AI for Upstream Oil & Gas

RAPIDMINER

Table of Contents

The Promise of Data Science for the Oil & Gas Industry

There's never been a greater incentive for oil & gas companies to become more efficient and make smarter process decisions. As the amount of data that's gathered through drilling operations increases, so does the motivation to use it—the fluctuating cost of commodities, increased competition, and ever-growing need to minimize environmental risk are all evidence of this.

Traditional analytics techniques no longer help to create meaningful process changes—at least not on their own. While they can help to understand what has happened in the past by analyzing historical data and providing intuitive visualizations, when it comes more challenging tasks such as finding new drilling locations, knowing what has happened isn't enough. Each location comes with its own unique geological properties, and given how high the cost of drilling is, looking in the wrong place has expensive consequences.

In the case of exploration, combining historical data with seismic data for a specific region of interest helps to not only predict the right places to drill, but also which reservoirs will produce the greatest yield.

To be clear, this doesn't mean that historical data is useless—just that it's incomplete. When used to power more forward-looking techniques like data science, it can help to uncover insights that you wouldn't find otherwise. For example, in the case of exploration, combining historical data with seismic data for a specific region of interest helps to not only predict the right places to drill, but also which reservoirs will produce the greatest yield.

While embracing predictive analytics will require a shift in your thinking, it won't require you to drastically reshape your teams or organizational structure. In fact, many oil & gas companies are already using the tactics described later in this paper to get the most value from their existing resources.

In this guide, we'll discuss how oil & gas companies can make data-driven decision making the norm by making data science more accessible. We'll also explore how the upstream segment can leverage data science to create efficiencies in the key areas of drilling and asset maintenance, and explain how those efficiencies have helped companies lower their costs and maximize profits.

Making Data Science Accessible to the Upstream Segment

Even if you recognize that data science has the potential to help transform drilling operations, you may be wondering how to get started. If so, you're not alone—between the sheer volume of data that's collected through daily operations and a lack of in-house data science experts, it can be difficult to envision a path to success that doesn't involve throwing out your existing playbook.

While this is understandable, the truth is that organizations who are finding success with data science aren't changing the way they do things from the ground up—they're using all the information that they're already gathering to modernize processes in a way that helps them compete more effectively. Every company has different levels of data science expertise, ways of going about their business, and technology at their disposal. The key to success with data science is to embrace those differences and create an approach that works for *you*.

To make this more concrete, it's helpful to understand the difference between "going deep" and "going wide."

The Trouble With Going Deep

"Going deep" means treating data science like a specialized function. Companies who take this approach isolate analytics teams from the rest of the business, give them the freedom to explore innovative new ideas, and expect that they'll eventually develop a groundbreaking solution. This almost never happens. A much more common outcome is that research is conducted for research's sake, without much regard for business impact. While you may bring a few models into production this way given enough time, true transformation will elude you for three key reasons.

Lack of business context

Data science teams need **clearly defined business objectives** and a deep understanding of the data they're working with to develop the right solutions. Upstream activities like exploration are complex, and learning how to interpret the data they produce isn't easy—especially when you need to understand the data well enough to create useful predictive models.

Even the data scientists who are capable of learning the nuances of key business processes won't do so overnight. That's why at best, isolating data science teams from the rest of your business will delay the creation of useful models. At worst, it'll prevent those teams from being able to create and deploy models at all.

Distrust of models and their predictions

It's no secret that many decision makers don't inherently trust AI solutions. In fact, a recent study found a strong correlation between a lack of success with AI and low levels of trust. This problem is exacerbated when business stakeholders aren't involved in projects from an early stage.

To sign off on deploying a model and making decisions based on its predictions, leaders within your organization will need to understand how and why it works. If you're unable to explain this or clearly communicate the monetary value of predictions, it's very unlikely that you'll be able to get a model deployed regardless of its accuracy.

Limited long-term value

The data science lifecycle doesn't end with the deployment of a model. For a solution to produce an appropriate return on investment, it needs to be carefully monitored over time. When models are left alone post-deployment, they will eventually become less accurate and make flawed predictions, which presents tangible risk for organizations making decisions based on them.

Organizations who isolate their data science efforts are especially susceptible to this. Ongoing model management typically isn't the top priority for data scientists—they're more focused on applying innovative techniques to new business problems. Naturally, when their focus shifts to the creation of new models, the ones you have in production can become an afterthought if you don't have a process in place for long-term monitoring.

The Case for Going Wide

"Going wide" describes a holistic approach to enterprise data science. Organizations who take this approach don't treat data science as a specialized function—instead, they aim to align their **people, processes, and technology** in a way that facilitates impactful projects.

The "people" part of this is especially critical. Without involving the right stakeholders and helping them develop the skills to meaningfully work with data, your projects won't generate the desired results. By contrast, if you make **upskilling** a priority, you help business experts build their data literacy, enable data scientists to learn from others' work, and promote collaboration between the two groups. When data and insight flows freely throughout an organization, everyone wins.

Of course, to adopt this type of approach, you need to have the right technology in place. After all, regardless of how well a business expert understands data science concepts, it's not feasible to expect that they'll learn Python well enough to create truly differentiated solutions. By leveraging an end-to-end, code-optional data science platform like RapidMiner, you ensure that a conceptual understanding is enough for non-coders to start contributing to projects.

In this next section, we'll explore how Wipro, one of RapidMiner's key strategic partners, leveraged our platform to build scalable and reusable data science solutions that have helped to transform key processes for organizations operating in the upstream segment of the oil and gas industry.

