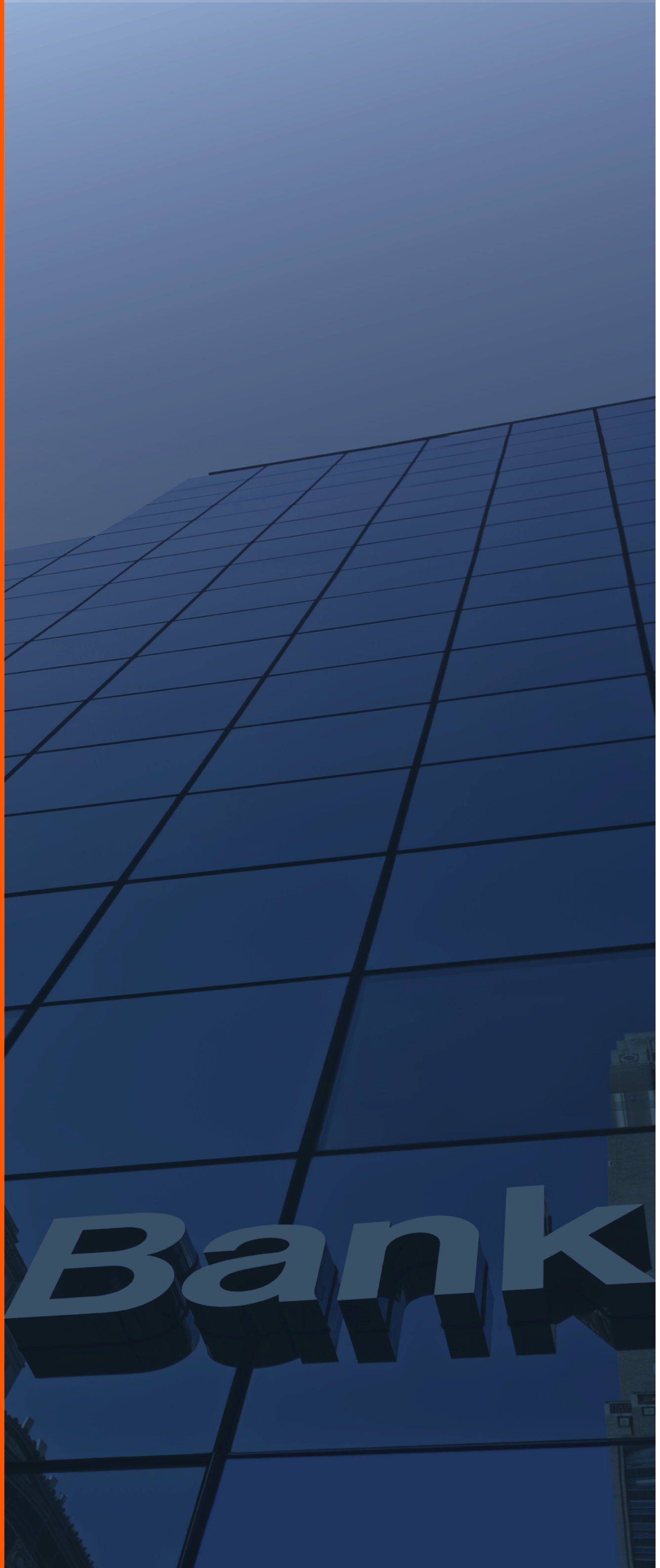


INDUSTRY OVERVIEW

How Banks Are Winning With AI and Machine Learning

DataRobot

Introduction

Making predictions has been a part of the banking industry since its inception. The art of underwriting and pricing loans has always required estimation of the default risk of a borrower and of the loss amount in the event the borrower defaults (taking into account loan structure, collateral, and other risk mitigants such as guarantees). This required deep expertise and long experience on the part of loan officers and credit officers. In community banks, intimate client relationships and long history with clients allowed credit decisions to be largely subjective based on the reputation of the borrower.

More recently, spurred by the Basel frameworks (the original, Basel II, and now Basel III) for measuring capital adequacy, sophisticated analytical methods were developed for quantifying borrower default probability and expected loan loss. These included both instrument level calculations as well as portfolio level aggregation methods which took into account factors such as default correlation.

In many institutions it was this need that drove the risk function to spearhead the first data warehousing projects and to sponsor the acquisition of the first statistical modeling tools. As the de facto centers of excellence for quantitative analysis, these groups began to be called on for help quantifying other types of risk (fraud, money laundering, market risk, operational risk, prepayment risk, and cyber risk to name a few). Eventually many were called on to assist business lines modeling prospecting, client attrition, and everything in between.

