BANK OF ENGLAND

FINANCIAL CONDUCT AUTHORITY

Machine learning in UK financial services

October 2019

Contents

Executive summary

Machine learning (ML) is the development of models for prediction and pattern recognition from data, with limited human intervention. In the financial services industry, the application of ML methods has the potential to improve outcomes for both businesses and consumers.(1) In recent years, improved software and hardware as well as increasing volumes of data have accelerated the pace of ML development. The UK financial sector is beginning to take advantage of this. The promise of ML is to make financial services and markets more efficient, accessible and tailored to consumer needs.(2) At the same time, existing risks may be amplified if governance and controls do not keep pace with technological developments. But the risks presented by ML may be different in each of the contexts it is deployed in.(3) More broadly, ML also raises profound questions around the use of data, complexity of techniques and the automation of processes, systems and decision-making.(4)

The Bank of England (BoE) and Financial Conduct Authority (FCA) have a keen interest in the way that ML is being deployed by financial institutions. That is why we conducted a joint survey in 2019 to better understand the current use of ML in UK financial services. The survey was sent to almost 300 firms, including banks, credit brokers, e-money institutions, financial market infrastructure firms, investment managers, insurers, non-bank lenders and principal trading firms, with a total of 106 responses received.

The survey asked about the nature of deployment of ML, the business areas where it is used and the maturity of applications.(5) It also collected information on the technical characteristics of specific ML use cases. Those included how the models were tested and validated, the safeguards built into the software, the types of data and methods used, as well as considerations around benefits, risks, complexity and governance.

Although the survey findings cannot be considered to be statistically representative of the entire UK financial system, they do provide interesting insights.

The key findings of our survey are:

- **ML is increasingly being used in UK financial services**. Two thirds of respondents report they already use it in some form. The median firm uses live ML applications in two business areas and this is expected to more than double within the next three years.

- **In many cases, ML development has passed the initial development phase**, and is entering more mature stages of deployment. One third of ML applications are used for a considerable share of activities in a specific business area. Deployment is most advanced in the banking and insurance sectors.

- **From front-office to back-office, ML is now used across a range of business areas**. ML is most commonly used in anti-money laundering (AML) and fraud detection as well as in customer-facing applications (eg customer services and marketing). Some firms also use ML in areas such as credit risk management, trade pricing and execution, as well as general insurance pricing and underwriting.

(1) Carney, M (2018), 'AI and the Global Economy'.
(2) Carney, M (2018), 'AI and the Global Economy'.
(3) www.fca.org.uk/news/speeches/future-regulation-ai-consumer-good.
(4) Proudman, J (2019), 'Managing machines: the governance of artificial intelligence'.
(5) In this report the term application means the integrated whole of a ML application, including data collection, feature engineering, model engineering and deployment. It also includes the underlying IT infrastructure (eg data storage, integrated development environment). A ML application could include multiple models and ML algorithms. ML applications should be seen as separate if they fulfil different business purposes or if their set up / components differ significantly.

- **Regulation is not seen as an unjustified barrier but some firms stress the need for additional guidance on how to interpret current regulation**. Firms do not think regulation is an unjustified barrier to ML deployment. The biggest reported constraints are internal to firms, such as legacy IT systems and data limitations. However, firms stressed that additional guidance around how to interpret current regulation could serve as an enabler for ML deployment.

- **Firms thought that ML does not necessarily create new risks, but could be an amplifier of existing ones**. Such risks, for instance ML applications not working as intended, may occur if model validation and governance frameworks do not keep pace with technological developments.

- **Firms validate ML applications before and after deployment**. The most common validation methods are outcome-focused monitoring and testing against benchmarks. However, many firms note that ML validation frameworks still need to evolve in line with the nature, scale and complexity of ML applications.

- **Firms use a variety of safeguards to manage the risks associated with ML**. The most common safeguards are alert systems and so-called 'human-in-the-loop' mechanisms. These can be useful for flagging if the model does not work as intended (eg in the case of model drift, which can occur as ML applications are continuously updated and make decisions that are outside their original parameters).

- **Firms mostly design and develop ML applications in-house**. However, they sometimes rely on third-party providers for the underlying platforms and infrastructure, such as cloud computing.

- **The majority of users apply their existing model risk management framework to ML applications**. But many highlight that these frameworks might have to evolve in line with increasing maturity and sophistication of ML techniques. This was also highlighted in the BoE's response to the Future of Finance report.(6) In order to foster further conversation around ML innovation, the BoE and the FCA have announced plans to establish a public-private group to explore some of the questions and technical areas covered in this report.

(6) Bank of England (2019), 'The Future of Finance — our response'.

1 Introduction

1.1 Context and objectives

The UK economy is increasingly powered by big data, platform business models, advanced analytics, smartphone technology and peer-to-peer networks.(7) At the same time, innovation in the financial sector is dramatically changing the markets we regulate(8) but also the way in which we regulate.(9)(10) As an industry, financial services are (and will always be) very data-reliant. Hence, this new data-driven economy goes hand in hand with fundamental changes to the structure and nature of the financial system supporting it.(11) And ML is a principal driver contributing to this new finance.(12)

ML has wide-ranging applications in financial services and, when combined with increasing computational power, has the ability to analyse large data sets, detect patterns and solve problems at speed. The use of ML has the potential to generate analytical insights, support new products and services, and reduce market frictions and inefficiencies.(13) If this potential is achieved, consumers could benefit from more tailored, lower cost products and firms could become more responsive, learner and effective.

It is important that regulatory authorities understand ML; including the current state of deployment, maturity of applications, use cases, benefits and risks. This was the motivation behind the BoE and FCA joint survey, which was carried out during the first half of 2019. The objective was to gain an understanding of the use of ML in the UK financial sector. The results, together with ongoing dialogue with the industry and other authorities, both domestically and internationally, will help identify where there are policy questions that need to be answered in the future, in order to support the safe and productive deployment of ML within the financial sector.

This joint BoE-FCA report is the result of the analysis of the responses to the survey and presents:

- a quantitative overview of the use of ML across the respondent firms;
- the ML implementation strategies of firms that responded to the survey;
- approaches to the governance of ML;
- the share of applications developed by third-party providers;
- respondents' views on the benefits of ML;
- perceptions of risks and ethical considerations;
- perspectives on constraints to development and deployment of ML; and
- a snapshot of the use of different methods, data, safeguards performance metrics, validation techniques and perceived levels of complexity.

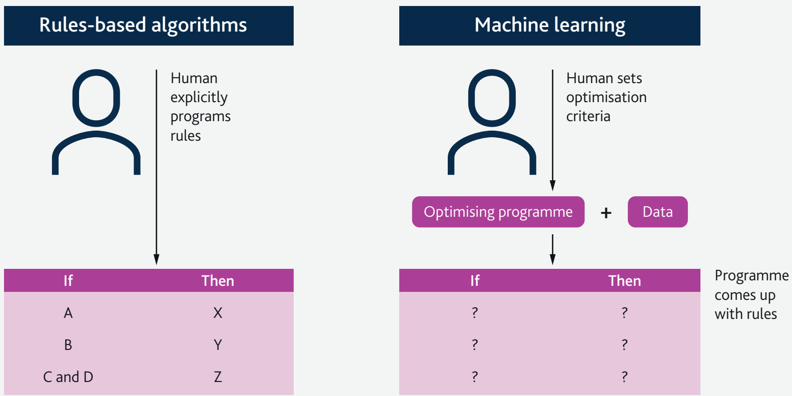
(7) Carney, M (2019), 'A platform for innovation — remarks'.
(8) www.fca.org.uk/news/speeches/innovation-hub-innovation-culture.
(9) www.fca.org.uk/news/speeches/financial-conduct-regulation-restless-world.
(10) Chakraborty, C and Joseph, A (2017), 'Machine learning at central banks', *Bank of England Staff Working Paper No. 674*. Turrell *et al* (2018), 'Using online job vacancies to understand the UK labour market from the bottom-up', *Bank of England Staff Working Paper No. 742*. Proudman, J (2018), 'Cyborg supervision — the application of advanced analytics in prudential supervision'.
(11) See Mnohoghitnei, I, Scorer, S, Shingala, K and Thew, O, 'Embracing the promise of fintech', *Bank of England Quarterly Bulletin*, 2019 Q1.
(12) Carney, M (2018), 'AI and the Global Economy'.
(13) www.fsb.org/wp-content/uploads/P011117.pdf.

Box 1
What is the difference between artificial intelligence and machine learning?

Artificial intelligence (AI) is the theory and development of computer systems able to perform tasks which previously required human intelligence.[(1)] AI is a broad field, of which ML is a sub-category.

ML is a methodology whereby computer programmes fit a model or recognise patterns from data, without being explicitly programmed and with limited or no human intervention. This contrasts with so-called 'rules-based algorithms' where the human programmer explicitly decides what decisions are being taken under which states of the world (**Figure A**).

Figure A Machine learning algorithms make decisions without being explicitly programmed



Many ML algorithms constitute an incremental (rather than fundamental) change in statistical methods. They introduce more flexibility in statistical modelling. For instance, many ML models are not constrained by the linear relationships often imposed in traditional economic and financial analysis.

However, over the last decade, computing power and the amount of data processed has grown exponentially. This has allowed ML models to become an order of magnitude larger and more complex than more traditionally used techniques. As a result, ML models can often make better predictions than traditional models or find patterns in large amounts of data from increasingly diverse sources.

(1) www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/.

The report closes with a non-exhaustive selection of case studies, describing a sample of typical use cases, including:

- Anti-money laundering and countering the financing of terrorism
- Customer engagement
- Sales and trading
- Insurance pricing
- Insurance claims management
- Asset management

1.2 Methodology

When designing the survey, the BoE and FCA considered the Legislative and Regulatory Reform Act 2006 principle that regulatory activities should be carried out in a way which is transparent and proportionate.

In total, 287 firms received the questionnaire and 106 submitted responses. The BoE surveyed 58 dual-regulated firms[14] and received 47 (81%) responses.[15] The FCA surveyed 229 FCA-regulated firms and received 63 (28%) responses.

The BoE selected firms with the aim of surveying each type of BoE and Prudential Regulation Authority (PRA)-regulated firm. This sample was determined to cover a significant share of BoE and PRA firms. It also included several firms that are small in terms of their market share but were considered to be advanced in the use of ML and therefore of interest for horizon-scanning purposes.

The FCA sample was built according to the following criteria. Sample selection reflected the need to represent firms that, due to their size and the number of customers, have the potential to affect the highest number of consumers, or are more likely to be anticipating future trends in the market, thus affecting consumers in the future. To meet these two objectives, for each FCA supervised sector, the FCA selected a sample of 'large firms' (among the largest sector firms in terms of income). Further, for each sector the FCA selected a sample of 'fast growing firms' (the sector firms with the highest income growth rate). This was judged to be the best way to get both an accurate snapshot of the state of ML at firms affecting a very large number of UK consumers, and a glimpse of where the market is heading.

Overall, the combined sample is skewed somewhat towards larger firms. In addition, it can be surmised that some firms did not respond to the survey because they have no ML applications and therefore the responses lean more towards firms that currently use ML. Therefore, the sample and survey findings should not be seen as representative for all types of firms or the entire UK financial services industry. The findings presented in this report should instead be considered as a snapshot of ML adoption. Our hope is that this will serve as a benchmark for future research and will stimulate debate.

The case studies presented in the Appendix were selected based on the number of responses received, ie we selected the most common use cases reported by participating firms.

The results presented in this report are anonymised and aggregated with the respondents grouped into the sectors listed in Box 2.

All charts in this report are based on data from the BoE and FCA survey.

Box 2
Sector classification used in the report

Sector	Type of firms included[1]
Banking	Building Societies, International Banks, Retail Banks, UK Deposit Takers, Wholesale Banks.
Insurance	General Insurers, Insurance Intermediaries, Life Insurers, Personal and Commercial Lines Insurers.
Non-Bank Lending	Debt Administrators, Credit Brokers, Crowdfunders, Debt Purchasers/Collectors, Lifetime Mortgage Providers, Consumer Credit Lenders, Motor Finance Providers, Non-bank Lenders, Retail Finance Providers.
Investments and Capital Markets	Alternatives, Corporate Finance Firms, Fund Managers, Principal Trading Firms, Wealth Managers and Stockbrokers, Wholesale Brokers.
Payments, Financial Market Infrastructure (FMI) and other	Credit Reference Agencies, Custody Services, E-money Issuers, Exchanges, Financial Market Infrastructure, Multilateral Trading Facilities, Payment Services Firms, Platforms, Price Comparison Websites, Providers of Credit Information Services.

