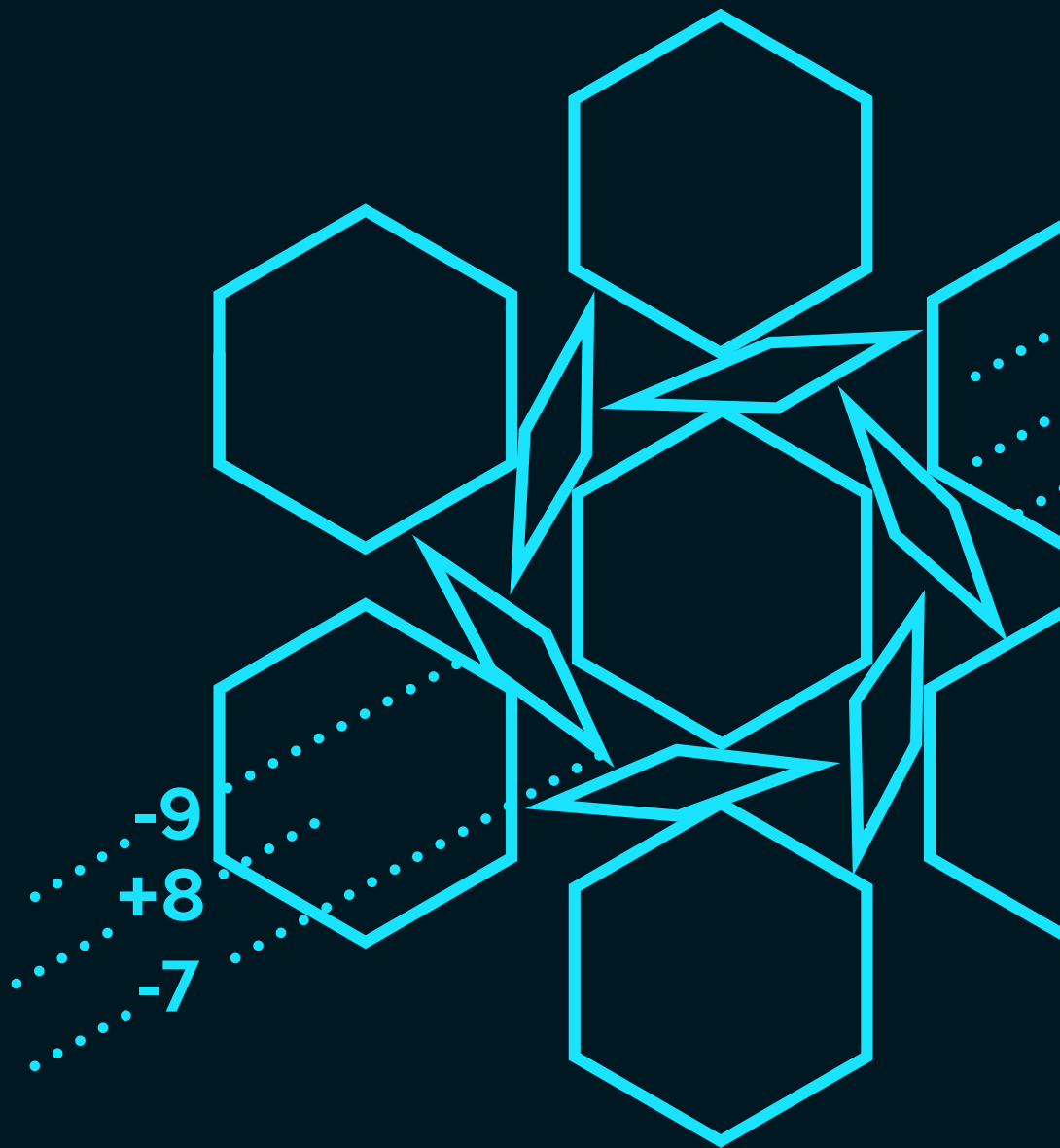




DATA SCIENCE FOR BANKING AND INSURANCE

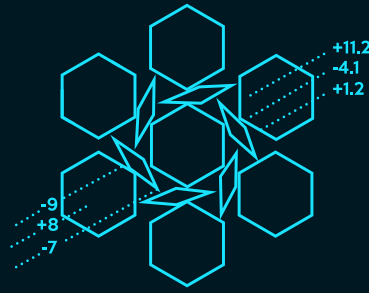
Surviving and Thriving in the Era of Internet Giants
and Financial Technology Startups



WHITE PAPER

www.dataiku.com





S U M M A R Y

3 INTRODUCTION

4 PART ONE: THE THREAT

5 The 1990s Internet Gauntlet: Evolve or Perish

5 New Challenges: GAFA, Fintech, and the Internet of Things

10 The GAFA/Fintech Edge: Big Data & Algorithms

13 PART TWO: THE SOLUTION

14 Harnessing Unique Assets & Data Science to Compete

16 Data Science in Action: Use Cases in Traditional Banking and Insurance Companies

17 *Interlude : How Quantum-Based Solutions are Going to Change Banking and Insurance*

23 *Interlude : 3 Challenges to Address in Banking/Insurance Data Projects*

26 PART THREE: TAKING ACTION

27 The Right People

29 *Interlude : The Digital Era and Data Science: What Does It Spell for the Future of the Actuary Profession?*

31 The Right Data

32 The Right Technology & Process

36 CONCLUSION

37 END NOTES

38 APPENDIX

INTRO DUCTION

Over the course of many centuries, the banking and insurance industries have developed processes, products and infrastructures that have shaped the economic history of humankind.



But now, **they are threatened with extinction by challengers** who appeared on the world stage a mere **couple of decades ago**, and some who emerged just a scant few years ago, but who **nonetheless are already rewriting the rules of financial services**. These challengers include **Internet-era giants like Google, Amazon, Facebook, Apple, Baidu and Alibaba**; nimble startups like Credit Karma, Lending Club, Square, Lemonade, TransferWise and GoFundMe; and even, through the Internet of Things, wholly **unlikely competitors like manufacturers of consumer and industrial goods**.

Banks and insurance companies **can fight back by accelerating the digitization path** they have been on for some time, and **enriching it with the tools of the newcomers' trade** - namely, data science, big data and algorithms. As they do so, they should also make maximum use of their unique assets, including talent with much sought-after expertise in mathematics and statistics, deep subject matter knowledge **solely lacking in many data science endeavors, a massive, largely untapped reservoir of customer data**, and a network of physical branches and offices that can deliver a human edge in the quest for meaningful, **multi-channel and multi-sensory customer experiences**.

However, **success depends on the speed with which traditional banks and insurers respond to these new challengers**, in their skillful exploitation of their competitive assets, and in **assembling the right people, data, tools and processes to get the job done**.



The disruptive threat is real, but the battle is not lost.





PART ONE: THE THREAT

4 PART ONE: THE THREAT

5 The 1990s Internet Gauntlet: Evolve or Perish

5 New Challenges: GAFA, Fintech, and the Internet of Things

GAFA Titans Enter Financial Services

The Rise of Fintech

The Internet of Things: Coming Between You and Your Customer

10 The GAFA/Fintech Edge: Big Data & Algorithms

Achieving Uncommon Results with Unconventional Data

Growing Markets through Underserved Populations

Gaining Breakthrough Efficiency with Machine Learning

Change is not Optional

A- THE 1990S INTERNET GAUNTLET: EVOLVE OR PERISH



In a prescient 1995 speech, Hugh L. McColl Jr., then CEO of what was soon to become Bank of America, encouraged members of the Banking Administration Institute to adapt to the nascent Internet age, or perish:

“As every schoolchild knows, the dinosaur didn’t survive the Ice Age...

It’s not that he lacked the capacity to evolve. He just didn’t have the time. Unlike the dinosaur, bankers can see the changes ahead. We have a choice in the matter. The dinosaur never did.”

Hugh L. McColl Jr.

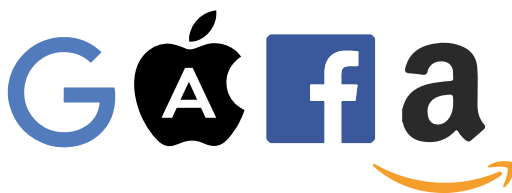


The financial services industry – including banking and insurance companies – largely heeded that warning, and **over the next 20 years began adapting legacy systems and paper processes to the new digital era.**

It was a difficult process but one that was beginning to yield returns in operating efficiency and new convenience and value for customers. **The global financial crisis of 2007/2008, however, slowed this digital transformation as financial institutions reallocated resources to recovering from their losses and adapting to a more stringent regulatory environment.**

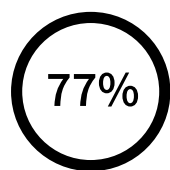
B - NEW CHALLENGES: GAFA, FINTECH, AND THE INTERNET OF THINGS

1 • GAFA TITANS ENTER FINANCIAL SERVICES



Now, in 2016, as the banking and insurance industries have by and large **recovered from the crisis, they find themselves facing new challenges born of the Internet era.** First and most significantly, the giants of the digital age, like Google, Apple, Facebook and Amazon (or GAFA) in Western markets, and Chinese powerhouses like Baidu, Alibaba, Tencent and Xiaomi (BATX) in Eastern markets, are beginning to make direct inroads into banking and insurance markets.