Today, you would be hard pressed to identify a line of business or function in a bank that doesn't have multiple needs for predictive analytics. But as business and functional heads have become more aware of the enormous potential of data and analytics, the need for more data, better modeling capabilities, and the capacity to turn data into operational insight has exploded. All banks are realizing that they must find new ways of capturing, organizing, and making data available, and must up their game with new tools and techniques for learning from their data and imbedding data-based capabilities into products, services, client interactions, and operations.

In this white paper, we will survey the landscape of use cases we see banks all over the world pursuing, delve into case studies which describe how organizations large and small are leveraging automated machine learning and, finally, will prescribe some simple rules for spotting high-value use cases in your own organization.

Today, you would be hard pressed to identify a line of business or function in a bank that doesn't have multiple needs for predictive analytics.

Why is Data Science So Hard?

Modelers used to have to content themselves with linear and logistic regression techniques. Analytics, until not too long ago, largely entailed looking for good predictors by trying lots of different regression models and seeing what data was the most predictive, then engineering clever data features that combined multiple data values to try and make predictions better — effectively building an equation to best describe a business problem. Whilst many modern machine learning techniques had been formulated (for instance, neural networks date back to the 1950s), they were not viable for actual business applications — the sheer amount of computing power needed was only available at a prohibitive cost.

Those days are gone. Cheap commodity hardware and the rise of open-source technologies mean that machine learning no longer faces the technological barriers of yesteryear. Modelers can now crank through an enormous amount of data and let the computer do the hard work of finding the best predictors. The machine "learns" how to make predictions based on the data you provide. In the last decade or so, the sheer number of different machine learning models that can be brought to bear to glean insight from your data has exploded. Regression models have waned; now there are neural networks, random forests, support vector machines, and gradient boosted trees just to name a few. But this has given rise to a whole new set of challenges.

First, it is a practical impossibility to know which of the myriad available modelling algorithms and technologies will give you the best result given the data you have and what you are trying to predict. So data scientists have to try a lot of them; not even the most experienced data scientists can know a priori which model will work the best on a new use case. Given practical time and budget constraints and an enormous backlog of demand, data scientists usually have to rely on a few models and technologies they know extremely well. It would be a superhuman data scientist that knew the inner workings of all the different algorithms available and which will work best in any given situation.

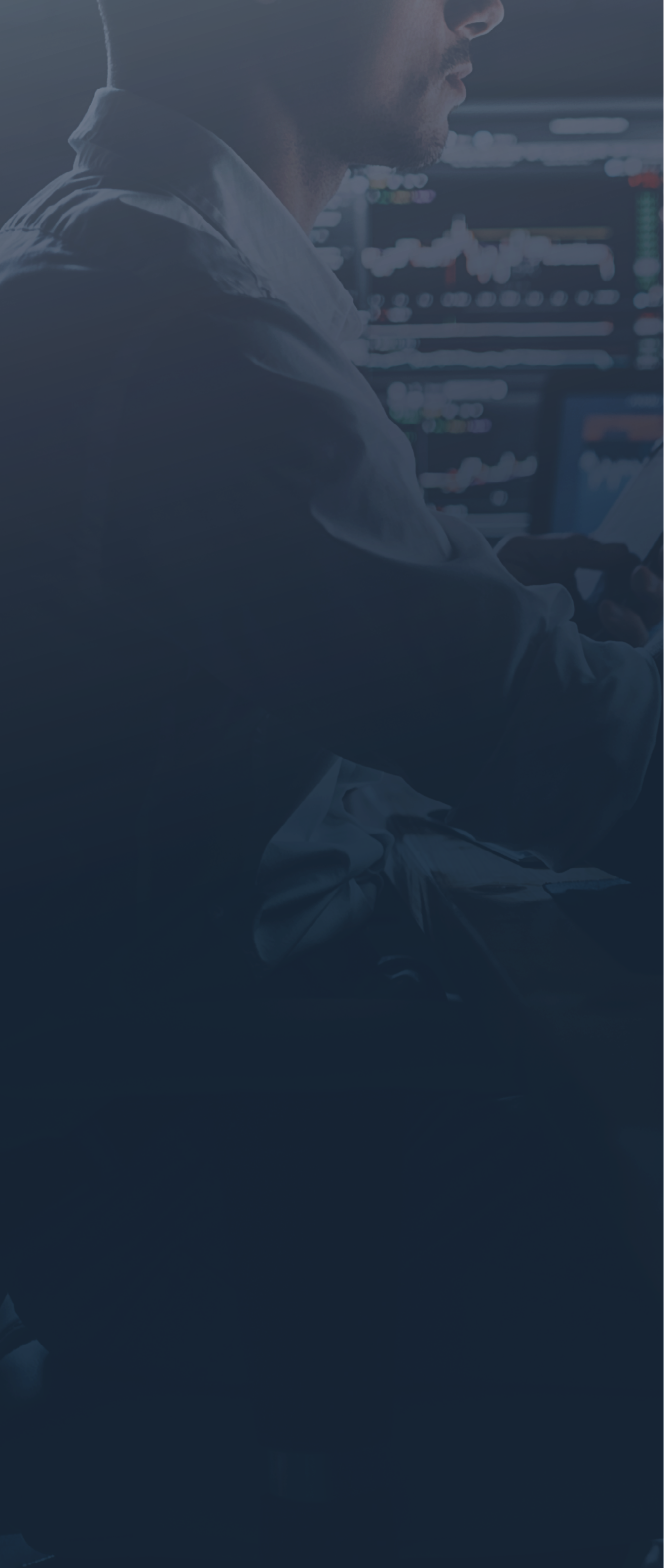
Second, it's not enough to know that a model works well. With skeptical users, never mind regulators, the ability to explain how and why the model works is critical. Which data is important and when (all of the time, some of the time, only rarely)? Can you explain the reasons for a specific prediction? Black box models that work brilliantly but without insight are of little value, even if you could get them past regulators — which you can't.

