dataiku

AI In Banking
Challenges, Solutions,
& Steps To Get Started Now

Introduction

Data has always been the foundation of the banking industry. What has changed in recent years, of course, is the amount of data available and the speed at which it is processed as well as the need to quickly respond to market changes. New technology gives banks the power to collect, store, and analyze exponentially more information than was imaginable not too long ago. In the wake of Fintech, banks already know that to succeed in today's ecosystem, they must use this wealth data at a massive scale to continuously innovate.

Global business information provider IHS Markit predicts[1] that the business value of artificial intelligence (AI) in banking will reach $300 billion by 2030.

Yet most struggle amidst complexities of the data itself, regulations, and more, to get AI initiatives off the ground. But they don't have to. Today's banks seeing success with AI initiatives:

1. Bring the idea of "doing AI" down from a pedestal and instead break it down into what it really means for them (which isn't always a sexy app or sleek chatbot).

2. Realize that Enterprise AI is a journey - a series of steps and gradual competencies to work up to over the next several years. One of those steps is the gradual improvement of messy internal processes so that entire teams or divisions take steps toward using data and AI to work smartly, efficiently, and within regulatory standards.

3. Get started now, because waiting a few more years to dive in will mean pushing the timeline of the Enterprise AI journey even further, while competition from other more agile companies (whether fintech, GAFA - Google, Apple, Facebook, Amazon - or traditional players) moves in.

[1] https://technology.ihs.com/612413/artificial-intelligence-in-banking-report-2019

4. Build on foundations. Many banks today are intimidated by the idea of AI even though they've already been doing it - or at least some of it - for years (and in some cases, decades). Quants, algorithmic traders, risk analysts, fraud analysts, pricing teams, the list goes on - these people and teams already form the building blocks of an Enterprise AI strategy. Successful banks build upon this already-existing framework.

"As AI spreads through the enterprise, solutions will emerge that empower developers with the specialized skill sets and tooling required to take on more AI workload and solve key challenges. IT leaders must follow the trend of AI democratization to drive their teams to create AI solutions."

Gartner Predicts 2019: The Democratization of AI[2], Magnus Revang et al., 29 November 2018 (report available to Gartner subscribers).

Ultimately, Enterprise AI for banks means turning data from the cost center it is today in to a revenue stream - a source of efficiency and a wealth of information that can be used to provide fundamental value to the business. And with the right approach, all of this can be achieved by leveraging the hard work that's already been done.

[2] https://www.gartner.com/doc/3894386

About the
White Paper

THIS WHITE PAPER WILL FOCUS ON:

- Defining what it actually means to leverage AI in banking.

- Uncovering the top challenges to AI adoption in the sector (and how to address them).

- Taking a look at the top initiatives banks can begin now in order to get started in AI

- Organizational structure(s) for enabling AI.

- Use cases for AI in banking (and how to choose where to start).

- Success stories from banks who are already leveraging AI

About the Authors

Jason Moolenaar | Jason has more than 20 years of experience in financial services and technology. After beginning his career as an equity sales trader with Deutsche Bank focused on hedge funds, Jason continued on as a Junior Portfolio Manager for a boutique asset management firm in Sydney, Australia. He subsequently spent the majority of a nearly nine-year career at Bloomberg in the electronic trading group focused on developing and selling technologies for connectivity, execution, complex event processing, pre/post-trade execution analysis, and automated trading. Jason also led financial services sales at various early-stage startups focused on applying AI and machine learning to a wide array of use cases for both buy- and sell-side institutions. He currently leads financial services sales across the Eastern United States and Canada at Dataiku.

Hursh Rughani | Hursh holds both undergraduate and postgraduate degrees in information management and has more than nine years of experience working in the data analytics space, including guest lecturing on analytics in the enterprise at the University of Sheffield. Starting his career off in consulting focusing on financial services, Hursh spent most of his time working with investment management companies. At M&G Investments (Prudential), he worked closely with the company's Analytics Centre of Excellence as well as projects within risk, finance (AUM and FUM analysis), and operations. At Tableau, Hursh managed 15 strategic enterprise clients within the financial services space. Today, he leads financial services sales in the United Kingdom and Ireland at Dataiku in order to bring Enterprise AI strategy to enterprise banking customers and prospects.

Pierre Ménard | Pierre has 14 years of experience in technology, mainly dealing with financial services customers. He started his career at a French IT consulting firm in its dedicated Banking & Insurance unit. Today, Pierre leads the financial services team in France at Dataiku. He has been at Dataiku for more than five years, which means he worked with its very first customers and has been able to see the expansion and deployment of more advanced use cases over time.

Top AI Challenges
(& How to Address Them)

Most banks today are at the beginning of their AI journey, and as previously mentioned, they fundamentally have a lot of the pieces in place. Yet moving forward is intimidating, often because businesses are paralyzed by the perceived complexity of AI or intimidated by the (false) notion that they must hire entire new teams to get there.

Here are the top five challenges banks face when it comes to implementing Enterprise AI as well as suggestions for addressing, or at least smartly approaching, each challenge.

1. DATA ESSENTIALS (ACCESS, WRANGLING, ETC.)

One of the biggest roadblocks to Enterprise AI in banking is not a question of putting machine learning models into production or even of creating the models themselves. However, simple data management is essential to enabling the organization to leverage data from the bottom up, democratizing data use across teams and roles.

The good news is that it's not a unique issue - a cross-industry survey revealed that no matter what the business, baseline things like cleaning and wrangling data as well as connecting to data sources always ranked as the top challenge for participants. And for the record, that's good news because it's a problem that many have already solved before, which means banks don't need to reinvent the wheel to find quick solutions.

And in general, the answer lies in tooling - having a central (yet controlled) place where data can be accessed and used, all in a user-friendly interface that isn't just meant for data scientists, quants, or other technical profiles. Critically, the system for accessing and cleaning data should not rely on underlying architecture. Meaning no matter how many different places data is currently stored - or will be stored in the future - staff won't have to constantly change their day-to-day processes or tools to adapt.

Ideally, data access and wrangling isn't just happening in a one-off ETL (extract, transform, load) tool, either; it must be incorporated into downstream systems so that, when relevant and appropriate, technical teams can take over the work of analysts and easily apply predictive or machine learning techniques to bring more value.

2. THE REGULATORY ENVIRONMENT

The elephant in the room when it comes to AI in banking is, of course, the regulatory environment. The European Union's General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), New York Department of Financial Services Cybersecurity Regulation, third-party risk management (TPRM) expectations, the SEC disclosure guidance - the trends in banking compliance regulations when it comes to data are complex, and only becoming more so each year..

Bank Resilience and Stress Testing	Shadow Banking and Funds	Consumer and Investor Protection	Systemic Risk	Market Integrity
ECB Stress Testing CCAR AQR FDSF	UCITS IV, V & VI AIFMD FSB Review Shadow Banking	RDR DFA Title X MIFID II/MIFIR UK CASS | MCD MMR | PRIPs	FSB Guidance on RRP | BRRD UK RRP Guidance Living Wills - DFA.II SRM	FATCA Data Protection MLD4 Benchmarks MAD/MAR

Banking Industry Regulatory Landscape

Conduct and Culture	Regulatory Reform	Capital and Liquidity	Derivatives	Trading
Compensation DFA: Say or Pay	Liikanen FTT | SSM CRAs | CDSs Banking Reform Act	Basel III CRD 4 (CRR & CRD) DFA Collins Amendment COREP/FINREP IFRS 9 BCBS Review of Securitisation	EMIR CCPs Dodd Frank Act: Title VII DFA: Swaps Push Out	HFT DFA Volcker Rule Short Selling JOBS Liikanen

"Regardless of what definitive changes lawmakers and regulators might make, banking organizations should continue to drive effectiveness and efficiencies across their risk and compliance programs so they can meet applicable laws, regulations, and supervisory expectations."