In particular, GAFA and their overseas counterparts have been incrementally testing the waters with **offers that include online and mobile payment, money transfers, personal lending, account and savings management, peer-to-peer lending (crowdfunding), insurance and currency trading.**



More than three-in-four (77%) financial services executives are concerned about losing customers to companies like Apple, Google, Amazon, Lending Club, etc., who could offer alternative financial services.¹

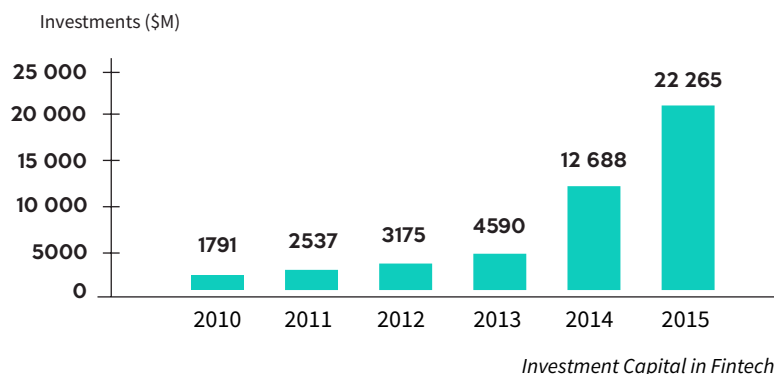
While the results have been mixed to date for GAFA, with more advanced market penetration by BATX players, the US digital powerhouses have shown a firm commitment to growing their footprint in financial services. This can be seen in initiatives like Financial Innovation Now, an “alliance of technology leaders” including Amazon, Apple, Google, Intuit and PayPal, that is working to “modernize the way consumers and businesses manage money and conduct commerce,” and advocating for regulatory policies that will better support “new marketplace innovators in financial services.”

They are also investing in financial technology (fintech) startups in this new marketplace. These startups, like the GAFA giants who make up one segment of their backers, are cherry picking high-volume financial services that are tailor-made for the online and mobile world into which they were born.

Company	Theme	Initiative	Launch Date
	Payments	Google Wallet app	2011
	Payments	Google Compare	2012
	Payments	Android Pay	2015
	Payments	Apple Wallet (previously Passbook)	2012
	Payments	Apple Pay	2015
	Payments	Messenger Payments	2015
	Lending	Amazon SMB Lending	2012
	Digital Currency	Amazon Coins	2014
	Lending	Amazon Store Card & Card Comparison	2015
	Insurance	Amazon Protect (Product insurance)	2016
	Payments	Amazon Payments	2017
	Payments	Check-out by Amazon (B2B e-commerce solution)	NC

Banking and Insurance Initiatives from GAFA

2 • THE RISE OF FINTECH

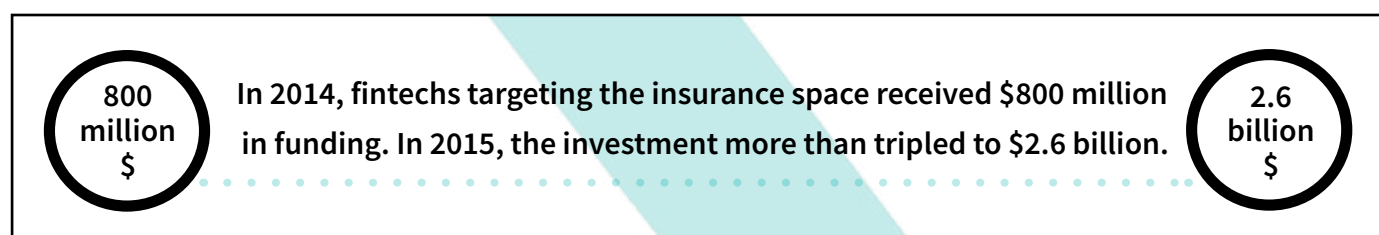


With investment capital that **reached \$22.3 billion in 2015** (an increase of 75% over 2014), fintechs are powering into a broad range of banking services, including virtual banking; personal and small business lending; financial advising and investment brokering; credit scoring; currency trading and money transfers; equity crowdfunding; payment processing and more.²



While fintech companies began as (and remain) disruptive challengers to traditional banks, **banks have countered by partnering with, acquiring and investing in these upstarts** (banking institutions accounted for 38% of total fintech investment capital in 2010, growing to 44% in 2015).³

FINTECH IN INSURANCE: 'INSURTECH' STORM SURGE LOOMING



For the insurance market, it is a slightly different story. Google and Amazon have made very limited direct forays into insurance, and the market to date had seen relatively low fintech (or 'insurtech') investment, though that is changing and the number of insurtechs is increasing.

There are currently, for example, quite a number of peer-to-peer insurers, such as Friendsurance, Lemonade, InsPeer, and InShared. The oldest and best established of these, Friendsurance, piggybacks on social media big data infrastructures to enable customers to create a circle of both real and virtual 'friends' to share the costs of small claims and deductibles, with traditional insurers covering large claims. Another startup in development, Teambrella, will power its peer-to-peer service by Bitcoin in a bid to make insurance "fair and transparent."

It will work by allowing each member of a 'team' to deposit funds into a special personal Bitcoin wallet, with claims mutually reimbursed only if both the submitting teammate and three out of eight semi-randomly selected teammates sign for it.

ON-DEMAND INSURANCE: LURING CUSTOMERS WITH AMAZON-STYLE 1-CLICK EASE

Other innovative insurtech service include item/event specific, on-demand coverage. Trōv, for instance, provides what they call “smart insurance”, that allows people to insure “just what they want, when they want, and for just as long as they want” - entirely from a mobile app. Specifically, the app collects data about a customer’s possessions and provides machine-learning enhanced risk pricing for single item coverage, of any duration (down to seconds) and at premium levels that can scale down to pennies, with chat robots handling claims. The result, according to Scott Walchek⁴, Trōv CEO and Founder, is a great customer experience:



“Insurance today is bogged down by heavy process and forms, often requiring the need to talk directly to a person. By moving the entire process to the phone we’re making getting insurance as simple as a ‘1-click’ Amazon purchase. What’s more, claims can be as simple as a quick text message exchange with reimbursement or shipping of a replacement item happening in minutes - instead of days or weeks.”

Scott Walchek Trōv CEO and Founder

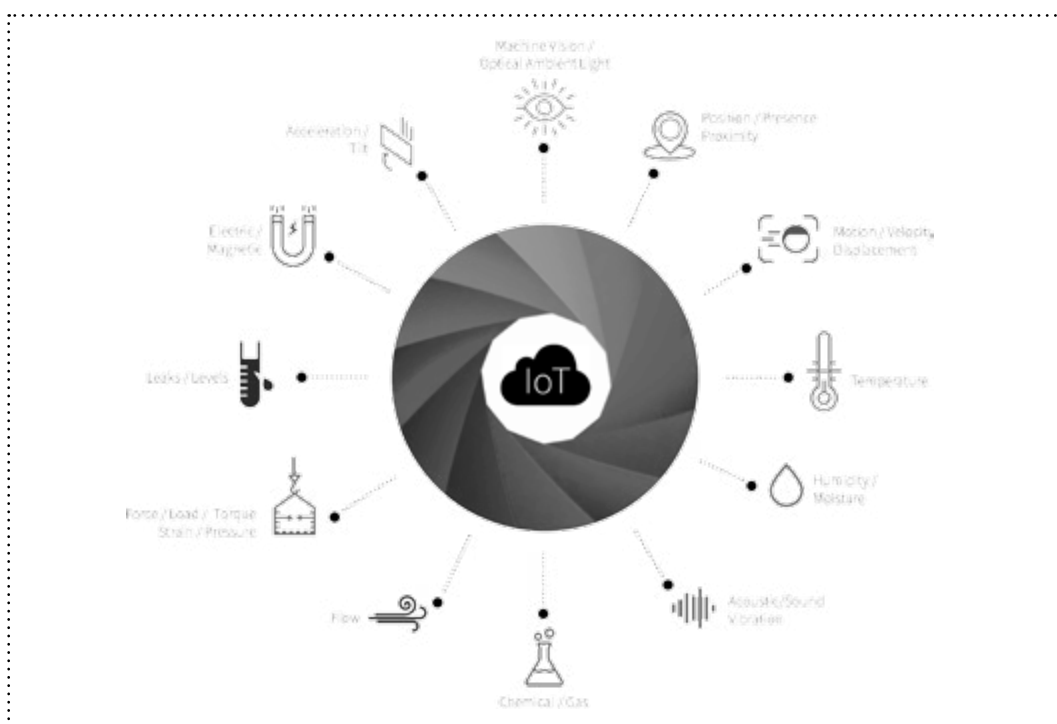
In spite of this wave of insurtech innovation, arguably the more immediate challenge to the insurance industry is taking shape within the Internet of Things, which is also encroaching on banking services.

3 • THE INTERNET OF THINGS: COMING BETWEEN YOU AND YOUR CUSTOMER

In the Internet of Things (IoT), billions of sensors, computer processors and communication devices are being embedded in or attached to every kind of ordinary ‘thing’ imaginable – human beings (‘wearables’), mobile phones, tennis shoes, water pipes, grapevines, cattle, toasters, street lamps, etc. - to share data across mobile networks and the Internet. It is a revolution that may see up to 30 billion smart and connected devices in service by 2020 (not counting the most ubiquitous ‘smart’ device: the smartphone).

While the types of data collected and the uses for it are wide-reaching and highly diverse, the functional role is largely one of remote monitoring, analysis, and/or control (think of self-driving cars and home patient care).

At present, industries like manufacturing, health care, retail and security are leading the IoT market, but financial services are poised to be transformed by the IoT revolution as well.



IoT Can Deliver Astonishing Variety of Data

MANUFACTURERS SEEK NEW LINES OF REVENUE

For example, companies that manufacture consumer and industrial goods **are looking to piggyback service offerings on top of their smart and connected products**, which they can now monitor and control in real-time.