Third, there are almost always trade-offs that need to be carefully weighed and considered. Perhaps a model is very good at sniffing out positive outcomes, missing very few, but at the cost of a high number of false positives—a classic problem in fraud detection and AML. Or a model is exceedingly accurate but can't perform at the speed needed to support actual business operations. Perhaps a model works very well for one location or customer group, but because of behavioral differences works less well in others.

Fourth, even the best model can't correct for errors, gaps, or biases in your data. Nowadays, it's cheap and easy to store data; that wasn't always the case. This means that you may have a limited data history to work with; is that good enough? Maybe you have data for one geography but are considering expanding into another — will a model trained on the one work for the other? Are all of the data features used to create your model appropriate — can you defend use of all the data for decision making purposes, even if the model finds that age, for example, is a good predictor? What if your history has embedded biases? If human-originated biases are reflected in the training data, this bias will be reflected in the resultant model.

Finally, will your model continue to perform as customer behaviors evolve?

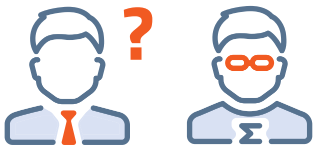
There are also more practical implications to consider. More models mean more overhead. More testing, more tuning, more documentation, more governance. More models to be maintained and monitored in production. Where are more granular models justified, and where not? And, if you develop many models you may quickly discover other bottlenecks upstream in data acquisition, cleansing, preparation, and governance, or downstream in systems testing, application integration, deployment and monitoring.



The Data Science Talent Merry-Go-Round

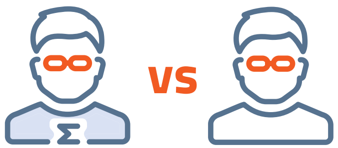
Not only is data science and predictive modelling hard, it's getting harder to find people who are good at it. There is a worldwide shortage of skilled practitioners: the demand is growing exponentially — much faster than the supply. Small and mid-size banks, in particular, are finding it difficult to attract and build a critical mass of talent, sometimes questioning whether it might even be cost prohibitive to do so. But even major global players struggle to find enough of the right calibre of data science talent to deliver on the promise of the AI revolution. As a result, the backlog of data science projects is growing, with executives increasingly frustrated that they are being neglected and that the time to value for these projects is increasing; conversely, data scientists are overworked and demoralized by their ever-growing project backlog.

Three factors exacerbate this.



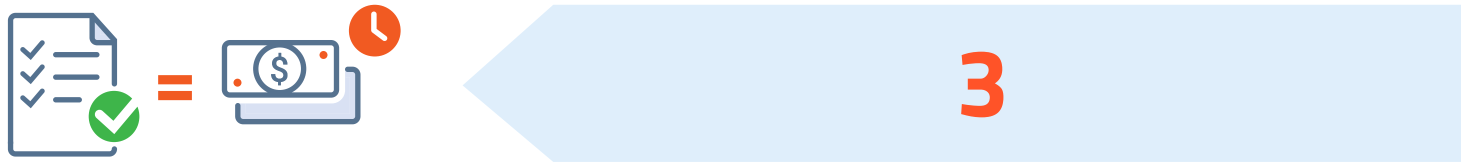
1

First, many bank executives privately think data scientists are challenging to work with. They may find data scientists hard to understand, may be intimidated by the perceived difficulty of the math behind the modeling techniques, and may have trouble articulating business needs in the form of an actionable problem that data scientists can deliver on. It can be tremendously difficult for the layman to predict which data science efforts will ultimately pay off. Almost every executive can tell a tale of a data science project that absorbed enormous time and money and came to nothing.



