EBOOK

AI in Insurance: Top Use Cases, Challenges, and Trends

Introduction

Combining mathematics and data analysis in insurance for prediction is not new but has perhaps come into the spotlight with recent worldwide buzz surrounding data science and machine learning. And in the past few years, the pace and scale of data science, machine learning, and ultimately AI adoption has accelerated to enhance or reinvent the processes core to the insurance business.

"AI will have a highly disruptive impact on insurance claims management, leading to cost savings of almost $1.3 billion by 2023, across motor, life, property and health insurance, up from $300 million in 2019."

- Source: Juniper Research

This growth of data science, machine learning, and AI in insurance is driven by a variety of factors, including:

- The breadth of use cases that can be developed using machine learning techniques, particularly those that go well beyond traditional uses of data by actuaries.
- Disruption from fintech (insurtech) and the resulting need to retain and attract clients through a better customer experience.
- Reinforced client price sensitivity.
- The potential business impact in using AI for more and larger use cases.
- Increased pools (and subsequent hiring) of machine learning and AI talent.
- The need to model evolving - and ever more complex - risks.
- The need for increased profit and loss management in a long-term, long-yield environment.

This white paper will explore some of the up-and-coming use cases, challenges that traditional insurance companies face in implementation of those use cases, and ways to address those challenges to be successful in the race to AI. It will also explore trends in how successful companies are executing on AI initiatives.

[1] *https://www.juniperresearch.com/press/press-releases/bank-cost-savings-via-chatbots-reach-7-3bn-2023*

AI in Insurance: High-Value Use Cases

"As AI becomes more deeply integrated in the industry, carriers must position themselves to respond to the changing business landscape. Insurance executives must understand the factors that will contribute to this change and how AI will reshape claims, distribution, and underwriting and pricing. With this understanding, they can start to build the skills and talent, embrace the emerging technologies, and create the culture and perspective needed to be successful players in the insurance industry of the future."

- McKinsey, Insurance 2030: The Impact of AI on the Future of Insurance[2]

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

- Increased Revenue (Marketing)
- Mitigating Risk
- Cost Optimization (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 insurance industry. This section takes a (non-exhaustive) look at some of today's most high-value use cases.

[2] *https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance*

INCREASED REVENUE (SALES AND MARKETING)

Customer retention and churn prediction: A 2019 study by TechSee[3] revealed that 50 percent of insurance customers actively search for an alternate insurer at renewal. And even though more than half (54 percent) of insurance companies made efforts to keep those customers, their efforts were largely unfruitful. Ultimately, it costs more to gain a new customer than to retain an existing one, so the return on investment (ROI) opportunity to optimize customer retention is huge.

That leaves huge opportunities for AI not to simply predict possible churners, (as some customers will leave no matter what) but to take things one step further with a technique called uplift modeling. Since marketing efforts will not change the mind of every potential churner, uplift modeling is a second prediction after the initial prediction that identifies potential churners likely to respond positively to marketing messages.

dataiku

If Treated

If Not Treated	If Treated: +	If Treated: -
+	**Sure Things** Those that would have had a positive response either way and thus represent wasted marketing costs.	**Do Not Disturbs** Those who would have had a positive response but are then negatively impacted by marketing and thus should not be targeted.
-	**Persuadables** The target group - those that would have had a negative response but are then positively persuaded by marketing.	**Lost Causes** Those that would have responded negatively with or without marketing and thus represent wasted marketing costs.

AI-powered customer acquisition: Businesses can develop machine learning-based systems that help sales prioritize their work by assigning an individual probability of conversion to each prospect, whether that prospect is an individual or a group.

One insurance company working with Dataiku did this by first looking at data on existing clients (specifically, their cost of acquisition and lifetime value). They then used this analysis to establish "look alikes" for each prospect - that is, an existing customer who has similar characteristics and therefore will likely mirror the future actions of the prospect.

The end result of this system is a tool available for sales that allows them to more effectively prioritize their prospects by providing two pieces of information to consider: likelihood of conversion and likelihood of recuperation of acquisition costs. The team also created an interactive map containing this data so that any travel to visit prospects could be maximized by visiting other promising prospects nearby.

[3] *https://techsee.me/wp-content/uploads/2019/10/2019-Insurance-Churn-Survey.pdf*

Optimal conversion: Predicting who will convert from a quote to a policy is a delicate balance, but being able to do it accurately allows for optimized spending on marketing to reach customers that are most likely to buy. The sheer number of factors and amount of data at play make it a great use case for AI.

Many insurers want to focus their efforts on optimal pricing algorithms instead of optimal conversion; for example, AXA used machine learning[4] to predict if a driver may cause a large-loss case during the insurance period. However, this is an extremely challenging use case for AI not only because of regulatory constraints, but technical ones as well.

In AXA's case, they used a Random Forest - a common machine learning algorithm that is very accurate for certain use cases - and achieved only 40 percent accuracy. In order to get to nearly 80 percent accuracy, AXA had to develop a much more complex deep learning (neural-network) model.

Improved customer service, driven by machine learning-powered customer segmentation: This category is broad, but important, as increasingly, it's the customer experience that provides the "stickiness" needed to retain clients. AI-powered innovations that range from applications delivering personalized offerings to internal recommendation engines allowing representatives to offer relevant services can help drive additional revenue.