(1) Listed alphabetically and based on BoE, PRA and FCA classifications.

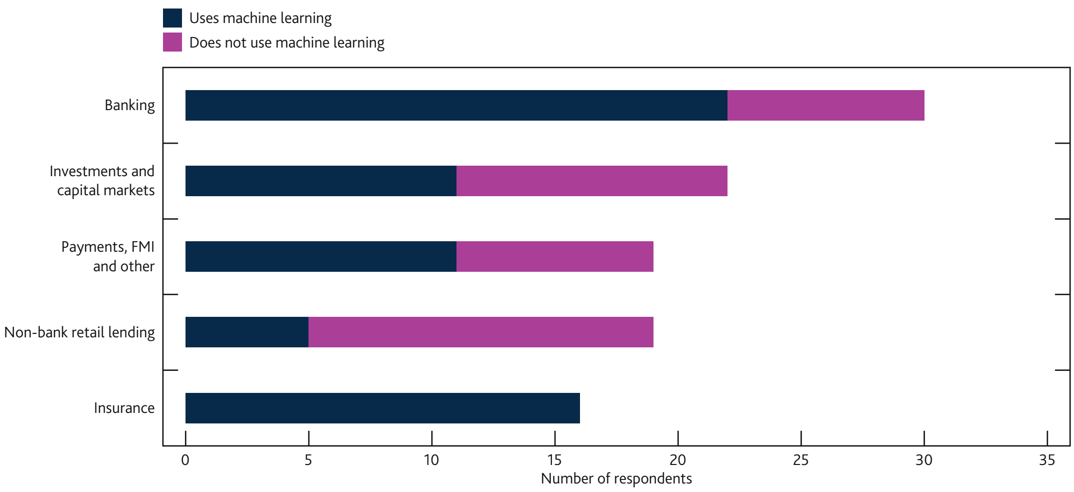
(14) Regulated by both PRA and FCA as well as Financial Market Infrastructure firms, which are regulated by the BoE not PRA.
(15) In addition, four BoE/PRA-regulated firms did not submit complete responses because they do not have any ML applications.

2 The state of machine learning adoption

2.1 Machine learning is already being used live by the majority of respondents

ML is increasingly being adopted in UK financial services, according to our survey. Two thirds of respondents report they already use ML live in their business (**Chart 1**), albeit many only have a limited number of use cases. ‘Live’ in this context means that it is used to support client interaction, business decisions or transactions. Reported use cases range from equity trading, where firms use ML to optimise order-routing and deal execution, to AML where firms use ML to analyse millions of documents for ‘know-your-customer’ checks, to insurance, where firms use ML to estimate more personalised risk premiums.

Chart 1 Two thirds of respondents have live machine learning applications in use



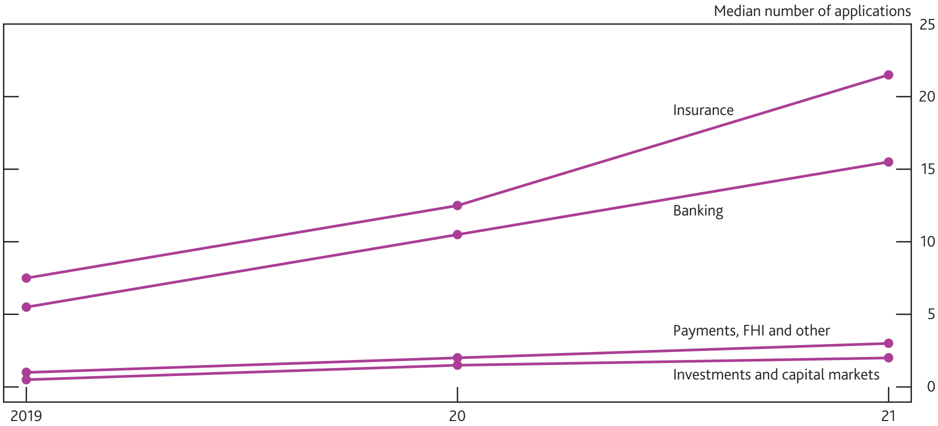
The median firm uses ML in two distinct areas. To illustrate, the median firm may have one application in, say, credit scoring and another one in, say, compliance. There is a significant spread around this and, at the more advanced end, 15 firms (14% of respondents) have more than 10 distinct live applications.

Insurance and banking are the sectors in our sample with the most live cases (**Chart 2**). The median insurance firm has 7.5 live applications and the median banking firm has 5.5. This is partly driven by the fact that the insurance and banking sectors in our sample feature a bigger share of large firms, as highlighted in Section 1.2. Larger firms may possibly be more advanced in their ML deployment due to benefits of scale, access to data, ability to attract ML talent, or greater resources. However, more research would be needed to shed light on the specific reasons for sectoral differences.

Looking to the future, respondents expect significant growth in the number of live ML applications. The median respondent expects their number of ML applications to more than double over the next three years (**Chart 2**). For banking and insurance the expected growth is bigger still, with firms in each sector expecting their number of ML applications to almost triple, to 15.5 and 21.5 respectively. This underlines growing interest in ML and the prospect of increasing use across the financial sector in coming years.

Respondents' predictions reflect the fact that firms report a growing number of ML applications in development that may be ready to go live in coming years.(16) As shown in the next section, roughly, for any six applications firms use, four additional ones are already being developed.

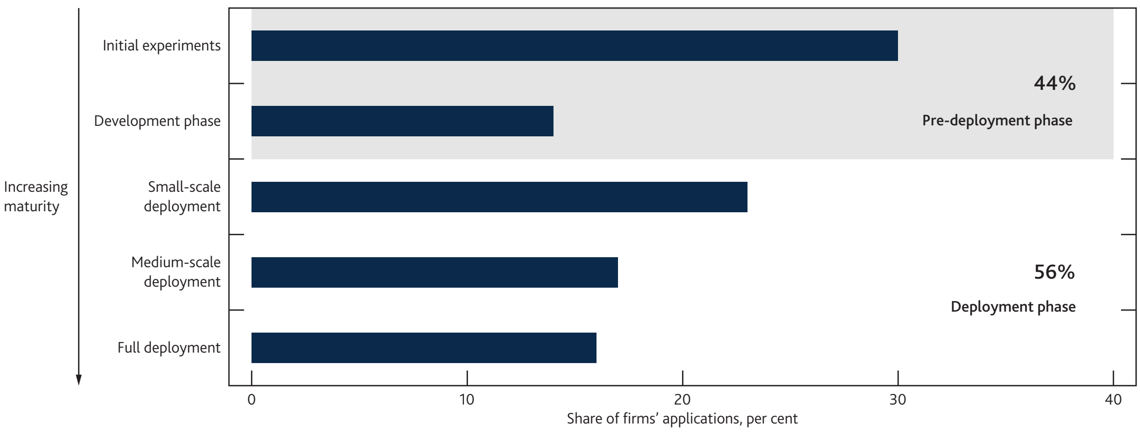
Chart 2 Respondents expect significant growth in use of machine learning over the next three years



2.2 In many cases firms' deployment of machine learning has passed the initial development phase

To better understand how respondents are developing ML, we asked firms to indicate the maturity of their ML applications across five distinct categories (**Chart 3**). In many cases, firms' ML applications have passed the initial pre-deployment phase — which includes proof of concept and research and development — and entered the deployment phase — where the application is used live within the business. Of the total number of ML applications reported by firms, almost two thirds (56%) are live (**Chart 3**).

Chart 3 For any six applications firms use, four additional ones are under development[(a)]



(a) Small-scale deployment refers to 0-30% of a business line; medium-scale deployment refers to 31-60% of a business line; full deployment refers to 60-100% of a business line.

2.3 Respondents identify a broad range of use cases

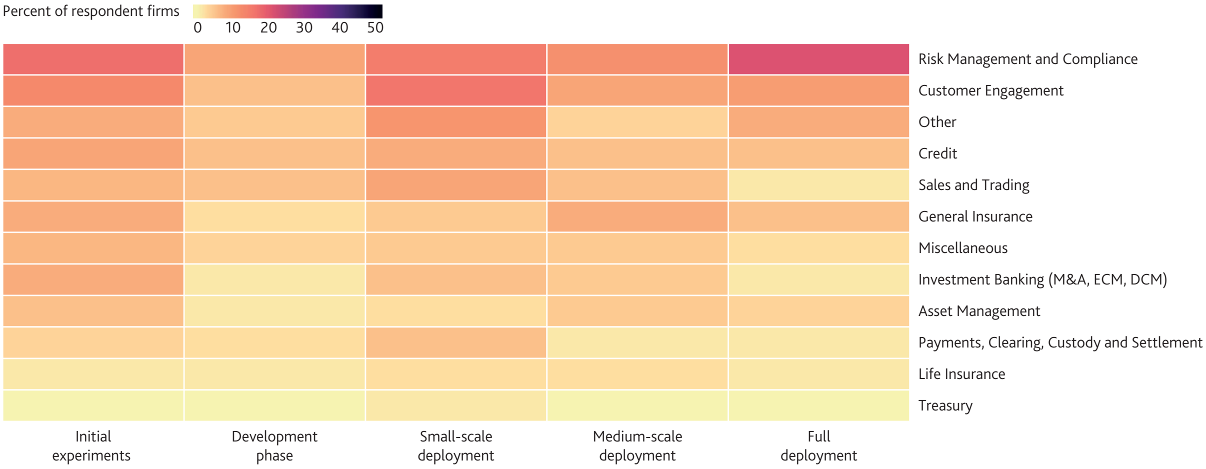
Respondents use ML in a wide range of business areas. **Chart 4A** presents a heatmap, showing what share of firms in the overall sample have at least one application in a given business area. It highlights that back-office functions, such as risk management and compliance see the most frequent use cases at the moment, which include, for instance, AML and fraud detection. However, ML is also increasingly being applied to front-office areas, like

(16) While keeping in mind that many proof of concept and research and development projects will not make it to the deployment stage.

customer management as well as sales and trading. Overall, the business areas with the most frequent and mature levels of ML deployment are: risk management and compliance; customer engagement; credit; securities sales and trading and general insurance.

Chart 4A The most frequent and also mature use cases are risk management and compliance, and customer engagement[(a)]

Firms with at least one application as a percentage of all respondent firms



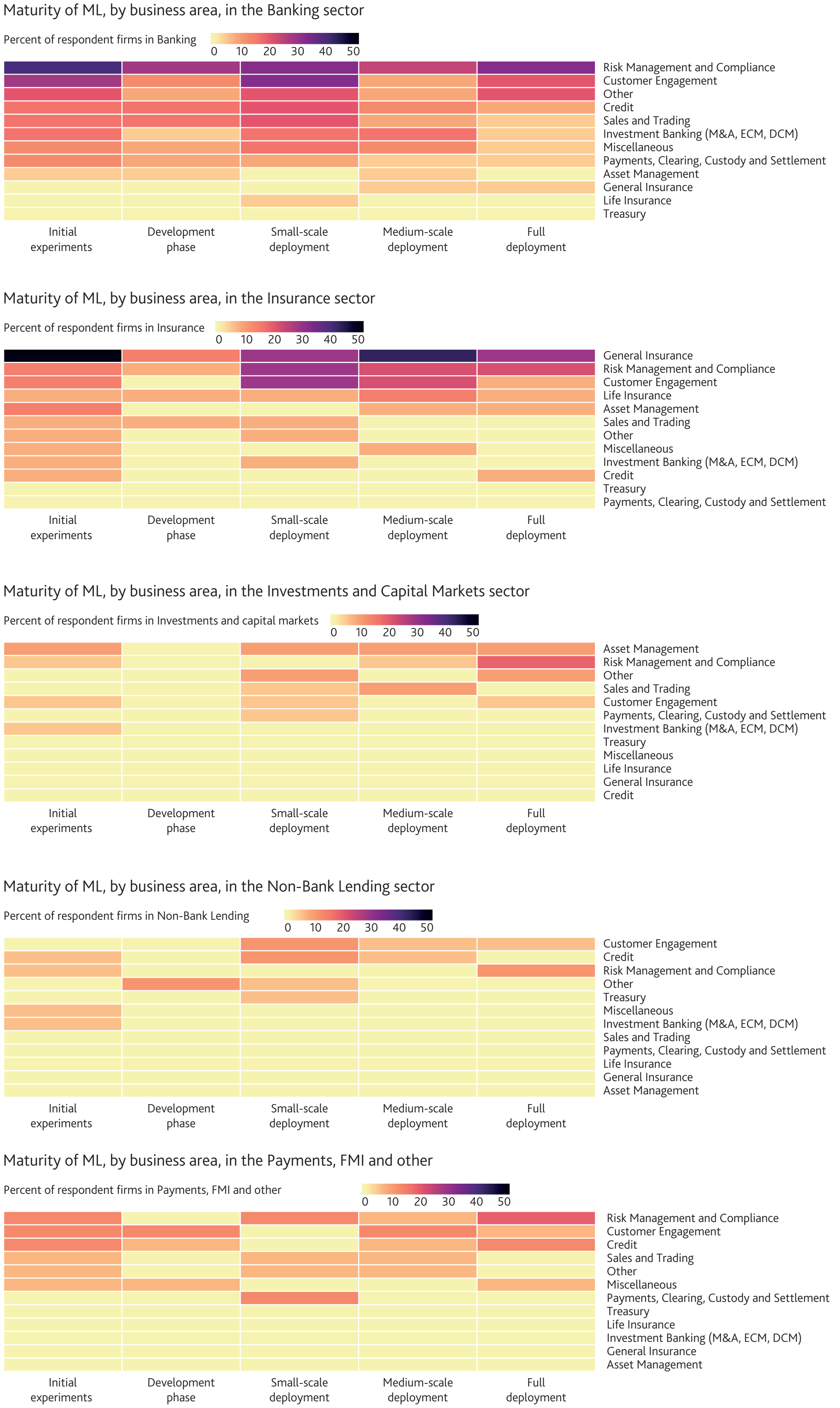
(a) Small-scale deployment refers to 0-30% of a business line; medium-scale deployment refers to 31-60% of a business line; full deployment refers to 60-100% of a business line.

