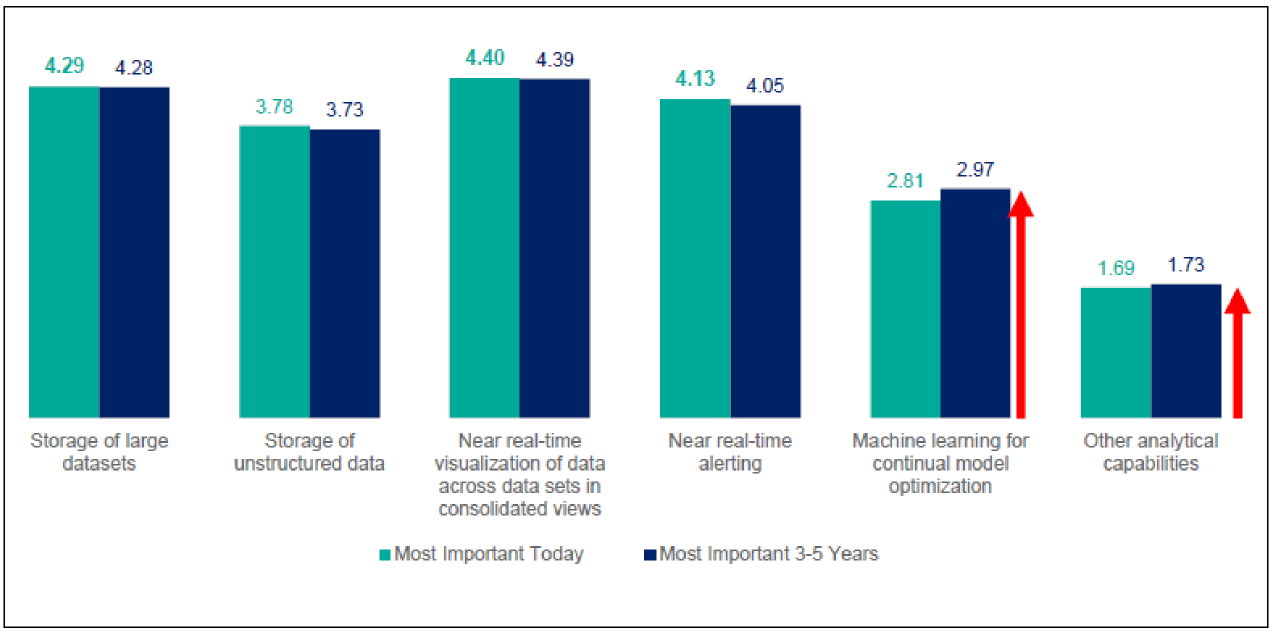

Figure 2: Analytical capabilities that are most relevant to oil and gas in the next 3-5 years (Accenture and Microsoft, 2015)

The aforementioned trends were captured despite an oil and gas downturn. This indicates a significant level of enthusiasm for big data and analytics despite economic challenges. The startup space has also been leveraging this enthusiasm as young companies specializing in the analytics space are growing rapidly.

Building Enterprise Solutions from the Ground Up

While startups can indeed offer value through unique competitive advantages, there is also much value to be found with internal efforts. Therefore, oil and gas companies have put forth a large effort to better collect, streamline and organize one of their biggest assets, their data. Much of this effort has gone into not only creating large compendiums of historical data, but also into the speed at which that data can be accessed and utilized. The industry as a whole has become increasingly adapted to the "3Vs" of big data: volume, velocity and variety. Generally speaking, the difference between big data and more traditional forms of business intelligence is inference. Whereas business intelligence utilizes descriptive statistics and other methods to measure things directly, big data allows users to infer relationships (regressions, nonlinear, causal, etc.) between inputs and forecast events. At its most complex levels, big data lends itself to the usage of advanced analytics including artificial intelligence, machine learning, and deep learning. This is becoming more and more possible as automation and increases in sensor equipment in the field provide data at speeds and quantities never before possible.