Deloitte, 2019 Banking Regulatory Outlook[3]

And that only covers regulations touching the issues of cybersecurity and privacy. Once combined with additional regulations surrounding finance crimes protection, the Financial Accounting Standards Board (FASB) standards, and much more, the outlook becomes complex quickly.

Indeed, more so than any other industry, banks must put concerns of transparency and reproducibility at the forefront. In fact, this challenge alone carries such heavy consequences that banks often (incorrectly) assume that they won't be able to implement AI processes at all.

And to be sure, black-box machine learning (that is, models that simply spit out a prediction or a result with no visibility at all into how the decision gets made) is not within the realm of possibility for banks. Models that cannot be explained simply do not have the level of reproducibility of results that regulatory standards demand.

It's important to realize that what makes machine learning models accurate is often also what makes their predictions difficult to understand: they are very complex. But it's not impossible to build very good white-box models. So in order to comply with regulatory requirements, banks will nearly always need to build models that are inherently interpretable.

[3] https://www2.deloitte.com/us/en/pages/regulatory/articles/banking-regulatory-outlook.html

But the silver lining there is that it is a tradeoff, and it's not impossible to build very good white-box models. So in order to comply with regulatory requirements, banks will nearly always need to build models that are inherently interpretable (which is subtly different than building black-box models and then retroactively attempting to explain them).
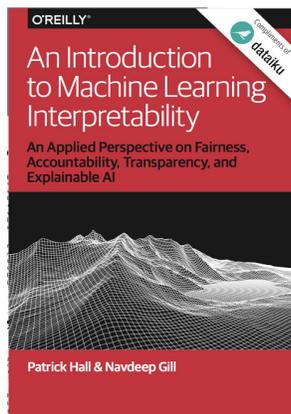
Aside from the building of models themselves to be interpretable, there is the larger question of governance itself. In today's landscape with data privacy laws cropping up worldwide, it's not uncommon to see data or analytics teams in banks become paralyzed by uncertainty in how to navigate this new world.

Under data privacy regulations, clearly, working directly with personal data is extremely limited, and working with anonymized data - while an interesting option if effective - is incredibly difficult (not to mention resource-intensive) to actually do correctly. So what other options are there for banks to work with data in an increasingly regulated world (and one that will only continue to be more regulated, not less)?

Pseudonymization is the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information. While this clearly means that pseudonymized data is still personal data (since it is not anonymized), it does provide some additional freedom for banks to work with data provided that they have specific, defined projects with controlled access and a clear data retention policy.

With some expertise and education of people, careful governance processes, and the right tools (for example, choosing an AI platform that allows for complete model interpretability), the regulatory challenge is not impossible to surmount.

FURTHER READING

data iku

EXECUTING DATA
PRIVACY-COMPLIANT
DATA PROJECTS

A Guide for Data Teams

A GUIDEBOOK BY DATAIKU
www.dataiku.com

O'REILLY®

Compliments of dataiku

An Introduction
to Machine Learning
Interpretability

An Applied Perspective on Fairness,
Accountability, Transparency, and
Explainable AI

Patrick Hall & Navdeep Gill

3. MODEL RISK MANAGEMENT (MRM) AND VALIDATION

In a topic intrinsically intertwined with regulatory requirements, the struggle between velocity and proper processes for model risk validation can be debilitating for the progress of AI initiatives. Of course, the challenge is a complex one because mitigating risk is the top priority, which means that it cannot be sped up in a way that compromises the quality of the validation itself.

But there are still improvements that can be made and ways to address this challenge; namely by introducing consistency and reproducibility into the process both of validations and the final pushes to production.

For example, it is often the case that the model risk validation team(s) look at models from different organizations or groups across the company, each of whom have their own individual processes and send the models in different formats, containing different information, etc. That means for each review, the model risk validation team loses time in trying to get their bearings and figure out what it is that they're looking at. Similarly, without a consistent system or process by which models are delivered, the next step (deployment to production) also becomes complicated and time consuming.

Instead, banks which are able to quickly move through the model risk validation stage have everyone across the company working in the same tool so that once models need to be validated, the team evaluating their risk knows what they're dealing with, how to find the data sources the model is built on, it's clear what data transformations were done, etc. From there, the tool where the models were delivered and validation completed is also ideally the same tool where deployment to production happens.

Getting everyone on the same page and in the same tool ensures a faster process start-to-finish that can ensure models make it into production in a matter of weeks (not months or years as can be the case with many banks today).

FURTHER READING

WHY ENTERPRISES NEED DATA SCIENCE, MACHINE LEARNING, AND AI PLATFORMS

A WHITE PAPER BY DATAIKU
www.dataiku.com

4. HIRING

Ask any major bank today what their biggest challenges are when it comes to progressing in the AI space, and hiring will be top of the list. And for good reason - it's certainly a critically important element to an AI strategy. Yet there are also lots of misconceptions and missteps made when it comes to hiring data and analytics talent.

Taking a quick step back, it should come as no surprise that actually, hiring for data and analytics talent is an issue for most companies. Usually for smaller companies, it's because they don't have the budget to attract top talent, but for bigger companies, it's more often that they don't have hiring committees that know what they're looking for when it comes to data scientists or what the goals and needs of the organization are with regards to data strategy.

In fact, more often than not, companies go about hiring a data scientist before they even have a project or goals in mind. This doesn't make sense for several reasons, not least of which because there isn't just one kind of data scientist, and so hiring the right one with the right skills depends heavily on use case.

But for banks, hiring is even more of a challenge because on top of facing some of the aforementioned issues, it's also difficult to find people who are both cutting-edge when it comes to AI technologies as well as who have a deep understanding of the industry and its regulatory restrictions and requirements. How will one bank find an entire team of data scientists with these skills, much less ensure that they can retain them (especially given that data scientist turnover is notoriously high)?

The answer to this challenge lies, once again, in data democratization. It doesn't make sense (financially or in terms of risk) to search for hundreds of unicorn data scientists and building teams from scratch. Instead, today's successful banks leverage the talent of the hundreds of thousands of staff who already have the business knowledge. As previously mentioned, many banks overlook the fact that they already have lots of data talent throughout the organization and that the challenge is less one of hiring and more of training and education.

One final unique challenge that banks face when it comes to hiring is the immense uptick in hiring of compliance staff that has become commonplace given the need to meet strict regulatory standards. This is something that many banks have discovered simply cannot continue - it's not scalable, and there will be a certain tipping point where without the right tools enabling efficiency, more compliance staff will not mean more compliance. Again in this case, the challenge of hiring is less about doing better or more hiring and instead about tooling, enablement, and introducing efficiency with existing staff.

5. ORGANIZATIONAL CHANGE

Obviously, transforming into a data-driven bank and one that leads the way when it comes to AI initiatives isn't easy, so one of the challenges that can't be ignored is an underlying resistance to change, especially addressing the fear of automation and employee hostility toward changing roles. The negative attention that AI garners in the press today is palpable, and organizations that think they can solve resistance to change by telling staff to ignore it, or worse, by turning a blind eye toward their concerns, are naive.

The fact is that there is a very real fear among the public and likely among staff that jobs will be automated away and that they will be fired. But that doesn't mean that banks need to shy away from introducing new tools and automation, which as detailed already, are critical to protecting the organization against larger risk, not the last of which is human error.

