datacamp

Digital Transformation in Finance: Upskilling for a Data-Driven Age

1 The state of data science and machine learning in finance

The financial data deluge

Over the past decade, big data and digital technologies have disrupted industries and consumer behavior alike. IDC and Statista estimate that the volume of data generated yearly rose from two zettabytes in 2010 to 59 zettabytes in 2020, marking a thirtyfold increase in data generated in the past 10 years alone (Statista). This data deluge is only expected to grow, with projections predicting 149 zettabytes produced yearly by 2024.

> A zettabyte is equivalent to about 250 billion DVDs.
>
> The Guardian

While various industries are vying to take advantage of the data deluge with business intelligence, data science, and machine learning, the financial services industry is best equipped to benefit from big data. Data is at the heart of the financial services industry—across retail banking, investment banking, and insurance. Financial services organizations produce and store data on their customer transactions, detailed customer profiles through compliance processes, insurance claims, stock market exchanges, and more. The amount of data generated is astounding: The New York Stock Exchange alone produces one terabyte of trade data daily (Investopedia).

We've already seen fintech startups take advantage of shifting consumer behavior and the financial industry's data deluge. Digital banks such as N26, Revolut, and Monzo abandoned the brick-and-mortar model and opted for a purely digital banking experience, relying on data to improve user experience and automate workflows (Revolut). Klarna, Europe's most giant fintech unicorn, provides interest-free installment options with automated approval or rejection using machine learning (CNBC). The data deluge has not only opened up space for disruptive, innovative services—it's opened the door for data-enabled digital transformation across the industry.

Data-enabled digital transformation

Disruptive digital-first startups across all industries have prompted many incumbents to invest heavily in digital transformation. The financial services industry is no exception. An Accenture and Oxford study in 2018 found that 87% of retail banking executives have developed a long-term plan for technology investment and digital transformation (Accenture). This is especially true in the COVID-19 economy, which has moved consumer purchasing online and accelerated digital transformation programs across all industries.

> **"We've seen two years' worth of digital transformation in two months. From remote teamwork and learning to sales and customer service, to critical cloud infrastructure and security."**
>
> **SATYA NADELLA, CEO OF MICROSOFT (APRIL 30, 2020)**

This acceleration is exceptionally pressing in the financial services industry. A recent study from the Economist Intelligence Unit cites that 45% of banking executives believe building a "true digital ecosystem" is the best strategic response to the pandemic. In the same survey, 66% of respondents believe that new technologies such as machine learning and artificial intelligence will bring the most significant impact to the banking industry by 2025.

Taking an example from the ground, the urgency of using contact-less financial tools ushered an 84% increase in Citibank's daily mobile check deposits and a tenfold increase in activity on Apple pay (Forbes). This has prompted Jane Fraser, president of Citigroup and CEO of its consumer bank, to declare, "Banking has changed irrevocably as a result of the pandemic. The pivot to digital has been supercharged. [...] We believe we have the model of the future—a light branch footprint, seamless digital capabilities, and a network of partners that expand our reach to hundreds of millions of customers."

The success of such digital transformation programs pivots on the seamless integration of digital technologies with data-driven insights. Without a data-enabled digital transformation, financial services organizations wouldn't be able to examine the data produced by digitization, iterate on their digital services and processes, and automate services and workflows using machine learning.

Leaders are charging ahead

The good news is that leaders in this space are already taking their data science and machine learning initiatives off the ground. JPMorgan Chase developed "COiN"—a machine learning-based contract intelligence tool to automate legal document analysis (Harvard Business School). This has saved JPMorgan Chase about 360,000 hours spent by loan officers yearly performing routine tasks like interpreting loan agreement documents.

Wells Fargo layered machine learning capabilities on top of its mobile banking application. This generated a slew of convenient predictive analytics use cases for consumers, such as personalized financial advice, better financing management, personalized financial insights on consumer behavior, and timely payment reminders based on historical data.

On the insurance side, Liberty Mutual has developed an application that lets automobile owners instantly estimate damages to their vehicles using their smartphone cameras. They're able to leverage thousands of historical car crash images to accurately predict damages and repair costs. AXA UK recently piloted a machine learning solution to help its claim handlers better process automobile claims, and enable a smoother claim journey for customers.