The increasing prevalence of advanced analytics and artificial intelligence in the oilfield is starting to change the culture of how we as an industry operate. However, as with any major culture change, there are a number of commonly associated misconceptions. One area of contention is often the thought that new analytics practices are an attempt to usurp subject matter experience as the only way to solve problems and make decisions. In truth, artificial intelligence doesn't seek to undermine existing practices. Rather, it seeks to strengthen subject matter knowledge by melding new tools and understanding into existing practices. There are definitive weaknesses in artificial intelligence and deep learning that can only be overcome through human intervention because, while models such as neural networks or decision trees may be able to identify hidden patterns that a human may not be able to decipher, they are incapable of using reason. The only way to obtain plausible results deployable in the real world is for subject matter experts and analytics experts to collaborate in order to glean insight in previously unattainable ways.

Even to those who understand this interplay between more traditional knowledge and the benefits of analytics, artificial intelligence, and machine learning, implementing these techniques has its own set of challenges. In a world where data is sparse, unconnected, and siloed, utilizing said data is extremely challenging. This is what makes the ability to automate, share and connect quality data paramount to success. This type of data connectivity allows users to turn quality data into actionable intelligence. It also helps promote collaboration between the field, information technology (IT), subject matter experts (SMEs), data scientists, and others. This atmosphere of teamwork and communication is what grants companies the ability to leverage their assets and compete at the highest level. The rest of this paper will discuss a few examples of what such collaboration with high quality data can accomplish.

Text Mining in Drilling Operations

Oil and gas operations produce colossal volumes of data, from drilling reports to sub-second seismic data to other multi-disciplinary information on completions design, reservoir depletion, etc. Much of this data is text-based and stored in either a structured format, such as a column in a SQL database, or in an unstructured format, such as a vendor pdf report. Over the last two decades, statistical methods have been developed to categorize and process this data via text mining. Text mining holds significant implications for the oil and gas industry as it will allow for further quantitative data collection and provide context to associated numeric data. Furthermore, text mining enables improved competitor intelligence gathering through analysis of readily available public data such as quarterly financial reviews and social media posts. These can provide invaluable insights into current trends in the oil and gas sector and gauge the industry atmosphere. However, the process of text mining can be challenging as it requires a secure process to access trusted data, procurement of statistical tools and skills to apply towards the text, and, most importantly, a multi-disciplinary collaboration between technical groups and SMEs. This section will focus on the process by which our Advanced Analytics team - in collaboration with the IT, Data Management and the Drilling Operations (DO) teams - were able to derive meaning from structured drilling reports in a SQL database.

This text-mining exercise was an extension of a separate data mining project in which the DO team wanted to analyze historical non-productive time (NPT) events that occurred across all US drilling operations. After the Advanced Analytics team delivered a dashboard capable of tracking NPT, the DO team proposed to extend the collaborative effort with the objective of providing context to their NPT statistics using text mining. Each instance of NPT in the SQL database had a corresponding comment section in which the field personnel could enter a description of the NPT event. The DO team wished to collect and analyze these comment fields to find the relationship between certain words and NPT. Drilling engineers met with a data scientist from the Advanced Analytics team to provide a list of words commonly associated with negative drilling outcomes, such as troubleshoot and repair. They also brainstormed synonyms and colloquialisms for these terms so that an accurate count of these word-occurrences could be obtained. The data scientist met with IT and data management to receive guidance about drilling data integration and appropriate SQL querying protocols. Occurrences of the DO-specified words were aggregated and analyzed across key factors such as time, location, and rig performance. The SAS software suite was used to visualize the prominent relationships between DO-specified and general words in the NPT comments. The results of this project were not only used to improve the DO team's understanding of their data, but also enabled QA/QC review of the NPT data within the SQL database and prompted the creation of data entry standards by the DO and data management teams.

NPT Details

Start Date	End Date	Dur (Net) (hr)	Accountable Party	Com
4/29/2012 03:00	4/29/2012 10:00	3.00	HELMERICH & PAYNE INC	WAIT ON ELECTRICIAN FOR TOPDRIVE, TROUBLE SHOOT TOP DRIVE & CHANGE OUT ONE OF THE TOP DRIVE & WRANGLER SENSORS.