Instead, it means accepting and facing this challenge through education. Not only education about why automation is important, but how humans fit into the process and what their role will be going forward.

Data Scientist vs.Data Analyst -
What's the Difference?

For banks looking to hire data talent, the question of data scientist vs. analyst is a common one. Yet as the industry progresses, the question is actually becoming increasingly more complicated. Here's why.

Education

Being a data scientist (or not) vs. data analyst isn't a simple question of fulfilling a specific degree or area of study. For example, many who started out in fields like statistics, quantitative analysts, or actuarial science are no doubt doing work that overlaps with today's data scientists, and some have even taken the full leap to data science.

Similarly, data scientists can also have backgrounds in areas of study like computer science, and many (about half, according to one publication[4]) have PhDs. Increasingly today, one must also consider the role of the thousands of online data science or data education courses and whether that qualifies someone as an analyst or data scientist in and of itself.

Analysts may or may not be formally educated as such and can have diverse backgrounds and areas of study, including risk, general business, and finance, but potentially also liberal arts. And, of course, there is the pervasive rise of the so-called "citizen data scientist" that seems to saddle the two worlds of analyst and scientist, further complicating the delineation.

Does one specific area of study or the learning of one skill push someone from analyst to data scientist? Or vice-versa for lack of knowledge in any particular area?

Well, no - especially considering the fact that certain experience or skills (for either a data scientist or analyst) might matter for one team or business line within the bank, but not another. For example, hiring someone to create dashboards and a churn prevention campaign for the marketing division is a totally different skillset than hiring someone who will be building an anomaly detection system to improve cybersecurity. However, both might be data scientists.

Skills

One of the most common stake-in-the-ground, blanket characteristics that many try to use to distinguish between data scientists and analysts is that data scientists are technical, and analysts are not. And while it's true that generally data scientists probably are more proficient in coding environments overall, there are plenty of analysts out there who also are technical and are comfortable with at least one coding language.

It's also worth saying that even data scientists who have all the technical skills in the world have little value in companies if they don't also have communication skills and/or business acumen - that is, the ability to connect with those who know the business best in order to actually use those technical skills to provide real value.

With the rise of automated machine learning, analysts - or other profiles that lack deep knowledge of machine learning and/ or statistics - will only expand their utility further down the data

---

[4] https://www.kdnuggets.com/2018/05/simplilearn-9-must-have-skills-data-scientist.html

pipeline. But, of course, these developments won't render data scientists obsolete, either. Their skills around machine learning interpretability and model maintenance are generally areas that analysts do not possess.

When talking about skills for either profile, data cleaning and exploring simply must be mentioned, as they are probably the two skills (along with the ability to question data and tie data projects to business need) that are critical for data scientists and analysts. Given that the bulk of any data project will be getting the data into the right shape to apply machine learning (or whatever the ultimate goal might be), it's a cornerstone of both the positions.

Like education, it's clear that the skills required to be a data scientist vs. analyst are not black-and-white. There is plenty of overlap between the two - probably more so than there are specificities to one role or the other. So data scientist vs. analyst - where does that leave us?

The Bottom Line: Who Should I Hire?

Ultimately, the range (and overlap) of skills between the role of data scientist and analyst means that the two are more like sliding scales than two separate buckets.

That means banks looking to hire for one role vs. the other should take a critical look at their current staff, their needs, and what the end goal is for the data team or data projects. The best solution would be not to hire at all, but rather to move someone who already has the skills into the project (quants and actuaries, for example, can be great options).

Proper tooling can also uplevel staff, making them more efficient and productive whether they are a data scientists or an analyst. For example, a data science, machine learning, or AI platform can aid business people to work with analysts, analysts to work with data scientists, and to bring it full circle, data scientists with IT or data engineers.

Use Cases: Data Science, Machine Learning, and AI in Banking

Overall, use cases for data science, machine learning, and AI fall into one of five categories:

INCREASED REVENUE (MARKETING)	AVOIDING RISK OF NON-COMPLIANCE (RISK)	DECREASED COSTS (OPERATIONS)	COMPETITIVE EDGE (INNOVATION)

SPEED-TO-VALUE & TEAM EFFICIENCY (ORGANIZATION)

These categories are broad, which means there is no shortage of use cases when it comes to data science, machine learning, and AI in the banking and financial services sector. The larger challenge is often not finding use cases, but choosing the ones that will prove to be most successful as a data-driven initiative based on potential impact see Best Practices: Choosing a Use Case (Where to Begin?).

However, this section takes a comprehe nsive (though non-exhaustive) look at both top use cases and up-and-coming use cases today.

INCREASED REVENUE

Revenue Attribution: In a desire to get closer to their customers, consumer banking institutions have been instituting the concept of Book of Business, which entails IT teams creating rules to allocate every customer to one or more branches in their footprint. In theory, this allows branch managers and staff to focus on delivering services to these clients. However, the set

of rules created to allocate customers can significantly impact the compensation of branch personnel, making this process incredibly contentious and disruptive. AI can specifically be used in this scenario to create clear, interpretable, machine learning-based Book of Business models with which bank staff are comfortable.

Customer Churn: Whether at a larger scale in retail banking or a more personalized approach in wealth management, effectively identifying customers at risk of leaving and then automating a system to take action on those at-risk customers is the perfect space for AI. To enhance churn detection systems, data science approaches had first allowed to exploit AI more comprehensively the structured systems can pull data from the different silos and signals from a wide variety of sources to come up with accurate predictions, a higher capacity to anticipate potential churners and to provide field teams with new contextualized information to help them take appropriate actions. More recently, Computer

Vision, Natural Language Processing techniques and speech to text technologies empowered new AI initiatives to identify and route automatically and immediately dissatisfactions detected in mails, e-mails and calls, to address them faster and better retain customers, which can then automatically send a message reaching out to the customer or notify the wealth manager to take action.

Recommending and Upselling: Providing the right message to a client (whether in retail banking or wholesale service) through the appropriate channel at the right time has always can make the difference in upselling a service that a client needs. For financial services (and retail banking in particular) in which marketing actions are culturally product-oriented, AI is at the heart of the transformation toward a customer-centric business organization. Using AI to look at large swaths of diverse data to determine what products or services should be recommended to someone at a certain moment based on the last interactions can be hugely advantageous to banks that take advantage of this form of hyper-personalization. This is a very large field that encompass projects like scenarized multi-touch point campaigns, real-time recommendation engines, and life detection events.

Insurance: In addition to fraud detection and customer retention/churn (which have already been addressed in this section), banks that offer insurance services can also benefit from machine learning-based systems to improve optimal pricing and conversion, claims triage, and claims forecasting. AI for pricing optimization allows insurance lines-of-business to dynamically monitor the marketplace and adjust their prices based on patterns that are being detected in the marketplace. For claims triage, not being able to accurately isolate claims that warrant fast settlement or on the other hand warrant deeper investigation can cost millions and mean that some claims are significantly overpaid - developing a machine learning-based system that is more sophisticated and nuanced than a rules-based system can make this kind of granularity a reality. Finally, AI can also help insurers accurately - and in many cases automatically - predict the number and the size of claims, giving insurance executives more accurate loss predictions.

REDUCED RISK

Cyber Security: According to Forbes[5], the typical American financial services firm is attacked a staggering 1 billion times per year - that's more than 30 attacks per second. Keeping up with the rate and sophistication of attacks requires the most cutting-edge systems, and those are machine learning- and AI-based. Developing a machine learning-based anomaly detection system allows for the use of wide and varied data sources that are essential to finding needle-in-the-haystack anomalies. The very nature detecting intrusions into systems or malware means fraudsters are specifically and deliberately trying to produce inputs that don't look like outliers. Adapting to and learning from this reality is critical, and it's something that can only be achieved with AI.