These pilot projects only show that the financial services industry is approaching an inflection point where winners will capitalize on data science and machine learning to set themselves apart from the competition. But although the opportunity landscape is uniquely rich for the financial services industry, so are the challenges in operationalizing data science and machine learning at scale.

The data science challenges facing the financial industry

1. Data Quality and Infrastructure

A significant challenge for any organization trying to operationalize and scale data science and machine learning is enabling a modern, secure, centralized, and discoverable data infrastructure (DataCamp). This is especially true for large financial services organizations, which contain many siloed departments hosting their proprietary data on legacy systems. Disparate and outdated data infrastructure creates misalignment on data governance and data quality and limits the use of rich datasets in analytics and machine learning. It also weakens trust in data as data scientists and analysts do not know what data exists within their organizations.

Many organizations have tried to bridge the data access gap by migrating to cloud storage services. While compliance is a crucial challenge for financial services organizations when migrating to the cloud, many cloud service providers are responding with public, private, or hybrid cloud solutions tailored for financial services (IBM).

For example, the United Kingdom's central bank, the Bank of England, partnered with Cloudera to store their data in a data warehouse, enabling analysts and economists to access the data they need quickly. HSBC recently partnered with AWS to leverage cloud storage, compute, and machine learning services in an effort to modernize data infrastructure.

2. Regulation and Governance

One of the unique challenges the financial services industry faces when operationalizing data science is the regulatory and risk landscape.

For starters, already established compliance requirements touching on risk, culture, resilience and stress testing, operations, and more make the financial services industry one of the most regulated sectors operating today (Dataiku). The regulatory landscape is only becoming increasingly complex with data protection laws like the European Union's General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), and more.

These regulations will further shape financial services organizations' data science and machine learning plans, determining which use cases can or cannot be operationalized. Beyond data protection, the space of machine learning and AI governance is still nascent. Governance models are still being developed to ensure that machine learning models are interpretable, not biased, and compliant with industry standards (Ernst & Young).

3. Skills

Enabling data infrastructure and reaching consensus on regulatory frameworks are essential to driving a data-enabled digital transformation in finance. But the most significant threat to financial services digital transformation prospects is the data fluency skill gap.

In fact, the PwC 2019 yearly banking trends report focused solely on the skill gaps in this space. The report asked over 235 Banking and Capital Markets CEOs about the skill gaps within their organizations. Eighty percent of respondents believe that skill gaps in data fluency and artificial intelligence represent a threat to their organizations' growth prospects. More than 60% of CEOs believe that these skill gaps hurt their ability to innovate and provide better customer experiences.

Merely hiring skilled data workers isn't enough. There's a shortage of talent that understands the complexities of the financial services industry and a shortage of data science talent in general. As demand for data scientists across industries grows and organizations take longer to fill data science positions, it's become nearly impossible to find hundreds or thousands of data professionals with domain expertise in financial services.

> Data fluency is the ability to **understand** data, **communicate** insights from that data, and ultimately to make more **informed** decisions. It's about empowering employees with the skills to drive better **business insights**, faster. It includes a spectrum of data skills, from data literacy to data science.

Beyond hiring, upskilling is the only way forward

What's the best way to address the data fluency skills gap? Digital Banking Report found that almost half of financial services executives believe that upskilling is the best strategy to tackle skill gaps.

Ernst & Young found that developing programs to upskill front line units, risk management functions, and internal audit is not only crucial in driving digital transformation but also for tackling challenges in setting up good data science and machine learning governance models.

Regulators also echo this notion. In a January 2020 report on big data and analytics in banking, the European Banking Authority found that "From the front line to the boardroom, an institution runs a risk if employees do not have a sufficient understanding of the strengths and limitations of the Big Data & Advanced Analytics enabled systems they work with." They stressed this risk by pointing out that "There is currently a shortage of skills, as it is not easy to find human resources with all the necessary knowledge and experience (e.g., in data science, business, IT, statistics)."