These services include **predictable offers like remote diagnostics and repair and proactive maintenance**, but interest is growing in unconventional areas as well, like insurance (for the product itself or the environment it monitors, like crops or manufacturing lines) and, to a lesser degree at present, **provision of loans, leases or purchases based on equipment usage data**.

Consider, for instance, the way telematics has enabled manufacturers like Rolls-Royce, GE and Pratt & Whitney to lease their aircraft engines to airlines on a flight-hour basis (with bundled proactive maintenance), **bypassing traditional finance and leasing processes and providers**.



GAFA & FINTECH SEE ENORMOUS POTENTIAL IN IOT

Recognizing the market opportunities in IoT, GAFA is also investing aggressively in the domain, with **Google spending more than \$3 billion to acquire just one company – smart thermostat maker Nest Labs** – in one of the four major verticals into which its capital is flowing: **the connected home, wearables (like smartwatches), automotive (connected cars), and robotics**. And, the connected home, wearables and the smart car markets are all prime sectors for a new generation of car, home and life insurance products **based on real-time monitoring and the collection and analysis of behavioral data**.

It is unclear at this time whether Google will **enter the insurance market directly** through these domains, or will simply **monetize the rich new streams of data it enables**, including acting as a broker between insurers and consumers, as is happening now with Nest insurance partners.

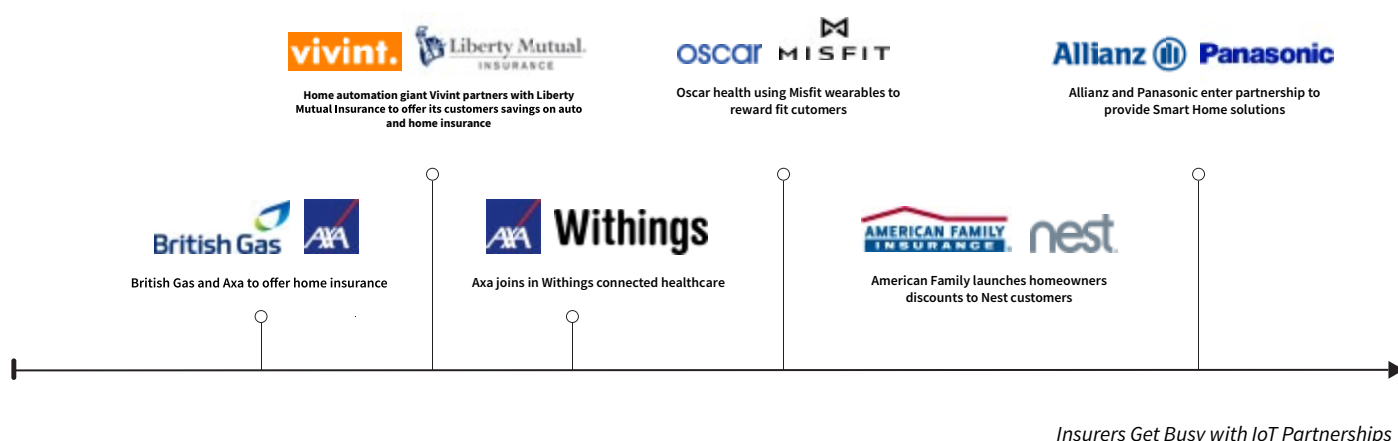
Apple is likewise operating in a broker mode at present, **entering, for example, into home insurance partnerships in collaboration with Evrythng, a maker of an IoT platform for smart home products**. A similar path seems likely for Amazon as its Dash button evolves from a one-touch ordering device to a controller for managing smart-home features.

Insurance companies are not, however, partnering just with GAFA and its partners or acquired companies. **They are also partnering directly with start-up device makers like Water Hero and Gem, for example**. These two companies make devices that enable **continuous monitoring of home water pipes and water flow**.

They support instant alerting when anomalies are detected, and in the case of Water Hero, **enable an automatic shut-off of water flow**. With roughly one-third of all household claims related to water leaks, the appeal of such technology to insurers is obvious.

The question remains, however, **who will own the customer relationship?** The race is afoot for customer proximity, and it is challenging the banking industry as well. **The engagement of the banking industry to date in the IoT domain has mainly involved the use of smartphone technology for authentication and localization, and internal, operational uses like monitoring inventories and offices/branches.** Fintech companies, however, lead most banks in exploiting smartphone-centered IoT technology to support innovative, customer-facing mobile banking and payment transaction experiences, disintermediating but not yet replacing banks.

If banks and insurance companies want to protect themselves against IoT-enabled disintermediation, or even direct incursions into core services, **they will need to better leverage IoT technology to drive innovation**, and they will need to appropriate the tools that are giving their new competitors an advantage in both IoT and non-IoT spheres: **big data and algorithms.**



C - THE GAFA/FINTECH EDGE: BIG DATA & ALGORITHMS

As digital pure players born of the Internet age, the economic capital of GAFA and fintech companies is based on using big data to big advantage through intelligent algorithms.

In the case of GAFA, they amass staggering amounts of data, and run algorithms against that data to produce real-time content, services, recommendations and ads targeted to the level of the individual, that quasi-magical ‘market of one’ that has helped shape today’s powerful and demanding customers. An example is Amazon’s use of big data mining and predictive analytics to deliver personalized purchase recommendations.

Fintech companies do this as well, using volumes of data that don’t match GAFA, but that do include data sets drawn from GAFA and non-GAFA big data producers (like government agencies and weather services). And, like GAFA, fintech companies have achieved compelling, personalized experiences through the innovative use of big data technologies, digital/mobile-first strategies, and algorithm-based data mining and predictive analytics.



1 • ACHIEVING UNCOMMON RESULTS WITH UNCONVENTIONAL DATA

For example, fintech start-ups (and BATX companies) **in the lending sector are using algorithms and unconventional sources of data, like social media activity and cell phone usage**, to more quickly and accurately assess identity, fraud risk, and credit worthiness, and to automate underwriting and loan origination processes.

In addition to being able to serve customers more quickly, cheaply, and accurately, **this recipe of diverse big data and advanced algorithms is also helping these companies expand markets by serving young or underserved populations with little to no credit histories** – a global population estimated at more than 2.5 billion.



Though these companies are generally protective of the precise algorithms and data sources used, lending start-ups like **Lending Club, Kabbage, LendUp, Affirm, and ZestFinance** all use this recipe of big data, algorithms and automated platforms to deliver **innovative and personalized customer experiences**, and to achieve better company financial performance (ZestFinance, for example, claims that it has achieved a **40% improvement⁵** in default rates over those associated with traditional industry scoring methods).

2 • GROWING MARKETS THROUGH UNDERSERVED POPULATIONS

As recounted in an overview of fintech innovation through big data and algorithms (“Big Data and Online Scoring: Fintech and Beyond”⁶), **people seeking loans can install InVenture’s mobile app, which then tracks how they use their phones, and this usage provides an effective alternative means of risk assessment.**

For instance, based on historical data InVenture has collected and analyzed, **sending too many text messages or having a low battery too often may be a sign that an applicant is a credit risk, while people who make more calls at night or use gambling sites (surprisingly) tend to be more credit-worthy.**

InVenture, for example, **is providing much-needed lending services to people in several countries in Africa where credit scores don’t exist and institutional lending is rare.**

While startup companies may be at the forefront of using unconventional big data sources and algorithms to bring innovation to insurance and lending, even traditional industry companies are finding machine learning-based strategies can bring enormous efficiencies to labor- and paper-intensive workflows like underwriting.

3 • GAINING BREAKTHROUGH EFFICIENCY WITH MACHINE LEARNING

After first digitizing the largely paper-based application process, **machine learning components were developed to automate the underwriting workflow.** These components included a fuzzy logic rules engine to encode underwriter guidelines, an evolutionary algorithm **to optimize the engine’s performance as guidelines and data evolved**, and a natural language parser to limit the amount of free text that required review by an underwriter. The system achieved full automation of underwriting for **19% of applications**, and provided pre-processing that freed underwriters to focus on **applications best benefitting from human review.** It also enabled the development of a **web-based customer self-service application.**⁷



For example, Genworth Financial, one of the world’s oldest and largest insurance providers, **began using machine learning to improve its underwriting process in 2002.** Specifically, Genworth developed an end-to-end system to automate the underwriting of Long Term Care (LTC) and Life Insurance applications.



Genworth succeeded with its early foray into algorithm-based underwriting in part **because the system designers implemented their solution in a series of intermediate steps so that actuaries and underwriters could build the necessary “comfort level” with the new approach.** This is a useful lesson. It is helpful with any algorithm-based automation system to give professionals the time and tools they need to understand and gain confidence in how it works and how it makes decisions. **However, the bottom-line imperative to embrace the use of algorithms and big data sources as rapidly as possible is clear.**

4 • CHANGE IS NOT OPTIONAL



If traditional banks and insurance companies **do not move quickly**, they risk **losing out on expanded market opportunities**, certainly, but more than that, they risk, at a minimum, **being reduced to transactional plumbing** for the “new marketplace innovators in financial services,” or, at worst, suffering death by a thousand small cuts as **profitable lines of business are cannibalized.**

The imperative for these companies to evolve is driven home forcefully by fintech venture capitalist **Mircea Mihaescu** in his **BankNXT** article “*The New Age of Algorithms*.”⁸ In it, he argues that although a sea-change in the way banks look at and use data is slowly emerging, they need to move more rapidly in embracing the “**new ‘algorithm-enabling’ technology**” **at the heart of fintech innovation.** As he points out, the advantages of algorithm-driven approaches to traditional challenges seem obvious, but too few are using them:



Mircea Mihaescu

“It seems obvious to do algorithmic lending - and no bank does it, only startups. It seems obvious to do deep learning on transactional data, and very few banks do it today, but many startups.