Fraud Detection and Prevention: Whether being used to detect ATM fraud, bad check writing, or insider threat, fraud detection is all about finding patterns of interest (outliers, exceptions, peculiarities, etc.) that deviate from expected behavior within dataset(s). Using multiple types and sources of data is what allows banks to move beyond point anomalies into identifying more sophisticated contextual or collective anomalies. In other words, variety is key. And when it comes to managing massive amounts of data in near real-time from many sources, machine learning and AI are essential - it's the only way to build ever more sophisticated fraud detection systems as fraudsters' methods evolve.

AML, KYC, and Wire Transfer Fraud: The value that AI can bring to anti-money laundering (AML), know your customer (KYC) and wire transfer fraud efforts is not only greater accuracy in detecting issues overall, but also - importantly - reducing the number of false positives and reducing overall the number of people doing manual work (particularly in investigating false positives). Because machine learning-based fraud detection systems are more agile and can learn patterns over time, they can continue detecting anomalies even as user behaviors change, which makes them more powerful and prevents costly (not only in terms of staff time, but also in terms of customer frustration and loyalty) manual review of false positives. This is especially important for US banks in the wake of Customer Due Diligence (CDD) Rule.

---

[5] https://www.forbes.com/sites/bhaktimirchandani/2018/08/28/laughing-all-the-way-to-the-bank-cybercriminals-targeting-us-financial-institutions/#57fe1d46e906

Trade Failure Prediction: When trades fail, it impacts client satisfaction and has regulatory implications, both of which hurt the overall business. This use case is a good one for predictive analytics because combining data from various sources and applying machine learning models can infer the likelihood of failures with new transactions. This would already be a good first step, but AI can also be introduced by operationalizing these insights and deploying them into a production environment to automatically flag and remediate potential failures.

Credit Risk & Loss Forecasting: McKinsey released a report in late 2018[6] detailing that increasing a credit model's predictive power by just one percent is in reach for most banks. And indeed, the way to do so is through machine learning, which can analyze more data from more sources, faster, to make credit decisions often better than a human analyst. Using a data science, machine learning, or AI platform to do so is even more ideal as it allows for the level of transparency required to ensure that models determining credit risk and loss are interpretable and not a black-box solution.

Cash Management Product Risk: Cash management is one of the most basic, fundamental, day-to-day activities of every bank, and that's exactly what makes it such a prime candidate for revolution by machine learning and AI. Most teams are already using predictive analytics to identify high-risk invoices, duplicate payments, etc. However, this process is extremely cumbersome (thus risky) as data is cleaned in one tool, given to data scientists who develop models in R or Python, and then communicate back to the business. So the key to this use case is optimization between internal stakeholders (like finance, internal operations, and other groups who are involved in payments and receivables) and data experts to increase the speed at which they accurately identify at-risk invoices, duplicate payments, etc., and address the potential issues.

DECREASED COSTS/SPEED-TO-VALUE & TEAM EFFICIENCY

Improvement and Automation of Processes: Improving or automating processes for ETL (extract, transform, load), data preparation, disparate data, etc., all bring organizational value through speed-to-insights and team efficiency. They are also quick, low-hanging-fruit-wins that can be tackled first across all groups and used as a proof point before taking on larger and more challenging use cases.

Replacing Any Rules-Based System: Using machine learning to replace any rules-based system being used throughout the enterprise, no matter what line of business, makes sense given that in order to be compliant, rule-based systems generate a high number of false positives. Machine learning systems can ingest and score alerts from rule-based systems, helping analysts identify risky alerts and accurately passing information across teams for the appropriate investigations.

Streamlining MRM: As mentioned in the challenges section of this white paper, there is a lot that data science, machine learning, and AI platforms can offer in terms of introducing consistency and reproducibility into the process both of MRM itself and the final pushes to production. If for each review the model risk validation team loses time understanding the system and framework different teams are using, banks' ability to accelerate the number of models in production will be limited. MRM requires a consistent system or process by which models are delivered so that the next step (deployment to production) can be completed seamlessly and risk-free.

Regulatory Reporting Automation : With the multiple regulations affecting Financial Services, there has been a multiplication of associated reporting demands. Because of that, a big part of such works are conducted in a manual

---

[6] https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/digital-blog/whats-the-value-of-a-better-model-more-than-you-may-think

way often finishing in consolidated spreadsheets. So, for reportings made in a regular way, poor efficiency results and an increasing part of time of some officers is dedicated to that activity. Also, manual process is putting at risk of mistakes and dependency to a few people mastering the treatments on data. In complement to Data Management and/or RPA process that are addressing part of the pains, machine learning approaches are used to better detect anomalies and to enhance more advanced automation by replacing some business rules.

COMPETITIVE EDGE

Analyzing Money Flows, Competitive and Benchmark Analysis: Predictive market modeling based on massive amounts of data from a wide variety of sources will make investors more informed and better able to serve their customers, with a wealth of knowledge at the tips of their fingers that even teams of humans couldn't hope to analyze by hand.

Evidence-Based Research: With the Markets in Financial Instruments Directive (MiFID) II forcing investors to unbundle trading fees and research costs, research departments need to develop new, unique research products to generate revenue. While t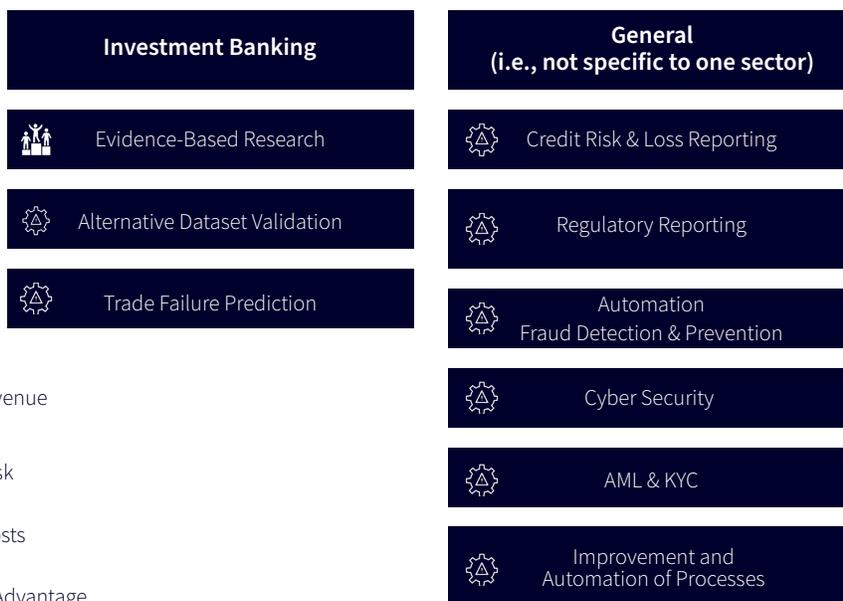raditional research analysts are very comfortable in spreadsheets, they typically do not have deep coding experience. At the same time, data scientists and quants have been incorporated into the research process, but what's lacking is the ability for data scientists, quants, and research analysts to put their heads and their skills together to collaborate and build unique analytics solutions. This is a prime opportunity for data science, machine learning, and AI platforms to fill the gap, allowing all three profiles to work on data the way they like (i.e., through code or not) in the same space.