Figure 3: NPT entry with DO-specified word captured in the SQL database

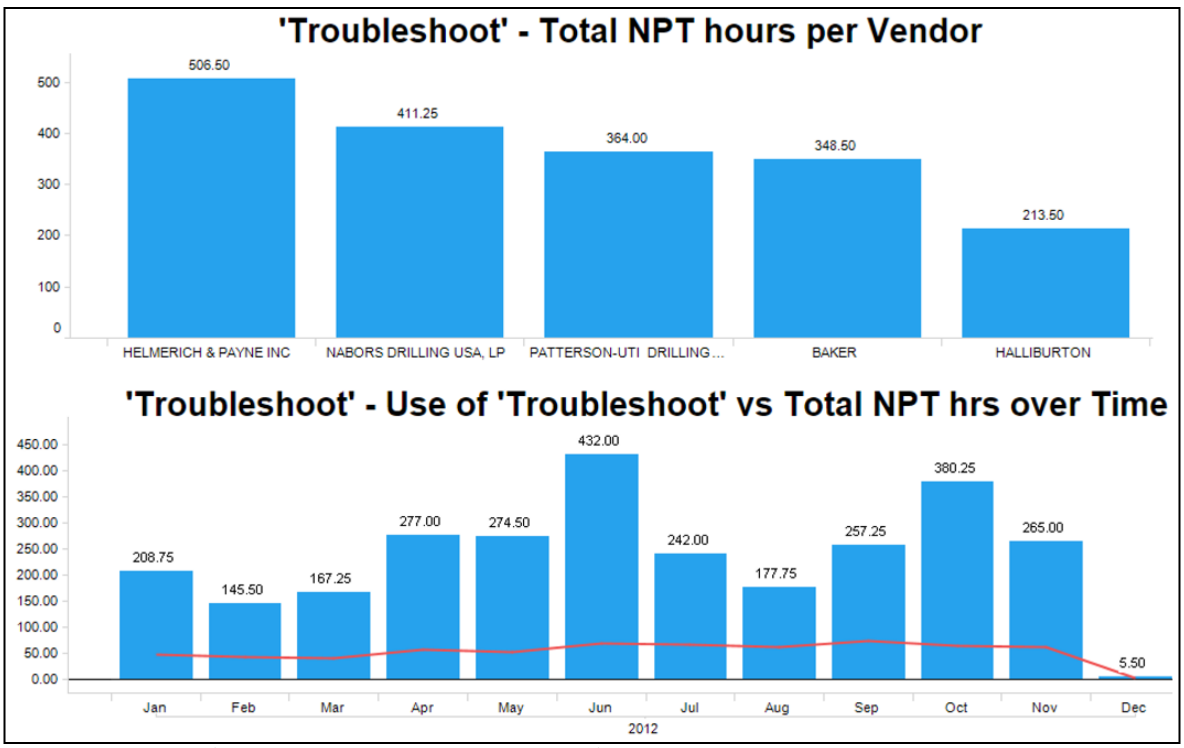

Figure 4: Example of text mining analysis results. All instances of NPT with comments containing the word 'troubleshoot' were aggregated by accountable party and time.