2

Second, the relationship between the "data people" and the "IT people" is often difficult, making the integration of analytics-driven insights into production applications painful and time consuming. Data people may be more interested in the process of building models than with their implementation and support. They may look upon traditional IT people with disdain, viewing them as too bureaucratic, process-oriented, and slow. IT people, on the other hand, may view the data people as "rogue IT" — building and even deploying solutions without rigor, proper governance, or controls. Both sides of the debate have some merit, but regardless of "who is right", these factors make the already complex and challenging task of bringing predictive analytics projects to the stage where they add tangible business value even more difficult.



Finally, regulators and auditors have set a high bar for documentary and testing requirements for most models, especially those that are high-risk (i.e. the ones that absolutely, cannot be wrong, due to the high cost of failure and/or regulatory consequences) or that become operationally critical. Extensive testing and documentation requirements often take significant capacity away from new initiatives and can be an inefficient use of expensive data scientist time.

As a result of these challenges, many banks find themselves relying on vendor models — and thus on expensive solution providers — to meet their needs. Often, these models are general-purpose, built to meet the needs of many different organizations, with differing scale, business mix, locations, and client types. These models cannot work well everywhere at once. Alternatively, banks may rely on consultants to build models for them. This may result in models more specifically suited to a particular business or location, but at very high implementation and support cost — and without building knowledge and expertise in the business.

CASE STUDY 1:
A LARGE NORTH AMERICAN BANK

OPPORTUNITY: In their fight against financial crime, a large North American bank experienced a high volume of false positive suspicious activity alerts — activity that, once investigated, proved to be harmless. Significant time and resources were required to sift through false alerts and find the few that required closer scrutiny. The inefficiency diverted resources from their primary mission of researching and identifying transactional patterns that could be evidence of money laundering.

SOLUTION: Using DataRobot, the bank trained a model using known outcomes — which suspicious activity alerts ultimately resulted in suspicious activity reports (SARs) being filed and which did not. The model was tuned to maximize negative predictive value, or to avoid false negative results (truly suspicious activity missed). The resulting model reliably separated alerts into populations likely to result in a SAR from those unlikely to result in a SAR.

RESULT: Sixty percent of alerts were correctly categorized as highly unlikely to result in a SAR, and through testing over several months were verified not to contain false negatives. The DataRobot model allowed financial investigators to focus efforts on alerts more likely to result in a SAR and to save hundreds of hours a month trying to manually separate out alerts unlikely to result in a SAR.

CASE STUDY 2:
WELLEN CAPITAL

OPPORTUNITY: Wellen Capital found that default rates were unpredictable and unintentional biases were affecting underwriting quality. Fortunately they had extensive historical data to learn from and deep expertise in their markets.

SOLUTION: Using DataRobot's cloud-based solution, Wellen Capital created sophisticated predictive modeling capabilities quickly. With DataRobot, Wellen Capital leveraged their historical default data and built granular models to predict default in each industry in which they operate, lowering portfolio default rates significantly.

RESULT: The models made credit decisions and risk-based pricing better, and underwriting and credit approval faster. Using DataRobot, Wellen Capital built and deployed sophisticated models — on par with larger, more sophisticated banks and fintechs — quickly and without enormous investment in data infrastructure and talent.

The Advent Of Automated Machine Learning

Automated machine learning (invented by DataRobot) solves many of the challenges described above and makes the others more manageable.

Automated machine learning:

1. Finds the best model for your particular situation through competitive elimination from an extensive resource library of common models—cutting hundreds or even thousands of hours off the time required to find the best model for your situation
2. Ranks the top performing models so you can evaluate and select from among the models best suited to your particular problem
3. Provides transparency into each model's use of data, telling you not just which data is most import, but when
4. Explains individual predictions, down to specific data features and their values
5. Provides diagnostics for understanding model accuracy using a variety of performance metrics
6. Provide tools for understanding and making tradeoff decisions (e.g. between speed and accuracy, positive versus negative predictive value, when and where additional models may be cost justifiable)
7. Automatically create most of the documentation required for model validation and model risk management reducing the opportunity cost of time spent on lower value activities (that data scientists, almost universally, dislike spending time on)
8. Reduce the cost, difficulty, and risk of deploying models into your production environment by providing minimally invasive deployment options such as code generation, API deployment, and deployment to Hadoop
9. Make it easier to monitor model performance and detect drift or performance degradation over time, alerting modelers to the need for retraining or creation of challenger models
10. Makes retraining models on new data and redeploying models into production simple, fast, and low risk

Where Banks are Getting High Value from AI and Machine Learning

There are hundreds of AI and machine learning applications in every function and business line in a bank. With automated machine learning,banks large and small around the world can drive revenue growth, differentiate themselves through superior client experience, reduce operational costs while improving quality, and improve risk management effectiveness and efficiency.