It seems obvious to do pattern extraction on IT logs to detect internal fraud, and only startups do it today. It seems obvious to correlate physical customer presence with physical payment data, and only startups do it today. It seems obvious to perform real-time spending analysis on transaction data, and only startups do it well today. It seems obvious to use blockchain technology to enable ‘no trusted third party needed’ transactions, and only startups do it today.”

Whether Mihaescu’s take is 100% accurate or not, traditional banks and insurers are without a doubt lagging behind their GAFA and fintech counterparts in exploiting big data and algorithms, but it is not too late, and if they can close the algorithm gap, they will be positioned to exploit wholly unique assets and mount a formidable competitive strategy.

PART TWO: THE SOLUTION

13 PART TWO: THE SOLUTION

14 Harnessing Unique Assets & Data Science to Compete

The Unique Assets of Bank and Insurers

Appropriating the GAFA/Fintech Data Science Advantage

16 Data Science in Action: Use Cases in Traditional Banking and Insurance Companies

17 Use Case: Enabling Personalized Policies & Premiums in the Auto Insurance Market

18 Use Case: Proactively Preventing Customer Defections in Banking

19 *Interlude : How Quantum-Based Solutions are Going to Change Banking and Insurance*

20 Use Case: Discovering New Customer Segments in Banking

21 Use Case: Automating Life-Event Marketing in Insurance

22 Use Case: New Efficiency and Accuracy in Detecting Fraudulent Insurance Claims

23 *Interlude : 3 Challenges to Address in Banking/Insurance Data Projects*



A- HARNESSING UNIQUE ASSETS & DATA SCIENCE TO COMPETE

1 • THE UNIQUE ASSETS OF BANKS & INSURERS



Traditional banks and insurers carry **both advantages and disadvantages** in competing in the new age of algorithms. Disadvantages include entrenched legacy technology and business processes, physical branch/office overhead, and an at-present unequal regulatory burden.

Moreover, in the case of GAFA, **financial services companies can never rival the multi-billion person user/customer base and attendant resources of the tech giants**, nor achieve their omnipresence in people's digital lives (it is estimated that people spend more than **half of their time online on one of GAFA's services**, and today's smartphones, whether iOS or Android, are rarely out of reach). However, traditional banks and insurers also possess unique competitive assets, namely:

- 1.** Deep historical troves of (largely untapped) customer data
- 2.** Physical touchpoints to complement and humanize digital ones for a unique multi-channel experience
- 3.** A higher level of trust among consumers for safeguarding personal data
- 4.** A deep reservoir of professionals with both extensive domain expertise and advanced quantitative skills



ASSET #1: UNTAPPED RESERVOIRS OF CUSTOMER DATA

With data mining and predictive analytics, the hidden value in the first of these assets - massive stores of day-to-day transactional customer data - **can provide a unique advantage in better understanding, predicting and delivering what customers want, while helping to better address risk, fraud and market uncertainties.**

ASSET #2: AN EXTENSIVE BRANCH/OFFICE NETWORK



The second asset, physical branch or office networks, can seem like merely a cost overhead, but it can play a vital role **in developing meaningful customer relationships as financial services become increasingly digitized**. As the online giants are learning, digital-only relationships have their limits: a binding customer experience is built on both physical and digital touchpoints.

Traditional banks and insurers are at an advantage here, if they make optimal use of their network to build relationships. Regional managers at companies like Bank of America and M&T Bank are, for example, seeing a real evolution as their physical branches morph into advising centers for customers, with one regional M&T manager noting a swing in service activity underway **from 80% transactions and 20% expert advice to 20% transactions and 80% expert advice**.⁹

DIGITAL GIANTS SEEKING PHYSICAL CONNECTION

Digital giants like Amazon and Google are joining Apple in developing retail outlets to nurture a physical connection with their products, people and brand:

Apple is expanding its base of sleek stores in China, India, and Latin America, while renovating U.S. stores with next-generation designs

Amazon plans to aggressively expand its base of pop-up stores in U.S. shopping malls to 100, and reportedly envisions opening as many as 2,000 Amazon Fresh-branded grocery stores across the US over the next 10 years based on the success of a 20-location pilot underway now

Google is testing the brick-and-mortar waters with a new pop-up store in New York City

Online banking veteran ING is complementing its digital-only ING Direct service with cafés that serve as “agency offices” complete with comfy sofas, affordably priced coffee, free WiFi, free conference rooms and other amenities (but without the traditional branch services that would trigger Community Reinvestment Act regulations)

ASSET #3: STRONGER CONSUMER TRUST



The existing physical touchpoints for traditional financial services companies can also serve to reinforce an important third asset: trust. While the final crisis did shake consumer confidence, individuals’ trust in traditional financial institutions remains strong. **According to an IBM survey, 70% of respondents indicate that they trust traditional banks more than non-bank competitors**, and when asked in another survey which institution they trust more to safeguard their personal information and privacy, consumers ranked traditional financial institutions higher by a wide margin over new online providers.¹⁰

ASSET #4: PEOPLE WITH QUANTITATIVE SKILLS & INDUSTRY EXPERTISE



Traditional financial service companies have a further advantage in having long employed professionals with advanced mathematical and statistical skills, providing them with a ready workforce of industry-savvy quantitative experts who can be trained to compete with GAFA and fintech on what has to date been their primary home field advantage: **an adroit use of big data and algorithms to create great customer experiences in the digital, and increasingly the physical, spheres**. In other words, banks and insurance companies are well-positioned to master data science.

2 • APPROPRIATING THE GAFA/FINTECH DATA SCIENCE ADVANTAGE

WHAT IS DATA SCIENCE ?

Data science is the collaborative discipline that is enabling the age of algorithms. **Specifically, it is the combination of people, data, tools and processes that GAFA and fintech use to transform statistical and mathematical skills, information technology** (especially big data technologies), and industry subject matter expertise into actionable insights and business innovation.



DATA SCIENCE IN FINANCIAL SERVICES

Specifically, **data science is enabling these companies to leverage data mining and predictive modeling to personalize offers, reduce risk, create disruptive new products, expand markets, minimize operating expenses, automate traditionally manual processes, and much more.** These would be very beneficial business enhancements for traditional banks and insurance companies too, and some are already using data science to achieve them. **Let's look at some concrete examples of this,** and then we will detail how organizations can best gain competitive competence in data science.

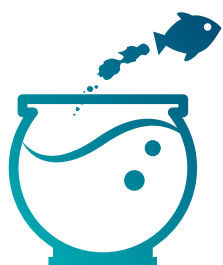
B- DATA SCIENCE IN ACTION: USE CASES IN TRADITIONAL BANKING AND INSURANCE COMPANIES





ENABLING PERSONALIZED POLICIES & PREMIUMS IN THE AUTO INSURANCE MARKET

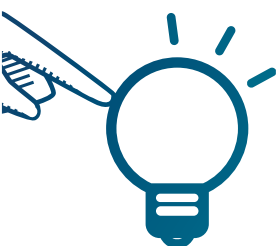
The Challenge



To balance the need for sufficient reserves with competitive pricing, insurance premiums have typically been assigned using historical data such as the driver's or vehicle's history and statistical probabilities associated with factors like age and gender to determine risk.

It's a strategy that works reasonably well, but it is a static snapshot of risk that cannot factor in an individual driver's behavior beyond prior reported incidents, limiting insurers' ability to accurately and continuously assess individual risk and deliver truly personalized products.

The Solution



IoT technology (or telematics) is now capable of **delivering visibility into that behavior through large streams of real-time, insurance-relevant data transmitted through onboard devices, like data on position, speed, acceleration, braking, etc.,** and perhaps, in the near future, DUI or traffic law abidance. One traditional insurance company that is **taking advantage of this new visibility is Allianz.** Allianz offers a car insurance policy that allows data tracking **through a mobile app bundled with a GPS-equipped dongle** customers can simply plug into a car USB port. The company uses the data to develop innovative, **personalized pricing models like pay-how-you-drive (PHYD) and flexible coverage options like pay-per-mile plans.**

Results

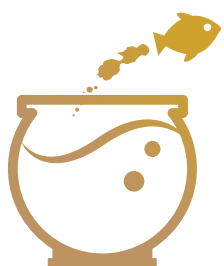


The new PHYD policies that **Allianz has been able to develop using these new data streams enable it to reward good drivers with lower premiums** (up to 40 percent for drivers aged 28 and under), and to better understand its customers, **decrease fraud, and encourage positive driving behavior** (reducing incidents and claims). In addition, the new data is enabling Allianz to offer expanded services, like breakdown assistance, stolen vehicle tracking and recovery, and emergency/accident assistance, which are based on the real-time detection of incidents and automatic initiation of response services. **As a result of innovations such as these, Allianz's 2015 total revenues reached a new high of 125.2 billion euros, up 2.4% over 2014, with 2015 operating profit up 3.2% to 10.7 billion euros.**





PROACTIVELY PREVENTING CUSTOMER DEFLECTIONS IN BANKING



The Challenge

Like many companies in many industries, **Bank of America has struggled to find effective ways of closing the back door, that is to say of identifying customers who may be on the verge of defecting to a competitor, and intervening to retain them.** The company turned to its data science team to explore new strategies for stemming such defections.



