TECHNOLOGY SPOTLIGHT

Automated Machine Learning: Fast-Tracking the Journey to Becoming an AI-Powered Enterprise

December 2017

By Jessie Cai and Chwee Kan Chua

Sponsored by DataRobot

The future of an artificial intelligence (AI)-powered world is already here. Be it the fast diagnosis of cancer, convenient identity verification via face recognition, or autonomous driving, AI is transforming the way enterprises and consumers alike work, live, and play. Machine learning – as an effective means to achieve AI – has been making a mark in the enterprise world, through its ability to solve complex business problems with pattern recognition by sifting through vast volumes of data. AI has attracted great interest from organizations due to the promises of higher productivity, reduced cost of labor, and shorter time to value generation. Underscoring this is a recent series of IDC Cognitive/Artificial Intelligence surveys conducted in the United States, Western Europe, and Asia/Pacific, which found that 67% of organizations globally have already adopted or plan to adopt AI in the next five years. This IDC Technology Spotlight highlights the latest trends in the adoption of AI/machine learning, and takes a look at how DataRobot's automated machine learning platform helps pave the way for organizations' transformation into AI-driven enterprises.

Harnessing AI to Unleash the Power of Data

The world is generating more data with every passing second – digital-native organizations have created waves of disruption across industries, and inspired tech entrepreneurs to turn more things into data-generating devices. For organizations established before the digital era, many have successfully digitized their major business processes and are faced with a massive collection of historical records. The natural next step is to explore viable means of exploiting and even monetizing data.

In Asia/Pacific, IDC research found that 21% of organizations have integrated data from multiple sources to derive business insights; 41% prioritized real-time decision support from big data; and 16% considered the creation of data products as the main goal of investment in big data and analytics initiatives for the next 12–14 months. For organizations in search of an effective means to better harness the power of data, AI is becoming increasingly important. According to the IDC Worldwide Semiannual Cognitive/Artificial Intelligence Systems Spending Guide Forecast 2016 H2, global spending on AI is expected to reach close to US$58 billion by 2021. The banking, financial services and insurance (BFSI), manufacturing, retail, public sector, and healthcare industries are anticipated to invest the most.

There are common grounds accounting for the strong demand for AI technology across industries:

■ **Versatility.** AI can serve many purposes in various business processes – sales and customer engagement, legal and human resources (HR) services, knowledge system augmentation, asset maintenance and management, etc.

■ **Adaptability.** AI capabilities, as compared to traditional modeling, not only provide real-time analysis and actionable insights to reduce tedious manual processes, but also constantly improve the accuracy through model validation, maintenance, and refinement.

■ **Ecosystem reach.** AI is one of the keys to connect to the growing ecosystems in the data economy, in which organizations thrive by exploiting massive amounts of data, timely and intelligently.

Reaping the Benefits of AI: An Industry Perspective

Machine learning as a cross field between computational statistics and computer science is a highly effective means to achieve AI. Machine learning algorithms can learn from data iteratively and allow computers to detect patterns without being explicitly programmed. Together with predictive analytics, they account for most of the intelligent applications in the business world, including recommendation system, customer profiling, anomaly detection, a large part of autonomous driving, and more. The following sections take a closer look at how AI/machine learning is leveraged in the BFSI and healthcare industries.

Adaptive Know Your Customer (KYC) for Revenue Generation and Risk Abatement in Banks

Bank customers are now digitally connected 24/7 via many alternative payment and financing channels. The recently introduced complexity has put compliance issues and system risks of banks under increased scrutiny. At the same time, banks need to innovate – often together with digital disruptors – to provide a seamless experience and increased convenience for their customers, in order to stay competitive. For both imperatives – risk abatement and revenue generation, the essence is about knowing the customer better by analyzing all data possible, and doing so continuously. Banks are in general faster to realize the effectiveness of AI technology, such as machine learning in extracting insights from big data. Compared with traditional methods, AI has the unique ability to help banks stay adaptive to the schemes of bad actors and customer preferences, both of which are continuously changing.

Common use cases of machine learning applications in banking include:

■ Fraud detection that makes use of advanced behavior analytics

■ Marketing intelligence including product and service recommendation

■ Automated due diligence and compliance checks

Insurers in Pursuit of a New Ecosystem-Based Business Model

Many insurance companies are exploring an ecosystem-based business model to digitally collaborate with partners in a consumer-centric ecosystem. AXA, for example, announced in 2016 about collaborating with Alibaba and Ant Financial Services to distribute AXA's insurance products and services through the latter two's e-commerce and e-finance platforms. Collaborations such as these will not only bring more customer preference data to insurers, but also require insurers' business operations to become more automated, so that incidents and claims can be processed faster, event triggers can be translated comprehensively and in real time into actionable marketing insights, and the value of the e-platforms maximized.