There are hundreds of opportunities to leverage AI and machine learning in every line of business and function in a bank.

	Business line specific	Sales, marketing, relationship and product management	Risk, credit, and pricing	Compliance and Cyber
CONSUMER BANKING				
Retail Banking	Deposit potential/stability	• Churn/attrition risk • Targeted marketing • Targeted offers • Call center optimization • Prospecting • Lead optimization • Profitability prediction	• Credit scoring & approval • Credit pricing • Predicting risk adjusted return (RAR) • Loss forecasting • Portfolio optimization • Collections optimization • Fraud detection and prevention • Targeted risk review	• BSA/AML/KYC • Fair lending • Client complaints • SR 11-7
Mortgage Banking	Prepayment risk			
Consumer Lending	Price elasticity of demand			
Private Banking (HNWI) and Wealth Management	Needs based recommendations/ Relationship deepening			• Cyber threat detection and attack recognition • Information security risk
COMMERCIAL BANKING				
Commercial Banking	• Loan pricing optimization • Collateral valuation • Small business pricing and RAR	• Prospecting • Relationship deepening • Pricing analytics • Deposit volatility	• Credit default and loss severity • Scoring and automated loan approval (small business)	• BSA/AML/KYC • Communications surveillance • SR 11-7
Treasury, Cash and Liquidity Management	Cash flow projections		Predicting relationship profitability	
Special Assets & Workout	• Optimising resolution strategy • Securitization valuation	Early warning of financial stress	Recovery potential estimation	
INVESTMENT BANKING ASSET MANAGAMENT				
Capital Markets/M&A	Deal finding/issuance prediction	• Prospecting • Targeted marketing/sales • Value added services/client behavior • Event prediction • Price elasticity (e.g. RFQs)	• Model risk management • Loss forecasting • Margin/credit approval • Scenario/stress testing • Pricing and valuation	• BSA/AML/KYC • Communications and trade/deal surveillance • SR 11-7
Sales and Trading **Investment Research**	• Margin management • Settlement optimization • Quantitative strategies & research • Trade recommendation • Sentiment modeling • Trade execution strategy • Smart order routing			
Asset Management			• Model risk management Scenario/ stress testing • Ex-post attribution analysis behavior	

CASE STUDY 3:
FREDDIE MAC

OPPORTUNITY: Freddie Mac found there were more opportunities for AI and machine learning than they could pursue. Their data science team was stretched beyond capacity and they were not able to meet the demand.

SOLUTION: DataRobot's automated machine learning platform made it faster and easier for Freddie Mac's experienced data scientists to build models, allowing them to deploy more AI solutions than they could before. DataRobot's ability to find the best model through competitive elimination helped Freddie Mac's data scientists create solutions in a fraction of the time it had previously taken. Freddie Mac described this as putting "AI on their AI". With DataRobot they were also able to leverage more diverse sources of training data, including unstructured text data.

RESULT: The productivity of Freddie Mac's data scientists with DataRobot was increased dramatically helping them satisfy more demand across their business without a proportional increase in cost.

Deployment Models

With automated machine learning you get the productivity of a much larger team of data scientists. Automated machine learning has the potential to increase the throughput of your data science team not by fractional increments but by an order of magnitude.

There are a few different approaches:

- Use of automated machine learning to quickly eliminate less promising modeling options so that data scientists can focus their efforts on the models most likely to yield the best results — making the data science process more iterative, agile and responsive
- Use of automated machine learning to develop lower risk models (e.g. for client offers, prospecting, back office productivity) so that data scientists can focus their efforts on the most mission-critical needs
- Leveraging automated machine learning to build many models with the same capacity previously required for a single model — improving model granularity where it matters most, e.g. for different products, regions, or client segments where behavioral differences matter

In addition, some banks deploy automated machine learning capabilities outside of formal data science teams. Often the deepest expertise in the data, and the best understanding of the business opportunity, sits in the business line (or function). With automated machine learning, banks can tap into that deep well of data and business expertise to augment their formal data science capacity. This is referred to as the "democratization" of data science.

CASE STUDY 4:
LENDING TREE

OPPORTUNITY: Lending Tree found that as their network of providers grew, providing more choices to borrowers was not good enough and they needed to provide more personalized recommendations, matching the most relevant providers to borrower needs more closely. Lending Tree was fortunate to have expertise in their data and their business to draw on. These resources, however, did not have formal data science training.

SOLUTION: With DataRobot, Lending Tree "democratized data science", training their data and business experts in the use of the automated machine learning platform for model development, evaluation, comparison and testing. This helped the business build and deploy sophisticated models, and over time increased business awareness of AI and machine learning potential. Lending Tree found that more high value use cases were being identified for AI and machine learning by business analysts after the deployment of DataRobot.

