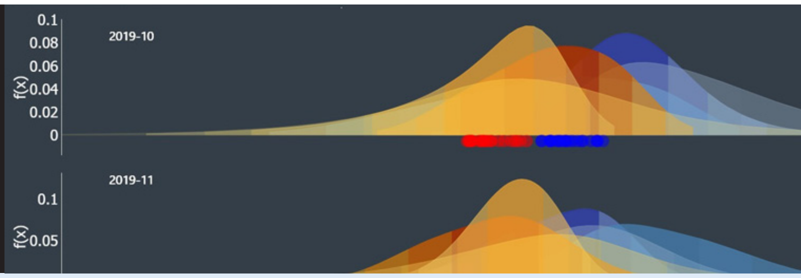
argusmedia.com

Argus-Vortexa Data Science Insights

Speaker Q&A



Cetin Karakus, Global Head of Quantitative and Analytical Solutions, BP. Cetin has two decades of experience in designing and building large scale software systems. Over the last decade, he has worked on design and development of complex derivatives pricing and risk management systems in leading global investment banks and commodity trading houses. Prior to that, he has worked on various large scale systems ranging from VOIP stacks to ERP systems. In his current role, he has been leading the efforts to build a state-of-the-art Big Data analytics platform through the joint efforts of multiple engineering teams. He works at an intersection of data, analytics and trading with the constantly shifting ground that is global energy markets. Cetin has a degree in Electrical & Electronics Engineering and enjoys thinking and reading on various fields of humanities in his ever diminishing spare time.

What are the benefits of predictive analytics and where is the best place to use predictive analytics in the oil and gas industry?

I would say it is the holy grail of data analytics.

Predictive analytics allows you to predict how the dynamics of the world you are operating in will play out in a sufficiently detailed and certain way, so you can act on this knowledge in advance. For example, if you are trading in financial markets, you could build up positions in instruments that are going to appreciate; if you are monitoring the weather you can warn people of hurricanes and other adverse conditions to allow them take precautions to save lives and property; if you are in healthcare, you can warn people of potential medical problems so they can find treatment before these potential problems become more serious; if you are in charge of an NGO with constrained finances, you can use predictive analytics to focus your efforts on the specific policy actions that will result in the best measurable outcomes for each dollar spent.

There are too many areas in which predictive analytics can be used in oil and gas industry to list them all here. But the following are a few key areas in which the effects of predictive analytics will be more visible.

1 Efficient exploration: Upstream exploration activities is very expensive and predictive analytics could help with directing work to the most prospective sites. Even after site selection and the decision to go ahead with development, optimal configuration of specific site plans — for example, where to drill wells in an onshore field — could benefit greatly from the power of predictive analytics.

2 Optimal maintenance: The oil and gas industry depends on a huge amount of physical infrastructure. This constantly needs to be maintained for safety and efficiency. Predictive analytics could help ensure the required level of maintenance is carried out at the lowest possible cost, keeping infrastructure such as wells, pipelines, refinery units and LNG tankers safe and functional, while avoiding waste and unnecessary replacements.

3 Supply and demand forecasting: Like any commodities, oil and gas markets are driven by supply and demand. Many forces come together to shape these dynamics, including domestic demand, import and export flows, disruption, capacity constraints, and local and foreign competition. Predictive analytics could help with forming high-fidelity supply-demand models and then using those models to gain insights into where markets are heading and the possible effects of future events — refinery shutdowns, for example.

What is needed to execute predictive analytics?

There are four key fields — data, infrastructure, modelling and experiments — and you need capable people operating in each of them.

Data: Data is the main ingredient and has to be collected, cleaned, normalised — put into a common form or shape — stored, secured and served up for downstream processes.

Petroleum
illuminating the markets

Infrastructure: You need a flexible computer infrastructure to process data, run models and simulations, and generate visualisations and reports.

Modelling: You need to build models that are sufficiently well parameterised abstractions of the reality and that can be calibrated with the new data. The right approach is to combine common sense with the scientific rigour of experts in the field.

Experiments: You need to build a highly iterative experimentation loop where forecasters can easily form new hypotheses, create experiments to test such hypotheses and run those experiments on a sufficiently large data set to either refine existing predictive models or create new ones.

What is the difference between artificial intelligence (AI) and big data?

Nowadays, people mean machine learning (ML) when they refer to AI. Roughly speaking, AI deals with creating capabilities that are not explicitly programmed when a system is built — giving a computing system the ability to potentially deal with situations that could not have been foreseen when it was built.

There are two major approaches to building this ‘smartness’ into a system.

1 Rule-based logical inference: You can build a database of facts and rules — for example, ‘all swans are white’, ‘swans can swim’, ‘dogs have four legs’ — and a logical inference engine can generate new facts from an existing compendium of facts. This is the basis of expert systems that became popular in the 1980s and early 1990s.

2 Data-driven learning: Rule-based systems require not only the initial seeding of rules, but continuous maintenance of those rules — amendments, additions and removals — throughout the system’s life cycle. Systems that rely on data-driven learning, on the other hand, extract the rules and facts from the data itself, so the system does not need an initial seeding of rules. It can learn such rules from the data itself. This process is dynamic, so as new data arrives, the system updates its rules to adapt to the ‘new reality’.

Data is the lifeblood of ML-based AI. While rule-based AI systems rely on hand-crafted and refined human expert knowledge, ML systems have to extract such gems of knowledge from the data itself and will need heaps of data to be able to discern the useful information buried among all the noise. That is where big data comes into the picture. ML relies on the availability of big data, which we can define as a large quantity of data often coming in different forms and potentially at high frequency — for example, sensor recordings arriving every few milliseconds. ML will not be viable without big data because it will not reach the required level of sophistication and high fidelity of predictions without the large datasets that can feed its underlying algorithmic machinery.