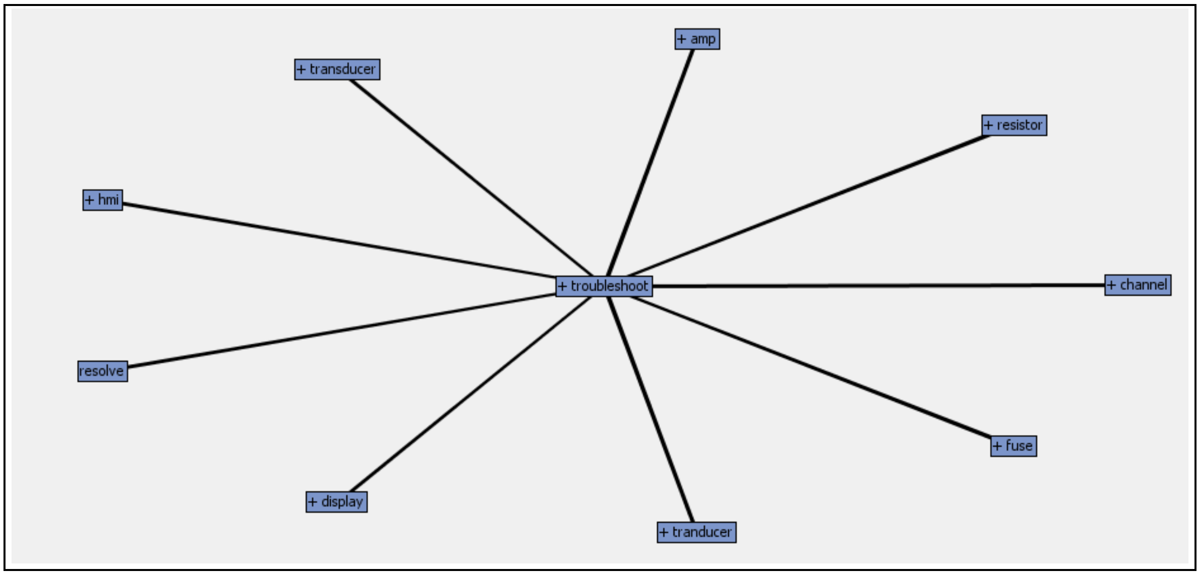

Figure 5: Example of a concept link diagram produced with SAS software suite

The biggest takeaway from this exercise was the importance of collaboration. The Advanced Analytics team could quickly and efficiently navigate through the complex SQL database because of communication with the IT team. Input from the DO team brought context to the data and allowed it to be more accurately assessed. An in-depth analysis of the NPT data identified areas of improvement in data collection, enabling the Data Management team to make actionable decisions. Some of these decisions included defining the minimum information required in all future NPT comments, discussing these new standards in training events for company men, and mandating the use of spellcheck before the inclusion of comments in the SQL database. Another key factor in the successful completion of this project was the data scientist's ability to address the needs outlines by the SME. Analytics professionals are often tempted to provide solutions to problems not identified by the SME resulting in in the dissatisfaction of all parties involved. When presented with a solution not in scope of the original problem, SMEs may feel that their needs are not adequately addressed, thereby leaving the analytics professional frustrated as their solution is not utilized by the SME. Heeding the input of the SME is crucial as it guides data scientists to focus on the most pressing issues and together they can implement sustainable change within the business.

Predictive Asset Failures

Recent decreases in drilling and completions activity have caused the industry's focus to shift toward improving operational efficiency in upstream production. Oil and gas companies have begun to utilize analytics in the production arena, allowing data to influence key operational decisions such as artificial lift method selection and artificial lift performance optimization. An electric submersible pump (ESP) is an artificial lift method used to lift large volumes of fluid from wellbores that are experiencing a decline in reservoir pressure. These high-powered pumps can be used in a variety of environments. However, installation and pulling operations for ESPs can be very costly. Additionally, these pumps tolerate minimal solids production. Therefore an excess of residual proppant can result in expensive ESP repairs. An understanding of the various factors contributing to ESP lifespan could lead to significant cost savings by minimizing repair and pulling costs as well as decreasing lost production. This section will illustrate the multidisciplinary approach taken to predict the duration of ESP function, also called the ESP lifespan.