Common use cases of machine learning applications in insurance include:

■ Adaptive premium calculation based on client behavior and preferences collected from edge devices

■ Automated appraising of road accident and medical claims to assist payout calculation

■ Marketing intelligence including product and service recommendation

Transforming Healthcare from Patient Engagement to Automated Claims Processing

AI systems can not only help healthcare providers to enhance patient experience, but also to address patient needs for better accessibility and affordability. These opposing realities will drive more providers to leverage AI to mitigate cost and improve quality of care.

Common use cases of machine learning applications in healthcare include:

- Mobile applications with AI embedded to enable remote engagement, pre-screening, and monitoring of patients' condition

- Prediction of patient readmission and prevention of sepsis

- AI tools to empower medical specialists, and provide better care faster by information augmentation

Strong Need to Take out the Technical Complexity of AI for Organizations

What sets AI apart from traditional analytical modeling is that it can learn from data on its own, freeing up manpower from designing and updating cumbersome business rules which can become outdated quickly. However, AI is still a complex technology which requires substantial onboarding efforts. Below highlights some of the primary challenges faced by organizations in adopting AI.

Lack of Skilled Workforce Is Hindering Wider Scale Adoption

Based on the recent IDC Cognitive/Artificial Intelligence surveys, very positive sentiment has been reported among existing users of AI worldwide. In particular, 77% of organizations in the United States have found existing AI adoptions meeting or exceeding expectations; and 85% of organizations in Western Europe anticipate spending on AI to increase further in the next two years.

However, there is presently a higher demand for data scientists and AI specialists than the market can supply. 73% of organizations worldwide said they do not have data scientists or AI specialists in-house; and 57% of organizations cited the lack of data and skill sets as a primary challenge in the adoption of AI technology, more so than any other inhibitors.

From Democratization of Data Science to Scaling-up Through Specialized Tools

Open source machine learning libraries and frameworks like Scikit-Learn, Tensorflow, and XGBoost can free data scientists from rewriting common algorithms. These have greatly contributed in the democratization of data science, allowing more data workers to try out predictive modeling and machine learning. However, this is not sufficient to bring AI technology to the mass market, largely due to the technicality involved and reusability issues incurred.

In Asia/Pacific, IDC found that less than 17% of organizations used any machine learning libraries and frameworks, despite having adopted or having firm plans to adopt AI technology. Even for organizations with in-house experts, there are hurdles in model building, integration, and deployment to overcome:

- The necessary data preparatory and exploratory process can be highly repetitive, laborious, and time-consuming, in turn delaying value delivery and undermining stakeholders' trust.

- Generating high-quality predictive models requires convergence of data science and domain expertise. Focusing heavily on the technicality of the modeling process might distract the workforce from solving the actual problem.

- AI models generated may not be ready for production, slowing the time to value.

- Due to frequent changes of market dynamics, existing AI/machine learning models only stay fit until the next batch of training data becomes available. Data scientists are under pressure to deliver models and updates at a speed the business desires.

The reality is AI as a complex technology will not best scale by human learning. A more effective path forward is to have access to specialized services and tools that remove the technical complexities and cover all the needs from data preparation, feature engineering, model selection, and visualization. IDC believes automated machine learning tools will lend organizations an edge, especially when faced with a shortage of data scientists or AI specialists.

Considering DataRobot's Automated Machine Learning Platform

Founded in 2012, DataRobot is a Boston-based technology company offering an automated machine learning platform for users of all skill levels to build and deploy predictive models quickly and accurately. DataRobot provides robust and user-friendly modeling, and allows users to focus on real-world problems without worrying about the technicality of the modeling process. The platform makes the community of top data scientists more accessible, creating opportunities for businesses, big and small, to leverage best-of-breed modeling capabilities. Presently, users of DataRobot have generated more than 200 million predictive models through the fully automated model building tool in which curated machine learning algorithms are embedded and updated continuously.

DataRobot's automated machine learning platform offers the following key benefits:

- Requiring **zero coding**, the tool is extremely easy to use. Users are only required to bring the training data.

- **High-quality outcome** is guaranteed by applying stacked algorithms curated by top data scientists from the best of open source communities.

- Being a **turnkey solution for end users of all levels**, the platform is able to create pervasive impact for organizations of all types.

A Zero-Coding Tool that Delivers Quality Results

Ease of use and quality results are the two key requirements for AI to run in the enterprise world and to create the much-needed process automations.