[4] *https://cloud.google.com/blog/products/gcp/using-machine-learning-for-insurance-pricing-optimization*

Enhanced client advisory on financial protection services: In some markets, insurers play a key role in providing long-term financial protection and retirement schemes to customers. In this field, fueling deep understanding of clients in recommendation engines capable of adjusting to a wide range of financial profiles help insurers to deliver much more tailored financial advisory. Such recommendation engines, developed as proprietary tools by the insurance companies, enable them to position themselves as long-term advisors to their clients, offering appropriate recommendations for risk-adjusted solutions.

MITIGATING RISK

Claims forecasting and prediction: In the age of AI and algorithms, older modeling techniques fail to incorporate the wide variety of data sources needed to produce forecasts precise enough for the modern enterprise. Traditional claims reserve estimates don't look at the individual characteristics of policyholders, which effects predictability of future claims.

AI-based systems can use machine learning to take into account many more patterns in data than a human could, including those individual characteristics for better accuracy. And because AI-based systems can easily scale with automation, they can predict payments at an individual policy level, not just at the group level.

Regulatory reporting automation: Insurance companies worldwide have to deal with a slew of regulatory reporting standards that are both time consuming and risky if done incorrectly. The very act of having a centralized AI platform in which all data projects are built and stored is, in and of itself, a step in the right direction for smoother regulatory reporting.

With all actions logged and - in the case of Dataiku - a transparent view of all data pipelines, regulatory compliance can be monitored more real-time to ensure that there are no surprises if an audit arises. In case of an audit, AI itself can help automate manual work like validation of customer data, customer data security operations, and more.

COST OPTIMIZATION (OPERATIONS)

Underwriting: Machine learning is well suited for underwriting, identifying patterns in diverse data sources (from imagery to credit bureau data) to create a more tailored risk assessment.

For example, unstructured data can be incorporated to improve underwriting decisions. Public satellite or private imagery can be quickly and automatically analyzed to confirm risks on a property or site, while data from IoT devices (combined with other publicly-available information on individuals) can be instantly parsed for much more accurate - and timely - assessments of coverage.

Of course, insurance rating laws require rates and rating factors that are not excessive, inadequate, or unfairly discriminatory. AI has the potential to reinforce current underwriting approaches if developed with a strict white-box approach with full understanding and traceability of factors and resulting outputs (read more about this in the section **Responsible AI & the Insurance Industry**).

An Inside Look at Complex Risk Modeling

ABOUT THE AUTHOR

Dr. John Kelly is a founding executive, CTO, and lead architect for Envelop Risk's core cyber-risk model. He was previously the Director of Analytics at QxBranch, a quantum computing software firm acquired by Rigetti Computing in 2019. Prior to QxBranch, Dr. Kelly served as the Technical Lead for Corporate Data Analytics for Lockheed Martin, where he worked directly with Envelop Risk Director and former Lockheed Martin Global CTO Dr. Ray Johnson. Dr. Kelly's background is primarily in machine learning and signal processing, and he holds a PhD from Carnegie Mellon University in Electrical and Computer Engineering.

Predicting cyber risk is an incredibly challenging problem. Cyber risks are complex – they're sentient, constantly evolving, high tech, and a global threat at a massive scale. They've also become critical; beyond the basic business risks like product viability, competition, and financing, cyber now ranks alongside natural disasters as an existential threat.

A fascination with the structure of cyber risk led to our founding Envelop Risk. Our team's interdisciplinary background in national security, machine learning, computer science, insurance underwriting, simulation modeling, assured software development, and economics - combined with a passion for solving seemingly impossible problems - made this project an irresistible challenge.

In taking on this challenge, we immediately face a data paradox. Historically, a common refrain in the industry is "where do you get the data?" There are very good reasons to ask this question; there is not sufficient, reliable, structured data to create a conventional statistical model with predictive power. Further, significant amounts of data are proprietary and internal to firms and cannot be accessed ethically. On the other hand, the cyber ecosystem is a natively data-intensive digital realm. Massive amounts of relevant technical, economic, legal, and intelligence data are available on global cyber activity.

While Envelop Risk cannot specifically predict that company X will lose Y amount to cyber risk in year Z, we can characterize what is happening globally to a high degree of accuracy, make probabilistic predictions about companies or portfolios, measure how accurate we believe those predictions to be, identify and assess correlated risks, and seek to anticipate immediately how the overall cyber economy will respond to changes in conditions.

Certainly, the massive amounts of data we obtain and analyze to do this wouldn't have been computationally possible even ten years ago, and it would not be possible to model without machine learning technology. We've used Dataiku DSS from the very beginning to accelerate development and stay focused on improving our predictive power.

Below, we discuss the macro of our systems, our integration of human and machine learning (Augmented Intelligence), our modeling and simulation approach, how we use the tool in underwriting, and a note about the future.

Augmented Intelligence for Cyber Risk Modeling

Envelop's core risk modeling system consists of a large number of smaller models, components, and datasets that seamlessly flow together to produce reported outputs. This flow has five main sections: data ingest, feature engineering, modeling, simulation, and reporting.

All sections include a feedback loop between respective expertise and data. Expertise informs what data to obtain, and post analysis helps differentiate between signal, noise, new knowledge, and spurious correlation. Human expertise is also deployed as part of the calibration process as well as to parameterize areas where we know the data is insufficient or lagging, and experts can provide useful estimates. These results feed the next round of investigation.