RESULT: DataRobot helped Lending Tree become an AI-enabled enterprise, by empowering their valuable business analysts with sophisticated model development and deployment capabilities.

Maturity and Adoption

Some banks are rushing to adopt automated machine learning, while others are taking a more cautious approach.

We think that automated machine learning represents a once in a generation opportunity to differentiate your organization and achieve break away performance relative to your competition.

There are a few different approaches:

- Banks that do not adopt AI technologies like automated machine learning will lose ground to their competition and ultimately may not survive.
- Banks that overlay automated machine learning on their existing business models will not only survive but will outperform their competition.
- Banks that redesign their products, services, and business models to take full advantage of the power of automated machine learning will crush their competition and dominate their industry segments and markets.

There are a number of factors which seem to influence the openness of a bank to automated machine learning. Some of these are obvious, such as an organization's willingness and ability to change, an organization's current level of maturity in data and analytics, and how open minded the organization's executives are to potentially game-changing advances in technology.

These factors, however, don't guarantee success. For example, some organizations that are well advanced in data and analytics are actually the most resistant to change — the functions now entrenched and defensive of changes to the status quo. In some cases, it is the less advanced practitioners who see automated machine learning as an opportunity to leapfrog their competition and are moving rapidly to adopt the new capabilities. Some smaller banks (correctly) see automated machine learning as a way to level the playing field with their larger, more sophisticated competitors. But we have seen global banks too that view automated machine learning as the much needed game changer they've been looking for. In the latter cases especially, active sponsorship from the highest level of the organization is the determining success factor.

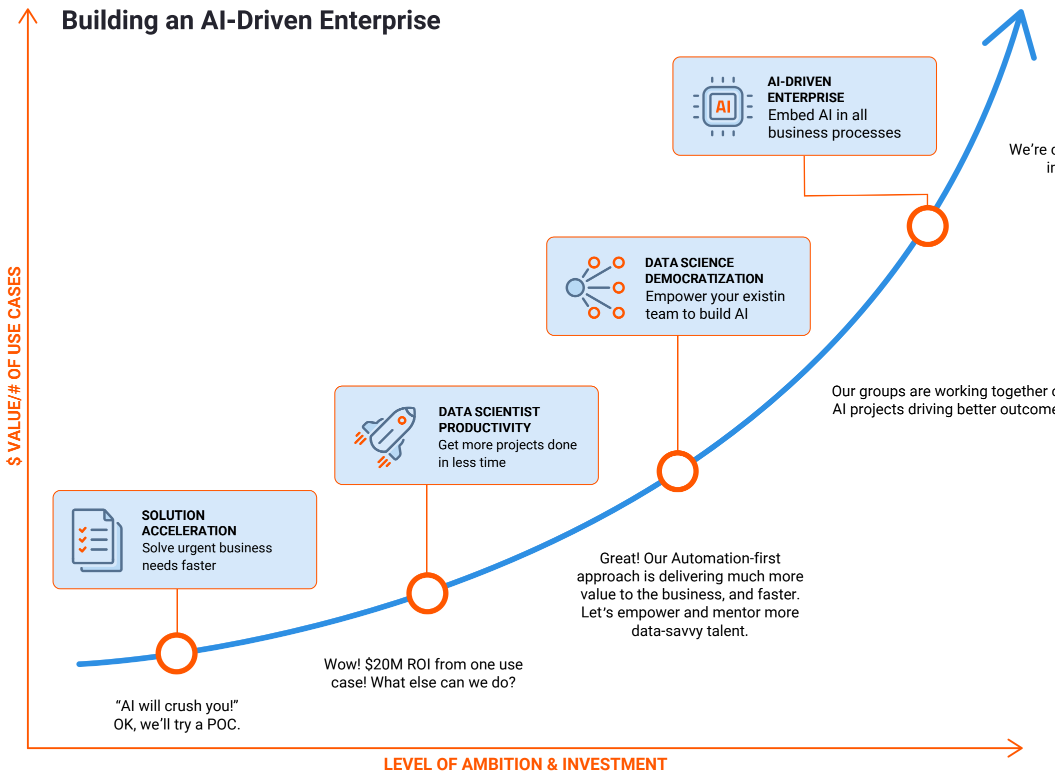
CASE STUDY 5:
BULGE BRACKET INVESTMENT BANK