Zero Coding Required, Just Bring Your Data

DataRobot's automated machine learning platform allows users of any background to build and deploy advanced analytical models. The interface is simple and user friendly. Zero coding is required and users just need to bring along the data and specify the prediction target. The rest of the process, including data partition, standardization, optimization function selection, feature engineering, and model ranking, can go autopilot simply by clicking on the "Start" button. A REST API of the most suitable predictive model will be generated at the end.

Users can also manually select metrics for optimization from a list of options available. Likewise, for data partition and feature engineering, more experienced users can choose to manually overwrite the default settings, and run the model training process many times to explore different possibilities.

The generated API will be ready to be deployed onto a prediction server, or to be included in application packages with another click of a button. This effectively addresses the model integration issue, accelerating time to value by the consolidation of the necessary steps toward productive model consumption.

Quality Results from Curated Algorithms

DataRobot's automated machine learning platform acts as a curator that applies stacks of open source machine learning algorithms, such as Spark, H2O, Scikit-Learn, and TensorFlow, to the data set uploaded. These algorithms, called the "workers" in the DataRobot platform, are automatically recommended according to the characteristics of the given data set – number of variables, types of variables, sample size, number of missing values, etc. Noteworthily, the

recommendation itself is powered by machine learning and continuously improves itself with every new release.

The recommended workers run in parallel on the training data. When these algorithms/workers are all done with the training, models are generated, ranked, and presented in a leaderboard, showing alongside a few key validation and evaluation results such as gain, lift, confusion matrix, and predictive power. The platform would choose the top ranked model for API generation – users can of course choose to drill down the evaluation reports for more details, and decide whether they want to select another model, rerun with manual feature engineering, or perhaps go further back to find alternative data to enrich the training set.

To ensure users have access to the best of open source, and continuously so, DataRobot's team of data scientists keeps close watch on the development and refinement of open source algorithms, putting these to test in solving actual problems and working together with the developer team to make timely updates to the platform.

Farmers Insurance Group is a customer that is using DataRobot to tell signals apart from noises hidden in a massive pile of client data that keeps on accumulating. Acknowledging that every data scientist is biased, the in-house data scientist team has adopted a practical "let's test it" approach, making use of the curated workers of the DataRobot platform to ensure that they are leveraging the best model for each and every project.

Ramping up Data Utilization for Organizations

The increasing volume of data assets in many organizations creates much urgency to accelerate data utilization. For organizations having difficulties in delivering models at a pace meeting business needs, it is only a matter of time for them to start finding ways to become more productive. Reducing or avoiding recreating codes for different data sets, for example, can be done by specialized tools like automated machine learning platforms. Not only does it free data workers from repetitive code creation, it also empowers business users of all levels, with or without a data science background, to become autonomous in model building. For both types of workers, such tools can help them to stay focused on the discovery and delivery of insights – instead of working on the technicality details – to get more done in a much shorter time.

Turnkey Solution for Users of All Levels

While DataRobot offers a turnkey tool to automate the AI model training process based on machine learning, it does not work like a black box where no knowledge of the reasoning process is made available to the users. For non-technical users, this can translate to learning opportunities that do not distract them from being productive. For users with stronger analytical skills, their modeling productivity will be substantially boosted. The fact that all algorithms are gleaned from open source libraries and frameworks – and therefore totally traceable and referenceable – renders a high level of trust to the models generated for all users, from apprentice data workers, seasoned data scientists, and software engineers, to business analysts and business executives.

Having an easy-to-use, outcome-guaranteed, and transparent-by-design tool is key in reaching out to bigger user communities and widening value generation. Data workers' user experience is becoming increasingly important. The right tool can trigger a mindset change toward data-driven decision making across business units, and unleash the creativity of those that understand the business problem and have access to the relevant data asset – both of which are critical to speed up data utilization.

One of DataRobot's users is Traveloka, an online travel booking platform in Southeast Asia. Traveloka's growing popularity was creating challenges for its team of data scientists, who had to keep up with the demand of updating predictive models for their customers searching flight and hotel choices. Even with a small number of models, it was highly time consuming for the team to tune for optimal performance. By adopting the DataRobot platform, Traveloka's data scientist team was able to deliver accurate predictive models at a fast pace. The team was also able to effectively communicate with their management with the help of clear visuals.

A Flexible Tool for All Types of Organizations

The DataRobot automated machine learning platform, which works on cloud and on-premises, enables the democratization of data science, making machine learning available for all types of organizations. Apart from speed and transparency, flexibility is provided by DataRobot's automated machine learning platform. Organizations can make agile use of their data asset, and pay only by the computation resource consumed. Smaller companies with their own training data in store can have high-quality predictive models generated without having to invest heavily for infrastructure setup and advanced modeling capability, or having to worry about the lack of skilled staff.