We think of our system as an augmented intelligence system. By this we mean the team plus the tool work together to continually increase our ability to predict risk. The technology continues to sharpen its accuracy, and at the same time, experts use and refine the tool to leverage these data-driven insights to develop their own understanding. Over time, the augmented intelligence system - technology plus humans – progresses to keep pace with the evolving cyber threat.

Where do you get the data?

Our system consolidates data from a multitude of proprietary, public, and hybrid sources to construct a comprehensive view of the global cyber landscape. The system processes substantial amounts of data to properly prepare for use in modeling and simulation. It conducts quality assurance checks and operational readiness reviews to ensure that the data is reliable and stable.

We constantly evaluate and deploy new data sources based on reliability and predictive power. Each type of data requires a slightly different process in terms of assessment and potential value. We take an outside-in approach when gathering information to maximize our initial returns on gathering company data and then continually refine our results, focusing on:

1. Threat landscape, including threat intelligence and historical claims, losses, and portfolio performances.
2. Company profile data that could map the company to the threat landscape.
3. External cyber posture data (data we can legally collect without the company's participation) that informs the likelihood that attacks from those threats cause losses.
4. Internal cyber posture data that further refines the loss likelihood.

At a conceptual level, the goal is to map cyber threats to the cybersecurity controls that can thwart them and evaluate cybersecurity controls from our data sources.

Modeling & Simulation: How We Account for Uncertainty

Modeling the complexities of cyber risk at a high level or trying to find a single distribution that fits the expected cyber losses of a company would be an intractable undertaking that would produce questionable results. By breaking risk down into its constituent components, we can better validate each individual piece and then produce high-level output by simulating the relationships between those components.

Our models do not predict a set value; rather, they parameterize a distribution that is carefully chosen to best reflect the behavior of a specific component of cyber risk. Using a Monte Carlo simulation that draws from each of the distributions allows the system to better account for - and integrate - uncertainty in the final output.

Individual data models predict frequency, severity, insurance terms (when unknown), attack parameters, coverage area breakdowns of losses, effect of attack parameters on severity and coverage areas, portfolio performance, catastrophic scenarios, and correlated risks. Each model uses historical data and human expertise to predict parameters for a distribution that reflects the behavior of that component of risk. These feed the Monte Carlo simulation to produce probability distributions for overall expected losses and breakdowns of those losses.

Drilling down to the components of risk allow us to better understand the relevant data for each component and to fit well-studied statistical distributions that best reflect the known behavior of each component. When the nature of one of those components evolves, those changes can then quickly and accurately propagate to our final system output without needing to adjust other system components.

Underwriting First

Envelop's primary business is cyber reinsurance, which we offer through an agency agreement with MS Amlin. Demand for reinsurance is strong in the market, as insurers seek to capture market share in a growing market while protecting against risk aggregations. The strength of our capability has enabled us to bring significant capacity to the market and to deploy it intelligently to help our clients meet their portfolio needs.

Very early in our development, we prioritized the user experience of the underwriter, because the marketplace of insurance itself is a time-tested economic mechanism for risk transfer and diversification, led by disaggregated, competitive underwriting decisions. End results of this marketplace include more efficient allocation of capital, diverse product offerings, rapid innovation, and protection against systemic biases.

Unlike many new entrants, we built our system as a proprietary, internal capability that we do not license externally. We focus our resources on prediction and seek to equip our underwriting team to develop and execute efficient strategies. Because it is internal, our underwriters have full transparency into what's driving the results, allowing us to integrate their experience and knowledge, provide feedback and guidance to customers, track trends, and shift focus to particular market segments, products, or strategies. An internal capability also allows us to execute technical partnerships and evaluate market strategies that would be impossible with a third-party capability.

The model is used in all phases of underwriting, including capital allocation, strategy development, structuring, pricing, portfolio construction, and reporting.

Forward-Thinking Underwriting

The complexities inherent in cyber risk can be found in most risk areas, and while initiating any analytics capability from scratch is a time and resource-intensive undertaking, these modeling principles will be applied in all domains, new and existing. The capabilities afforded by combining data science with enterprise-grade software development, global expertise, and intelligence gathering will be required across business lines over the next decade.

Use in an insurance application requires forward-thinking underwriters who can conceptualize numbers at scale, think abstract, understand what software and data can do, and deploy these capabilities effectively in competitive, high leverage environments. The world is changing.

- Dr. John Kelly | CTO and Lead Architect

envelop

[Continued from Page 7]

Claims processing: The potential for AI to improve the claims processes is massive because it not only promises reduced costs from eliminating inefficiencies, but also increased customer satisfaction that has the opportunity to increase revenue.

Today's cutting-edge insurtech and - more recently - traditional insurance businesses are increasingly looking to AI for faster triage (e.g., larger claims with more uncertainty can be started more quickly by specialized teams, letting smaller claims be closed out even faster). Plus, technologies like deep learning (specifically natural language processing, or NLP) and computer vision move automatic processing into the realm of possibility, allowing businesses to move away from time-consuming, manual processing.

In fact, automation and advanced prioritization have the potential to touch nearly every part of the claims process, from intake to assessment to settlement. As an added bonus to speed, automation also can ensure fewer human errors and easier auditing. This is not to say that claims automation removes humans from the loop entirely; rather, much like fraud detection, it allows them to be leveraged more smartly only in cases where a human is truly essential. For example, claims with missing data might be routed to a human who can handle the case (and bonus: a bot could see how the human resolves the missing data and learn for future cases).