Alternative Dataset Valuation in Research: In the past few years, research teams have started to have access to and are leveraging unique datasets to create new insights, helping them generate alpha. But with the increase in data available for sale, research teams need a way to quickly analyze that data to identify cleanliness and value before purchasing. Machine learning, particularly via a data science, machine learning, and AI platform, is a good use case here, as it allows research teams to quickly upload data, identify missing values, join new datasets, and run automated machine learning models to determine the predictive value of the data.

Asset Management

- Analyzing Money Flows
- Competitive & Benchmark Analysis

Insurance

- Optimal Pricing & Conversation
- Claims Forecasting
- Claims Triage
- Cash Management Product Risk
- Customer Retention and Churn Prediction

Commercial Banking

- Customer Retention and Churn Prediction
- Recommendation and Upselling
- Reduction of Unjustified Blocked Payments
- Revenue Attribution

Use Cases for Data Science, Machine Learning and AI in Banking

Investment Banking

- Evidence-Based Research
- Alternative Dataset Validation
- Trade Failure Prediction

General (i.e., not specific to one sector)

- Credit Risk & Loss Reporting
- Regulatory Reporting
- Automation Fraud Detection & Prevention
- Cyber Security
- AML & KYC
- Improvement and Automation of Processes

Legend:

- Increased Revenue
- Decreased Risk
- Decreased Costs
- Competitive Advantage

Get Started:
Top Initiatives to Begin Now

Given the immense challenges that stand in the way of the path to Enterprise AI, it's clear that there is a lot of work to do. But banks that have found success are able to start small and work their way up. That means often times the first steps aren't sexy use cases or impressive technological feats in machine learning, but incremental changes that will set them up for future success. This section will look at first-step initiatives for people, processes, and technology that other banks have taken.

INCREASED REVENUE

Better leverage business people (whether analysts in risk, pricing, etc., or even business people without the "analyst" title) instead of continuing to hire new data staff without a plan or agenda. After all, this staff already has deep knowledge of business systems, needs, and the regulatory environment.

The options for today's modern banks are either to continue to hire new data and analytics specialists (who are generally difficult to find, expensive to hire, and hard to retain) or make everyone into a data specialist. Large enterprises outside of the banking word, as well as more and more financial institutions, are starting to realize the value of the latter alternative of data and analytics democratization.

Get staff out of spreadsheets, which contribute massively to security concerns, inaccuracies, and inefficiencies with versioning issues, lack of logs or rollback, and more. Because if complying with regulations and reducing risk are top concerns (and they are), then spreadsheets are the anthesis.

But beyond the obvious problems, moving all staff working with data out of spreadsheets and into a centralized data tool also provides efficiencies and opportunities. For example, with the right setup in terms of governance and permissions, analysts can access data directly without having to go through IT teams, work with it in one place, create visualizations or other outputs, and share access to those outputs directly with others who need the information.

"The Self-Serve Data program at GE Aviation was born out of a conversation in a conference room. The idea was that you would never be able to hire enough data professionals to meet the data demands of the business, so instead, why not turn the business into data professionals."

- Jonathan Tudor, Senior Manager, Self-Service Data and Analytics at GE Aviation
Read more on GE Aviation's data transformation

How I Became a Data Scientist: Tales of a Former Trader

By Alex Hubert

Data science has created a new capacity for powerful analysis by traders that, so far, few have taken advantage of. While quants have been doing data science for awhile now, traders have been largely relying on those quants to execute and have not widely made the shift to developing data science skills themselves. But why?

In the extremely competitive world of finance, it is the traders that are closest to the market. And in fact, compared to quants who are often working from a much more theoretical point of view, traders are a much better candidate for incorporating data science into their work.

Unfortunately, for the most part, they haven't yet. This is a shame because the advantage of diving in to data science is a much better understanding of underlying models. Again, in such a competitive industry, one cannot thrive simply by blindly relying on information (models) handed over from a quant. It's time for traders to get into the data themselves.

Before making the leap into my current position as lead data scientist at Dataiku, I was a trader at Schneider in the United Kingdom. Now more than ever, traders needed to find new competitive edges in an industry that has become entirely dominated by technology - machine learning (ML) and, ultimately, AI have a huge opportunity to provide that edge.
While this may seem scary to the unprepared trader, the good news is that it is never too late to begin the transition towards more data science-oriented strategies. And frankly, the general aversion toward ML and AI isn't unique to traders - employees across industries who don't fully understand how they work (and how they can work together with these technologies) often aren't willing to take the plunge.

But by understanding the ways in which ML and AI are transforming the industry, and taking a few simple steps to shift your mindset, anyone can begin to leverage tools like ML and algorithmic trading.

Why Machine Learning in Trading?

ML is able to use real-time data from unstructured and structured sources to find underlying patterns and trends that might otherwise have been hidden. While high-frequency, algorithmically-determined trading has always been around, traders were often limited by what they could do in spreadsheets.

Data science offers a way for traders to incorporate new and meaningful sources of data at scale and in real time. And the automation of these processes allow for even further improvements in reusability and productivity.

Changes in the market are notoriously hard to source because they can come from anywhere - data science techniques alleviates this by allowing traders to combine data from a huge variety of sources and examining massive amounts of data on past market

Alex began his career as a trader in the city of London, and shifted to become a data scientist after four years. He has worked on a wide range of use cases, from creating models that predict fraud to building specific recommendation systems. Alex has also worked on loan delinquency for leasing and refactoring institutions as well asmarketing use cases for retailer bankers. Alex is a lead data scientist at Dataiku, located in Singapore.

activity. Aside from the combining of data from various sources, one of the largest benefits of data science in trading practices is reproducibility. Quickly and reliably, information can be distilled that would be impossible for even the most diligent of traders to find before.

All of this brings up a very important question, which are the regulatory and security aspects. If using the right tools, combined with good data governance practices, data science should be a very white-box process. All activity on a data project is logged, and it's clear what data sources are being used for which data projects, how the data is being manipulated, etc. Therefore, the use of data science should be very manageable when it comes to things like audits or regulations.

Making The Transition To A Data Scientist

The typical profile of the data scientists is often well aligned with the profile of a trader -- and this is likely why the two fit together so well.
Some skills that are generally common to both
industries include:
* An understanding of how to manage risk;
* An ability to recognize patterns in large datasets;

* A knowledge of how to build relationships with stakeholders; and
* An understanding of financial math, including asset pricing theory, probability and statistics.

Some nice-to-have skills to work on developing for data science that will aid the transition:
* A coding language will be a plus
* Data science tools that will provide some structure

Data science is an extremely dynamic field, so the best policy for any data science oriented trader is simply to stay up-to-date with the cutting edge of the field.

Clearly developing skills in data science will have long-lasting impact on the career of every trader. While automated trading has almost reached the point of ubiquity, ML and AI have the potential to completely change what it means to be a trader.

Working with data outside of spreadsheets also provides the opportunity for smart automation not only for arduous ETL or data preparation, but actions farther down the data pipeline as well.

Enable people to work safely with sensitive data, which is indeed delicate, but not impossible. The idea that no data team can use or process personal data in any way, shape, or form following GDPR, CCPA, or the implementation of any other similar data regulation is a myth. In fact, most regulations (GDPR included) list specific stipulations under which personal data may be processed.

So yes, the ways in which personal data can be obtained and used are limited, and organizations certainly need to be extremely prudent at every level. But it's simply not true that data teams (including analysts as well as data scientists) are completely blocked from using personal data.