The long-term success of data-enabled digital transformation for financial services organizations depends on data fluency upskilling.

2 How DataCamp partners with the finance industry to achieve data science objectives

DataCamp has partnered with hundreds of financial services organizations to help them achieve their data science and data fluency goals. For example, we **partnered with AXA XL** to upskill data scientists and actuaries on natural language processing and predictive analytics. These skills enabled AXA XL to extract essential information from digital documents and automate processes.

Bloomberg used DataCamp as part of a blended learning environment to teach data analysis with Python and empower employees to write data-driven financial news stories.

On the banking side, DataCamp partnered with a major global retail bank to transition their risk analytics department from SAS to Python, reducing dependence on licensed legacy software and focusing on future-proof open-source Python packages like pandas and scikit-learn.

Taking a broader look, the use cases in financial services can fall within three categories: Cost Reduction, Revenue Maximization, and Risk and Governance. The following section contains some use cases for data science and machine learning in the financial services industry and how DataCamp can help.

Cost Reduction

Robotic Process Automation (RPA) is a form of business process automation that combines rule-based systems with machine learning to automate repetitive tasks. Traditionally, robotic process automation tackled repetitive and simple rule-based tasks, such as data entry and payroll, using in-house or vendor-purchased software. Usually, RPA tools are programmed to replicate mouse clicks and triggers to automate repetitive tasks. They typically require traditional programming skills to develop. But when combined with machine learning, financial services organizations can even automate more complex processes that require structured and unstructured data. For example, machine learning-enabled RPA systems can automate complex KYC (know your customer) processes, route customer emails, and more. More importantly, RPA allows for more consistent, quick, and efficient data generation (Venture Beat). This data can be leveraged in business intelligence and machine learning applications.

Commonly used technology

Relevant Tracks

R Programmer

12 courses

Python Programmer

16 courses

Relevant Courses

Data Manipulation with R

Developing R Packages

Cleaning Data in R

Supervised Machine Learning with R: Classification

Python for Spreadsheet Users

Data Manipulation with Pandas

Cleaning Data in Python

Software Engineering for Data Scientists in Python

Supervised Machine Learning with scikit-learn

Chatbots have long been one of the most apparent machine learning and data science use cases across many industries. According to IBM, organizations worldwide spend around $1.3 trillion on 265 billion customer service calls yearly. Financial services organizations comprise a large portion of these since they provide complex products, from insurance plans to bank accounts. Chatbots automate customer service calls through a chat or voice-based conversation with a virtual agent. Using natural language processing techniques like sentiment analysis, topic modeling, and building dialogue systems, financial services organizations can automate simple customer service requests, provide context for live agents, and improve customer experience. An example of this is Bank of America's chatbot, "Erica," accessed from their digital banking app. Erica automates routine customer service requests like providing quotes and placing trades for investment accounts, providing useful account information, managing lost or stolen cards, and more.

Commonly used technology


Relevant Tracks

Natural Language Processing in Python

6 courses

Text Mining with R

4 courses

Relevant Courses

Introduction to Natural Language Processing in R

Sentiment Analysis in R

Topic Modelling in R

Introduction to Natural Language Processing in Python

Sentiment Analysis in Python

Building Chatbots in Python

Business Intelligence combines business analytics, data manipulation, and visualization to help organizations make more data-driven decisions. Business intelligence tools such as Tableau and Power BI offer financial services organizations a non-coding approach to create dashboards that unify disparate data sources, visualize trends, analyze key performance indicators, mine for patterns in large swaths of data, and discover any pain points in processes or services. Moreover, BI dashboards can be built for free with Python and R as they offer a plethora of open-source data visualization and dashboarding packages like Python's Bokeh, and R's Shiny packages to get dashboards quickly off the ground. Oftentimes, all of this is glued together with SQL queries on the back end collecting and organizing relevant data to display. This translates into less time spent on manual reporting (e.g., regulatory reporting), better-enforced data governance, and a clearer view of the organization's data (Zirous).