Fraud detection and prevention: Insurance organizations are all exposed to fraud risks, whether dealing with false claims, false billings, unnecessary procedures, staged incidents, withholding of information, and much more. This industry must be on the cutting edge of technology to stay ahead of fraudsters and reduce losses.

With limited resources on fraud investigation teams, every investigation into a case ultimately identified as low risk is wasted time. Hiring more staff to conduct these manual audits is an expensive and inefficient option - instead, the key is optimizing that team's work by using AI to detect fraudulent activity with a higher degree of accuracy. With detailed, specific small data from patients and providers feeding into these large data sets for analysis, audit teams look only at the highest-risk cases and can therefore detect more fraud.

Accurately Identifying Fraudulent Claims

Santéclair, a health network (part of Allianz), found fraudulent reimbursements stemmed both from opticians as well as patients, but they didn't have a system in place that allowed them to effectively analyze the right data and that would adapt with increasingly sophisticated fraudsters. Instead, they relied on "if-then-else" business rules to identify likely fraud cases, which resulted in the manual audit team spending their time on too many low-risk cases. With the increase of reimbursement volume (more than $1.5 million a year), they needed to improve their efficiency and productivity.

Leveraging Automation and Advanced Machine Learning

Santéclair identified these high-risk cases using Dataiku by:

- Outsmarting fraudsters with advanced machine learning algorithms that continually update and automatically learn or retrain using the latest data so that any new fraud patterns are immediately identified and audited. Dataiku handles the entire workflow, from raw data to exposing the predictive model to the operational applications.
- Automatically combining hundreds of variables from different datasets, including patient/prescriber history, interaction graphs, prescription characteristics, and other contextual data.
- Allowing teams to develop their data science skills through Dataiku's collaborative, easy-to-use interface.

Saving Customers Money with 3x More Effective Fraud Detection

Due to the comprehensive solution developed with Dataiku, Santéclair and Eulidia have:

- Enabled fraud detection teams to target actual fraud cases three times more effectively.
- Reduced time-to-market for similar projects by making a POC in a few weeks and then industrializing the project within a few months with a low impact on the IT team, thanks to the production-ready component in Dataiku.
- Saved their customers a lot of money by decreasing fraudulent behaviors in the health network and excluding the fraudsters from the network.
- Saved time with a model automatically updated and monitored along the way to prevent drifting of performance with little human supervision

"In less than a year, Santéclair has developed an unprecedented fraud detection system using Dataiku that allows our company to handle a growing volume of invoices and control costs. By choosing Dataiku, Santéclair was able to internalize its data skills and pursue additional analytics projects."

- Jocelyn Philippe, Head of Partnerships and Development at Santéclair

Improvement and automation of processes: There are still plenty of manual business processes in the industry outside of claims processing that can be improved through robotic process automation (RPA) and AI. One major example is policy management, but even smaller customer-side improvements (like streamlining the application process) can save time and reduce the chance of human error.

AI Meets Mail Processing:
AI for Insurance Admin Tasks

BEHIND THE MODEL

Léo Dreyfus-Schmidt is a mathematician and holds a PhD in pure mathematics from University of Oxford and University of Paris VII. After five years focusing on homological algebra and representation theory in Paris, Oxford, and the University of California - Los Angeles, he joined Dataiku where he has been developing solutions for predictive maintenance, personalized ranking systems, price elasticity, and natural language applications.

Even as we continue to reach new technological milestones and solve the world's most demanding problems, many insurance companies are still confronted with the oldest of administrative nightmares: piles and piles of physical mail.

Head of Dataiku AI Lab Léo Dreyfus-Schmidt offered a scalable solution to the eternal problem of mail processing by using AI and deep learning techniques to solve the four major problems of mail processing, driven by a real use case for an insurance company:

1. Distinguishing if a letter is handwritten or typed
2. Parsing the text from a typed letter
3. Detecting words in handwritten letters
4. Extracting meaning from the images of words

Their goal was ultimately to deliver a production-ready tool that could be used to automatically sort any letters received and send them to their appropriate departments. Traditionally, this would have to be done by hand -- an expensive and time-consuming task.

The first challenge that the data team had to overcome was a very heterogeneous data set. While initially expecting to receive a pretty even mix of handwritten and typed letters, the actual training set contained a mix of letters, envelopes, forms, leaflets, and other forms of written documents.

With the 200,000 unlabeled images that they received, they went through the long process of labeling every document by type using a webapp they built. This allowed them to begin building their deep learning model on a large training set of data.

The Model

In a process that involved constructing a vector representation of the document images using an autoencoder (a process explained more thoroughly in the talk on the subject[5]) and running a Random Forest machine learning model on the dataset, the team was able to successfully distinguish hand-written documents from typed ones.

Extracting text from images now accurately identified as being typed was the first (and perhaps most straightforward) step. They used an open source OCR (Optical Character Recognition) engine called Tesseract to do this.

Then came the hard part -- the handwritten letters. This process involved using computer vision techniques to detect paragraphs on the page and then to detect words from those paragraphs. They then stacked two common layers of deep learning techniques to learn and read the visual characteristics of those words.

Using some open-source datasets (and some augmented versions of those datasets) as the training set, they were then able to create a deep learning model in Dataiku that was able to identify the meaning of those handwritten words with fairly high confidence. Once this step was completed, the team had an operational method of extracting meaning from all of the incoming documents.