Widespread use in back-office areas partly reflects the fact that this type of activity is performed by most types of firms; while for instance, not all firms in the sample would be expected to undertake insurance activities or investment banking. In addition, AML and fraud detection are well established use cases because the need to connect large data sets and undertake pattern detection is a set-up that lends itself well to ML.[(17)] It is noted that treasury management (which is an activity conducted in most firms) is not yet an area where ML applications are commonly in use.

Overleaf we break down the most common and mature use cases by sector. The charts show that banking and insurance have a relatively higher share of mature use cases than other sectors. The charts also highlight that, in banking and insurance, use cases are spread across most areas of the business. In banking, following risk management and compliance, customer engagement is the area with the second most use cases. And, for insurers, general insurance distribution and underwriting have more use cases than back-office functions.

(17) www.iif.com/Publications/ID/1421/Machine-Learning-in-Anti-Money-Laundering and www.accenture.com/_acnmedia/pdf-61/accenture-leveraging-machine-learning-anti-money-laundering-transaction-monitoring.pdf.

Chart 4B Banking and insurance have the most mature cases, across a range of business areas[(a)]



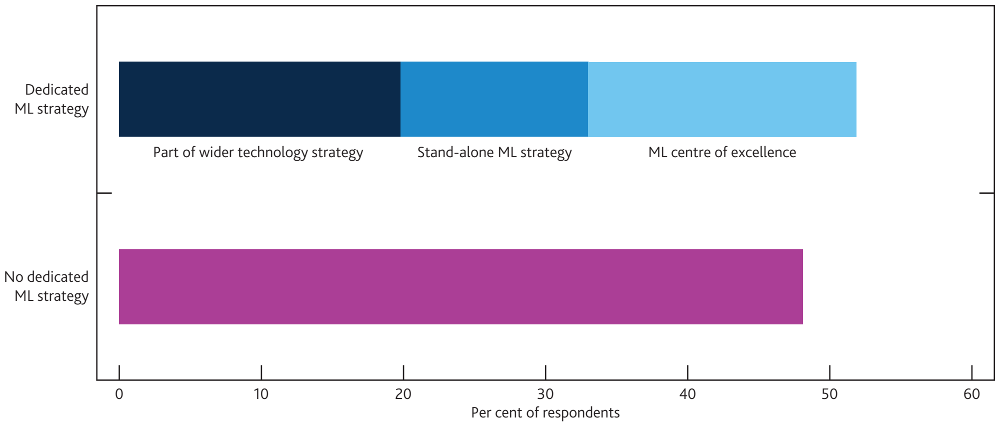
(a) Small-scale deployment refers to 0-30% of a business line; medium-scale deployment refers to 31-60% of a business line; full deployment refers to 60-100% of a business line.

3 Strategies, governance and third-party providers

3.1 The majority of respondents have a dedicated machine learning strategy

ML is emerging as a strategic priority for many of the firms in our sample. Currently, 52% of respondents have a dedicated strategy for research, development and deployment. Firms highlight three types of approaches (**Chart 5**): 19% are establishing or already have a dedicated centre of excellence that works to promote ML deployment across the organisation. Whilst 13% of respondents identify ML as important enough to develop a stand-alone firm-wide ML strategy. Furthermore, 20% of firms include ML as part of their overarching innovation or technology strategy but have not set up dedicated structures to promote it independently. Finally, the remaining 48% of respondents say they do not have a dedicated ML strategy. This includes firms that do and do not use ML.

Chart 5 The majority of firms have an explicit strategy for machine learning



Amongst respondents, the insurance (81%), banking (67%) and investment and capital markets (45%) sectors have the highest proportion of firms with a ML strategy. On the other hand, only 37% of payments, FMI and other firms and 28% of non-bank lending firms have a ML strategy.

Some smaller banks and a number of firms from all sectors report they do not have a strategy despite using ML. Several reasons were cited, including that the level of ML is sufficiently small that it does not justify a specific strategy, and ML, as with other technologies, is used to support specific business areas and their respective strategies. Many of the firms that do not use ML report that it is not a priority given the size, scope or focus of their organisation.

3.2 The majority of users apply their existing model risk management framework to machine learning[18]

Of the respondents that use ML, more than half (57%) say their applications are governed through their existing model risk management framework or enterprise risk function, including all three lines of defence.[19] Furthermore, 12% of ML users are establishing specialist committees to advise the respective governance bodies and risk management functions on ML-specific questions, and some have created ML principles that are embedded in the governance framework. Four firms also say they are in the process of establishing a ML ethics function that would address the particular ethical issues raised by ML models and the use of new data sources.

Several firms highlight the need for their risk management frameworks to evolve given their increasing use of ML, for instance, to address challenges related to the explainability of ML models[20] and potential model drift (where model outcomes change over time due to new or different data). Firms note that explainability plays an important part in ML model development, standards and governance procedures. With regard to model drift, some respondents highlight the need for model lifecycle management platforms to enable continuous monitoring of model performance.

Several respondents recognise the importance of ensuring employees at different levels of their organisation have the right knowledge and skill sets to understand the functions and implications of ML. They said this could include embedding individuals with ML expertise within the model risk management and data governance functions. Another aspect of this was making arrangements for senior decision makers to be informed by subject matter experts or to undertake training to ensure they understand the technical aspects of ML as well as the potential legal, regulatory and ethical considerations.

A quarter of ML users highlight data-related challenges and mention specific governance, risk management and control functions to deal with these. This includes assessing data sources that are used for modelling purposes in order to detect and address biased or incorrect data, as well as ensuring appropriate sign-off for access to specific data sets when testing and deploying ML models. From an organisational perspective, several firms said ML falls under both the model risk management and data control frameworks.

In Box 3, we highlight some theoretical implications that an increased use of ML could have for BoE, PRA and FCA supervisors.

3.3 Only a small share of machine learning applications are implemented by third-party providers

The majority (76%) of ML use cases are developed and implemented internally by firms, with the remaining 24% implemented by third-party providers (**Chart 6**). However, firms told us they often use off-the-shelf ML models, open source software and ML libraries developed by third-party providers, which are then further developed or adapted to specific use cases and deployed internally. Respondents from the non-bank lending sector have the highest use of third-party ML applications (36%), which may be because the average size of the firms in this sector in our sample was smaller and, therefore, they may have less capacity to internally develop applications. Or it may be due to the relative ability of third-party providers to integrate products into these firms given their processes and architecture.

(18) It is important to note that this report does not assess the adequacy of governance frameworks in relation to the use of ML.
(19) Often referred to as the 'three lines of defence', each of the three lines has an important role to play. The business line — the first line of defence — has 'ownership' of risk, whereby it acknowledges and manages the risk that it incurs in conducting its activities. The risk management function is responsible for further identifying, measuring, monitoring and reporting risk on an enterprise-wide basis as part of the second line of defence, independently from the first line of defence. The compliance function is also deemed part of the second line of defence. The internal audit function is charged with the third line of defence, conducting risk-based and general audits and reviews to provide assurance to the board that the overall governance framework, including the risk governance framework, is effective and that policies and processes are in place and consistently applied. See www.bis.org/bcbs/publ/d328.pdf.
(20) Bracke, P, Datta, A, Jung, C and Sen, S (2019), 'Machine learning explainability in finance: an application to default risk analysis', *Bank of England Staff Working Paper No. 816*.

Box 3
Algorithm complexity, supervision and governance

Supervisors like the BoE, the PRA and the FCA are technology neutral. That means, in principle, they do not require or prohibit the use of particular technologies. However, the EBA Guidelines on Information and Communication Technology (ICT) Risk Assessment[1] highlight that the 'depth, detail and intensity of ICT assessment should be proportionate to the size, structure and operational environment of the institution as well as the nature, scale and complexity of its activities'. So, while it will always depend on a multitude of factors whether a ML application poses a meaningful prudential or conduct risk, ML use can alter the nature, scale and complexity of IT applications and thus, a firm's IT risks. There are three dimensions to this (all of which we asked about in the survey):

- **ML applications are more complex.** ML models are often very large, non-linear and non-parametric. This makes it harder to comprehensively understand their properties and to validate them. This means certain forms of risk-taking could go undetected. This type of complexity can constitute a significant change to existing systems.

- **ML uses a broader range of data.** ML applications may often use entirely new types of complex, including unstructured, data. For instance, this could be data from news sources, satellite images or social media.

- **ML systems are larger in scale.** ML systems increasingly consist of a multitude of interacting components. This can make it harder to validate if they always interact as intended. In many cases, this change is incremental.

Chapter 5 explains in detail the various aspects of how ML can make systems more complex and how different types of data are being used.

However, the deployment of ML could also reduce risks. For instance, ML has the potential to reduce human bias, support the identification of market abuse practices, increase the effectiveness and efficiency of fraud detection and AML processes, as well as lead to better risk assessment and management.[2]

(1) https://eba.europa.eu/documents/10180/1841624/Final+Guidelines+on+ICT+Risk+Assessment+under+SREP+%28EBA-GL-2017-05%29.pdf/ef88884a-2f04-48a1-8208-3b8c85b2f69a.
(2) www.iif.com/Publications/ID/1421/Machine-Learning-in-Anti-Money-Laundering.

Firms also sometimes rely on third-parties when it comes to the underlying platforms and infrastructure, such as cloud computing. Overall, 22% of ML applications are run on the cloud, highlighting the link between in-house development of ML applications and running of these systems on internal servers (**Chart 7**). It is important to note, this figure differs by sector and non-bank lending firms have the highest share (39%) of applications run on cloud, which again may reflect the higher use of third-party ML applications.

Data from third-party sources

In addition to internal data, firms use data collected by third-parties in 40% of use cases. This includes data from different industries and non-traditional data sets (eg information about consumer characteristics for credit scoring, or information about automobiles for insurance pricing and claims processing), which can be combined with existing data to generate new insights, better predictions or more customised products.