Last year, the production Operation Excellence (OE) team came together with the Advanced Analytics team to collaborate on a proof-of-concept exercise to examine ESP failures. While much academic work has been done focusing on ESP life prediction, few professionals have incorporated predictive modeling into their efforts. Guo et al. built a support vector machine model to predict ESP failures using electrical and frequency data (Guo, 2015). Gupta et al. provided an analytical framework for real-time ESP predictive modeling (Gupta, 2016). The project presented in this paper approached this problem in a slightly different manner, using a combination of historical static data and summarized time series data to create a predictive model to forecast ESP lifespans as well as identify their key drivers. A panel of production engineers specializing in artificial lift met with data scientists from the Advanced Analytics team to identify all the potential factors that affect ESP function. The resulting list included static metrics from several areas such as wellbore geometry and completions design, as well as real-time production metrics such as casing pressure. Data collection required collaboration with the Data Management team to identify the location of static variables within several SQL databases. The Automation team was also consulted to help access artificial lift metrics collected from the field in real time. Historical data from 51 wells were used to create predictive models resulting in ESP lifespan forecasts which were later compared to the actual ESP lifespans. The predictive models appraised in this exercise included stepwise multiple logistic regression, decision tree, and high performance random forest (HP Forest). These were not the most complex models available, however they each have relaxed requirements in regards to data size and were most compatible for this dataset. Additionally, the modeling exercise provided a ranking of the included variables' influence on ESP lifespan. Model accuracy was evaluated using average squared error, the difference between the predicted value of the target variable and its actual value. The HP Forest model was found to have the lowest average squared error of all the models considered and was chosen as the champion model. The HP Forest model is an ensemble model composed of a number of unique decision tree models. The champion model had an average residual of 5.06 days and an average root square error of 31.48 days. This means that, on average, the model predicted ESP lifespan to within approximately five days of the true ESP lifespan and 90% of the model's predictive error was within +/- 30 days of the true ESP lifespan. It bears noting that the model tended to underpredict on ESP lifespans greater than one year, but this could be due to a shortage of long-lived ESPs in the training dataset. Many key drivers in the model measured ESP cycling, instances where the ESP is temporarily shut down and restarted, as well as completions design.

This exercise was instrumental in introducing analytics into the production operations environment. Historically, most production groups had relied on business intelligence and descriptive statistics to implement reactionary changes and thus they were somewhat hesitant to implement novel analytical methods. The OE team was crucial to bridging this gap. This group was created to enhance operations by utilizing new technology and techniques in the production arena. More than simply operational expertise, the OE team is staffed with professionals who embrace data-driven decision making and are willing to invest in change. The collaboration between the OE team and the Advanced Analytics team has grown beyond this project and has allowed for data science to be more readily accepted by other production teams.