- Léo Dreyfus-Schmidt ,PhD | Mathematician

data iku

[5] *https://www.brighttalk.com/webcast/17108/331676*

[Continued from Page 14]

COMPETITIVE EDGE (INNOVATION)

Seamless customer experience: The idea of a better customer experience by weaving AI throughout processes - from acquisition all the way through claims processing, prevention actions and advisory - in order to retain clients has already been well-covered throughout this section on use cases.

However, it's worth mentioning again the competitive advantage that a truly seamless customer experience can provide. In addition, as more processes become automated and as client information is leveraged in a refined manner, client service operators can switch from being task managers to giving a more human, tailored experience to customers, providing client experiences capable of competing with promises of new fintechs.

Product development: Machine learning and AI open up opportunities for new products that were previously impossible, perhaps because of risk or cost. For example, private insurance coverage for some very risky exposures are unavailable today for lack of a dynamic model that can actually accurately represent the possible losses. More sophisticated and complex models can incorporate data from more sources than ever before to accurately estimate loss outcomes and probabilities.

Improved agility around data and AI also offers insurance companies the opportunity to better adapt to new client demands and emerging risks. For example, the rise of environmental concerns and increase in climate change-related risks, which demands adaptation both in pricing models and products with new types of data lacking maturity of traditional KPIs, is a key area where mastery of data science-related techniqueswill come as an essential differentiator.

Trends in AI for Insurance

Understanding how AI can be valuable in different parts of the business is easy; but actually executing on any of the use cases described in the previous section requires calculated coordination between people, processes, and technology. As more and more insurance companies start their journey in building Enterprise AI, there are a few trends emerging as best practice:

People

Growing collaboration between data scientists and actuaries: Though the two are currently largely working on different projects, the crossover is growing. Finding ways (likely via technology) to encourage and facilitate their collaboration is paramount to the success of insurance companies in AI initiatives. Especially considering that actuaries largely outnumber data scientists in most traditional insurance organizations, getting the two to work toward a common goal is critical.

The center of excellence model: More and more companies in financial services at large are putting in place AI center of excellences to support business adoption.That's not to say that AI projects stay siloed in one team; however, there is a need for one central driver and owner of AI efforts who can then set the framework enabling business units and data scientists deliver impactful results.

GO FURTHER: UBS on How to Build a Data Science Service Center of Excellence

Processes

Self-service analytics: There is an increasing need, in parallel to a center of excellence, to enable line-of-business professionals or analysts to access and work with data to generate insights - predictive or not - and data visualization with little direct support from data scientists, IT, etc. That's where self-service analytics comes into play.

GO FURTHER: Get the White Paper: Enabling AI Services through Operationalization and Self-Service Analytics

Getting staff out of spreadsheets: One quick win on the path to Enterprise AI in the insurance industry is making a concerted effort to get staff out of spreadsheets. Not only are they error-prone, but they make regulatory reporting and data privacy standards a nightmare. While not a sexy use case, leveraging an AI platform to consolidate all efforts of working with data is a huge organizational win, with multiple applications across all functions.

GO FURTHER: The Guidebook to Going From Excel to Dataiku

Technology

The ability to work with unstructured data (i.e., text, images, etc.): The year 2019 was a landmark one for the field of natural language processing, more commonly referred to as NLP. Cutting-edge techniques once mainly restricted to the research area are now becoming much more mainstream and translating into real-world business applications. With the amount of automation to be done with documents in the insurance industry, having the technology and skills to execute on NLP is critical.

GO FURTHER: Get the White Paper: **Get up to Speed With NLP**

Data science, machine learning, and AI platforms These tools allow for the scalability, flexibility, and control required to thrive in the era of Enterprise AI because they provide a framework for data governance, efficiency, reproducibility, automation, responsible AI, operationalization, collaboration, and more.

GO FURTHER: Get the White Paper: **Why Enterprises Need AI Platforms**

Bringing Insurance into the Age of AI:

On Growing Data Science Collaboration for 5x Faster Time to Production

- 33k employees globally (16k in the United Kingdom)
- The UK's largest multi-line insurer
- Global data science practice - called Quantum - has more than 700 data and analytics professionals
- The Customer Data Science Team is the company's customer-first data center-of-excellence and is made up entirely of data scientists

"As a customer data science team, we're always looking at how we can make things better for customers. And happily, that also tends to drive profit."
- Tom Spencer, Head of Customer Data Science | Aviva

Analysts and actuaries have been around practically for centuries to bring mathematical models to the world of insurance. But data science, machine learning, and AI have the potential to take it one step further.

Even so, getting these initiatives off the ground has generally not been easy for most in the insurance business. The industry is often characterized as traditional and slow-moving, or worse, one that is not as customer-centric as it should be. However, there are some working to - and succeeding at - bucking these trends to bring data science to the forefront of insurance, transforming the way the business works with data for the better.

Aviva's Keys to Success

Aviva, the United Kingdom's largest multi-line insurer, developed their Customer Data Science Team around two years ago to complement their already existing (and robust) global data practice. One of the Customer Data Science Team at Aviva's most celebrated projects is ADA (Algorithmic Decision Agent), Aviva's personalization AI, which helps the company be more specific and relevant to its customers. The AI, built using Dataiku, helps the company understand its customers better and delivers tailored marketing experiences based on their needs.