Chart 6 Most machine learning applications are implemented internally

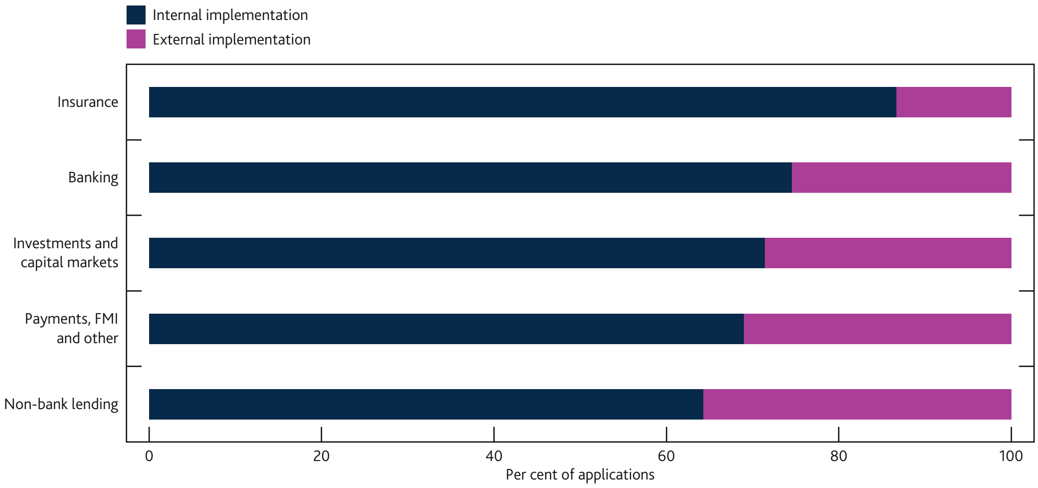
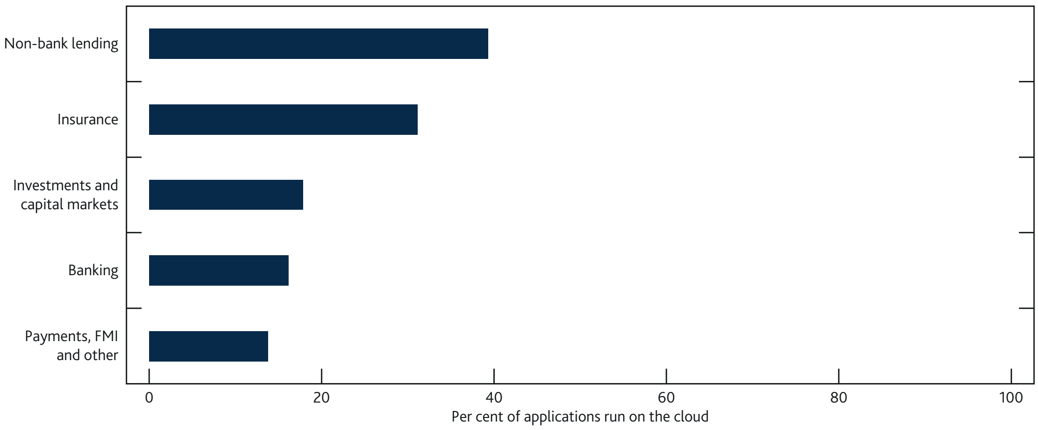


Chart 7 Most machine learning applications are run on internal servers and not on the cloud

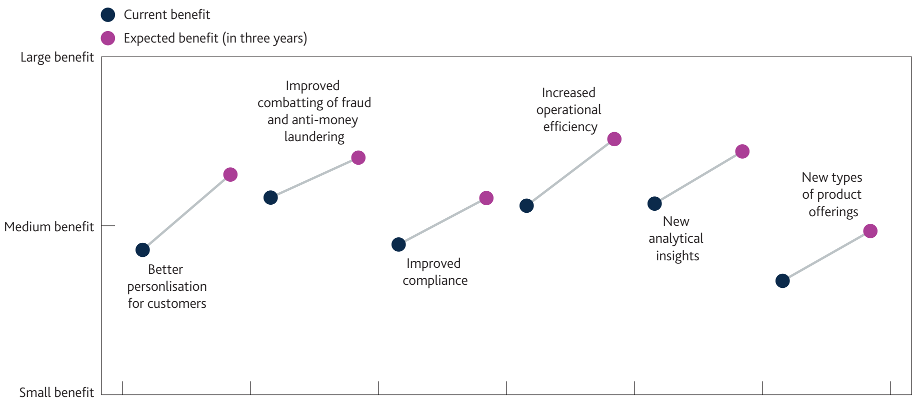


4 Firms' perception of benefits, risks and constraints

4.1 Respondents already see benefits from machine learning and expect these to increase

Respondents in all sectors think ML already benefits their business. Furthermore, and in line with firms' expectation that the number of ML applications they use will grow, respondents estimate the benefits will increase significantly over the next three years (**Chart 8**). The survey asked participating firms to score some of the current benefits of using ML applications, from small benefit to large benefit.(21)

Chart 8 The highest perceived benefits are in fraud detection and anti-money laundering, followed by operational efficiency gains and new analytical insights



Firms currently consider improved AML, fraud detection and overall efficiency gains (with the associated cost savings) as the biggest and most immediate benefits of using ML. There is a correlation between these benefits and the high number of ML applications in AML and fraud detection (**Chart 4**). Moreover, some firms mention they use ML in business areas where they identify clear efficiency gains and cost savings because they can persuasively demonstrate the benefits relative to traditional techniques. However, firms expect that increased benefits will also come from better personalisation of products for customers, new analytical insights and improved services over the next three years, all of which they consider could be revenue-generating (**Chart 8**).

4.2 Firms recognise model validation and governance need to keep pace with machine learning developments

Respondents recognise a range of risks that might arise from the application of ML in financial services. The survey responses suggest that ML applications can increase the technical complexity of models, and thus risk management and controls processes will need to keep pace. Firms do not think the use of ML necessarily generates new risks. Rather, they consider it as a potential amplifier of existing risks.

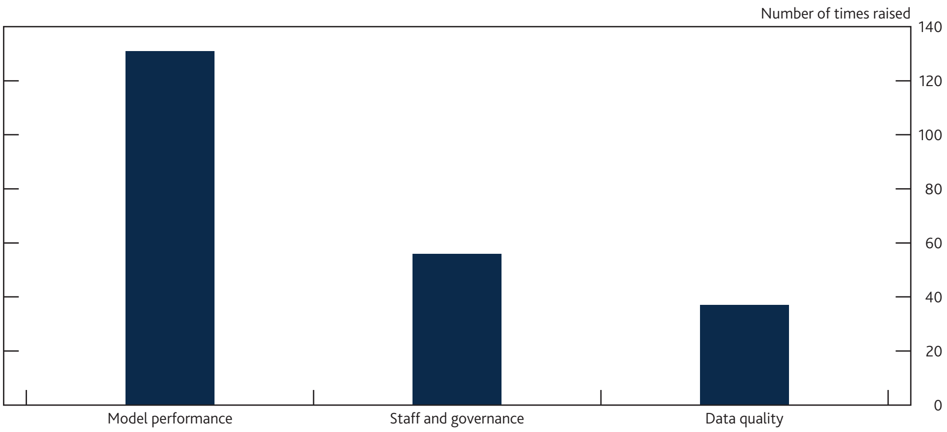
(21) Small benefit was allocated a score of 1, medium benefit was 2 and large benefit was 3.

Respondents explained that risks could be caused by a lack of ML model explainability meaning that the inner working of a model cannot always be easily understood and summarised. This forms part of more general questions around validating the design and performance of ML models. Another concern raised by firms is that models may perform poorly when applied to a situation that they have not encountered before or where human experience, institutional knowledge and judgement is required.

Firms also mention potential risks associated with data quality issues (including biased data). As firms note, these risks can have a negative impact on consumers' ability to use products and services, or even engage with firms. This can, in turn, damage the firm's reputation and lead to operational costs, service breakdowns and losses.

Overall, respondents think the top five risks that might occur because of ML applications relate to: lack of explainability; biases in data and algorithms; poor performance for clients/customers and associated reputational damage; inadequate controls, validation or governance; and inaccurate predictions resulting in poor decisions. In **Chart 9** we summarise these into three overall categories: model performance, staff and governance, and data quality.

Chart 9 Firms consider issues related to model performance the biggest amplifiers of existing risks



Firms highlight that there are a number of ways these risks could be managed, including through sound model validation and implementing safeguards. For example, certain methods can help mitigate risks when ML models do not work as intended, whilst others help identify potential errors and risks during the development phase. These are summarised in **Figure 1** and explained in detail in the following chapter.

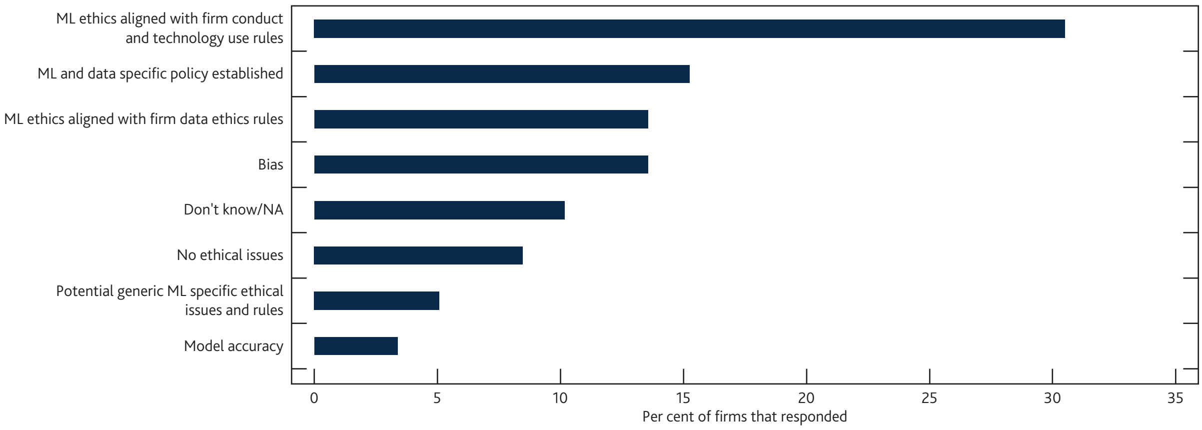
Figure 1 Examples of risks and possible ways to address these

	Model validation	Staff and governance	Data quality
Example	'Black box' ML models are harder to explain and make decisions outside their original parameters	Staff may be insufficiently trained to understand and address risks related to ML models	Poor quality data, limited training data or biases may produce unintended and negative results
Possible ways to address the risk	Evolve model validation approaches — see section 5.4	Ensure employees have the right skill sets — see section 3.2	Apply data quality validation framework — see section 5.2 and 5.4
Safeguards	Alert systems, 'guardrails', human-in-the-loop before execution, kill switches and back-up systems — see section 5.5		

The survey also included a question on firms' perception of potential ethical issues arising from the deployment of ML applications (**Chart 10**).

Firms interpreted this question in different ways. Some respondents understood this question to be about individual ethical issues, while others instead focused on how the firm is dealing with the potential ethical implication of the application of ML in financial services. The emerging picture represents again a wide range of opinions about how risk and harm might derive from firms applying ML.

Chart 10 Firms' perception of possible ethical implications arising from machine learning deployment(a)

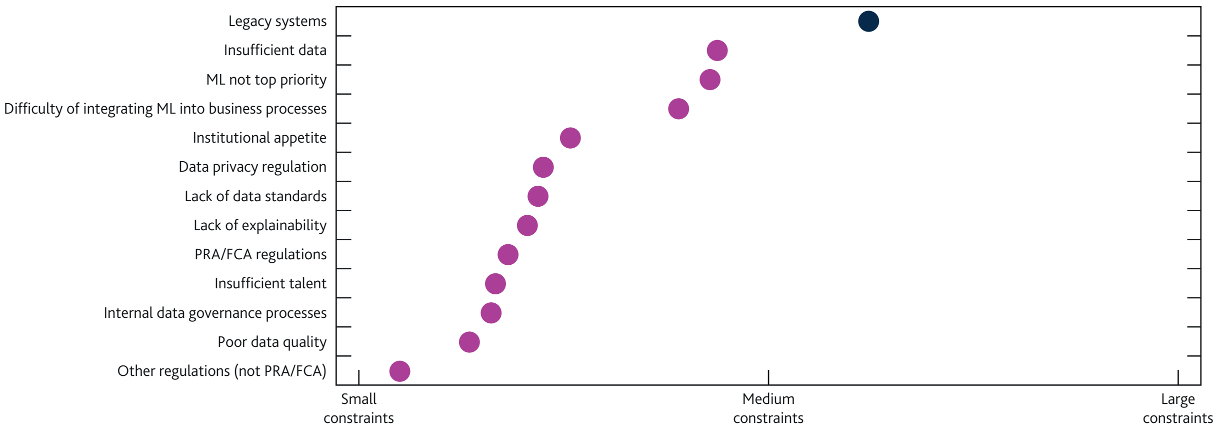


(a) This chart does not include all responses. It only shows the survey responses for firms using ML.