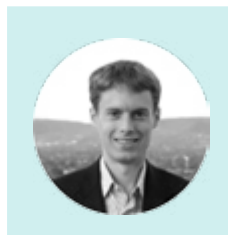
The Solution

The data scientists used transaction histories **to develop new behavioral models of mortgage customers and cardholders who are at risk of switching.** The scientists then worked with colleagues to develop a recommendation system based on these models that could push retention offers to at-risk clients at their next contact with the bank, whether online, in a branch or via a call center.

Results



While the KPIs for this transformation are private, Bank of America was well pleased with the results of the project and other data science initiatives, including the use of algorithms and big data **in enhancing their risk management capabilities, with reports suggesting that they reduced their loan default calculation time by around 95%.** Accordingly, Bank of America has expanded and strengthened its data science capabilities **through a consolidated analytics group with matrixed reporting lines to multiple business functions and units.**



HOW QUANTUM-BASED SOLUTIONS ARE GOING TO CHANGE BANKING AND INSURANCE

by John Kelly, Director of Analytics at QxBranch

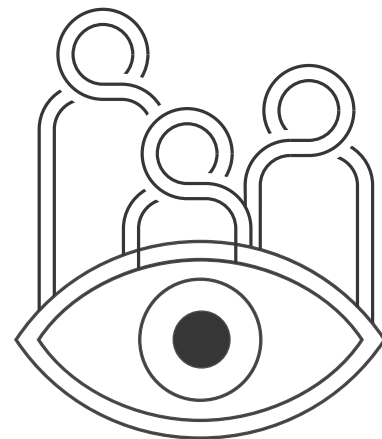
Data Science and Finance: a Challenging Relationship

Data science is revolutionizing banks and financial institutions. The emergence of machine learning as the dominant approach to analyzing data represents a relatively rapid enhancement in capability and in the skills base and management structures required to implement cutting-edge solutions. Finance firms are seeking external help to identify and exploit new opportunities.

The sensitivity of much of the data, the regulatory requirements, the sheer volume of high resolution information, and the criticality of the systems makes this area among the most challenging in which to operate. At QxBranch, our team prides itself in delivering state-of-the-art solutions that navigate this complexity seamlessly.

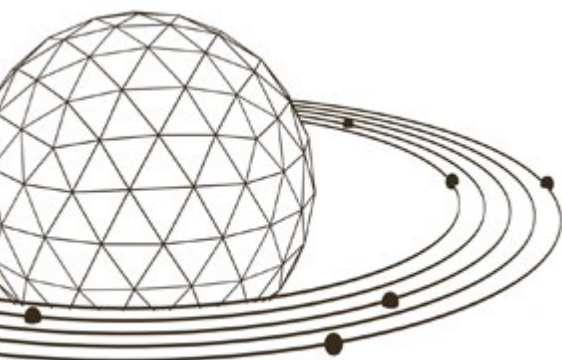
How Big Data Help Banks Know their Customers Better

As you might expect, the finance industry benefits from advanced analytics when it comes to risk management, trading, and economic analysis. Advanced analytics is also delivering outsized value in monitoring behavior for fraudulent, malicious, or otherwise illegal transactions. As in other industries, advanced data analytics applied to customers can demonstrate detailed insight — defining and profiling banking customers, illuminating behavior, and understanding their ‘journey’ through products (credit cards, home purchases, investments, etc.). This helps banks customize products to their clients’ needs, offering better service and optimizing customer lifetime value and retention.



Understanding Risk with Machine Learning

QxBranch sees significant potential for applying advanced data analytics to insurance products, a major area of activity for the company. Advanced models are particularly well suited to risk areas evolving dynamically that do not necessarily behave according to the same principles over time - such as the case with cyber insurance. We develop some of the most complex solutions within a straightforward, end-to-end analytics framework - Dataiku.



Quantum-Based Solutions: The Next Step

QxBranch is a leader in developing quantum computing software. Quantum computers will revolutionize every area of the economy in the coming decades and many of its initial applications will be based on machine learning.

QxBranch works with leading firms in finance (and other verticals) to characterize how the technology will affect their business and competitive landscape.

Dataiku will integrate into quantum software seamlessly. Quantum calculations need to run on pre-processed and cleaned data. For such computations, we adopt Dataiku to ingest, clean, and prepare the data at hand. From there, we input a custom python module that calls a quantum function. The calculations run on the quantum computer and probabilistic results are then re-injected into Dataiku for analysis and visualization.

QxBranch, an advanced data analytics firm that operates globally, applies expertise in systems engineering to develop complex, robust solutions to data-focused problems for investment banks, hedge funds, and insurance firms.

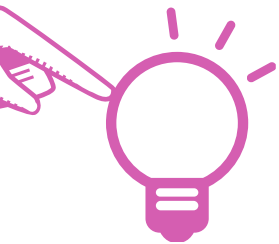


DISCOVERING NEW CUSTOMER SEGMENTS IN BANKING



The Challenge

A large retail banking group felt like the segmentation they used to target sales and marketing campaigns to existing customers could be improved, but **their analysts seemed to have refined customer segmentation as much as was possible using the bank's CRM systems.**



The Solution

The company decided to try a strategy similar to the one Bank of America deployed to help close the back door: **mining their large archives of transactional data to develop new models of customer behavior.** Specifically, the insurer combined transaction data and CRM data and used mining algorithms to surface previously undetected correlations between behaviors and profiles.



Results

The new segments identified as a result of these correlations are now being used by marketing and sales **to guide offers and promotional campaigns, with a three-fold improvement in conversion rates.** Pleased with the results, the company has now tasked their data scientists with analyzing weblogs and other indicators of online behavior **to optimize the digital and omni-channel customer experience.**

AUTOMATING LIFE-EVENT MARKETING IN INSURANCE



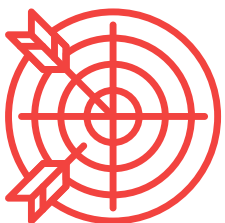
The Challenge

A larger insurance company found it challenging to identify when major events like the birth of a child, a marriage or divorce, a car purchase, the sale of a home, etc., were occurring so they could offer customers and prospects products tailored to these events in a timely manner.



The Solution

The company's data science team combined CRM data, contracts data, weblogs and social media data, and analyzed the aggregate data to develop predictive models of when these events were occurring or likely to occur, and enabled this intelligence to be transmitted to sales and marketing.



Results

The project has delivered an enhanced ability to detect and target customers based on life events, enabling the development of new event-themed, personalized campaigns that can be automatically triggered in real-time. The company is now expanding their data science projects to include detection of fraudulent claims and the development of new personalized products and services based on individual customer behavior.



NEW EFFICIENCY AND ACCURACY IN DETECTING FRAUDULENT INSURANCE CLAIMS



The Challenge

A large provider of supplemental insurance wanted to find more effective ways of detecting **potentially fraudulent claims for optical care, which they estimated cost them 15€ million annually**. They only had sufficient resources to follow up **on 10% of the claims identified as potentially fraudulent**, so they wanted to find a way to ensure to the highest possible probability that the claims referred for investigation **were indeed those most likely to be fraudulent**.



The Solution

The insurer's data science team combined siloed internal data on prescriptions, patients and care providers and tested numerous fraud detection algorithms against this aggregate data. **Having found a model that proved reliable, they set up a real-time API based on that model to guide the automatic routing of claims within the claims management systems based on their likely validity.**



Results

The company reports that the new system of identifying potential fraud has proven itself **to be three times more effective at fraud detection than their legacy approach**, and they anticipate even better performance as the model is continuously improved through the flow of real-time data.



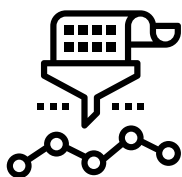
BlueDME

3 CHALLENGES TO ADDRESS IN BANKING/INSURANCE DATA PROJECTS

by Julien Cabot, CEO at BlueDME

«Since 2011, I have had the opportunity to participate in a number of Data Lab projects in banking and insurance sectors. These years of experience have taught me that, in order to succeed, it is necessary to address three major challenges that go beyond the simple preparation of a Data Lab project.»

1. Accessing and qualifying data are key factors of success since, without data, no approach is possible, however brilliant this approach may be. **Moreover, 60% to 70% of the effort that goes into a project is linked to the qualification and preparation of data**, necessary tasks yet ones which add little value. Given the hourly cost of Data Scientists, effectiveness and productivity in the processes of collecting, researching and preparing data are fundamental. Any tool and approach that helps to accelerate this phase quickly proves worthwhile.



2. The capacity to integrate results from Data Science work into operational processes, whether in the form of pre-calculated indicators or prediction models that are usable in real time is essential. In effect, the analytical study leads to new transformations of data and “mathematical objects” – in the form of prediction models – that operational information systems are still able to integrate. However, it is often necessary to look to the study’s generation of real value.



3. Measuring the return on investment of Data Science projects is important. Owing to the exploratory nature of the projects, the exact estimation of the Return on Investment unit of each project is particularly complex. An approach that draws on a wide range of projects enables the projects that generate an important ROI to finance experiments that procure a more limited result, as is the case with Private Equity investment strategies.



Julien Cabot



In banking and insurance, innovative approaches are already being implemented in order to address these three challenges.



IDEA 1: A SEARCH ENGINE OF DATA DEDICATED TO PROFESSIONAL TEAMS



A search engine for data dedicated to professional teams in actuary, to those responsible for statistical studies, and to risk teams working with a Data Lake as well as alongside a Data Lab has many benefits. It facilitates the realization of a catalogue of internal and external data, the exploration of this data in a way adapted to different tasks, and the sharing of the most useful data within the banking-insurance field. The underlying concept of Blue DME's Data Exchange platform is the development of a collaborative approach to using Data Lake data, enriched by a Data Lab, between two professional domains.

