DEUTSCHE BUNDESBANK
EUROSYSTEM

BaFin
Bundesanstalt für
Finanzdienstleistungsaufsicht

Machine learning in risk models – Characteristics and supervisory priorities

Consultation paper

Table of contents

I. Machine learning is advancing

The debate surrounding the use of artificial intelligence and machine learning (collectively referred to as ML methods for short) has recently been gaining momentum, precisely in the area of financial services, fuelled by the availability of "big data" together with enhanced computing power. Standardised principles and procedures[1] have made the development of ML methods simple and accessible.

The use of ML methods can help to quantify risks more accurately and enhance process quality, thereby improving financial firms' risk management.

These issues have already been addressed by both BaFin and the Bundesbank in several publications. On 16 July 2018, BaFin submitted its report entitled "Big data meets artificial intelligence – challenges and implications for the supervision and regulation of financial services"[2] (BDAI report for short) for public consultation, and on 28 February 2019, it published an overview of the results along with an initial assessment.[3] The BDAI report demarcated the field of digitalisation and identified challenges. In March 2020, BaFin explained its misgivings about a general approval requirement for algorithms and outlined how ML can be embedded in the risk-oriented supervisory approach.[4] These considerations are fleshed out in the "Principles for the use of algorithms in decision-making processes" published in June 2021.[5] In its discussion paper "The Use of Artificial Intelligence and Machine Learning in the Financial Sector"[6] published in November 2020, the Bundesbank set out basic theoretical considerations on how to deal with ML methods in the context of prudential supervision, in which, for instance, it derived inferences on the intensity of supervision of ML and put into context the significance of the explainability of ML methods.

Supervisors and regulators across the globe are also looking into ML methods. Noteworthy are papers published by De Nederlandsche Bank, which has developed principles for the use of ML,[7] and France's ACPR, which centres on the explainability of ML.[8] The EBA[9] and the Bank

[1] These include DevOps, MLOps and software libraries which contain standard ML methods.
[2] BaFin, 2018, "Big data meets artificial intelligence", available online at: https://www.bafin.de/SharedDocs/Downloads/EN/dl_bdai_studie_en.pdf?__blob=publicationFile&v=11
[3] BaFin, 2019, "Big data meets artificial intelligence – results of the consultation on BaFin's report", available online at: https://www.bafin.de/SharedDocs/Veroeffentlichungen/EN/BaFinPerspektiven/2019_01/bp_19-1_Beitrag_SR3_en.html
[4] BaFin, 2020, "Does BaFin have a general approval process for algorithms? No, but there are exceptions", available online at: https://www.bafin.de/SharedDocs/Veroeffentlichungen/EN/Fachartikel/2020/fa_bj_2003_Algorithmen_en.html
[5] BaFin, 2021, "Big data and artificial intelligence – principles for the use of algorithms in decision-making processes", available online (but in German only) at: https://www.bafin.de/SharedDocs/Downloads/DE/Aufsichtsrecht/dl_Prinzipienpapier_BDAI.html
[6] Bundesbank, 2020, "Policy Discussion Paper, The Use of Artificial Intelligence and Machine Learning in the Financial Sector", available online at: https://www.bundesbank.de/resource/blob/598256/d7d26167bceb18ee7c0c296902e42162/mL/2020-11-policy-dp-aiml-data.pdf
[7] De Nederlandsche Bank, 2019, "General principles for the use of Artificial Intelligence in the financial sector", available online at: https://www.dnb.nl/media/jkbip2jc/general-principles-for-the-use-of-artificial-intelligence-in-the-financial-sector.pdf
[8] Autorité de contrôle prudentiel et de résolution (ACPR), 2020, "Governance of Artificial Intelligence in Finance", available online at: https://acpr.banque-france.fr/en/governance-artificial-intelligence-finance
[9] EBA, 2020, "Report on Big Data and Advanced Analytics", available online at: https://www.eba.europa.eu/eba-report-identifies-key-challenges-roll-out-big-data-and-advanced-analytics

of England[10] have likewise published their views of data-driven analyses and ML. The paper published by the Hong Kong Monetary Authority has placed a practice-oriented focus on the hurdles involved in implementing ML.[11]

This consultation paper builds on earlier national and international publications and makes connections between the prudential risks of ML and current supervisory practices. This approach not only simplifies the background but also follows the process that financial corporations undergo when introducing ML methods.