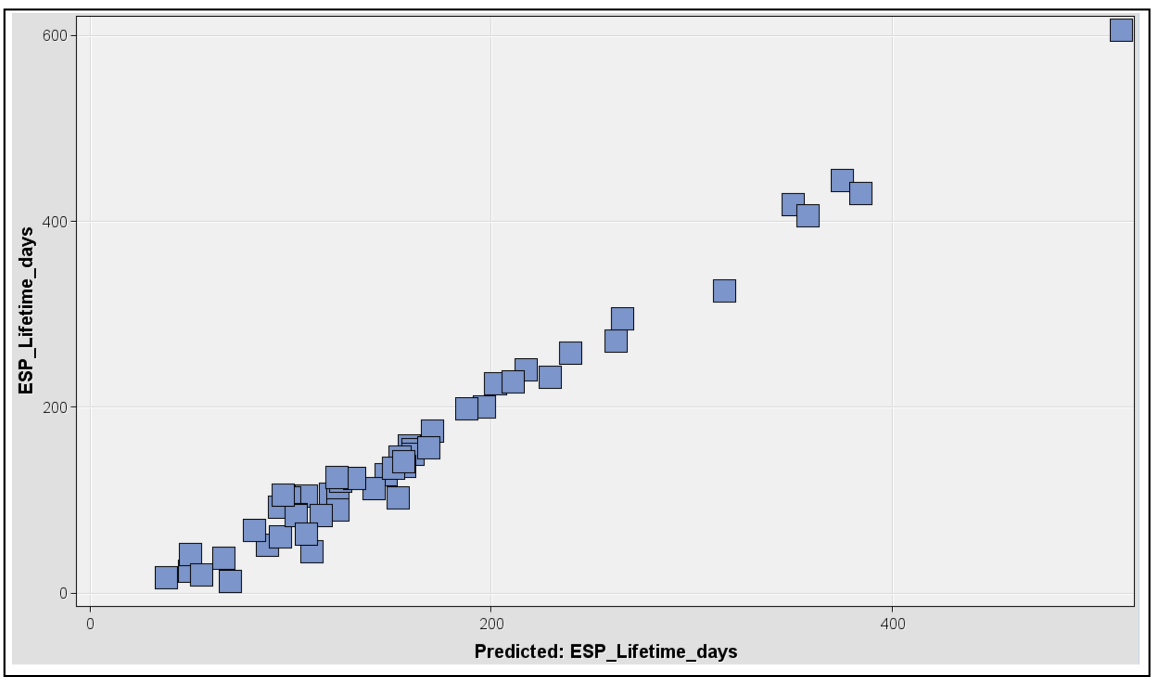

Figure 6: Comparison of predicted ESP lifespan vs actual ESP lifespan

Machine Learning in Completions

The speed and quantity of data associated with real-time completions operations presents both unique challenges and opportunities for analytics, especially when employing high-level artificial intelligence and machine learning. Without complete buy-in from all involved parties, it becomes virtually impossible for such solutions to become valuable and actionable. It takes close collaboration between analytics, IT, operations and others to succeed in bringing action to data.

Devon's real-time center is uniquely positioned to take advantage of advanced analytical solutions. Thusly, the analytics team was tasked with creating a new method of predicting screenouts before they happen. According to the Schlumberger Oilfield Glossary, a screenout is "a condition that occurs when the solids carried in a treatment fluid … create a bridge across the perforations or similar restricted flow area." (Schlumberger Oilfield Glossary, 2017). Less severe screenouts can sometimes be remediated without large expense or significant loss of production, but larger screenouts can require the use of a coil tubing unit or other more costly means of remediation. The ability to recognize screenouts before they happen could not only save lost production from screened-out stages, but reduce overall completions costs as well.

Data collection and processing can be particularly challenging with completions data. Completions personnel are becoming increasingly well-versed in the value of said data, both static and real-time. Static data, regardless of what database or storage solution a company has, will always be subject to the quality and accuracy of its entries. Careful quality checks will need to be performed on any and all data, often with the help of subject-matter experts. A data scientist will often be unable to perform proper checks and filtering without domain knowledge. These difficulties can be resolved over time with the creation of and adherence to strict data entry standards. However, real-time data offers even larger challenges. While drilling operations use WITSML (wellsite information transfer standard markup language) to set guidelines for data transmission, completions contractors and operators generally do not adhere to one particular set of standards. Before even considering artificial intelligence, the mechanisms for collecting and storing such data need to be decided upon and obeyed long enough to create a data store worth utilizing.

The Automation team at Devon worked alongside the Completions team to come up with a solution to not only organize and store real-time data, but to integrate it with other data sources and provide a data store to be utilized by other processes. This real-time data (labeled in Figure 5 as "Completions Data") flows out of the frac van through a communications skid and field network before landing at an interface service in the Devon data center. The universal file loader (UFL) processes the serial stream according to configuration imposed by personnel in Devon's real-time center. Upon conclusion of a job, the communications skid is free to either follow that frac van to the next job or be utilized elsewhere. This data could now be collected and stored in an accurate, comprehensive, and timely manner. This novel data collection process was a collaborative effort between the Advanced Analytics, Automation, and Completions teams. It also required adjustments to data collection practices in the field and was necessary to facilitate proper data preparation prior to data mining and modeling.

Once the data were organized and easily accessible, data mining could begin. It's important to use an effective modeling cycle with the help of a data mining procedure such as SAS Enterprise Miner's SEMMA process. Through a unique combination of attributes acquired through aforementioned data automation processes and calculations provided by completions experts, various models could be created, assessed, and operationalized. Of significant note is the "explore" phase of the SEMMA process. During this phase, one gleans understanding of the data through the discovery of variable relationships, distributions, and abnormalities through the use of data visualizations and statistical tests. Had the aforementioned efforts by Completions and Automation teams not taken place, not only would the data be essentially unattainable in a usable format, but it is unlikely one would be able to proceed past this point as large quantities of unwieldy, dirty data would prove unacceptable for modeling.