We talked to Aviva's Head of Customer Data Science Tom Spencer and Data Scientist Ayca Kandur about what makes their team 5 times faster at going from raw data to production:

1. **Good data.** Upstream to downstream, one of the most important contributors to great data science is great data. For Spencer and his team, that means not only high quality raw data, but having a staff that knows what data is, what it means, and where it comes from.
2. **Proper tooling.** When Spencer started building the Customer Data Science Team at Aviva, his first priority was getting great people, but a close second was getting the tools in place that would allow that team to work together and to be its most productive. Today, the entire data science team uses Dataiku for every step of the data pipeline, from connecting to data to data preparation, building models to deploying them to production, and everything in between.

"When we started building ADA ... it was taking us quite a long time with the existing legacy systems. But through using Dataiku and the API functionality, we reduced the amount of time from beginning to end to build a model and push out the model into the marketing channel."

- Ayca Kandur, Data Scientist | Aviva

3. **Staying grounded in results.** The team at Aviva recognizes that data science is neither fun nor useful if the business ultimately isn't actually using what they're producing, so they have a strong focus on pushing to production (not just playing in a sandbox) for real impact.
4. **Staying Agile.** Delivering value fast is also important to Aviva, which means the Customer Data Science Team strikes a balance between quick-and-dirty R&D and more structured push-to-production strategies.

"The answers often aren't from insurance, they're from other disciplines out in the world. Our job, then, is to partner with our colleagues that have deep insurance expertise for a data-driven outlook."

- Tom Spencer, Head of Customer Data Science | Aviva

Responsible AI & the Insurance Industry

Going from producing one machine learning model a year to thousands a day is well within the average insurance organization's reach, and operationalization has made it possible for a single model to impact millions of decisions (as well as people). Yet, despite the exponential increase in the amount of machine learning models in production, only a few companies have dedicated the time and effort to ensure these models are deployed responsibly.

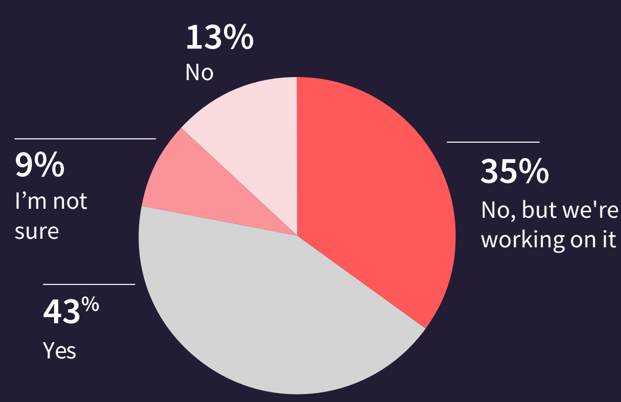
Responsible AI is perhaps even more important to consider for insurance organizations, as they must strike a delicate balance between efficiency and profitability as well as customer satisfaction, trust, and regulatory compliance. A responsible use of AI covers three main dimensions, which should all be considered when developing an organizational implementation strategy:

- **Accountability:** Ensuring that models are designed and behave in ways aligned with their purpose. This includes using white-box over black-box models when it makes sense, which is more inherent in the insurance world due to regulations but still can be a challenge when it comes to increasing model complexity (read more in White-Box vs. Black-Box Models: Balancing Interpretability and Accuracy) as well as taking the right precautions when it comes to model bias (read 3 Steps Toward More Ethical AI).

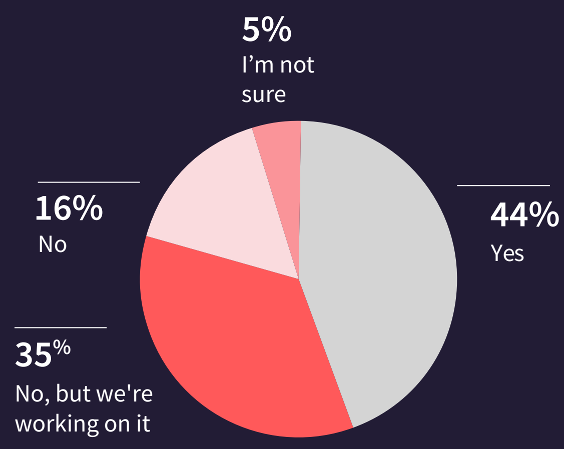
In a 2019 survey of more than 400 data professionals, Dataiku asked: Does your organization have processes in place to ensure data science, machine learning, and AI are leveraged responsibly and ethically?

The overall responses vs. those of respondents in the financial services industry were as follows:

Does Your Organization Have Processes in Place to Ensure Data Science, ML, and AI are Leveraged Responsibly and Ethically?