The focus of this paper is on solvency supervision, and specifically the application of ML methods in areas of particular relevance to supervisors. These include, on the one hand – as an exception to the principle that algorithms do not require supervisory approval – ML methods that are used in prudential inspections and approval procedures, and thus in internal models for calculating regulatory own funds requirements (Pillar 1) and, on the other hand, those that are used in risk management under Pillar 2.

In this context, consumer protection aspects and the ethical issues of ML play a relatively minor role and will therefore be disregarded in the following.[12]

There is no uniform definition of machine learning due, first, to the large number of different approaches and, second, to the lack of a clear dividing line to traditional techniques.[13] However, ML methods often have certain characteristics that are particularly strongly pronounced and which thereby set them apart from traditional techniques. The intended aim of this paper is therefore to identify such characteristic traits of ML methods which have relevant implications for supervisors and to come up with ideas on how supervisory practices could evolve in order to be able to respond to the risks involved.

This also poses the question as to whether not only supervisory practices but also the regulatory foundations themselves need to be reworked and whether or not it may be necessary to create a fundamentally new supervisory approach for ML methods. Without wanting to "jump the gun" here: since the current regulatory foundations are worded in a technology-neutral manner, they are largely transferable to the characteristics of ML methods, with only a few places where it might be necessary to adapt the regulatory foundations.

Although this paper refers below primarily to "banks", the characteristics and prudential implications presented here can, in principle, also be applied to insurers and other enterprises engaging in the development and implementation of ML methods for regulated financial services. This paper is intended as a consultative document in order to launch a discussion process with the industry: it contains blocks of questions, the responses to which are designed to advance supervisory practice.

[10] Bank of England, 2019, "Machine learning in UK financial services", available online at: https://www.bankofengland.co.uk/report/2019/machine-learning-in-uk-financial-services

[11] Hong Kong Monetary Authority (HKMA), 2019, "Reshaping Banking with Artificial Intelligence", available online at: https://www.hkma.gov.hk/media/eng/doc/key-functions/finanical-infrastructure/Whitepaper_on_AI.pdf

[12] Within supervisors' prudential mandate, these topics are covered by examinations of operational risk.

[13] Some known definitions created by regulators and developers are listed in the annex; their variety illustrates the difficulties involved in obtaining a precise taxonomy.

The paper is structured as follows. Chapter II identifies characteristics of ML methods which could be relevant to the design of supervisory practice. On that basis, Chapter III discusses potential changes to supervisory practices. Chapter IV summarises the key findings and Chapter V lays out a roadmap of the consultative period.

II. Characteristics of ML

This paper does not develop a universally applicable definition of ML methods. Its purpose is, rather, to create a boundary which is sufficient for supervisory purposes to assess models under Pillars 1 and 2. Depending on whether and which ML characteristics exist for a specific methodology to be examined and the extent to which they are pronounced, these characteristics will be discussed in terms of supervisory practice, inspection techniques and inspection intensity.[14]

1. Dimensions and examples

The characteristics can be grouped into the three dimensions of the **AI/ML scenario**, which is explained in greater detail in the Bundesbank discussion paper: [15]

The (1) **methodology and data basis** collectively describe the complexity and thus the model risk associated with the ML procedure. The (2) **use of the output** contains the importance of the procedure within risk management. The intensity of inspections is not guided by the distinction between (3) in-house development and **outsourcing** or the underlying **IT infrastructure.**

Table 1: Characteristics of ML methods

AI/ML scenario	Characteristics
1 Methodology and data basis	Complexity and dimension of the hypothesis space
	Complexity of training
	Adaptivity
	Data sources
	Data types
	Data volume
2 Use of the output	Importance in the model
	Area of application
	Degree of automation
3 Outsourcing and IT	Outsourcing