It has proved important that, when creating new solutions, all viable models and processes are considered. Target variables must be chosen appropriately and all types of machine learning evaluated for feasibility. SMEs need to be consulted in order to properly choose inputs and properly filter and transform data. To do this, the Advanced Analytics team has worked very closely with the Completions team to track these models and assess their efficacy. Efforts are currently under way to continuously improve predictive models as well as their real-time visualizations and processes for interaction with the field.

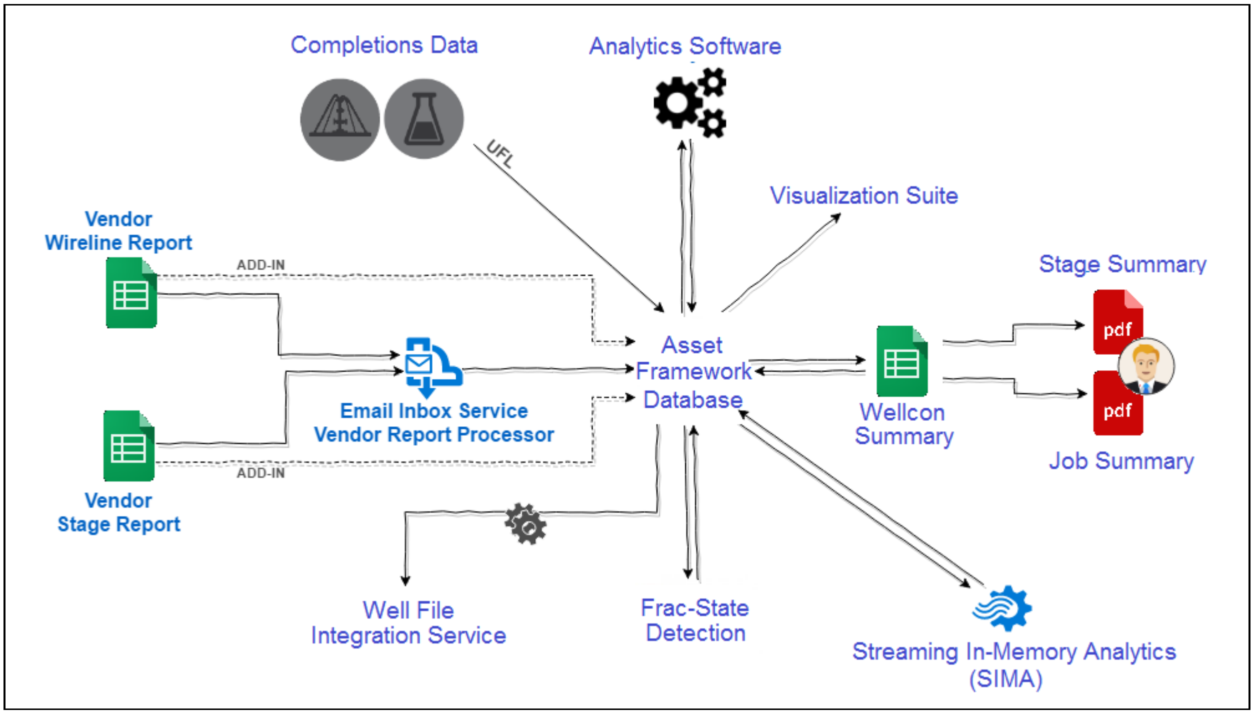

Figure 7: High level diagram outlining completions data flow

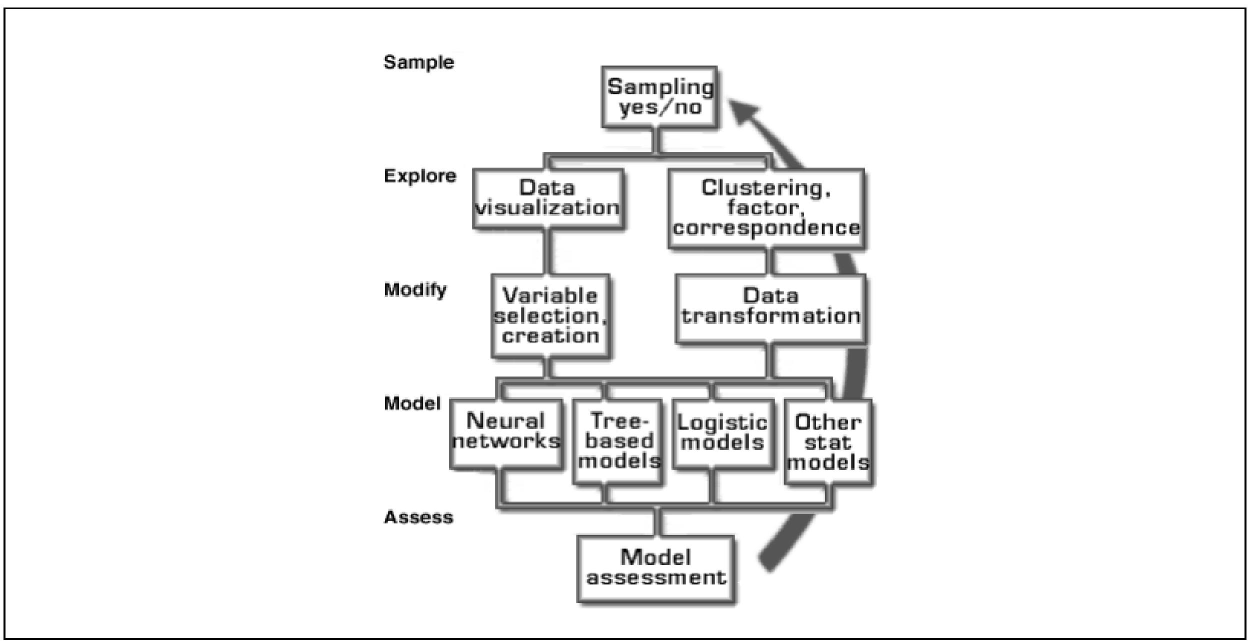

Figure 8: Overview of SEMMA process for data mining (SAS Institute, 1998)

Figure 9: Real-time screenout risk dashboard prototype

Conclusions

Fostering a culture of analytics and proving its value in an industry that's not always receptive to change can be extremely testing. The key to success lies not within siloed work efforts, but within constant communication and teamwork. When the human aspect of analytics is leveraged in close combination with artificial intelligence and machine learning, the results can be extremely powerful. With big data and analytics leading the way to the future, it's never too soon or too late to start becoming invested in the process and reap the benefits anywhere you can make smart, data-driven decisions.