One of the biggest advantages for banks in using data platforms is simpler compliance with data regulations. Data platforms allow for:

- **Personal data identification, documentation, and clear data lineage** - that is, they allow data teams and leaders to trace (and often see at a glance) which data source is used in each project.

- **Access restriction and control** - including separation by team, by role, purpose of analysis and data use, etc.

- **Easier data minimization** - given clear separation in projects as well as some built-in help for anonymization and pseudonymization, only data relevant to the specific purpose will be processed, minimizing risk.

- **Leverage people who already have stats and math skills**, bringing them into the data science age. From actuaries to quants, don't let these skills go to waste and instead work on education and transition into roles where their deep industry knowledge can be easily leveraged.

PROCESS

Plant the seeds for self-serve analytics, which in a general sense, is any system by which business or analyst staff can access and work with data to generate insights (predictive or not) as well as data visualization with little direct support from a data scientists, IT, or a larger data team (though the self-service platform itself should be supported by these personas).

Ideally, self-serve analytics would be established as a baseline throughout the entire organization. Often that's easier said than done - getting everyone to agree on what the system should look like, the functionality it should have, etc., can take months, but it's critical nonetheless.

Banks that have been successful in implementing a self-serve analytics program are the ones that get buy-in from different business lines from the ground up. That doesn't just mean getting business lines to sign-off, but really working with them from the beginning to understand their needs deeply and having them test tools and processes to determine what the best solution is.

Ultimately, only the stakeholders in different lines of business will be able to properly evaluate the functionality and ease-of-use of any self-serve data tools based on the skills of their teams. But beyond this, getting their input from the start means that come launch, those business stakeholders will already be invested and can help get the program off the ground.

Start building an efficient way to operationalize models, because in order to become an Enterprise AI player, it's not just about putting one model in production - it's about hundreds or thousands of models.

In any business, data teams face a number of challenges with machine learning model operationalization. Typically, to get the model deployed, coordination is required across individuals or teams, where the responsibility for model development and model deployment lies with different people.

Once in production and running, however, teams then face the problem of maintaining that model, tracking and monitoring its performance over time. In some cases, by the time a model is finally deployed it is already outdated as new, evolving data is coming in.

As the model is re-trained and the deployment is updated, keeping track of different model versions in various stages of development becomes an ever more demanding task - and this is just for one model. Now imagine if you had tens or hundreds?

On top of this, banks have the extra added challenge previously touched on regarding validation by risk teams, which further complicates and lengthens the model operationalization process.

The solution to all of this is preparing now for smooth, efficient, fast (yet complaint) operationalization, which means being able to seamlessly integrate end-to-end not only the production process, but data ingestion, preparation, and modeling. The ultimate goal to work toward is maintaining the consistency and coherency of an entire data science project, where final deployments can be traced right back to the initial datasets and processes and actions that took place within the project.

TECHNOLOGY

Even when talking about people and processes to bring AI to financial organizations, it's difficult to do so without bringing up technologies. While not the only piece of the puzzle, it's true that technology tends to be the cornerstone that enables changes for both processes and people.

Ultimately, banks need to choose the tools and technology that are right for them - that is, the ones that will allow them to future-proof their technology investment, ensuring it serves their needs now but also that it will serve needs in the future and not become quickly obsolete. Here are the first steps as well as longer-term considerations for developing a technology stack that works for the business as well as for the end users

Leverage open source, because there's no question that open source technologies in data science, machine learning, and AI are state-of-the-art and that banks adopting them signal that they're dynamic and future-minded. In fact, the bleeding edge of data science algorithms and architecture is only about six months ahead of what is being open sourced.

For example, looking at a tool like TensorFlow (which is a library for building and training neural networks) there is simply nothing else like it on the market right now. It was open sourced by Google (specifically by their Google Brain project) in late 2015, which means that anyone using it is using the same standard that Google is using for their neural network development.

However, there is a caveat: open source isn't that simple to use, as it often lacks a layer of user friendliness, which can limit its adoption to only the most technical members of an organization. So adding some sort of packaging or abstraction layers that make open source more accessible is a must in order to ensure that open source tools remain an asset and not a burden (as well as to maintain high data governance standards and ensure regulatory compliance).

Despite this caveat, it's worth stressing that open source is worth it. Because finding, hiring, and keeping data scientists (i.e., people with a background in machine learning) is hard. Open source very much acts as "technological honey" that banks can use to attract the best data talent. After all, data scientists want to hone their skills on tools that will become more widespread in coming years.

In addition, it's important to remember that talent is not uniform. Some machine learning experts code in Python, others in R -- banks that leverage open source tools can have both kinds of experts. But those using a proprietary solution with its own particular language and learning curve to adopt can only recruit people who know (or are willing to learn) this particular solution.

Seriously consider the consequences of build vs. buy and the associated ROI. It's no surprise that the age-old question of make vs buy has made its way into the AI space. And building one's own platform or tool for the production of AI systems seems tempting at first for many reasons. Namely, organizations that go this route are feel that a home-grown solution would be more custom, controlled, and secure as well as less costly and (bonus) - no vendor lock-in.

However, it's worth noting that building one's own tools - and specifically one's own AI platform - has drawbacks as well:

- Maintenance: Building an AI platform will undoubtedly require continual resources, even after the launch. Not only to maintain existing features but also to fix any bugs that arise plus add new features to keep pace with the change of today's AI technologies.

- Continual innovation: Along the same lines, building one's own AI platform isn't a one-shot deal. The field is constantly evolving, which means an AI platform will need to constantly adapt in order to avoid becoming quickly outdated and defunct. It will not only need to adapt, but adapt quickly in order to keep pace with technology and competitors.

- Cost: Building includes not just the initial cost of development, but also this continual maintenance. That means instead of being a one-time cost, the home-grown AI platform actually continues to be a growing cost center (consider hiring the engineers to maintain, turnover in the department, and more).

Commercial solutions certainly have drawbacks as well. Not choosing the right vendor to provide AI solutions can mean being locked in to a tool that is slow to innovate, doesn't provide the right security controls, lacks features the business needs down the line, and a host of other possibilities.

Choose technology vendors smartly; for example, steer clear of those that don't allow the use of open source technology, that make users learn an entirely new language (which means slow onboarding and barrier to entry), or that locks the business into using one kind of data storage, one kind of architecture, etc.

When making the choice to buy instead of build AI platforms, it's critical to look for an AI platform or vendor that is:

- Flexible and robust: From allowing contributions and use by multiple profiles (quants, data scientists, actuaries, analysts, and more) to offering a myriad of connectors to data sources and other business-critical systems, look for an AI platform that is technology- and profile-agnostic. Flexibility also means the ability to incorporate cutting-edge open source tools (as mentioned above), allowing banks to leverage the latest and greatest as well as attract the best talent.

- Secure, yet transparent: Any tool banks use should allow AI processes to be transparent enough to comply with responsible AI standards as well as any specific regulations, yet secure enough to ensure that the company isn't being put at risk.
- Devoted to innovation: AI platform vendors should be committed to moving quickly with the world of AI technology. That means having a large research and development (R&D) staff that will allow them to do the heavy lifting in terms of incorporating the latest trends, features, etc., into the platform.

Don't cause an IT nightmare by cobbling together a host of different tools throughout the data pipeline. Using one tool among analysts for data preparation, another among data scientists for building models, and yet another for validating and deploying those models into a production environment not only is inefficient overall in terms of time spent building data pipelines, but it leaves lots of room for error from an IT perspective, which means increased risk of data loss, security issues, and more.