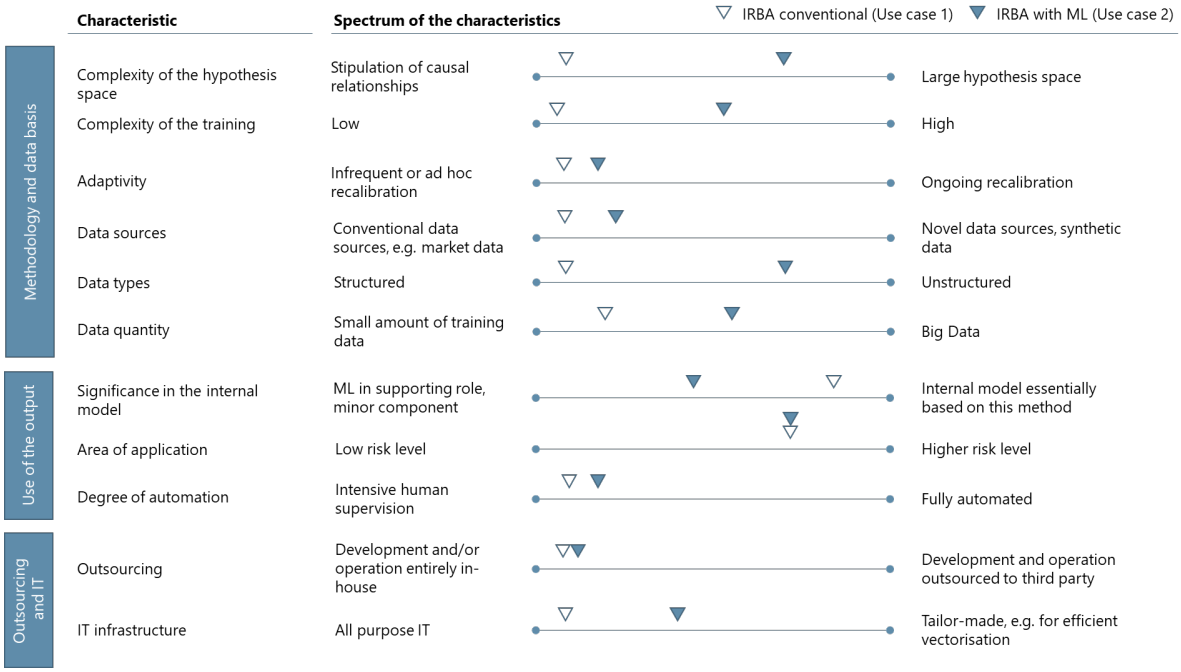
[14] This approach is reminiscent of the "duck test": see https://en.wikipedia.org/wiki/Duck_test.
[15] Bundesbank, 2020, "The Use of Artificial Intelligence and Machine Learning in the Financial Sector", online at: https://www.bundesbank.de/resource/blob/598256/d7d26167bceb18ee7c0c296902e42162/mL/2020-11-policy-dp-aiml-data.pdf

ML methods and conventional approaches form a continuum. The figure below assigns characteristics to two fictitious rating system use cases.

Use case 1: A conventional IRBA rating system uses a logistical regression to estimate borrowers' probability of default (PD). A small number of variables are entered into the model. The data are structured.

Use case 2: A multilayered neural network is used to estimate the PD for an IRBA rating system. The model is rarely retrained. The dataset comprises many variables, both structured and unstructured data. The results from the neural network, together with the results of conventional models, feed into the estimation of PD.

The characteristics presented here are typically more strongly pronounced in the neural network. Other ML methods or areas of application, for their part, have their own profiles.



2. Methodology

2.1 Complexity and dimension of the hypothesis space

Models in Pillars 1 and 2 reflect a causal link assumed by the modeller between input (such as market and portfolio data) and output (such as prices of financing instruments or risk metrics). This link is also referred to as a hypothesis. The basic eponymous feature of machine

learning is the fact that the algorithm used only contains a small number of assumptions regarding the structure of the problem to be described. This structure is given, for example, in the form of fixed rules.[16] Alternatively, parameters describing this problem structure are also learned using the data themselves. The hypotheses available for adjustment are therefore much more varied, and the hypothesis space is not completely controlled by analytical techniques, especially as its selection is not primarily motivated by causalities.

As in many more conventional methods, the modeller only specifies the space of all hypotheses applicable tothe method (hypothesis space). This space is to be searched, and the learning method selects a specific hypothesis (in the form of a set of model parameters) via optimisation. The hypotheses can have any level of complexity. The modeller defines the potential hypotheses through the selection of the ML method (e.g. neural network, random forest, k-nearest neighbours) and their specification (model design, hyperparameters).

Simple hypothesis spaces also occur in linear or logistical regressions, which are present in all conventional internal models.

A deep neural network (DNN) constitutes a much more complex model structure. Here, the hypothesis is mathematically expressed through the composition of non-linear and linear maps so that the hypothesis space has a considerably larger dimension. The functional relationship between input and output learned from a DNN can generally no longer be understood by means of a simple description using mathematical formulae, which is referred to as a "black box" characteristic.

The limited transparency of the model's behaviour has consequences for the model development, model validation and the implications for the significance of the underlying data. It also gives rise to additional challenges with regard to the explainability of model results in order to ensure a sufficient understanding of the functional relationship and to justify the applicability of the ML methods internally and externally.