References

1. Mitchell, Julian. 2016. "This Startup is Using Big Data and Smart Tech to Save the Oil Industry". Forbes. https://www.forbes.com/sites/julianmitchell/2016/10/29/this-startup-is-using-big-data-and-smart-tech-to-save-the-oil-industry/#5dcaa262190f.
2. Accenture and Microsoft. 2015. "Digital Energy Survey 2015 from Accenture and Microsoft". https://www.accenture.com/us-en/insight-digital-energy-survey-2015-accenture-microsoft.
3. Guo, D., Raghavendra, C.S., Yao, K.T., Harding, M., Anvar, A., Patel, A., 2015. Data Driven Approach to Failure Prediction for Electric Submersible Pump Systems. Paper SPE-174062-MS presented at the SPE Western Regional Meeting, Garden Grove, California, 27-30 April, https://doi.org/10.2118/174062-MS (https://doi.org/10.2118/174062-MS)
4. Gupta, S., Nikolaou, M., Saputelli, L., Bravo, C., 2016. ESP Health Monitoring KPI: A Real-Time Predictive Analytics Application. Paper SPE 181009-MS presented at SPE Intelligent Energy International Conference and Exhibition, Aberdeen, Scotland, 6-8 September. https://doi.org/10.2118/181009-MS (https://doi.org/10.2118/181009-MS)
5. Schlumberger Oilfield Glossary. 2017. http://www.glossary.oilfield.slb.com/Terms/s/screen_out.aspx
6. SAS Institute. 1998. http://sceweb.uhcl.edu/boetticher/ML_DataMining/SAS-SEMMA.pdf