4.3 Constraints to deployment of machine learning are mostly internal to firms

Firms were asked to rank potential constraints that slow or stop them from deploying ML (**Chart 11**). The responses suggest the largest constraints are internal to firms. Aside from strategic decisions, namely ML is not a top priority, the three most cited are: legacy systems that are not conducive to ML, lack of access to sufficient data and the difficulty of integrating ML into existing business processes.

Chart 11 Legacy systems are the largest constraint to machine learning deployment(a)

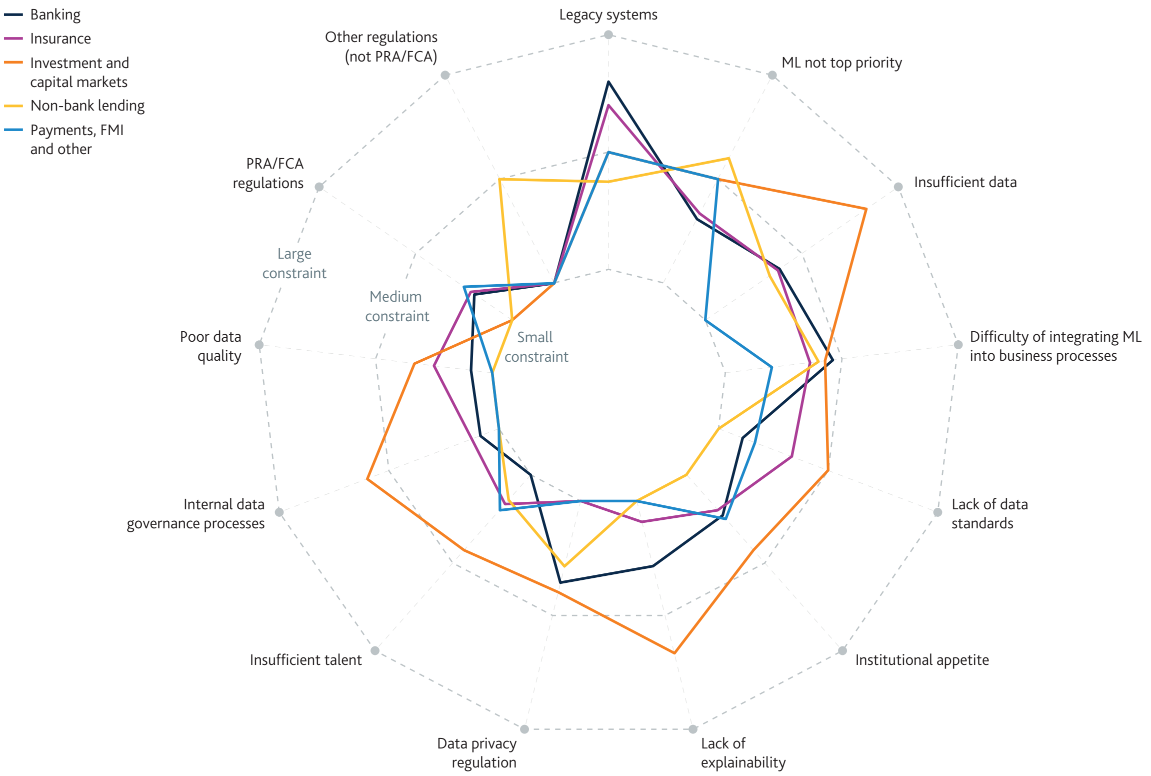


(a) Small constraint was allocated a score of 1, medium was 2 and large was 3.

However, it is important to note that, overall, respondents do not perceive there to be major constraints to ML deployment. The highest scoring constraint has been ranked only slightly above medium. This suggests firms do not consider the constraints, for example associated with older IT systems, to be insurmountable.

The ranking of constrains differs by sector, as shown in **Chart 12**.[22] Legacy systems are viewed by firms in all sectors as a major constraint to the deployment of ML applications, especially so in banking and insurance. This might be due to the sample being skewed towards larger and more established firms, which often cite legacy systems as a key barrier to innovation. Conversely, newer firms tend to have more agile IT architecture, which means they can to use ML applications more easily. Respondents indicate the difficulty of integrating ML into existing business processes as constraints of medium severity. Insurance firms and investment and capital market firms note the lack of data standards as a constraint.

Chart 12 Firms report constraints across a number of issues, but none are perceived as big



4.4 Regulation is not seen as an unjustified barrier

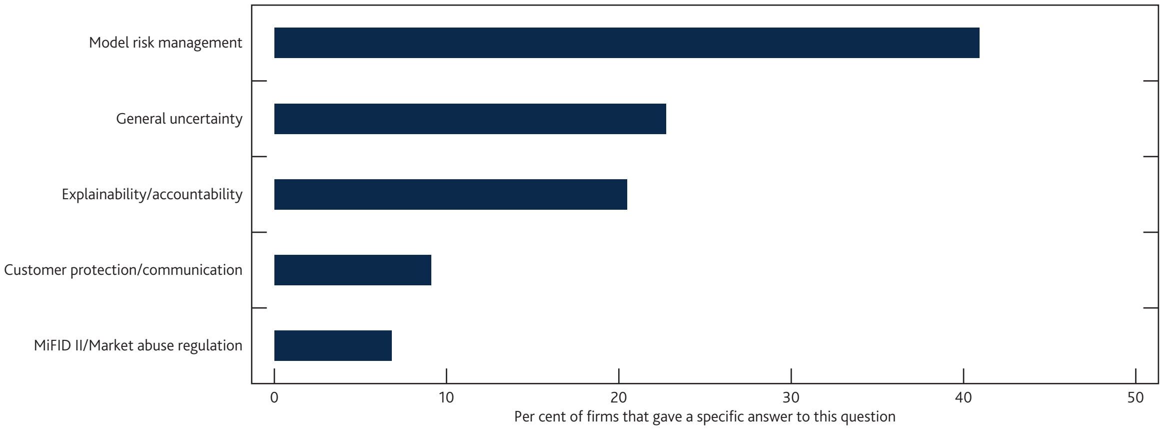
The majority of respondents (75%) do not consider PRA/FCA regulations to be an unjustified barrier when deploying ML. As shown in **Chart 11** and **Chart 12**, firms' stated that regulation is only a small barrier. It is also important to note that well-judged regulation is intended — by design — to be a barrier to certain practices in order to maintain financial stability or protect consumers. Therefore, this finding could indicate that in future regulation may need to be updated or adjusted to account for developments in ML.

Of the respondents that do consider PRA/FCA regulations to be a constraint, the most common issues cited are around model risk management and the need to adapt processes and systems to cover ML-based models. Some firms note the challenges of meeting regulatory requirements to explain decision making when using so-called 'black box' ML models (**Chart 13**)[23]. Also, some firms mention a lack of clarity and uncertainty around how existing regulations apply to ML, but did not further specify which regulations in particular.

Some firms stated that more clarity around ML deployment could serve as an enabler. Additional guidance could potentially help firms design controls, model risk management frameworks and policies for ML applications, as well as understand regulatory expectations for specific use cases.

(22) Small constraint was allocated a score of 1, medium constraint was 2 and large constraint was 3.
(23) Bracke, P, Datta, A, Jung, C and Sen, S (2019), 'Machine learning explainability in finance: an application to default risk analysis', *Bank of England Staff Working Paper No. 816*.

Chart 13 Firms identify model risk management as the one where regulatory constraints are most significant



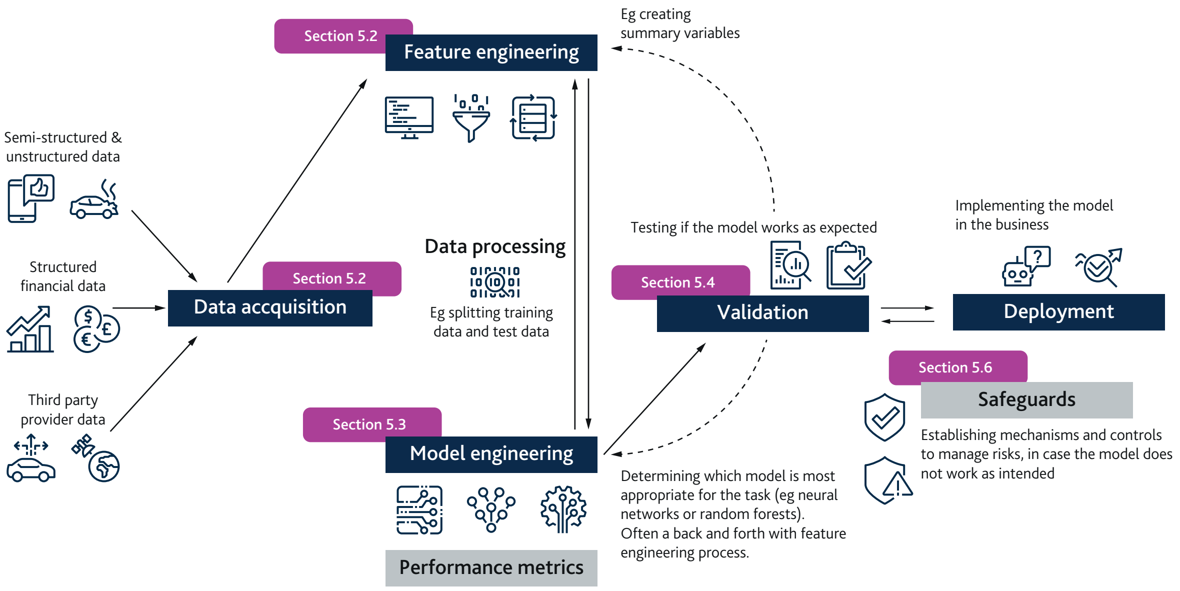
5 How machine learning works

5.1 Machine learning applications consist of a pipeline of processes

ML applications often consist of multi-step processes in which a number of distinct computational steps feed into each other(24) (**Figure 2**). The different stages of the ML pipeline are:

- Data acquisition and ingestion (section 5.2);
- Feature selection and engineering: Choosing the most relevant variables and creating derived ones (including for example through dimensionality reduction) (section 5.2);
- Model engineering and performance metrics (section 5.3): Model selection, optimisation of model parameters and model analysis (evaluation of model performance);
- Model validation (section 5.4): Testing if the model works as expected, which includes among other things the interpretation of how the model works;
- Deployment and safeguards (section 5.6): Implementing the model in the business and setting up safeguards to manage potential risks.

Figure 2 The machine learning pipeline



5.2 Data acquisition and feature engineering are evolving with the advent of machine learning

Different types of data

ML models learn from data of which there are three main types: (i) structured; (ii) semi-structured and (iii) unstructured.(25) **Figure 3** summarises the different features of these types of data.

In the case of structured data, each piece of information has a relatively narrowly defined meaning. For example, this could be data in standard relational databases and spreadsheets, such as a person's account balance.

(24) www.gartner.com/binaries/content/assets/events/keywords/catalyst/catus8/preparing_and_architecting_for_machine_learning.pdf.
(25) www.bigdataframework.org/data-types-structured-vs-unstructured-data/.

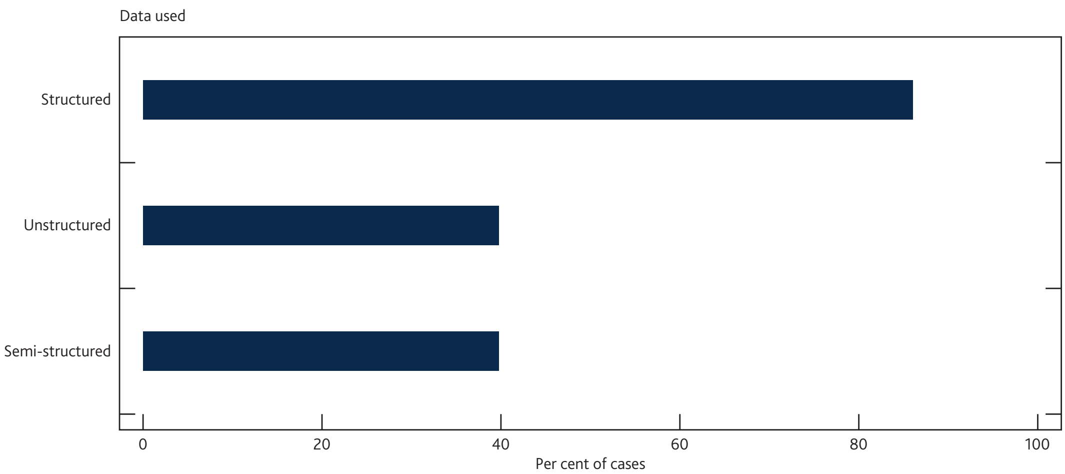
Semi-structured data are less pre-organised. For instance, the code behind a website structures contents into certain types of information (eg the sites' colour scheme or heading) but leaves room for less clearly pre-defined information.(26) Unstructured data has the fewest pre-defined fields. For instance, pixels in an image do not have a pre-defined meaning. It has to be inferred after the data is collected. It is this aspect that makes unstructured data harder to manage and analyse, requiring ML algorithms to extract (structured) meaning from the (unstructured) source.(27) This also makes data validation — making sure the data is accurate and reliable — more complex, as it may be ambiguous what the 'right' interpretation of the data is.