Covering data preparation, exploration, prediction, and visualization, DataRobot's automated machine learning platform can process many data types including categorical, numerical, natural language, and streaming events. Its active customers include start-ups, small and medium-sized enterprises (SMEs), and multinational corporations (MNCs) across many industries including banks, insurance companies, media and marketing agencies, healthcare providers, telecommunication companies, Internet companies, sports and entertainment companies, and manufacturers. The use cases that DataRobot users have worked on are highly diversified:

■ **Reducing churn for consumer-facing businesses.** Machine learning models based on customer profile and preference data can effectively micro-segment customers and identify the high-value ones who are at risk of leaving for competitors.

■ **Reducing cost of prospect acquisition for marketing functions.** Machine learning models can help marketers find and connect to customers who are more likely to use their products and services. The better the predictive model, the more cost savings can be achieved.

■ **Detecting fraud for banking, fintech, and insurance companies.** Lingo patterns of fraudsters, for example, can be extracted automatically for the operational team to route cases of questionable nature for detailed manual review.

■ **Identifying anomaly signals from streaming machine data.** Part failures can be better predicted and prevented when there is enough historical data covering common failure types.

Some other noteworthy examples of how users are leveraging DataRobot's automated machine learning platform include:

■ Crest Financial, a fintech company, that uses the DataRobot platform for their rapidly increasing predictive workload. Accurate predictive models have substantially fast-forwarded the approval process for thousands of customers in need of a small amount of financial aid.

■ Michigan Center for Integrative Research in Critical Care, which is working on a project called Analytics for Hemodynamic Instability (AHI). The center is leveraging the DataRobot platform to continuously analyze sensor data streaming in at 60 to 1,000 data points per second from patients under constant monitoring. Hemodynamic Instability incidences can be predicted in a more timely and comprehensive manner to augment caregivers' decisions. Previously, such decisions were limited by caregivers' availability to interact with patients and checking the monitoring reports. AHI is currently being clinically validated.

■ NECCO is a United States-based organization that works on foster care, adoption, and a variety of social services. The organization has been hugely successful in helping children in need to find matching foster families through the use of advanced predictive analytical models. NECCO achieved this without hiring any data scientists.

Opportunities and Challenges Ahead

Greater efforts may need to be made by DataRobot in reaching out to the larger market if organizations continue to view data scientists or AI specialists as the primary custodians of AI. The reality is that there may not be enough data scientists or AI specialists for every organization in the future. In the hunt for data scientists, organizations may have overlooked or are not aware that there are companies generating high-quality automated machine learning models in less than an hour without any data scientists in-house.

As highlighted earlier, 67% of organizations globally have already adopted or plan to adopt AI in the next five years, which suggests a strong demand to consume AI. This is also expected to translate to higher demand for self-service automated machine learning tools or platforms like DataRobot's, once organizations fully realize these tools offer them an effective means to come onboard AI, quickly and economically.

Essential Guidance

Automation of operations and processes powered by AI allow organizations to unlock value from big data, and to discover and react to triggering events and market dynamics much faster than before. Recognizing the technical complexity of AI and its many implications to existing business processes and IT systems, varied market responses are expected including quick welcomes, outright rejections, gradual conversions, and sudden churns. When talking about adoption inhibitors, IDC has observed that many times it is not about the value propositions being not strong enough, but more so the lack of clarity of a path forward for the enterprise to come onboard, with concrete results seen in an agile and progressive manner. IDC recommends the following for organizations in search of productivity gains and new growth to stay competitive:

■ Organizations should focus on solving real-world business problems pertinent to their core competency, and minimize attention on the technicality of the modeling process. Finding the right use cases and making agile use of the ever-increasing data assets so that everyone can become better at what they do, should be a matter of greater importance.

■ Investment decisions on AI technologies need to be carefully reviewed to consider factors including quality of result, time to value, data workers' user experience, model transparency, and flexibility. An easy-to-use and outcome-guaranteed tool can hugely accelerate value delivery, and trigger a mindset change toward a data-driven and pro-innovation culture.

■ There will not be enough data scientists or AI specialists for every organization, and it may neither be economical nor necessary to have any in-house. The best way for digital technology to scale up is not through human labor. Instead, it is more efficient to have access to ready solutions, or a solid and specialized platform or tool covering data preparation, prediction, and visualization, to get more done with less time spent.