When staff feels that processes are inefficient, that's where they begin to skirt around IT teams and build their own processes that they feel allow them to work faster (known as Shadow IT). Again, messy governance between lots of different tools is a quick path to non-compliance with regulations. Choosing the fewest possible technology vendors to accomplish the task(s) at hand will allow for easier governance. This is especially important as data efforts continue to grow and increasingly more day-to-day users of data come on board.

BGL BNP PARIBAS

Improving Fraud Detection
by Evangelizing Data Science

CHALLENGE: LIMITED VISIBILITY AND ABILITY TO HARNESS DATA PROACTIVELY

BGL BNP Paribas already had a machine learning model in place for advanced fraud detection, but with limited visibility into that model as well as limited data science resources, the model remained largely static.

Members of the business team were enthusiastic about updating the model but were stymied by lack of access to data projects as well as access to the data team to execute the required changes. The challenge was to harness a data-driven approach across all parts of the organization.

SOLUTION: EMPOWERING ALL EMPLOYEES TO BE DATA DRIVEN WHILE MAINTAINING HIGH SECURITY & GOVERNANCE

BGL BNP Paribas chose Dataiku Data Science Studio (DSS) to democratize access to and use of data throughout the company. In just eight weeks, BGL BNP Paribas was able to use Dataiku to create a new fraud detection prototype. And thanks to Dataiku's advanced, enterprise-level security and monitoring features, they were able to do all of this without compromising data governance standards.

The project involved data analytics and business users from the fraud department as well as data scientists from BGL BNP Paribas' data lab and from Dataiku. The collaborative nature of Dataiku and involvement of teams throughout the company allowed for the optimal combination of knowledge to produce an accurate model delivering clear business value.

IMPACT: COMPANY-WIDE FOCUS ON PROTOTYPING AND PRODUCTIONALIZATION

Dataiku's production features allowed for a smooth transition in BGL BNP Paribas' production environment, enabling the new fraud prediction project to show results very soon after the start of project. This, combined with Dataiku's ability to enable quick prototyping, allowed BGL BNP Paribas to quickly test new use cases in a sandbox environment, giving teams flexibility to evaluate new use cases in just a few weeks time to test the global approach and effect.

In turn, the success of the first fraud prediction project was the catalyst for company-wide change at BGL BNP Paribas. Following the completion of the project, there has been a marked shift in company culture to focusing on deployment, industrialization, and quick, easy prototyping when it comes to data products. In addition, because Dataiku is a tool for everyone and not just data scientists or analysts, there has been a shift to focus on using data in advanced analytics and machine learning throughout the company.

BGL BNP Paribas has already begun three additional data projects following the first fraud detection prototype and plans to continue to release new data products regularly to stay at the cutting-edge of the financial industry.

Marlette Funding,
The Best Egg Loan Platform

To Fraud Detection and Beyond
with Machine Learning

MARLETTE ⬡ FUNDING

- 420,000 customers and more than $7 billion funded   by switching to a machine learning-based model.
- US-based with 160 employees
- Six-person data team
- Improved their fraud detection capacity by 10 percent

Evgeny Pogorelov

Director of Decision Science

Marlette Funding, Best Egg Loan Platform

- Subject matter expert about the business and industry as well as the data engineering side
- Main goal: Enable the data scientists to do their jobs

Sami Bouguezzi

Data Scientist

Marlette Funding, Best Egg Loan Platform

- Primarily responsible for building machine learning algorithms for the company
- Main goal: Getting models he builds out of the sandbox and into production

The team improved their fraud detection capacity by 10 percent by switching to a machine learning-based model.

Despite the inherent challenges to ML-based fraud detection, the team at Marlette Funding not only were able to implement a system that caught more fraud than their previous system using fewer resources, but they also used the project as a catalyst for other ML projects to add more efficiency and a more data-driven approach throughout the whole of the business.

Marlette provides a suite of services through the Best Egg Platform for its bank partner, including capabilities that detect fraudsters - that is, people who have no intention of paying back loans. They previously built a statistical model in place that did this pretty well in catching fraud, but it resulted in lots of applicants being pushed to manual review that were not, in fact, fraudulent. This slowed the loan process for applicants and required resources to spend time reviewing non-fraud cases instead of higher-risk cases.

The team - Evgeny and Sami - decided to implement their ML-based fraud detection model using Dataiku from start to finish, as it would allow them to quickly get a model up and running in production and then move on to seamlessly build and deploy other high-priority projects (like pricing optimization or marketing mix optimization). Key steps in their process were:

1. Working hand-in-hand with the fraud operations team to gather data and understand requirements and goals of the project from the business side.

2. Gathering data available from a wide variety of sources (including their own website of course, but also from credit bureaus and vendors specialized in fraud detection data) to create one massive dataset.

3. Using Dataiku to manipulate the data, do feature engineering, and build the machine learning model (previously, the team was manually building models and using software only to deploy them, which was a challenge and resulted in delayed time to production).

4. Benchmarking the proposed model against the current strategy, where they saw it had potential for real business impact and performed 10 percent better than the simple statistical model.

5. Using the one-step API deployer released in Dataiku 5.0 to get the model into production and working on real data quickly for fast time-to-impact (again, this was a change from their previous process, which required weeks to deploy models in production and start to see results).

6. Evaluating the impact to customers and cost savings to the business to show the value of the model and garner support for their work to continue developing models for other parts of the organization.

This combination of the right processes, people working together, and technology allowed Marlette Funding to develop a best-in-class ML-based fraud detection model for the Best Egg Platform with a relatively small team that not only made a difference quickly, but can easily be monitored and updated to prevent it from drifting or becoming out-of-sync with business goals and changes.

Best Practices:
Choosing a Use Case (Where to Begin?)

After seeing all of the possible use cases and successes in the banking world, it's easy to get in over one's head and start dreaming up lofty ambitions. As discussed in the section *Top Initiatives to Start Now*, the best way to begin is with small victories.

But after those are established, it might be time to choose a first use case to tackle with machine learning. And with so many to choose from in the banking space, where to begin?

To hone in on good potential first use cases for an AI project:

- Start with a list of critical business issues from which to choose, possibly soliciting feedback and ideas from teams across the company for a variety of use cases. The reality is that some use cases don't work out; choosing multiple possible projects to start will better produce at least one successful project that can be used as a model for future data projects.

- However, make sure that the initial use cases are relatively small - biting off more than one can chew is a risky move. It's always possible to work up to larger use cases over time, but once a large use case fails, it can be difficult to backtrack and then get the support and resources to try other projects.

- Look at each use case on the list and determine:

  - What is the current process? For example, if one of the items on the list for the retail banking sector is "lower customer churn rates," how are churn rates currently being calculated, who is responsible, what prevention techniques are in place to prevent churn today, etc.

- Would the use of data in general or data science/ machine learning techniques specifically help with this business issue, and if so, how? For the churn example, maybe the answer is that yes, it would help by providing predictions of users likely to churn, which the marketing team can then target directly.

- Do we have the data to use this use case? When choosing a use case, you'll want to pick one for which data is already available. A requirement to collect data before beginning a first project is not ideal since it will significantly prolong the process.

- Where is the data stored and how can it be accessed? This can also potentially influence the decision if it will significantly slow or speed up the process.

- Are we willing to work on this use case with an external partner? Because you'll be working on the first use case with someone (vendor, consultant, etc.), the use case selected should be one that you're actually willing to have involvement from a third party.

- Will this use case help me make money, save money, or do something I can't do today? If the answer is "maybe" or "no," then scratch it off the list. The first use case selected should focus on opportunities with real and measurable results.

From there, the best candidate(s) for a first use case for machine learning projects will be executable in the given time frame, have clear deliverables, and can be put into production for visible business results.