Figure 3 Firms make use of three types of data

Data type	Description	Example
Structured data	• Highly organised • Data objects have fixed meaning • Eg Relational databases or data organised in tabular format	**Standard financial database** First name / Second name / Age / Account balance A / B / 57 / 334 X / Y / 28 / 5,536
Semi-structured data	• Less organised than structured data, some hierarchy (tags, structure) present • Some data objects without fixed meaning • Eg HTML, JSON, XML	**Website** <!DOCTYPE html> <html> <head> <title>Page Title</title> </head> <body> Your text / button here" <button>Your Text Here</button>
Unstructured data	• Least organised • Information that does not follow a pre-existing data model • Requires analytical techniques to transform it into meaningful information	**Images or text** 

According the survey responses, structured data is used in more than 80% of ML use cases (**Chart 14**). This is unsurprising given most financial data is structured, as historically other types of data have not been collected and it was difficult to process with traditional linear models, frequently used in finance.(28) However, firms also use semi-structured or unstructured data in more than two thirds of cases, often in conjunction with structured data.

Chart 14 Structured data sources are still most popular, but firms are increasingly using novel data sources(a)(b)



(a) Firms often use more than one type of data at a time which is why the percentages add to more than 100.
(b) The underlying data is based on the use cases provided by survey respondents.

(26) www.datamation.com/big-data/semi-structured-data.html.
(27) www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/The%20age%20of%20analytics%20Competing%20in%20a%20data%20driven%20world/MGI-The-Age-of-Analytics-Full-report.ashx.
(28) eprints.lse.ac.uk/63017/1/Kallinikos_New%20Games%20New%20Rules.pdf.

Unlike more traditional models, ML is capable of processing semi-structured and unstructured data. Hence, firms use ML to transform text and image data into interpretable information. This also means that previously less used sources are now being analysed for important and potentially profitable uses. The increasing use of unstructured and semi-structured data also raises new questions for firms, consumers and regulators alike.(29) For instance, it increases the importance of data validation, both before and after deploying ML applications live in the market (**Chart 16**). And raises questions around ethics, fair use and privacy.

Feature engineering

In several use cases, survey respondents use thousands of variables in their ML models. However, these variables are often part of the pre-processing phase, which includes standard data cleaning techniques (like dealing with outliers) as well as 'funnelling' numerous different variables into composite ones. Importantly, ML algorithms are used to perform this task, including dimensionality reduction methods and clustering methods, which we cover in section 5.3.

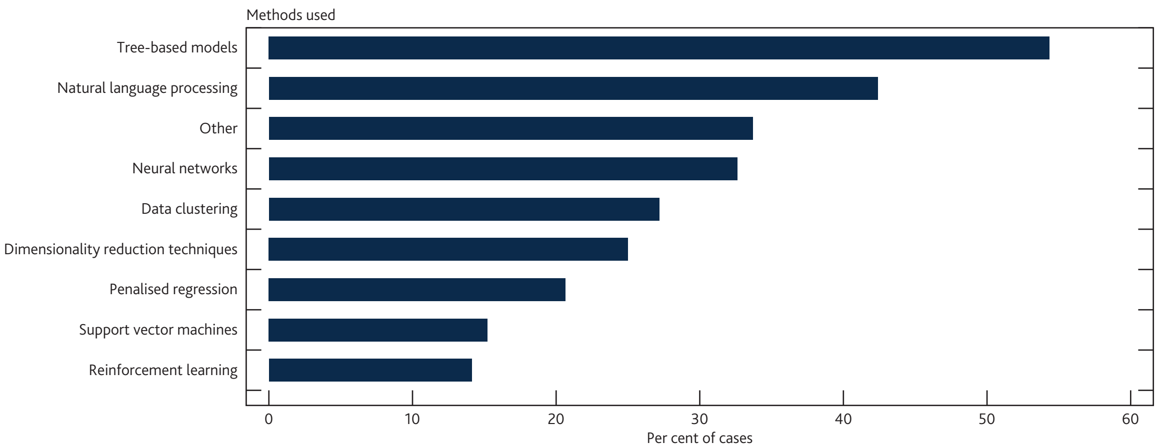
5.3 Model engineering and performance evaluation decide which models are deployed

Types of machine learning algorithms

Model engineering includes the selection of the most appropriate algorithm and training of the model, all of which is an iterative process (see Box 4 for an explanation of different ML methods). For instance, in some contexts — especially those where the amount of available data is limited — simple linear regression techniques may be most effective. In other contexts — for instance those where a large amount of complex, unstructured data are available — neural networks may be most effective.(30)

According to firms' responses, the ML methods most often used are on the more complex end of the current spectrum.(31) The most common ML methods are tree-based models; natural language processing approaches and neural networks (**Chart 15**). Models in the 'other' category included Bayesian approaches or image recognition.

Chart 15 Tree-based methods are the most popular techniques reported by firms(a)(b)



(a) Firms often use more than one method at a time which is why the percentages add to more than 100.
(b) The underlying data is based on the use cases provided by survey respondents.

(29) www.fsb.org/wp-content/uploads/P011117.pdf.
(30) www.imf.org/en/Publications/WP/Issues/2019/05/17/FinTech-in-Financial-Inclusion-Machine-Learning-Applications-in-Assessing-Credit-Risk-46883.
(31) www.d2l.ai/chapter_multilayer-perceptrons/underfit-overfit.html#model-complexity.

Box 4
Machine learning methods[1]

Penalised regression methods are standard regression methods, in which an algorithm picks the variables that are contained in the model. This is usually done by dropping variables that are not needed for prediction. These models are at the least complex and most interpretable end of the spectrum.

Tree-based models consist of a multitude of (often large) decision trees whose individual predictions are averaged. It works for both categorical and continuous input and output variables. Unlike linear models, tree-based models can map non-linear relationships.

Neural networks are algorithms modelled loosely on aspects of the brain's neurons, designed to recognise patterns and make predictions. Modern neural networks often involve estimating a large number of weights, which increase in number as more 'layers' are introduced.

Natural language processing involves the application of algorithms — often neural networks — to identify and extract the natural language rules such that unstructured language data is converted into a form that computers can understand.

Dimensionality reduction techniques reduce the number of variables under consideration by obtaining a set of principal variables. Approaches can be divided into feature selection and feature extraction.

Support vector machines (SVM) are supervised learning models that analyse data used for classification and (continuous) regression analysis. Given a set of training examples, each marked as belonging to two categories, a SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

Reinforcement learning methods are concerned with how virtual agents choose their actions in order to maximise a reward function as defined by a human. These methods do not require labelled input/output pairs and sub-optimal actions need not be explicitly corrected. Instead the focus is finding a balance between exploration (of uncharted territory) and exploitation (of current knowledge).

(1) James *et al* (2017), 'An introduction to statistical learning'. Goodfellow, I, Bengio, Y, and Courville, A (2016), 'Deep learning', MIT Press. Abu-Mostafa, Y, Magdon-Ismail, M, and Lin, H-T (2012) 'Learning from data: a short course'.

Firms often use tree-based approaches, such as 'random forests'. These consist of a multitude of (often large) decision trees whose individual predictions are averaged. These methods have been shown to be relatively successful for prediction in traditional financial data analysis contexts (such as price forecasting).[32] Natural language processing models are able to analyse unstructured text data, which lends itself well to customer service and insurance claims management use cases[33] (see the case studies in the Appendix for more information). Neural networks are used, among other things, to make forecasts based on historical information and find complex relations between variables. Most respondent firms' applications use, on average, a combination of three ML methods. In one use case, a firm uses eight separate ML techniques in a single application.

(32) www.researchgate.net/publication/333409685_Stock_Market_Analysis_A_Review_and_Taxonomy_of_Prediction_Techniques.
(33) www2.deloitte.com/us/en/insights/industry/financial-services/artificial-intelligence-ai-financial-services-frontrunners.html.

ML methods are more difficult to interpret than traditional linear regression models. The reason is that many ML models are 'non-parametric'(34), which makes them more difficult to explain — essentially more complex. As highlighted in section 4.2, firms think that this increased complexity makes model validation harder, which can translate into a potential risk. Validation methods, highlighted below, can address this, but new methods will likely be required, as ML techniques develop.

Performance metrics have multiple purposes

Performance metrics serve at least three purposes in the ML pipeline:

- They are used to pick the best model, which can be either a human led or automated process.
- These metrics are key for understanding how well the model is likely going to perform once deployed.
- They can be used to track the performance of the model over time. Checking the performance over time can be important for detecting structural changes that make the model less accurate.

5.4 Model validation is key to ensuring machine learning models work as intended

At the core of the ML pipeline is making sure that the application works as intended in practice. This is the issue of software validation. In **Table 3**, we use the aggregated survey responses to explain how firms do this in practice, with their own ML applications. Any of these methods might be used in the pre-deployment phase (where the application is being tested) or post-deployment (where the application is live in the market), as a way to continuously assess if the model works as intended.

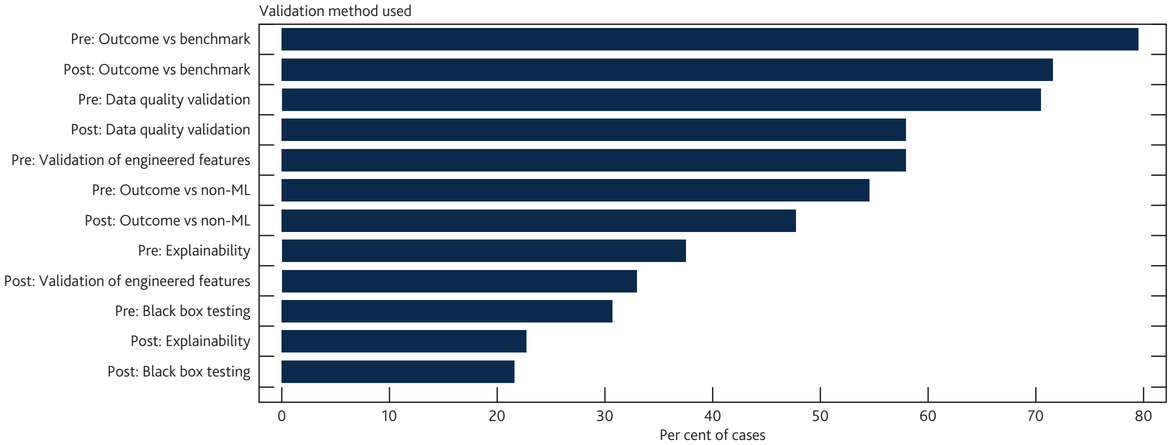
Table 3 Firms use a variety of model validation techniques to assess machine learning model robustness

Validation method	Description
Outcome monitoring against a benchmark	Decisions or actions associated with the ML system are monitored using one or multiple metrics. Performance is assessed against a certain benchmark value of those metrics.
Outcome monitoring against non-ML model/ A-B testing	Decisions or actions associated with the ML system are monitored using one or multiple metrics. Performance is assessed by comparing it to the performance of a separate, non-ML model. The same approach is used in A-B testing (also known as split testing).
'Black box' testing	Input-output testing without reference to the internal structure of the ML application. The developer 'experiments' with the model, feeding it different data inputs to better understand how the model makes its predictions.
Explainability tools	Tools aimed at explaining the inner workings of the ML model (going beyond input-output testing).
Validation of engineered features	Engineered features used in the ML application are scrutinised, including potential impacts on model performance.
Data quality validation	One or more techniques are used to ensure potential issues with data (such as class imbalances, missing or erroneous data) are understood and considered in the model development and deployment process. Examples of these include data certification, source-to-source verification or data issues tracking.