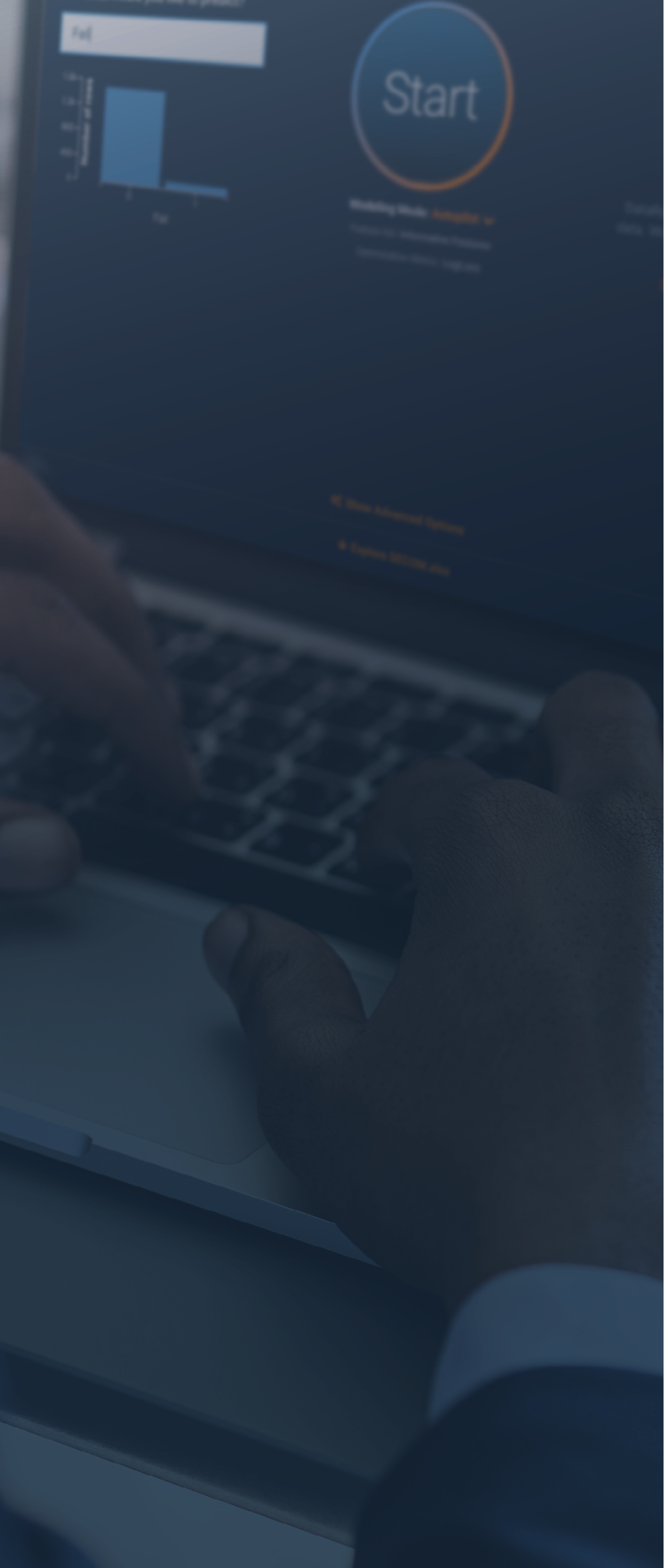
OPPORTUNITY: This large global investment bank asked DataRobot to help bond traders quote smart prices in response to incoming requests for quotations (RFQs). Previously traders quoted prices using a combination of regression models, trader experience, experience with the customer, and desk inventory. Smaller trades were priced automatically using a rules-based system. The bank was concerned that revenue was being "left on the table" both from lost trades and from won trades where more profitable prices could have been proposed without losing the business.

SOLUTION: The bank assembled a dataset containing the desk's RFQ history of prices quoted and the outcomes (whether the business was won or not) as well as metadata on the counterparties and products. From this training data, the bank used DataRobot to build a model to predict the probability of winning an RFQ for a given price. Now, when an RFQ is received, traders see a price elasticity curve showing the probability of winning the RFQ at different prices quoted, which they can factor into their decision on which price to quote. For smaller RFQs, DataRobot worked with the bank to determine optimal, profit-maximising probability levels to quote — fully automating the process.

RESULT: The DataRobot models outperformed the existing regression models significantly, and initial tests show this will improve the bank's win rate and improve margins.

At DataRobot see four discrete stages of maturity around automated machine learning: solution acceleration, data scientist productivity, data science democratization, and the AI driven enterprise. Initially automated machine learning can help banks develop and implement AI solutions faster. As the use of automated machine learning grows, exponentially increasing value can be delivered across a broader set of demands with the same size team. In the third stage, we see banks deploying automated machine learning in functions and business lines to tap into the deep data and business expertise embedded in these areas. Finally, the most mature banks are becoming AI driven — developing strategy for products, channels, and client experience around AI.