Organizational Structure
for Enabling AI, and the Role of Data Governance

CENTRALIZED VS. DECENTRALIZED MODELS

After taking the first steps toward Enterprise AI as outlined in the first section through people, processes, and technology, and seeing some success, the next logical question often becomes: how can the organization be ideally structured to ensure that the initiative continues to grow and that they can take on more and more AI use cases?

In general, there are two models:

Centralized, otherwise known as the "center of excellence" model, where data expertise lives with one central team for the entire organization, which is then contracted out to lines of business as a service to help get data projects off the ground. This model can work well because it ensures a centralized strategy that develops best practices for deployment throughout all lines of business. However, it means that the data expertswon't also be experts in the line of business (retail banking, insurance, investment, etc.), which places a larger emphasis on the need for collaboration with those experts.

Decentralized, where each business line has its own data resources dedicated only to that specific department. This model can work well because it means data resources are deeply embedded within the business and can have a broader understanding of the issues they face, therefore allowing them to develop more comprehensive and innovative solutions. However, it means that there is a larger risk of governance issues and previously-discussed "shadow IT" problems, as each data team develops its own processes and practices.

Each model has advantages and disadvantages, and the bottom line is that it's up to the organization to decide which works based on how the company is already organized, what the needs are for data projects, how solid communication is between lines of business, and more. The point is that which model the bank decides to use should be carefully considered for its advantages and possible future consequences, and the choice should be a conscious one - communication is key.

THE ROLE OF THE CDO VS. CIO

Regardless of the organization model employed, something that always comes up when it comes to organizational structure in the age of AI is the role of the Chief Data Officer (CDO) vs. a Chief Information Officer (CIO). It's true that they both govern information in some way, but tensions can arise if the responsibilities are not clear.

In general, the CIO should be more concerned with infrastructure issues and making sure that the company's infrastructure serves the business. That includes ownership over the data pipeline, ensuring that data is accurate and trusted. On the other hand, the CDO should be primarily responsible for establishing the organization's overall strategy for using data across the company.

Today, banks often have multiple CDOs or CIOs for different lines of business. This is a product of a decentralized organizational structure, which might mean a CDO and CIO heading up the data resources for a specific part of the company. In this case, the role of data governance and communication between these roles becomes critically important.

THE ROLE OF DATA GOVERNANCE

Of course, regardless of the organizational structure or the roles of CDOs and CIOs, data governance must play a central, overarching role. But how does it fit into the picture?

First of all, both CDOs and CIOs are instrumental in creating and upholding data governance strategies, which not only involve compliance with regulations, but also strategies for ongoing data quality, standards for data management and use, as well as plans for data architecture. Each facet of a data governance plan should be documented, and each line of business should be held responsible - regardless of whether the organization follows a centralized or decentralized plan for Enterprise AI.

Yet governance cannot remain only at the top: it must filter down through all other facets of the organization, down to individuals working with data. If data governance isn't made easy for those with day-to-day data roles, the reality is that it won't be followed. So building an infrastructure that enforces governance at ever level - such as through a data science, machine learning, or AI platform - is a critical piece of the equation.

Lessons from Fintech

Competition has become tighter in today's market thanks in large part to the rise of fintech players (that is, those that provide financial services through technology - think mobile payment apps and cryptocurrency providers). To make matters even more complicated, Forbes wrote in 2018[7] about the rise of techfin as well (which is a technology company that finds a better way to deliver financial products as part of a broader offering of services, e.g., Google or Apple's payment systems).

Obviously, these business are fundamentally different from traditional financial institutions. However, that doesn't mean that there's nothing to learn from their success. Two of the key factors in their rise, for example, have been:

1. Laser focus on the end customer, which has meant an increased focus on trust and transparency that has spurred the very growth of their user base. Traditional players can take a cue here and place an increased focus on interpretability, trust, and responsibility, especially when it comes to machine learning and AI.

2. Ability to use data to innovate, which is the result of agile teams and data democratization. Clearly, this is much easier to execute in a small company born in the age of big data, but larger, more traditional banks can take a cue by enacting simple principles like minimum viable product (MVP) for data projects.

[7] https://www.forbes.com/sites/jimmarous/2018/08/27/future-of-banking-fintech-or-techfin-technology/#12d2968e5f2d

Conclusion

Making the transition into the age of AI isn't easy, but it also isn't insurmountable; banks that take a step-by-step approach and set themselves up with the right infrastructure for people, processes, and tools can thrive. Those that don't adapt will find themselves slowly falling behind, losing business that favors newer technology and employees that want to expand their skills.

This white paper covered a lot of ground, so here are the top five overarching themes and takeaways:

1. Regulations are a unique challenge, but they don't have to hold banks back in the race to AI: While it's true that the ways in which personal data can be obtained and used are limited (and will continue to become more limited in today's strengthening regulatory environment), regulations don't stop people from working with data within the confines of a well-defined data governance strategy.

2. Topics like data trust, transparency, interpretability, and responsibility matter: In other words, they matter not only to customers, but also to regulators. By extension, that means they absolutely need to matter (and be top of mind in terms of developing initiatives and governance strategies in the coming years) to banks.

3. Innovation is critical on many levels: Beyond the obvious perks of bringing new business through new product development and more, it's essential to hiring and retaining top data talent. Leveraging technologies and processes that are cutting-edge (like open source) ensures that those working with data are challenged and continually developing relevant skills.

4. Banks don't need to start from scratch to start the journey to Enterprise AI: The fact is that they already have many of the pieces in place, including staff across roles and business lines that are already trying to use data to make decisions. That means the first steps are relatively low-hanging fruit - making processes smoother for those already using data in their day-to-day roles. More complex projects and use cases can bloom from there.

5. Data science, machine learning, and AI platforms are a clear win for banks: They can provide a platform for governance, including a transparent workspace from which to develop data projects within the confines of regulations. They can get staff out of spreadsheets and also data experts into open source. And perhaps most importantly, they generally reduce the risk involved in getting started with advanced analytics, machine learning, and AI projects.
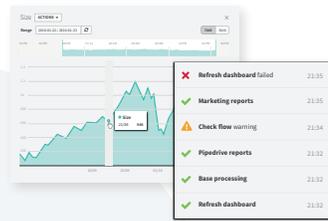
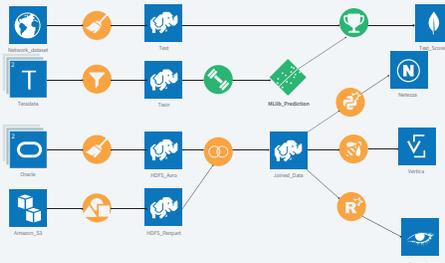# Your Path
to Enterprise AI

Clean & Wrangle

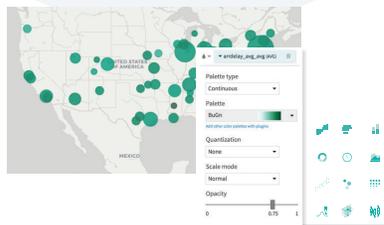### Monitor & Adjust

Build + Apply
Machine Learning

Deploy to production

Mining & Visualization

400+
CUSTOMERS

40,000+
ACTIVE USERS*
*data scientists, analysts, engineers, & more

Dataiku is one of the world's leading AI and machine learning platforms, supporting agility in organizations' data efforts via collaborative, elastic, and responsible AI, all at enterprise scale. Hundreds of companies use Dataiku to underpin their essential business operations and ensure they stay relevant in a changing world.

EBOOK