IDEA 2: REPRESENTATION THROUGH WEB SERVICES FROM NEW GENERATIONS OF PREDICTIVE MODELS

Representation through Web Services from new generations of predictive models – Random Forest, Gradient Boosting and more – thanks to modern solutions in Data Science such as **Dataiku DSS v3 is boosting efficiency and results**. The days of manually configuring logistic regression are coming to an end. The direct representation of binary prediction models enables the development of these advanced models to be simplified, whilst, above all, enhancing their performance.



IDEA 3: LOGICS OF SUCCESSFUL INTERNAL AND EXTERNAL MONETISATION.

The calculation of ROI produced by the result rendered by the predictive model in relation to the previous situation is limited to internal optimizations. The idea of developing new services based on data is becoming more and more important, making new high-margin revenues possible.

In banking and insurance, Data Lab projects are often the first step in a larger process, transforming the organization into a “Data-driven Business” model that goes much further than client awareness or risks.



AS CAN BE SEEN..

While Mircea Mihaescu is right that traditional banks and insurers lag behind their GAFA and fintech counterparts in exploiting big data and algorithms, **there are many leaders who are already taking the strides needed to close the algorithm gap**, and they are often doing so by leveraging one of their most valuable assets: their deep historical archives of customer data.

They are also leveraging their base of professionals with exceptional mathematical and statistical knowledge and industry expertise who can pivot with relative ease to using big data tools and techniques for advanced analytics, people like Christine Hofbeck, Vice President and actuary at Prudential, who has used her actuarial skills to segue into predictive analytics. As she has stated, the pivot takes a bit of effort, but the opportunities are vast:

“Price optimization. Lifetime value and retention models. Price elasticity. Risk selection optimization. Predictive underwriting. Enrollment optimization. Target marketing. Understanding drivers of policyholder behavior. Propensity to buy. Lapse and churn analysis. Likelihood to bind. Understanding market changes on customer behavior. Market segmentation. Broker segmentation. Finding hidden pockets of the population to develop niche products... I could keep going—the opportunities are endless.”¹¹

Getting on the right path to capitalize on these opportunities just takes pulling together the right combination of people, data, tools and processes.



PART THREE: TAKING ACTION

27 The Right People

A Chronic Shortage of Data Scientists

The In-House Strategy: Meeting the Need with Internal Talent

Developing In-House Talent

29 *Interlude : The Digital Era and Data Science: What does it spell for the Future of the Actuary Profession*

31 The Right Data

Reaping Value from ‘Digital Exhaust’

Look Beyond Conventional Sources

Take Advantage of Open Data

32 The Right Technology & Process

Accelerate and Secure Essential Workflows

Adopt a Platform Approach, and Choose Carefully

Establish Good Analytics Governance

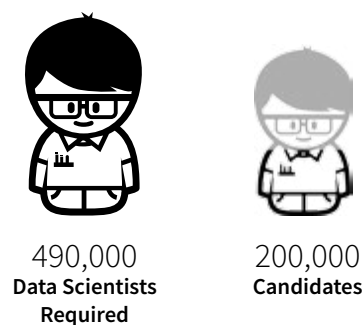
Think in Terms of Data Science Products and KPIs

A - THE RIGHT PEOPLE

1 • A CHRONIC SHORTAGE OF DATA SCIENTISTS

The ideal data scientist possesses advanced mathematical or statistical knowledge, solid computer programming skills, a knowledge of big data processing and storage tools and techniques, and subject matter expertise in the domain in which they are working.

The number of professionals who possess this diverse set of skills, knowledge and experience is at present small, **much smaller than the market demand**. As predicted by McKinsey & Company, by 2018, **the U.S. alone will see a demand for 490,000 data scientists against a pool of fewer than 200,000 candidates**.¹² While many universities have stepped up to fill that void through new data science programs, a chronic shortage will remain for some time.



2 • THE IN-HOUSE STRATEGY: MEETING THE NEED WITH INTERNAL TALENT

This shortage has led some **companies to adopt a strategy of helping existing talent**, like Christine Hofbeck, develop data science competence. For instance, Cisco Systems worked with two universities **to create distance learning education and certification programs in data science**, and more than 200 of their employees to-date have been trained and certified as data scientists.

In addition to simply helping to meet the need for data science talent, **this in-house approach offers the added benefit of having data scientists ready to go to work pre-equipped** with essential corporate knowledge and subject matter expertise. Fortunately, as the Cisco example shows, there are many resources available now to help companies succeed with in-house strategies. These include:



Numerous online, part-time and on-demand data science training programs offered through universities,



On-site training delivered by professional services companies and software providers,



Data science boot camps, and



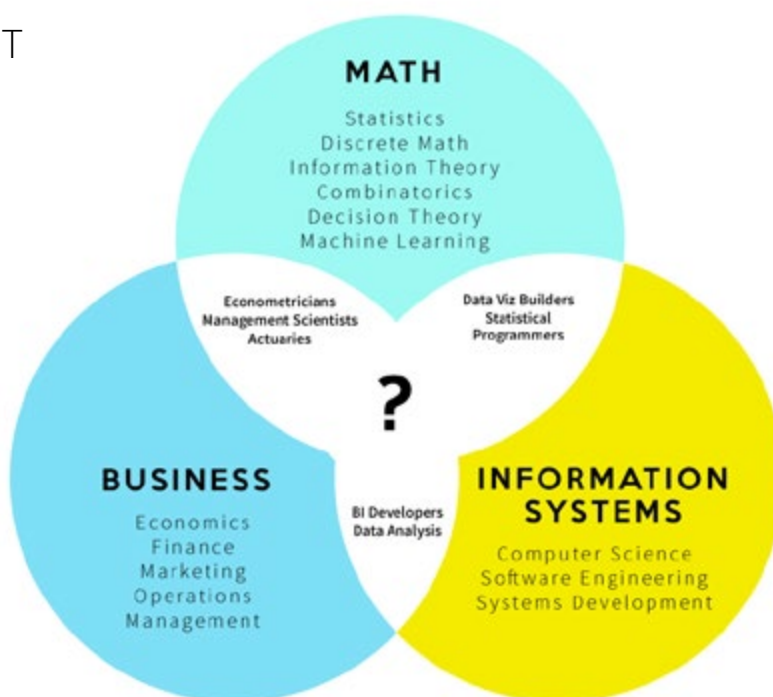
Development programs through professional associations, such as the Casualty Actuarial Society Institute, which recently added a certification in data science and predictive analytics for actuaries.

Banks and insurers are well-positioned to take advantage of resources like these and succeed with an in-house strategy.

3 • DEVELOPING IN-HOUSE TALENT

IDENTIFY CANDIDATES

As noted, most banks and insurers are already staffed with analysts and scientists with mathematical and statistical expertise, including many with Masters or PhD degrees. These professionals include actuaries, claims analysts, risk analysts, underwriters, economists, equity analysts, market analysts and many others **who could easily leverage their advanced mathematical and statistical knowledge and quantitative skills to understand and manipulate new types of machine learning algorithms**, and with relative ease extend their existing programming skills (VBA, SQL, SAS/JMP, etc.) with languages like Python or R that can help them execute these algorithms.



DO YOU HAVE IN-HOUSE STAFF WITH THESE SKILLS?

BUILD ON THEIR CORE MATHEMATICAL AND STATISTICAL SKILLS

The distance, for example, between using Excel and VBA for generalized linear modeling (GLM) in credit scoring, and using machine learning algorithms to enhance GLM results, or even using R and Support Vector Machine or Neural Network algorithms as alternative methods, is not that great, though the distance is relatively large between someone running a packaged algorithm they don't really understand and someone who can decipher the mathematical and statistical operations within an algorithm and can optimize or adapt it as needed – and do so in the context of their deep industry experience.

DEVELOP BIG DATA IT SKILLS

In addition to these core quantitative skills, most banking and insurance analysts are knowledgeable about information systems and often acquire and prepare data from diverse sources on their own, though they may be unfamiliar with big data management technologies. However, this, again, they can learn, and they are fortunate that big data technology has now matured to the point of being accessible to those with common IS skills. Specifically, most big data systems today feature interfaces and tools that make connecting to, transforming, querying, and visualizing data far simpler than was the case with first-generation big data technologies.

While acquiring these new or extended skills may be achieved with relative ease, those who participate should have a natural penchant for lifelong learning, for if there is one constant in data science, it is constant change, as is evidenced by the rapid, continuously evolving procession of big data management technologies, programming languages and algorithmic strategies employed within data science.

For the professionals who take on this challenge and commit themselves to the discipline, the rewards are many, including the ability to differentiate themselves professionally, attract premium compensation, and advance their careers. It can also relieve them of much of the drudge work associated with manual processes and calculations, freeing them to exercise more of their curiosity, creativity and judgment.



THE DIGITAL ERA AND DATA SCIENCE: WHAT DOES IT SPELL FOR THE FUTURE OF THE ACTUARY PROFESSION?

by David Dubois, President of the Institute of Actuaries



Actuaries: who are they?

Actuaries are the professionals who evaluate, manage and create models of risk in its economic, insurance-related, financial and social variations. They belong to professional associations, like the Institute of Actuaries, and share a broad foundation of advanced training in mathematics and statistics. This is in accordance with an international core syllabus established by the AAI (International Actuary Association), which is adapted for European associations by the AAE (European Actuary Association) and substituted by the associations of each country in order to take market specificities into account.