Commonly used technology

R python SQL Power BI tableau

Relevant Tracks	Relevant Courses
Shiny Fundamentals with R 4 Courses	Building Web Applications with Shiny in R Building Dashboards with shinydashboard
Data Visualization with Python 5 courses	Introduction to Tableau Analyzing data in tableau Introduction to Power BI
Fundamentals of Tableau (coming soon) 4 courses	Interactive Data Visualization with Bokeh Introduction to Data Visualization with Seaborn
SQL for Business Analysts 5 courses	Data Driven Decision Making in SQL Reporting in SQL

Claim triage and prediction are among the most exciting data science and machine learning use cases for the insurance industry. Every year, billions of dollars of auto, home, and health insurance claims are triaged and processed by analysts. However, this is a time-consuming and error-prone process. Analysts manually process claims that range from straightforward, quickly settled claims to complex claims that require thorough analysis. A report published by the American Medical Association estimated that the healthcare industry loses $17 billion yearly from claim processing errors (Carecloud). Using foundational skills in machine learning, topic modeling, and natural language processing, insurance providers can triage claims based on their complexity, enabling optimal workflows and better prioritization. Moreover, depending on complexity, they can also leverage machine learning to provide on-the-spot claim acceptance or rejection. For example, digital insurer Lemonade was able to process and pay a claim in three seconds with zero paperwork, perhaps breaking the world record for the fastest claim payment in human history.

Commonly used technology

R python™

Relevant Tracks

Natural Language Processing in Python

6 courses

Deep Learning for NLP in Python

3 courses

Supervised Machine Learning in R

5 courses

Relevant Courses

Introduction to Natural Language Processing in Python

Advanced NLP with spaCy

Feature Engineering for NLP in Python

Recurrent Neural Networks for Language Modeling in Python

Introduction to Natural Language Processing in R

Text mining with bag-of-words in R

Topic Modelling in R

Document digitization revolves around taking large swaths of paper documents and converting them into searchable, indexable, and analyzable digital documents. The financial services industry has long used paper-based processes to initiate KYC processes, open bank accounts, approve or reject insurance claims, and more. The vast amount of paper-work produced over the years can't be leveraged for practical data science and machine learning use cases since they need to be digitized. Using deep learning-based machine vision techniques, organizations can digitize documents like contracts, forms, and claims—and extract useful information that can be leveraged in other data science and machine learning applications.

Commonly used technology

python™

Relevant Tracks

Deep Learning in Python

5 courses

Image Processing with Python

3 courses

Relevant Courses

Image Processing in Python

Advanced Deep Learning with keras

Image processing with Keras in Python

Revenue Maximization

Customer Segmentation revolves around grouping and categorizing customers into different buckets or “segments” based on common patterns in their purchasing and usage behavior. While this is applicable across all industries, the rich complexity of products and services in the financial services industry makes it prime for customer segmentation. This typically involves using unsupervised machine learning clustering techniques. Organizations can then target customers with more personalized marketing campaigns, recommend and upsell relevant products, further entrench their relationship with customers, and ultimately maximize their revenue.

Commonly used technology


Relevant Tracks

Unsupervised Machine Learning with R

4 courses

Machine Learning Fundamentals with Python

5 courses

Marketing Analytics with Python

7 courses

Marketing Analytics with R

6 courses

Relevant Courses

Unsupervised Learning in R

Cluster Analysis in R

Unsupervised Learning in Python

Customer Segmentation in Python

Machine Learning for Marketing in Python

Customer Lifetime Value Forecasting centers around forecasting the total dollar worth a customer may provide for an organization over the course of their relationship. By leveraging supervised machine learning and predicting customer lifetime value, financial services organizations better understand the costs of acquiring and keeping customers. Moreover, forecasting customer lifetime value provides a snapshot of the different drivers behind churn (more on that in the next point) and retention. This allows financial services organizations to forecast loyal customers, reward them, and improve negative customer experiences (Qualtrics).