Using AI to Improve Accuracy in Oil Exploration and Drilling Asset Maintenance

As we mentioned at the outset of this guide, exploration for any Greenfield location is difficult and expensive. The development of new oil fields requires state-of-the-art technologies with huge investments over time. But the investment return is often unpredictable. Conventionally, decisions for exploration tasks are heavily driven by expert knowledge. But even for experts, most scenarios are not straightforward enough to be conclusive. This makes the entire exploration process time-consuming and expensive.

AI can help address this problem by leveraging a large amount of data from various sensors. This data improves decision-making by deriving insights that help companies better detect anomalies and make faster, more informed decisions. By predicting failures in assets such as pumps and drills before they happen, organizations can improve asset availability, reduce maintenance costs and enhance productivity.

The Major Problem in the Upstream Segment

The oil and gas industry has three segments: upstream, midstream and downstream. The upstream segment is the most capital intensive and brings the greatest uncertainty, which poses risks to investments. Upstream activities face the typical dilemma of **"one rock, two geologists, and three opinions,"** a common reference for uncertainty. Using AI in the upstream segment can reduce ambiguity through anomaly detection and fast-tracked outcomes.

As most players in the industry know, finding new drilling locations is complicated. Most locations come with their own set of challenges that make it difficult to accurately predict success. These challenges include difficult environmental conditions (e.g. in extreme climates like the Arctic), complex geological geometry (thin layers of oil-saturated rocks without any uniformity), locations beneath very thick layers of seawater and salt minerals, or poor permeability (such that the hydrocarbon is nearly immobile within the reservoir rock).

Wipro's Data Science group has developed a complete predictive asset maintenance workbench using AI to improve asset reliability and address the issues mentioned above. These predictive solutions predominantly leverage the end-to-end data science platform RapidMiner to quickly realize use-cases. Two of these new solutions, **Subsurface Classification** and **PCP Pump Failure Prediction**, reduce the cost of exploration and in well-drilling processes for greenfield developments.

Subsurface Classification

Drilling involves the use of multiple sensors to identify the subsurface that indicates the presence or absence of natural resources at the bottom of the drill. Geologists look for specific rock layers that have a high presence of carbon and hydrogen content, which indicate proximity to oil or gas reserves. However, sensor values are often so noisy that subject matter experts are the only ones who can understand the patterns well enough to best predict the subsurface type.

The AI-based subsurface classification solution learns from the past patterns of sensor values in a specific geography. This learning can later be applied to predict the subsurface type in any new oil field in a similar region. The AI/ML-driven solutions act as a smart cognitive recommender system to help the SME improve decision accuracy, which in turn reduces the cost of exploration. The classifier provides a strong accuracy of approximately 78-80% when used in adjacent locations. When tested for layers that are not adjacent, the classifier provides an even higher accuracy rate of 97-98%.

PCP Pump Failure Prediction

Progressive Cavity Pumps (PCP) are used to remove viscous oil, water and sand during drilling operations. A PCP pump is critical while drilling is in progress. Any unplanned failures can disrupt the operation and increase the costs of a drilling activity. The PCP pump failure prediction is a machine learning-driven solution that includes two models to predict PCP Pump failures days in advance.

The solution includes data from five pump vendors, has multiple combinations of pump configurations and features, and analyzes past maintenance records. The solution allows the drilling team to be prepared to proactively fix pump issues and avoid disruption in ongoing drilling.

The first model, the Random Forest classifier, detects pump failures with an accuracy of approximately 80% but can throw a false positive. To reduce those false positives, another model, Isolation Forest, is used, which is based on an anomaly detection technique. As such, a true failure prediction is considered only when both models predict the same outcome. The solution was able to provide approximately 95% accuracy and predict failure at least 2-3 days ahead of the reported failure date.

Predictive asset maintenance leveraging AI/ML tools and techniques can help mitigate risk and reduce the cost escalation that results from the uncertainties associated with oil and gas exploration.

To Wrap-Up

While heightened competition and a growing need to minimize negative environmental impact create a strong incentive for oil & gas companies in the upstream segment to become more data-driven, doing so isn't as simple as hiring a team of data scientists and giving them access to raw data.

By using a comprehensive data science solution like RapidMiner, your organization can fast-track model development without compromising the quality of the outcome––leaning on the domain expertise of a strategic partner like Wipro can help to further accelerate results by bridging the gap between process knowledge and data science knowledge.

By bringing in the right resources to make use of your existing people, expertise, and data, you can reimagine existing processes to improve reliability and predictability in exploration and drilling.

About the Authors

Manish Okhade

Practice Lead, Predictive Asset Maintenance, AI Solutions, Wipro Limited

Manish has more than two decades of industry experience in different domains like ER&D, digital and artificial intelligence. Manish has been involved with the development of Wipro's AI platform – Holmes – since its inception and developed multiple AI/ML use cases across various industries. Manish is presently leading the Predictive Asset Maintenance (PAM) Practice in Wipro's AI/ML solutions group. His current focus areas are building accelerators, thought leadership and forging partnerships with niche startups in the PAM domain.

Anurag Allena

Product Marketing Manager, RapidMiner

Anurag is a Product Marketing Manager for RapidMiner. His work involves developing a deep understanding of our platform, communicating its' value, and explaining the types of use-cases that enterprises can leverage it for.

For those driven to accelerate the pace of transformation, RapidMiner is the enterprise-ready data science platform that amplifies the collective impact of your people, expertise, and data for break-through competitive advantage. RapidMiner's data science platform supports all analytics users across the full AI lifecycle. The RapidMiner Academy and Center of Excellence methodology ensure customers are successful, no matter their experience or resource levels. Since 2007, more than 1 million professionals and 40,000 organizations in over 150 countries have relied on RapidMiner to bring data science closer to their business.