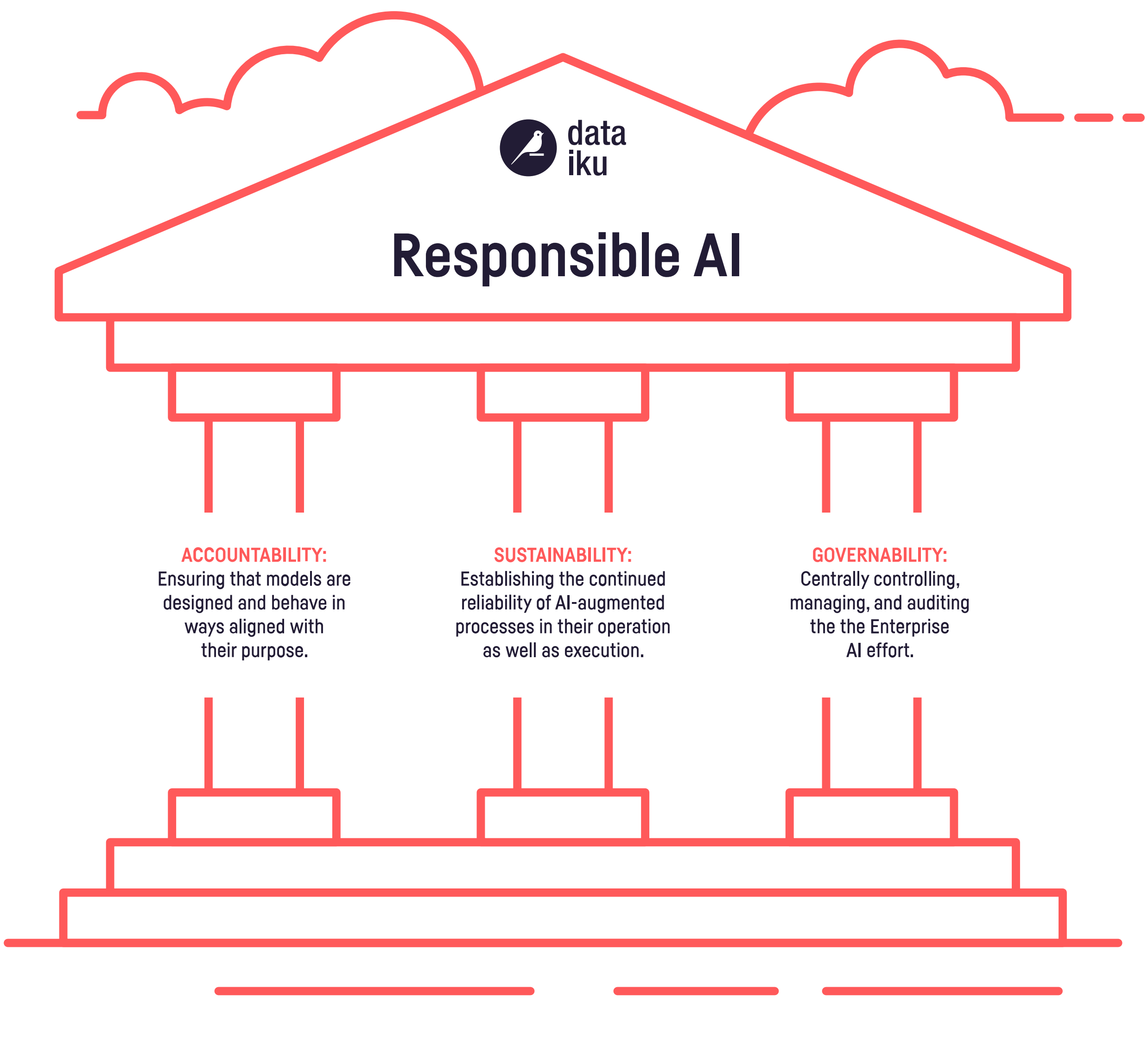
Respondents in the Finance Industry



- **Sustainability:** Establishing the continued reliability of AI-augmented processes in their operation as well as execution. This means that no matter what changes in technology or architecture lie ahead, they don't have adverse effects on existing models in production.
- **Governability:** Centrally controlling, managing, and auditing the Enterprise AI effort. This means expanding the idea of governance beyond the traditional IT sense and instead thinking about it more globally, from access and security all the way up to centralized model management (so-called MLOps).

Given these dimensions (which are each complex in and of themselves), it's clear that having a comprehensive strategy for responsible AI is not easy. So where can organizations begin?

Of course, it's important to empower people, hold them accountable, and develop concrete policies and processes around how responsible AI should be executed. But underpinning all of this is having the right technology, which can enable people and processes while at the same time delivering accountability, sustainability, and governability of data and AI projects. Dataiku, for example, is technology agnostic and built for best-practice methodology and governance throughout a model's entire lifecycle, including concrete features for responsible AI efforts like advanced, granular model explainability.



Challenges & Solutions to Get Started

Despite progress and trends moving the industry in the right direction, there are undoubtedly still challenges that can hinder progress as organizations move forward on the path to Enterprise AI. This section will address just a few and propose solutions:

Finding and hiring data talent: Hiring people with skills in machine learning and AI is extremely difficult across industries due to a shortage of talent and skyrocketing demand. But the insurance industry is better positioned than most to overcome this challenge via upskilling. With tens (possibly hundreds) of statistics-minded actuaries already on staff, providing the right tools to nudge them into the world of AI is a small step. Upskilling business staff is also a great way to fill talent gaps, and in many ways, it's necessary for insurance businesses that want to democratize AI. Ultimately, it's much more difficult to teach a pure AI talent the ins-and-outs of the business than to teach someone who knows the business like the back of their hand some basic skills for using data in their day-to-day work.

Legacy tools: While insurance is undoubtedly more advanced when it comes to technology than some of the financial services players on the banking side (just look at cloud adoption), there is still a lingering issue of legacy tools and systems. It becomes infinitely more difficult to upskill staff per the previous section if data tools are difficult to use, aimed exclusively at the coder population, or even if the user experience is constantly changing with each new technology that gets introduced. AI platforms like Dataiku can help resolve this challenge by being a unified visual abstraction layer for data projects, providing robust features no matter who the audience (coder or non-coder) and a consistent experience no matter what the underlying changes in technology.