The digital age starts to rattle the actuarial sectors

Actuaries are strongly present within insurance organizations (accounting for 49% of the membership of the Actuary Institute). Equally, they work as business advisors (18%), in banking (11%) and in reinsurance firms (10%). Faced with the inherent challenges brought by the digital revolution, these traditional areas of activity are undergoing a deep transformation, driving a progressive re-evaluation of how value chains are organized. The digital age is also repositioning the client and shifting the balance of power between client and supplier. By putting the client at the center of the production process, insurers and bankers alike are faced with strategically important stakes. They must successfully transform their organizations in order to make them more agile in a competitive and disruptive environment where each innovation generates new data that new competitors can exploit in order to quickly invent a new business model.

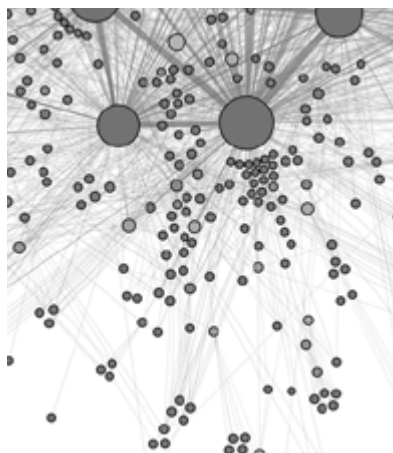
Data science starts to enrich the work of actuaries

Actuarial science, like data science, makes recourse to collected data, whose quality and relevance as a base for evaluating risks must be validated.

1 Owing to bias, input errors or incomplete data, the work of restoring and correcting this data presents a time-consuming and tricky step before the stages of creating models and evaluation. Faced with outdated information systems, the manner in which data is collected and consolidated directly influences the ability to use this data to create results. Data science is expanding the scope of data that can be taken into account.

2 Moreover, in an environment where risks are becoming increasingly complex and entwined, it is necessary to rethink the notion of risk and be wary of the reflexes and other casual links that center on experience and practice. The simple projection of statistical trends is insufficient.





Actuaries bring professional expertise to data science

The emergence of new business models and the influx of big data undoubtedly present new opportunities for the actuary profession. Data science presents an additional opportunity for the actuary discipline, and bestows it with a double expertise, that of “letting the figures do the talking,” and of professional expertise. Do the correlations revealed by algorithms fall under knowledge or simply Machine Learning? Is this aspect of predictability random or does it result from a strong law between two or several variables? How do we concretely use the results rendered from algorithmic processes?

As actuaries, we consider that there is not only confusion to be found in correlations, but an essential need for professional skills to complement new data technology.

We see data science as a powerful tool for making sense of many-sided and multi-supported data, which, until recently, we were not able to put to use due to the lack of powerful computer software.

It is therefore a question of operationally adapting to the lessons that we can learn from extracted data and, faced with this, exercise a critical eye in order to situate the analysis in a professional framework.

Banking and insurance will be revolutionized by the digital age, data science and IoT

Banking, as with insurance, is evolving in a complex regulatory environment. These regulatory and institutional domains are far from closed, and often succeed in implementing difficult innovations. The operational implementation of identified correlations often involves a complex process and necessitates the professional expertise possessed by actuaries.

1 - The digital age provides actuaries with a tool to fully express their creativity and develop their expertise at all stages of value chains: setting prices, accounting, administrative management, asset management, fraud detection, etc.

2 - Data science also opens up new areas relating to the segmentation and evaluation of revisited risks, those rooted in the intrinsic characteristics of the person being insured as well as on his/her behavior and lifestyle.

3 - Moreover, the fast diffusion and adoption of connected objects also supports the possibility of developing solutions for risk prevention and appreciation. Management teams must reinforce their flexibility and agility in order to provide the person being insured with an improved service that incorporates all the advantages brought by accumulated data.



No disruption without expert knowledge: how actuaries are essential to data science



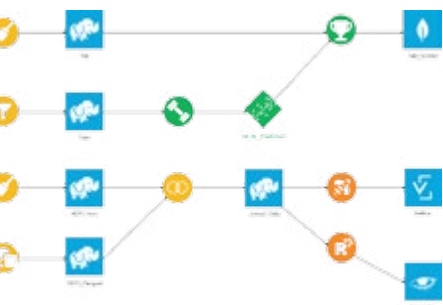
We are convinced that a high demand for professional knowledge in all its dimensions will result from this disruption. Identifying correlations by applying algorithms to big data is essential, but correlation does not signal causality. It is necessary that the execution goes through a skilful process and does not confuse information and decision, as tempting as it is to believe that algorithms reveal everything in an objective way.

This is why the Institute of Actuaries has mobilized itself, reinforcing the skills of its members by embracing big data, establishing the first training program in Data Science for actuaries in 2015 and developing a series of conferences about digital transformation within insurance and finance.

When dealing with data, it is not just about knowing the technicalities of the algorithms. Data science and knowledge of the technical scene provides actuaries with an opportunity to work beyond their traditional sphere of actuarial functions and risk management, and to excel within the transformation of their own profession. This is why the Institute of Actuaries has mobilized itself, reinforcing the skills of its members by embracing big data, establishing the first training program in Data Science for actuaries in 2015 and developing a series of conferences about digital transformation within insurance and finance.



B - THE RIGHT DATA



1 • REAPING VALUE FROM ‘DIGITAL EXHAUST’

While banks and insurers have a potential advantage over newer entrants in the historical customer data they possess, they need the proper tools to glean actionable insights from that data, and to be able to enrich it with external data. As fintech companies have shown, there is a wealth of digital “data exhaust” from social media and multimedia, smartphones, computers, and IoT and other consumer and industrial devices than banks and insurers can use and in certain cases already are using as a rich source for behavioral insights.

2 • LOOK BEYOND CONVENTIONAL SOURCES

For instance, if you are developing a fraud analytics tool for home insurance claims, you may want to integrate internal data from claims management systems, document management systems and third party statement with external data like weather, credit, and financial information to validate a claimant’s identity and reliability and the likelihood of a particular incident occurring at a particular place at a particular time (while preparing to one day use telematic data from smart home systems and blockchain authenticated transactional data to resolve much of the current ambiguity in claims).

An Endless Diversity of Data to Enrich Analytics

Customer Transactions, CRM Data, Weblogs, Clickstreams, Social-Media Profiles, Networks and Activities, Smartphone Usage, Telematics from Connected Devices, Demographic Data, Census Data, Public Health and Mortality Data, Weather Data, Climate Change Modeling Data, CCTV, Still Image and POS Data, Text Communications (Email, Text Messages), Call recordings, Political Donor Data, Criminal Justice System Data...

3 • TAKE ADVANTAGE OF OPEN DATA

Some existing sources of useful external data will be expensive to attain, but a tremendous amount of previously unavailable external data is available now as open public-sector data.

For example, the US and UK governments and the European Union have recently launched “open data” portals to make massive amounts of government statistics available, including for example health, education, worker-safety, and energy data. With much better access to third-party data from a wide variety of sources, and a unique internal trove of historical data, and the right talent, insurers and bankers are poised to gain a GAFA-style advantage in posing new questions about their customers and gaining a deeper understanding of their needs, desires, and habits - in the context of events - and in using those insights to craft experiences and offerings that are better tailored for specific customer segments, even down to the individual ‘market of one’ level.

However, it’s very important to ensure that the data used, and the way it is used, is legal and ethical, and that the quality is good enough to produce reliable analytical results. For issues such as these, it necessary to have the right technologies and processes in place.





C - THE RIGHT TECHNOLOGY & PROCESS

Now that you have the right data sources and the right people, it is important to equip your data science team with the right tools and processes to succeed. Here are four essential steps to take to achieve this and maximize the return on your data science investment:

- ☐ 1 Accelerate and Secure Essential Workflows
- ☐ 2 Adopt a Platform Approach, and Choose Wisely
- ☐ 3 Establish Good Analytics Governance
- ☐ 4 Think in Terms of Data Science Products and KPIs

1 • ACCELERATE AND SECURE ESSENTIAL WORKFLOWS

To succeed, the three essential workflows for data scientists need to be streamlined, automated, and flexible. Data scientists should be able to connect to and prepare data with ease from any data source, to execute modeling and predictions with their preferred language using the types of algorithms best suited to their needs, and to easily produce actionable results that can be consumed by diverse internal and external systems.



A - Data Preparation
& Enrichment



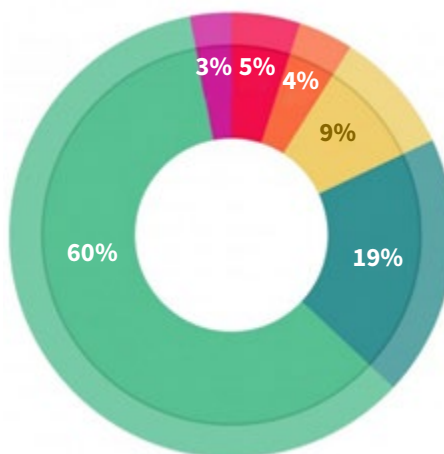
B - Modeling
& Prediction



C - Execution
& Deployment

A - Data Preparation & Enrichment

The data collection and preparation phase is the most labor-intensive one, consuming on average between 60-80% of a data scientist's time. It's critical therefore to select a tool that can automate or at least speed the extract, transform and load (ETL) workflows associated with data preparation.