Commonly used technology

Relevant Tracks

Supervised Machine Learning in R

5 courses

Machine Learning Fundamentals with Python

5 courses

Marketing Analytics with Python

7 courses

Relevant Courses

Supervised Learning in R: Regression

Machine Learning for Marketing Analytics in R

Supervised Learning with scikit-learn

Introduction to Linear Modeling in Python

Machine Learning for Marketing in Python

Customer Churn Prediction leverages customer usage and consumption data to accurately identify which customers are most likely to churn from a product or service. Financial services organizations can leverage supervised machine learning to architect churn detection systems that mine historical data, accurately predict most likely churners, and automatically target them with tailored offers and services based on their churn drivers.

Commonly used technology

Relevant Tracks

Supervised Machine Learning in R

5 courses

Machine Learning Fundamentals with Python

5 courses

Marketing Analytics with Python

7 courses

Relevant Courses

Supervised Learning in R: Regression

Machine Learning for Marketing Analytics in R

Supervised Learning with scikit-learn

Marketing Analytics: Predicting Customer Churn in Python

Machine Learning for Marketing in Python

Automated Product Recommendations have been a staple of the e-commerce and entertainment industries. Data science and supervised machine learning techniques—like market basket analysis and recommendation engines—allow organizations to provide better product recommendations for customers based on their purchasing behavior. Financial services organizations can then intelligently upsell and further entrench customers in their product ecosystem. This ultimately results in less churn, higher customer lifetime value, and a hyper-personalized experience for each customer.

Commonly used technology

Relevant Tracks

Supervised Machine Learning in R

5 courses

Machine Learning Fundamentals with Python

5 courses

Relevant Courses

Supervised Learning with R: Classification

Market Basket Analysis in R

Supervised Learning with scikit-learn

Market Basket Analysis in Python

Building Recommendation Engines with PySpark

Risk and Governance

Fraud detection is one of the most useful data science and machine learning use cases for financial institutions. A 2019 LexisNexis report found that fraud attempts have increased by a whopping 110% for financial services firms. This leads to rising costs both in transactions lost due to fraud, and labor and investigation costs. For every dollar of theft, retail banks incur $3.34 in total fraud costs. Moreover, a report drawn by the University of Portsmouth found that fraud cost financial services firms around $5.127 trillion in 2019 globally. Leveraging supervised and unsupervised machine learning with anomaly detection techniques is becoming imperative to combat fraud. Whether to address credit card fraud, fraudulent checks, insider trading, KYC, or wire transfer fraud, financial services firms are increasingly mining data using machine learning to flag high-risk transactions. For example, Revolut has employed machine learning to detect credit card fraud and address money laundering attempts.

Commonly used technology

Relevant Tracks

Supervised Machine Learning in R

5 courses

Machine Learning Fundamentals with Python

5 courses

Relevant Courses

Supervised Learning with R: Classification

Unsupervised Learning in R

Anomaly Detection in R

Fraud Detection in R

Supervised Learning with scikit-learn

Unsupervised Learning in Python

Fraud Detection in Python

Cybersecurity is a hot topic in finance. According to Forbes, financial services institutions are 300 times more likely to be the target of a cybersecurity attack. To put it in perspective, financial services firms incur one billion cyberattacks per year—or 30 attacks per second. Using the same anomaly detection techniques to combat fraud, financial services organizations can use machine learning to identify outlier behaviors and logs that are most likely to be cybersecurity attacks.

Commonly used technology

Relevant Tracks

Supervised Machine Learning in R

5 courses

Machine Learning Fundamentals with Python

5 courses

Relevant Courses

Supervised Learning with R: Classification

Unsupervised Learning in R

Anomaly Detection in R

Fraud Detection in R

Supervised Learning with scikit-learn

Unsupervised Learning in Python

Fraud Detection in Python

Credit risk modeling allows financial institutions to leverage data and machine learning to improve their credit risk forecasting. With these techniques, financial institutions can improve their predictive power, automate processes, and make more educated loan application decisions. They can also provide improved products and services for their customers by reducing time-to-value. It's important to ensure transparency for customers when determining credit risk and to safeguard against biased decisions (EETimes).