Breaking down silos: Today's enterprises tend to be siloed in many ways - from individuals to entire teams and potentially also data, it's difficult to get everything (and everyone) working together across these lines. Again, tools can help when it comes to breaking down the silos of data. But when it comes to people, the answer is collaboration. Insurance organizations should be looking for ways on AI projects that actuaries can work with data scientists and also with excerpts from the business side to come up with the best possible solutions to high-value use cases.


MEASURE SUCCESS
REASSESS GOALS

Conclusion

This white paper has covered at a very high level some of the use cases, challenges, and next steps for insurance, but obviously there are many other types of insurance business out there that might have different needs or use cases when it comes to data science, machine learning, and AI.

However, no matter the specific applications, the model for moving into this new era and for success remains the same:

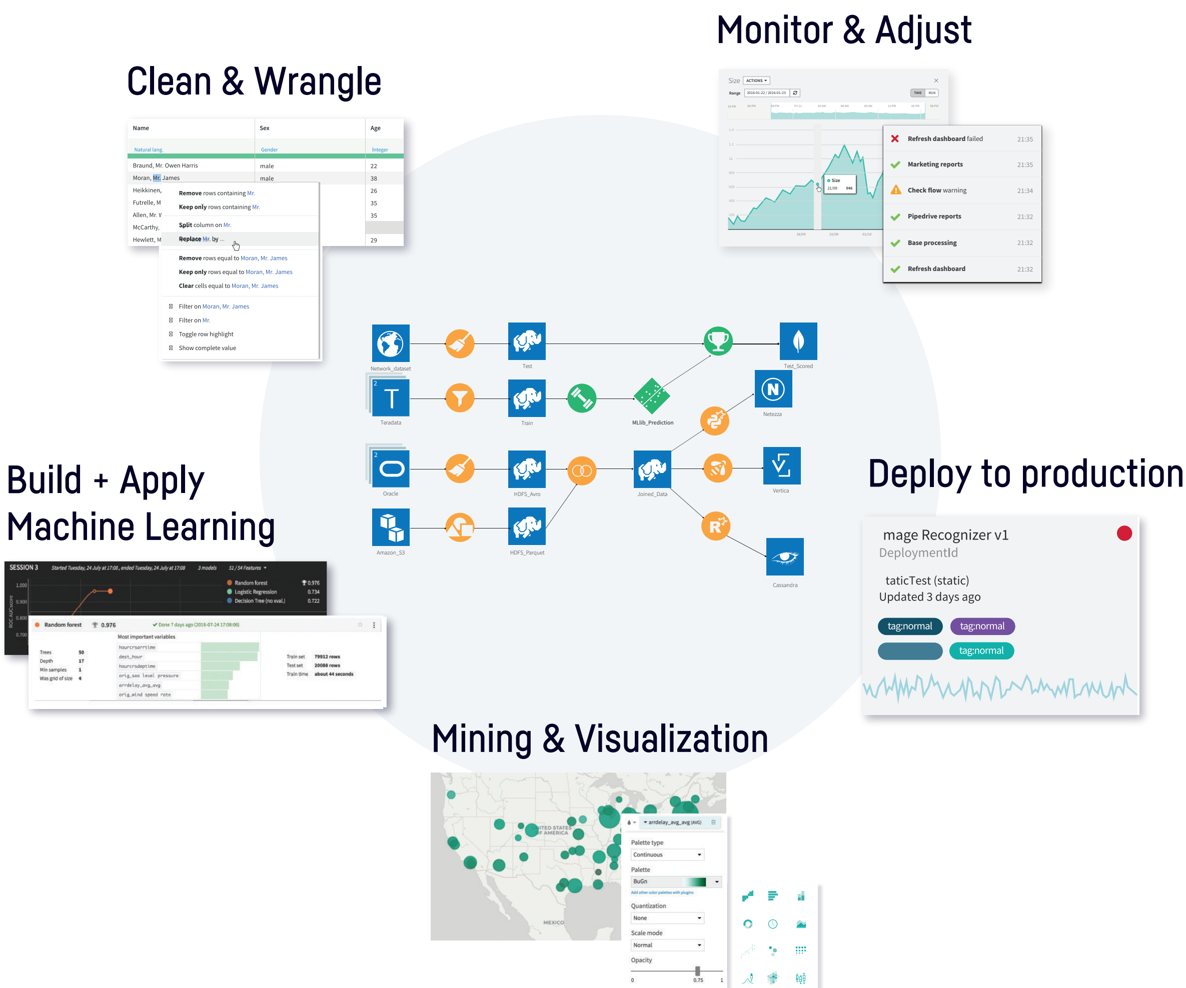
People: It's essential to start educating and upskilling staff on data science and machine learning technologies and initiatives. It's become increasingly clear that the only way to transform a business around data is for the initiative to be democratized; that is, not only supported from the top down, but also the bottom up.

Processes: One of the biggest challenges in democratizing working with data across the business is having the systems and processes to do so. Setting up self-service analytics systems[16] that allow people to access and use data in a controlled way is a good way to get started.

Technology: Of course, mobilizing people and creating processes are difficult to do without the right technology. Data science, machine learning, and AI platforms can facilitate the journey; for example, Dataiku does this at scale by making data accessible to a wider population within the enterprise, facilitating and accelerating the design of machine learning models, and by providing a centralized, controlled, and governable environment.

[16] *https://content.dataiku.com/ge-aviation-ssa/data-democratization-ssa?lb_email={{contact.email}}*

Your Path to Enterprise AI



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

www.dataiku.com