In **Chart 16**, we summarise which ML model validation techniques and frameworks are most frequently used (as described in **Table 3**). The most common method is outcome-focussed monitoring and testing against benchmarks, both before and after deployment. This enables firms to scrutinise how ML models would have performed historically in terms of profitability, customer satisfaction or pricing, for example. Data quality validation — including detecting errors, biases and risks in the data — is the next most frequently used method. Overall, these methods were used by two thirds of respondents. In about half the cases outcomes were benchmarked against a non-ML model. Explainability techniques(35) were used in less than half of the cases. However, many firms emphasise that validation frameworks still need to evolve to address challenges associated with the nature, scale and complexity of ML applications. Therefore the use of some validation techniques may increase in the future.

(34) In non-parametric models, the data is not required to fit a normal distribution and does not rely on numbers, but rather on a ranking or order of sorts. wwwf.imperial.ac.uk/~nsjones/TalkSlides/GhahramaniSlides.pdf.
(35) Bracke, P, Datta, A, Jung, C and Sen, S (2019), 'Machine learning explainability in finance: an application to default risk analysis', *Bank of England Staff Working Paper No. 816.* Joseph, A (2019), 'Shapley regressions: a framework for statistical inference on machine learning models', *Bank of England Staff Working Paper No. 784*.

Chart 16 Outcome-based validation methods are the most common[(a)(b)(c)]



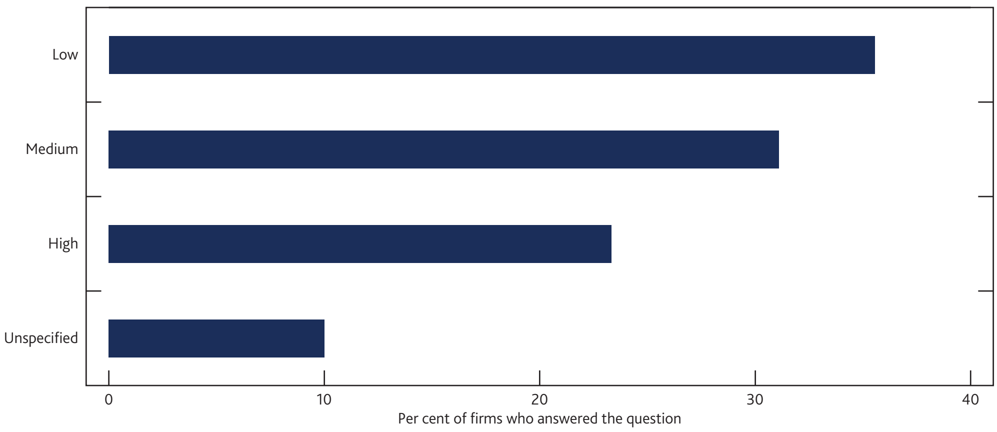
(a) 'Pre' indicates pre-deployment use of the validation method. 'Post' indicates post-deployment use of the validation method.
(b) Firms often use more than one validation method at a time which is why the percentages add to more than 100.
(c) The underlying data is based on the use cases provided by survey respondents.

5.5 Complexity can increase due to deployment of machine learning

We asked firms a range of questions about the complexity of their ML applications. We did so because complexity, in principle, can influence a firm's risk profile and therefore the supervisory approach (Box 3, Section 3.3).

Firms often mention that it is difficult to clearly define what 'complexity' means but attempt to make an assessment based on the number of components, number of data sources and algorithms in the ML model. Based on this, respondents gave their best estimates of the complexity which were then grouped in three main categories ('Low', 'Medium', and 'High'). In cases where ML models were provided by a third-party, firms stated that it made it more difficult to assess the degree of complexity. As summarised in **Chart 17**, more than half of applications are considered to be of medium to high complexity.

Chart 17 More than half of applications are considered to be of medium to high complexity[(a)]



(a) The underlying data is based on the use cases provided by survey respondents.

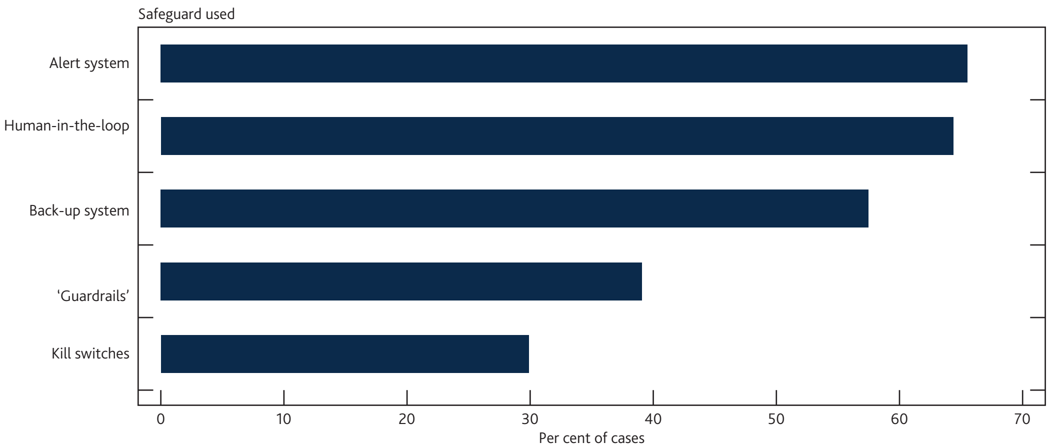
A minority — about one tenth — of firms report having ML systems that comprise at least three separate components. This can add to both the size of the overall ML pipeline (see **Figure 2** in section 5.1) and the complexity of the application. Such components can include separate data processing, analytics and decision-making engines. Generally these components are developed in-house and some others provided by third parties, which might create additional complexity when it comes to ensuring the smooth and robust interplay between different components of a ML system.

In addition, firms report using a large number of data sources and it is not uncommon to feed more than 100 different variables into a ML application.

5.6 Firms use a range of safeguards to address risks

Firms use a range of mechanisms and controls to manage the risks associated with ML applications. However, the additional complexity, issues with explainability and the continuous lifecycle of ML introduces new challenges, which require safeguards. Overall, the most common controls among survey respondents are alert systems and so-called 'human-in-the-loop' mechanisms (**Chart 18**). The former are systems that flag unusual or unexpected actions to employees. The latter are systems where decisions made by the ML application are only executed after review or approval from a human. 40% of use cases have 'guardrails' in place which switch off the model automatically if it produces undesired outputs. This is intended to mitigate the potential for model drift, which can occur as ML algorithms self-teach and make decisions that are outside their original parameters.

Chart 18 Alert systems and human-in-the-loop are the most common safeguards[(a)(b)]



(a) Firms often use more than one type of safeguard at a time which is why the percentages add to more than 100.
(b) The underlying data is based on the use cases provided by survey respondents.

6 Conclusion

6.1 Context

This joint survey and report constitute a first step towards deepening our understanding of the use of ML in UK financial services. This includes a deeper appreciation of the state of deployment, such as the specific business areas where ML is used and how mature it is, the different approaches to strategy and risk management, as well as the potential benefits, barriers and risks. The findings published in this report also serve as a basis for exploring technical issues, for instance around ML model validation and safeguards.

In addition to deepening our own understanding, feedback from participants suggests the survey was useful for firms and helped them better understand how ML is used within their own organisations. With this in mind, the BoE and FCA are considering repeating the survey in 2020 so we can continue to deepen our collective knowledge of ML and track its deployment in the UK financial system.

In conducting the survey and publishing this report, we are seeking to step up the dialogue with firms, academics and other regulators about how we can support the safe and robust use of ML in financial services. This is a research project and not designed specifically for policy development. However, the survey findings and subsequent dialogue can help provide a platform for identifying where regulation may help support the safe, beneficial, ethical and resilient development and deployment of ML both domestically and internationally.

6.2 What we have learnt

By carrying out the survey, analysing the results and engaging with firms, we found that:

- **ML is increasingly being used in UK financial services**. The majority of firms in our sample are making use of this technology. The expected doubling of the median number of applications further suggests that ML is likely to become an increasingly integral part of financial services.

- **Firms report they are benefitting from the deployment of ML**, in particular with regards to efficiency gains, better customisation of products and more effective combating of fraud and money laundering. If ML deployment can produce these reported benefits, it will help the UK financial system be more effective in serving the needs of the real economy, while at the same time delivering in the interests of consumers and ensuring market integrity.

- **In many cases, firms' ML deployment has passed the initial development phase**. We learnt from respondents that, in one third of use cases, ML is used in a relatively mature way. This means that firms' experience with this technology is developing. This mostly seemed to be in support of existing human decision making processes.

- **ML is moving beyond back-office operations and is being used in front office and customer-facing functions**. While ML is most frequently used in back office areas, there is a growing share of applications that have emerged in core business areas — such as credit risk, market risk assessment and in insurance underwriting.

- **While firms highlighted possible risks from the use of ML, they also indicated what could be done to address them**. Most often firms mentioned the need to conduct in-depth validation exercises to make sure complex ML models work as intended.

- **Firms are keen to improve their validation frameworks for testing ML models work as intended**. For instance, this includes approaches that help to make ML models more explainable.

- **Finally, some firms stated the need for additional clarity on how existing regulations apply to ML**. As with any technology that is new, it may not be obvious how existing norms and rules apply to it. Accordingly, several firms thought that regulatory expectation-setting on best practices around ML use would be helpful and could promote greater deployment.

6.3 Questions for authorities

The report findings are consistent with the view that ML will be an important part of the way financial services are designed and delivered in the future. As a general purpose technology, it will potentially be used in areas critical to financial markets and the safety, soundness and conduct of firms. But it will also likely be usefully applied in areas that are not critical from a financial regulation point of view. The task of the BoE and the FCA will be to continue monitoring the application of ML and identify ways to support the safe, beneficial, ethical and resilient deployment of the technology across the UK financial sector, as well as understanding its impact on the wider economy.

Firms are best placed to make decisions on which technologies to use and how to integrate them into their business. However, regulators will seek to ensure that firms identify, understand and manage the risks surrounding the use of new technologies, and apply the existing regulatory framework in a way that supports good outcomes for consumers.

This likely requires that regulators also engage with, and build up, an understanding of the technical aspects of ML, including those highlighted in this report. For instance, these could be issues around model risk management, which some respondents cite as a specific constraint to the deployment of ML. There are also questions regarding software and data validation, the governance, the resilience and security of ML applications within financial services, as well as potential ethical issues that arise from the use of ML and novel data sources. On the last point, we will continue to collaborate with the Information Commissioner's Office and other domestic authorities.

6.4 Next steps

This survey constitutes a first step towards better understanding the impact of ML on UK financial services and forms the basis for a conversation around how safe ML deployment can be supported going forward. As announced by Governor Carney in his Mansion House speech(36), based on the survey findings and with our increased level of understanding, we will explore potential policy areas relating to ML. In order to facilitate this dialogue, the BoE and the FCA have announced they will establish a public-private working group on AI to further the discussion on ML innovation as well as explore some of the questions above and technical areas covered in this report. We will also consider repeating this survey in 2020.

(36) Carney, M (2019), 'Enable, empower, ensure: a new finance for the new economy'.

7 Appendix — case studies

7.1 Purpose and background

The case studies included in this report are based on survey responses and offer a narrow perspective on how ML is applied in practice. These are intended as purely illustrative. They do not represent a view by the regulators on the rationale for using such technology, how it should be done or whether it is the right thing to do. No views are expressed as to the compliance with regulation, whether financial services or otherwise.

7.2 Methodology

The survey asked firms to provide information on two ML case studies within their organisation. These were broadly structured around the sections in Chapter 5: (i) description of ML application; (ii) data and methods; (iii) complexity; (iv) performance evaluation and testing and (v) safeguards. We did not ask for information on compliance with data protection law.

7.3 Anti-money laundering and countering the financing of terrorism

Description

Financial institutions continuously analyse customer data from a wide-range sources as part of their AML functions and their countering-the-financing-of-terrorism process. In the applications considered for this case study, ML is used at several key stages within the process to:

- Analyse millions of documents and check details against 'blacklists'[37] for the know-your-customer (KYC) checks before the on-boarding process begins.
- After this initial stage, banking firms are increasingly using ML to rate the likelihood of a customer posing a financial crime risk.
- As customers transfer money or make payments, firms use ML to identify suspicious activities and flag potential cases, so human analysts can focus on these specifically.

Data and machine learning methods

Given the high volume of text data involved in KYC checks and need to identify specific names, addresses, etc., most applications use NLP. Tree-based methods are also used to cross-reference the historical decisions made by analysts. It was mainly banking firms that provided examples of ML tools to monitor transactions and identify anomalies. These applications use structured payments systems data and tree-based models and neural networks, which are often developed in-house.