Commonly used technology

R python™

Relevant Tracks

Finance Fundamentals in R

6 courses

Applied Finance in R

7 courses

Finance Fundamentals in Python

6 courses

Applied Finance in Python

4 courses

Relevant Courses

Importing and Managing financial Data in R

Credit Risk Modeling in R

Importing and Managing Financial Data in Python

Credit Risk Modeling in Python

Portfolio management is the process of building and overseeing a selection of investments based on different appetites for risk. Using data analytics models in quantitative risk management, financial services organizations can minimize risk and maximize the reward for all their investment plans. This not only translates into less risk exposure but drives higher revenue and reduced costs.

Commonly used technology

Relevant Tracks

Finance Fundamentals in R

6 courses

Applied Finance in R

7 courses

Finance Fundamentals in Python

6 courses

Applied Finance in Python

4 courses

Relevant Courses

Importing and Managing financial Data in R

Introduction to Portfolio Analysis in R

Quantitative Risk Management in R

Equity Valuation in R

Importing and Managing Financial Data in Python

Introduction to Portfolio Analysis in Python

Introduction to Portfolio Risk Management in Python

Quantitative Risk Management in Python

By Sector

Retail Banking

Cost Reduction

Robotic Process Automation

Business Intelligence

Chatbots

Document digitization

Revenue Maximization

Customer Segmentation

Customer Lifetime Value

Customer Churn

Product Recommendation

Risk and Governance

Fraud detection

Cybersecurity

Credit risk modeling

Portfolio management

Investment Banking

Cost Reduction

Robotic Process Automation

Business Intelligence

Risk and Governance

Fraud detection

Cybersecurity

Credit risk modeling

Portfolio management

Insurance

Cost Reduction

Robotic Process Automation

Business Intelligence

Chatbots

Claim triage and prediction

Revenue Maximization

Customer Segmentation

Customer Lifetime Value

Customer Churn

Product Recommendation

Risk and Governance

Fraud detection

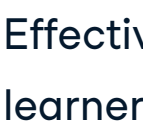
Cybersecurity

Portfolio management

3

DataCamp's proven learning methodology for building data science skills

Beyond courses and tracks, DataCamp's proven learning methodology provides a cyclical process for learning and retention. This learning methodology enables learners across the data fluency spectrum to assess their skills and identify gaps, develop a learning plan based on these gaps, practice skills, and apply them in a real-world setting. Experienced data scientists can upskill on new techniques in their target domain, and domain experts can learn the fundamentals of data literacy and data science.



Assess

Effective learning starts with understanding skill gaps and strengths. With DataCamp Signal™, learners can understand specific skill gaps they have across various topics and tools. From data literacy assessments like understanding and interpreting data to programming and machine learning assessments in R or Python, our 10-minute adaptive evaluations provide learners with personalized skill gaps and learning paths to address their skill gaps.

Learn

DataCamp's growing course library houses more than 350 expert-led, hands-on courses across various technologies and domains for all data skills and levels. Learners can hit the ground running with our learn-by-doing-approach—our bite-sized videos and interactive coding exercises allow them to start working with their preferred tool and topic right in the browser.

Practice

The next step in DataCamp's proven learning methodology is to practice all the information retained in courses. Using practice mode, learners can practice what they've learned with short challenges to test critical concepts. With over 3400 practice questions, learners can practice their skills across various technologies and topics. Our **mobile app** is the perfect way to practice and learn on the go.

Apply

Once skills have been assessed, cultivated through courses, and sharpened through practice, learners are ready to apply their skills in a project-based environment. With **DataCamp projects**, learners can solve a variety of real-world R and Python data science projects. Learners can opt for guided projects, where they can follow step-by-step tasks and receive helpful feedback as they apply their newfound skills. They can also opt for unguided projects, which are open-ended, offering a variety of possible solutions and a live-code-along video to follow how an expert data scientist would approach a solution.

This entire learning experience is easy to implement and manage for teams of any size, with an administrator dashboard that allows custom learning paths based on roles and departments, advanced analytics and insights to measure the impact of online learning, and seamless SSO and LMS integrations. Teams benefit from our Customer Success Managers, who partner with organizations to accelerate learning adoption and provide valuable recommendations to help achieve organization-wide data fluency. We have more than 6 million learners around the world—and we're just getting started.

Close the talent gap. Visit **datacamp.com**.

datacamp