What data scientists spend the most time doing

- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%
- Building training sets: 3%

The tool should feature:

- ☐ 1 Native, wizard-based connectors for a wide variety of data sources
- ☐ 2 Visual processors for code-free data review and cleansing
- ☐ 3 The ability to filter and search data as easily as in Excel
- ☐ 4 The ability to easily execute mass transformations, including built-in processors for common transformations and macros for custom processing
- ☐ 5 The ability to easily adapt transformation as data evolves

This last feature is essential as the data within the world of advanced analytics is never static, and technologies and approaches that do not address this changeability will always eventually be rendered irrelevant.

80%

Data preparation accounts for about 80% of the work of data scientists (and 76% view this as the least enjoyable part of their work).

B - Modeling and Prediction

As in the data preparation phase, openness, automation and flexibility are essential to workflow support for the modeling and prediction phase. An effective tool will therefore offer features such as:

- ☐ 1 Step-by-step visual guides to speed model-building
- ☐ 2 Native support for common machine learning libraries, with API access to specialized ones
- ☐ 3 The ability to code directly in common languages like Python and R
- ☐ 4 Visual feedback to assess performance, compare results, and optimize models
- ☐ 5 White box transparency and reporting to help analysts understand various interactions and the impact of choices made

C - Execution & Deployment

As the end goal of data science is to produce actionable, usable results, this last workflow stage should:

- ☐ 1 Automate the process of producing not just data visualizations, but consumable, standards-based APIs
- ☐ 2 Enable data scientists to package an entire workflow, including data and models, as a single deployable and reproducible package
- ☐ 3 Support versioning and rollback within a captured workflow

2 • ADOPT A PLATFORM APPROACH, AND CHOOSE CAREFULLY

Being able to rollback to a prior version, or easily produce deployable and reproducible packages, is almost impossible without using a single data science platform for all project phases from data collection to modeling to production.



A - Maximum Efficiency



B - Support for Collaboration



C - Knowledge Capture

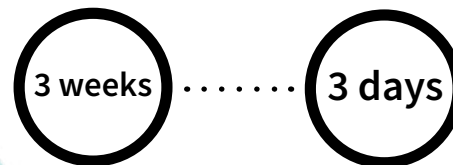


D - Openness is Key

A - Maximum Efficiency

But adopting a single platform rather than using ad hoc tools has many other benefits in addition to rollback and deployment. A significant one is the elimination of lost time, including time lost in switching between various data sources, systems and tools, and time lost in getting up to speed on prior work as scientists move between projects (as is common).

One large insurance company saw **the average onboarding time for a data science project** drop from **3 weeks to 3 days** once they switched from ad hoc development to the use of a unified data science platform.



B - Support for Collaboration

A platform strategy also enables teams to work collaboratively, with for instance, a data engineer collecting and preparing data and an analyst running the models, or a junior data scientist working alongside and learning from a more experienced peer. This is essential as data science teams are typically diverse, with each member having unique strengths and weaknesses.

C - Knowledge Capture

A platform strategy also provides a 'lifecycle' approach that enables an essential knowledge base of actionable intellectual property assets to be built, similar to the way product lifecycle management (PLM) platforms evolved in manufacturing to avoid reinventing the wheel by capturing, storing and reusing knowledge from design conception through manufacturing and on to the final disposition of goods.

D - Openness is Key

However, a platform approach will ultimately prove constricting in the fast-evolving world of data science unless it is open. Openness and vendor/technology agnosticism means equipping data scientists with a framework for collecting, analyzing, and visualizing data that will remain as constant as possible even as underlying data sources, programming languages, big data storage and management technologies and algorithms inevitably evolve.

3 • ESTABLISH GOOD ANALYTICS GOVERNANCE

Another advantage of a platform approach is that it makes it easier to put into place strong 'analytics governance.' As in PLM, good governance is critical to ensuring the validity of results and creating a 'safe data space' where privacy, confidentiality and security are rigorously protected. Governance in data science is important to producing high-quality, reproducible analytical projects, but it is something with which many industries struggle.

Banks and insurers, however, possess long and rigorous experience in developing organizational structures, governance frameworks and rules, which can give them an advantage in being able to scale and sustain the initial competitive advantage their data science initiatives can bring. However, regardless of the specific technical framework or governance strategies adopted, banks and insurance companies should aim to build their own internal data analytics capabilities, assets and IP, without relying on a third party provider, though the initial transformation around tools, processes, people and data is one that can be orchestrated and enabled by partners who can assist in laying the groundwork for data access, collaboration, and knowledge capitalization, and assist in deploying the tools and infrastructure needed to ensure sustainable data science competency.

4 • THINK IN TERMS OF DATA SCIENCE PRODUCTS AND KPIS

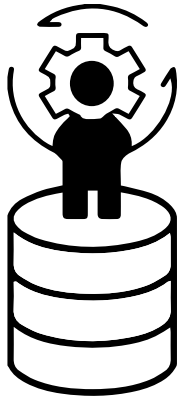
Finally, companies need to be diligent in designing projects that seek to solve a clear business problem, even if some resources are set aside for pure data exploration. This can help avoid one of the prime pitfalls in data science initiatives: becoming mired in too many experimental projects, with exploratory modeling never passing over into the generation of usable data science ‘products’.

In the end, predictions and prescriptions are only useful if they are available (for example, through APIs) for integration into external applications, like customer self-service web applications, mobile applications, or internal executive decision-making systems; or, at a minimum, are presented in visual form to guide stakeholder decision-making.

Going a step further in identifying and tracking KPIs to assess the business usage and value of data science projects and products is a final, invaluable means of ensuring that data science projects are well-grounded, and to correct course as needed to be sure that the transformative power of algorithm-driven innovation is harnessed to maximum effect.



CONCLUSION



Equipped with the right people, processes and tools, traditional banks and insurance companies can not only avoid the fate of becoming the back-end plumbing for GAFA and fintech challengers, they can appropriate the advantages of these newcomers and merge them with their own to become the new marketplace innovators of the 21st century.

Banks and insurance companies have the people they need. They have unique data assets and a unique physical presence to build on. They have an advantage in organization, governance and trust. All that's left is to realize that the age of big data and analytics is upon us, and to seize the opportunity to thrive in the new age through the mastery of data science.

- If industry veterans do so, they can both flourish in and shape the future of banking and insurance and what might that future look like?
- Will we linger over a latté in a cozy bank café while both human and robot agents equipped with artificial intelligence ply us with prescient advice?
- Will biometric authentication be the norm, with a touch of a finger or a glance of an eye making wallets, clicks and cards obsolete?
- Will we exchange virtual currencies, payments and policies across a secure global blockchain backbone?
- Will the lines between peer, agent, customer, broker, banker, underwriter, etc. disappear in a world where people and objects are connected to one another in a seamless Internet of Everything?

It's hard to say, but it's a future the banking and insurance industries are well-positioned to help write. _____



End Notes

- 1 2015 Makovsky Wall Street Reputation Study, <http://blog.ebiquity.com/2015/06/2015-makovsky-wall-street-reputation-study-research-conducted-by-ebiquity>
- 2 Accenture analysis on CB Insights data, in “ Fintech and the evolving landscape: landing points for the industry,” Accenture, 2016, http://www.fintechinnovationlablondon.co.uk/pdf/Fintech_Evolving_Landscape_2016.pdf
- 3 Ibid.
- 4 Scott WALCHEK
<http://www.the-digital-insurer.com/dia/trov-worlds-first-demand-insurance-things/>
- 5 40% improvement in default rates
<http://fortune.com/2015/12/01/tech-loans-credit-affirm-zest/>
- 6 <http://fintechranking.com/2016/03/17/big-data-and-online-scoring-fintech-and-beyond/>
- 7 Ibid, Makovsky, 1.
- 8 <http://banknxt.com/49972/new-age-of-algorithms/>
- 9 See «Banks balance technology with brick-and-mortar,» Rochester Business Journal, July 22, 2016, https://www.cnbank.com/uploadedFiles/CNB_Site_Home/Your_Bank/News/Articles/2016/RBJ_07%2022%2016.pdf
- 10 «Customer Loyalty Statistics: 2015 Edition,» Access Blog, Brandon Carter, on Dec 31, 2015, <http://blog.accessdevelopment.com/customer-loyalty-statistics-2015-edition#bank>
- 11 <http://theactuarymagazine.org/breaking-the-mold/>
- 12 <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>





ABOUT DATAIKU

Dataiku strives to be the acknowledged advanced analytics leader and preferred software solution in helping organizations succeed in the world's rapidly evolving data-driven business ecosystem. Guided by the belief that true innovation comes from the effective combination of diversity of cultures, of mindsets, and of technologies, Dataiku's purpose is to enable all enterprises to imagine and deliver the data innovations of tomorrow.

ABOUT DATAIKU DSS (DATA SCIENCE STUDIO)

Dataiku DSS is a collaborative data science software platform that enables teams to explore, prototype, build, and deliver their own data products more efficiently. It is an open platform designed to accommodate rapidly evolving programming languages, big data storage and management technologies and machine learning techniques, and is conceived to accommodate the needs and preferences of both beginning analysts and expert data scientists. It also uniquely support:

Collaboration

Collaboration features make it easy to work as a team on ambitious data projects, to share knowledge amongst team members and to onboard new users much faster. You can add documentation, information or comments on all DSS objects.

Reproducibility

Every action in the system is versioned and logged through an integrated Git repository. Follow each action from the timeline in the interface, with easy rollback to previous versions.

Production Deployment

DSS lets you package a whole workflow as a single deployable and reproducible package. Automate your deployments as part of a larger production strategy. Run all your data scenarios using our REST API.

Governance and Security

DSS helps you create clearly defined projects and make sure your data is organized. And with fine grained access rights, your data is available only to the right persons.

Try Dataiku DSS for free by visiting www.dataiku.com/try