Performance evaluation and testing

ML-based models can handle larger volumes of data and can yield lower false positive rates compared to individual analysts and traditional systems.[38] To quantify these rates, firms use outcome monitoring against benchmarks during both the development and deployment phases of KYC and alert processing tools. Where ML applications automate more of the decision-making process, explainability becomes more of a priority for firms. Respondents say they break down the unsupervised learning procedure of neural networks in order to justify why a particular customer or transaction is flagged.

(37) www.oecd.org/countries/monaco/list-of-unco-operative-tax-havens.htm.
www.fatf-gafi.org/countries/#high-risk.
(38) www.iif.com/Publications/ID/1421/Machine-Learning-in-Anti-Money-Laundering.

Complexity
Respondents pointed to the management of feedback loops as the most complex aspect of KYC solutions. For transaction monitoring, the main complexity issues arise from the management of IT infrastructure and the oversight of data pathways and validation. From our sample, we also see that tools of a high technical complexity often combine a range of ML methods to draw insights on customers. The input data is of all structures, and the explainability of the 'learning process' is of great interest to firms deploying such tools.

Safeguards
For the KYC tools, we observe that human analysts continue to play a decisive role in the process. Once alerts are raised, analysts can narrow their focus to these more relevant sources. At the more advanced end, tools have the capacity to output a 'next step' for the analyst, who may agree or disagree with the decision. Firms say this helps improve the performance of the model because the system will adapt and refine its options on further use depending on the human decision. In transaction monitoring, firms use less interpretable ML methods(39) and all five safeguards to mitigate the corresponding risk.

7.4 Customer engagement

Description
A typical ML customer engagement application enables customers to contact firms without having to go through human agents via call centres or customer support — these systems are more commonly known as 'chatbots'. Firms report these applications can reduce the time and resources needed to resolve consumers' queries.

In addition, ML can facilitate faster identification of user intent and recommend associated content, as well as transfer the consumer to a human agent as and when they are better placed to deal with the query.

Data and machine learning methods
Firms use a combination of unstructured, semi-structured and structured data, which reflects the type of customer engagement methods like live chat, phone calls and online forms. NLP is typically used in customer engagement applications, as it allows to analyse and extract data and information from vast amounts of text.

The whole process from development to deployment is often managed in-house. The development and the model is in some cases outsourced to third party providers, this includes the development of the underlying platform and infrastructure.

Complexity
The complexity of customer engagement applications lies mainly in the emulation of a human agent behaviour with the aim of containing and fulfilling the user's request. The ML models learn from the feedback of previous interactions with customers, collecting historical data and establishing a classification of customers' preferences, reactions, patterns and behaviours.

Performance evaluation and testing
The performance of these applications tends to be measured based on the reaction times or success rate of customer engagements. For example, the number of users that have their request answered successfully using the chatbot without the need for a referral to a human agent.

7.5 Sales and trading

Description
According to the survey responses, ML use cases in sales and trading broadly fall under three categories ranging from client-facing to pricing and execution.

(39) Bracke, P, Datta, A, Jung, C and Sen, S (2019), 'Machine learning explainability in finance: an application to default risk analysis', *Bank of England Staff Working Paper No. 816*.

- For client-facing activities, firms use ML to increase speed and accuracy of processing orders. For instance, firms may use NLP to process quotes received from clients, allowing for shorter response times.
- In pricing, ML models combine a large number of market time-series to arrive at an estimate of a short-term fair value.
- In execution, ML applications evaluate venue, timing and order size choices. Within this, ML may also be used for intermediate steps of the process; for instance, for calculating the probability of an order being filled given the available characteristics of the order. Firms use ML techniques to determine order routing logic, this is often contained within systems called smart order routers or broker/algo wheels. This can include the evaluation of venue, broker and execution algorithms, as well as determining the timing, price and size of particular orders. Within these, ML may also be used for intermediate steps of the process; for instance, for calculating the probability of an order being filled given the available characteristics of the order and prevailing market conditions.

Data and machine learning methods
Data used in these cases is still largely of a traditional, structured type, such as financial time series that is also used for non-ML models. Some firms also use unstructured data, such as text data, which can be used in the context of estimating prices in illiquid markets.

Respondents often use tree-based approaches, such as ‘random forests’ and claim they are successful at generating better predictions, such as price forecasting. However, the size and complexity of these models makes it difficult to explain exactly how they work and what the key variables are that drive predictions. Regression techniques with ML elements continue to be popular in this type of use case, and provide a relatively higher degree of explainability.

Performance evaluation and testing
Firms use a range of methods to validate ML applications. Next to standard predictive accuracy metrics, most common are outcomes-based tests that compare the ML application’s outputs with those of a benchmark or a more established model. A-B testing is a related approach that is also popular, involving a forensic comparison of ML and non-ML model outputs. Explainability methods are less frequently used. This is in line with many current approaches for validation (eg back-testing), which also are outcomes based.

Complexity
As in our overall sample, the complexity of applications in this field varied widely. Some are complex because they use advanced and hence complex ML approaches, such as reinforcement learning. Others have less complex techniques, but the applications consist of a number of separate but interacting components, increasing the systems’ overall complexity. In one case the application consists of more than 30 components. Others still highlighted the complexity of the data processing part of the ML pipeline as a source of medium complexity. Yet, several firms report limited complexity of their ML applications, eg when based on a small model, in a fairly narrow area.

Safeguards
Firms reported having alert systems in place highlighting when the ML applications are engaging in unexpected behaviour. Others highlighted that there continues to be a human in the loop, overseeing key decisions made by algorithms.

7.6 Insurance pricing

Description
The majority of respondents in the insurance sector use ML to price general insurance products, such as motor, marine, flight, building and contents. More specifically, firms’ use ML applications for risk cost modelling and propensity modelling within the price optimisation process.

- For risk cost modelling, firms use ML to analyse new data sources, such as geospatial data, and build the underling risk cost models to gain an understanding of the expected claims cost of an underwritten policy. This information is used in live rating, but also for technical pricing.
- For propensity modelling, ML can be used to predict product add-on selections, customer demands and estimated future claims costs, which can influence renewal premiums offered to existing policyholders.

Data and machine learning methods

Based on the survey responses, firms combine existing structured data and new data for pricing. Structured data includes internal written customer policy and quote data, as well as external databases (eg DVLA for motor insurance). In more than 70% of cases considered for this case study, the respondents used data collected by third parties, such as price comparison websites or industry-specific data providers (eg flight routes).

Tree-based methods were popular in this type of application, where multiple inputs are analysed to create an aggregated single risk price. Risk cost modelling applications in our sample consider a suite of decision trees for different aspects of the process, such as frequency or loss. In some cases, the ML application builds the underlying cost model (using non-linear methods), with multiple models aggregated into a single risk price. For propensity modelling, firms sometimes used tailor-made ML models to capture the key drivers of each customer route and project them forward into the future.

Complexity

The firms we surveyed thought that ML models, such as tree-based ensemble methods, can be more complex than generalised linear models. However, there were disparate views about the level of complexity introduced to the overall pricing process through ML. Some firms said the number, structure and complexity of features is similar to existing linear pricing models. Other firms think additional complexity is introduced by the number of different models that need to be aggregated up into a single risk price, as well as the need for real-time latency in order to adhere to price comparison website requirements or reflect certain features such as varying policy exposure periods.

Performance evaluation and testing

Firms use a variety of methods to validate ML applications both during development and deployment phases. Many of these are outcome-based tests to compare performance to existing linear models and benchmarks. In every use case, firms test the data quality during the development phase to avoid overfitting, bias and discrimination. The ML performance and predictive accuracy is also measured during the development phase by using suitable outcome quality and error metrics. Once implemented, firms continue to measure the ML performance within pricing as part of their general processes to monitor product performance.

7.7 Insurance claims management

Description

Out of the firms we surveyed in the general insurance sector, 83% use machine use ML for claims management. According to respondents, there are two key applications within the claims management process:

- ML applications analyse photos and unstructured data sources to extract the relevant management information from the raw data and predict the estimated cost. The ML application then uses historical data to compare this to the predicted total loss cost, and then make a decision as to which is the correct route for claims handler to follow.
- ML applications use predictive analytics to target claims that have a high likelihood of customer dissatisfaction or complaint, in which case they are flagged so a human can monitor the claim and intervene if required.

Data and machine learning methods

Firms use a combination of structured, unstructured and semi-structured data sources for claims management. The structured data is largely internal information from claims systems, such as notification forms and incident types, and policy details. The unstructured data is often submitted by the policy holders and ranges from images to location and sensor data depending on the type of claim. Firms combine these data sources with third party data from different industries, such as auto repair costs, to assist with the follow up information. Firms also analyse free text claims that are filled in by consumers or claims handlers.

Given the variety of data sources, survey respondents also use a range of ML methods. The most common are tree-based methods, including random forest and gradient boosted tree, which are used to assess the impact of different inputs. Firms also use NLP to review claims, verify policy details and pass them through a fraud detection algorithm before sending wire instructions to the bank to pay for the claim settlement. Some firms also use optical character recognition to analyse image data and handwritten claims notice documents.

Complexity

The majority of firms consider the employed ML methods to be complex, especially when tree-based models and neural networks are used. Some firms also noted data gathering was a complex task. However, almost all firms agree the application itself is relatively simple given they can monitor the performance on a regular basis. Firms are able to query a database, pre-processes data, obtains predictions from the application, and writes them back to the database. Similar strategies assess simple data manipulation, scoring of models, response and logging of all data used in call.

Performance evaluation and testing

Firms tend to track and evaluate the ML performance by using a range of metrics, such as volumes of decisions and total loss percentage over time. The performance is compared to traditional models. For predictive claims, the predictions are tested against the actual performance. This occurs either on a case-by-case basis or sample data and the overall goal is to prevent over-fitting and model degradation.

7.8 Asset management

Description

To date, ML often plays a supporting role in asset management only. Systems, as described below, are often used to provide suggestions to fund management. This equally applies equally to portfolio decision-making or trade execution.

ML applications are used for a range of processes within asset management:

- Analysing large amounts of data from diverse sources and in different formats.
- Digesting large selection of inputs to assist in establishing a fair market price for a security.
- Supporting decision-making processes by linking data points and finding relationships across a large number of sources.
- Sifting through vast amounts of news feeds and extracting useful insights.

Data and machine learning methods

Typically, asset management applications use structured and unstructured data, often collected by third parties. Neural networks are often the preferred methodology of choice. Although depending on the use case, a combination of different methods are deployed.

Reliance on third party providers can be limited but models and the underlying infrastructure or platforms are sometimes purchased from external providers.

Complexity

Given the objective of analysing vast amounts of data to offer a simplified framework for the user, the applications considered for this case study can be considered of medium complexity. Applications using multiple components often present higher levels of complexity, for instance, due to combining different data sources in a single environment.

Performance evaluation and testing

Firms validate applications in a broad range of ways, often combining a mix of methods, both during the development stage and the deployment phase of the application. The performance is evaluated at every stage, by testing the model against historical data and real-time performance relative to simulated results.

Safeguards

All applications we considered for this case study have back-up systems and human in the loop safeguards. As noted above, these applications are aimed at providing a set of suggestions to fund managers, with a human in charge of the decision making and trade execution.

Acknowledgements

The authors of this report are Carsten Jung (BoE), Henrike Mueller (FCA), Simone Pedemonte (FCA), Simone Plances (FCA) and Oliver Thew (BoE).

We are grateful to Andrew Bailey, Sir David Ramsden, Sam Woods, Chris Woolard, Lyndon Nelson, Tom Mutton, Nick Cook, Samantha Emery, Louise Eggett, Katherine Browne, Ashley Collins, Eir Nolsøe, Kitty Eyre, Hugh Miller, Giorgio Saladino, Irina Mnohoghitnei, Theo Bourgery, Laura Navaratnam and Simone Bebbington for their useful comments.

We would like to thank Khushali Shingala, Ewen Henderson, Alice Parker, Eryk Walczak, Melvin Quimis, Arjun Ahluwalia for contributions regarding data analysis and visualisation. And we would also like to thank Lindsey Fowler, Helen Martyn, Wayne Chapman and Pete McCabe for their tireless work in getting the report ready for publication.

We are also grateful to colleagues from across the BoE, FCA and PRA for their input, including the supervisors of surveyed firms for their support throughout the process. Finally, we would like to thank all of the firms that participated in the survey for their input.