How to Get Started

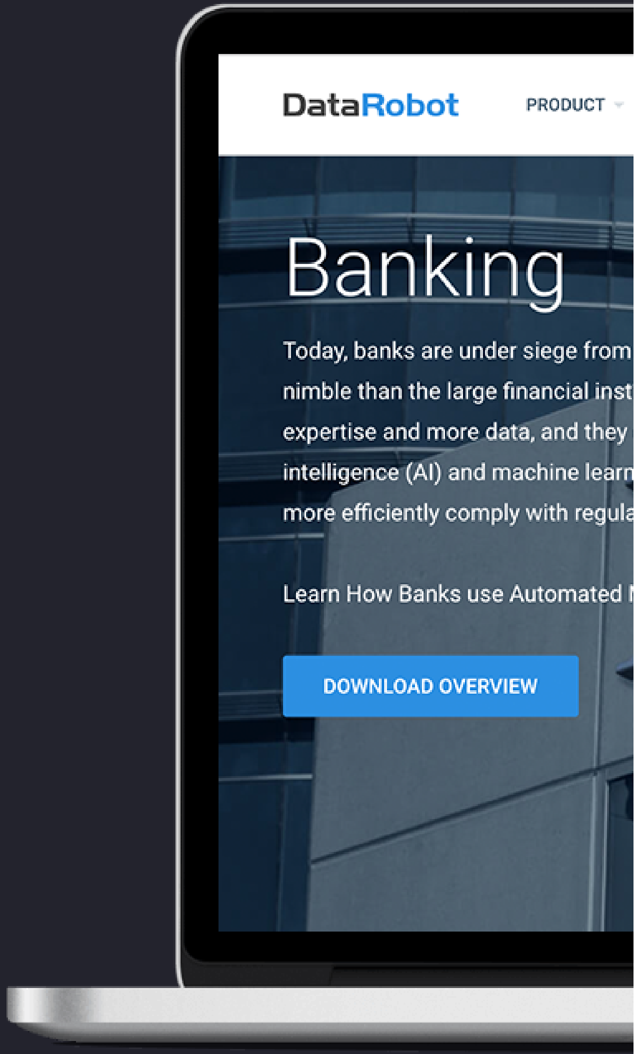
First, educate your leaders. Automated machine learning is simply too significant, too important an opportunity, too strategically imperative to leave to the CIO, CTO, Chief Data Officer or Chief Analytics Officer alone. This must be on the executive agenda and needs to be discussed at the highest levels of your organization.

DataRobot offers an "AI for Executives" class where we explain the concepts and terminology without intimidating technical jargon, talk through practical uses of AI and machine learning relevant to your business mix, discuss the characteristics of a good use case and review a few simple rules to follow to identify high potential use cases for your organization, and then challenge executives to identify a few use cases where better predictions would be of business value.

WHICH USE CASES ARE THE MOST PROMISING?

The best application of AI and machine learning will be different in every organization.
Generally, the best use cases exhibit the following characteristics:

- A definite outcome or event to be predicted can be identified. If the outcome to be predicted is unclear, more thought is needed.
- There is clear and quantifiable value to making better predictions — whether it be probability of borrower default, identifying the highest value prospects, or differentiating financial crime from innocuous transactional activity. The best use cases have indisputable business value that can be quantified and measured after the fact.
- The data needed to train the model is readily available and sufficiently high quality. A project that requires massive up-front investment in data acquisition, cleansing, or preparation is not the place to start.
- The path to operationalization is clear. In some cases the way to insert a predictive model into a business process or operating model is clear and low risk — in other cases major process re-engineering may be needed. Better to start with use cases where significant change to the business is not a prerequisite.
- A willing and able business sponsor has self-identified. In almost every case communication, training, and change management are critical to successful adoption and value realization. A project with a willing business sponsor will always be more successful than one where a new capability is thrust upon an unenthusiastic leader.



Conclusion

The age of automated machine learning is here. Banks that embrace it, take full advantage of it, and think through how to introduce it into every part of their organization will crush their competition. Banks that dither, delay, or avoid learning about this technology will lose ground and ultimately may not survive.

DataRobot is the inventor and category leader of automated machine learning. Having been the first mover, DataRobot has developed the most robust end to end solution. Let our team of banking industry experts show you how you can achieve your goals and beat your competition with automated machine learning. Leverage our experience working with banks across the globe on applications of machine learning in virtually every business line and function.

Regardless of the size of your bank, or your relative sophistication in data and analytics, we can help you leverage your data assets for competitive advantage.

For more information on DataRobot, or to schedule a demo, visit: **www.datarobot.com**

Contact Us

DataRobot
One International Place,
5o Andar Boston,
MA 02110
www.datarobot.com
info@datarobot.com

powered by aws

DataRobot