2.2 Complexity of training

Determining a specific hypothesis occurs in a learning process referred to as "calibration" or "training".

One main feature of training newer ML methods is the high number of calculations due to the large number of function arguments and model parameters, the complex sequence of nested calculation instructions, and iterative procedures. This causes challenges to arise in technical implementation, such as in the availability of necessary hardware resources or the numerical stability of the calculation method.

In general, there is no unique solution to an optimisation problem involving non-linear functions in high-dimensional spaces. The selection of an optimum as a specific hypothesis depends on the training algorithm and may also depend on randomness. Choosing different optima in successive training sessions can affect the stability of the ML procedure.

[16] These do not include expert systems that, for their part, optimise the weightings of "if-then" rules.

As the complexity of training increases, these optima come under closer supervisory scrutiny.

2.3 Adaptivity

Some ML methods are designed to be adapted to new data very frequently or even on a virtually continuous basis. This leads to a blurring of the distinction between model development and model operation, and between model maintenance and model changes subject to supervisory assessment. This distinction is particularly relevant under Pillar 1 if model changes require supervisory approval. Furthermore, there is the question of to what extent it is possible to validate these models and reproduce model output. Ensuring the continuous, adequate quality of data likewise plays an important role. As the adaptivity of procedures increases, a clear differentiation between model maintenance and model changes becomes more important from a supervisory perspective.

3. Data basis

The increasing use of large data volumes is a distinguishing feature of current developments at banks, but it is also a feature of complex or new ML methods in the narrower sense.

ML frequently builds on a larger number of **data sources** and a network of different data sources. Synthetic, i.e. artificially generated data are also used, as are unstructured data (**data type**). ML methods can often handle a large number of input parameters in the model, giving rise to a large **data volume**.

The performance of ML methods is determined not least by the volumes of data available for training, by the veracity of the data and the data quality. The dataset is thus coming under greater scrutiny from supervisors.

4. Use of the output

4.1 Relevance in the model

The ML method can have different roles within the entire model. It can be integrated into a model as a supporting component, e.g. to prepare data, or as a sub-component, e.g. as a module of a rating scheme. Sometimes the ML method is also the central component of a model. Or it is used outside of the internal model, e.g. as a challenger tool or as a proxy for the “real” model within specific areas of application.

As the importance of the ML method increases within and alongside the model, the more intensely it comes under supervisory scrutiny.

4.2 Area of application

The area of application outlines which results the ML method provides and to what extent these are included in the enterprise's business processes. Examples of areas of application may include internal models, early warning systems for risks or credit ratings. The greater the impact of the area of application on the risk situation, the stricter supervisors set the requirements for the ML method.

4.3 Degree of automation

The degree of automation can be divided into algorithm-determined and algorithm-based processes; the associated operational risk can be divided accordingly. Algorithm-determined processes refer to largely automated processes that use ML results and thus entail greater risks where ML methods are insufficiently monitored. By contrast, algorithm-based processes rely on human-controlled processes to a greater extent (which in turn has its own risks).

5. Outsourcing and IT infrastructure

Outsourcing to specialist service providers and the use of a specific IT infrastructure are also typical for ML methods. Fintech and BigTech firms offer modular systems to create ML methods and provide an IT infrastructure tailored to high levels of performance. Difficulties may arise if ML methods are integrated into so-called "legacy" IT infrastructure.

The approach supervisors take to assess these service relationships does not change with the use of ML methods; the relevant regulations such as the prudential requirements for IT (*Bankaufsichtliche Anforderungen an die IT*, or BAIT) also cover these use cases. This is why Chapter III does not discuss these aspects specifically.

Questions on Chapter II:

a) Do you think it is appropriate to forgo a strict definition of ML methods and instead take an application-based approach and gear supervisory and inspection practice to the individual characteristics of the methods used?

b) What other characteristics of ML methods do you believe could be important for supervisory paractice or for internal model governance?

c) In your opinion, which characteristics do not belong in this overview?

d) In which relevant areas of application do you employ ML methods or where do you intend to implement them?

III. Supervisory approach

As already mentioned in the introduction, ML methods do not generally require new supervisory practices. This chapter draws supervisory conclusions for the characteristics of the ML methods described in Chapter II. It also outlines to what extent adjustments to supervisory practice are required in certain areas. The supervisory approach is defined by the principle of proportionality and using the characteristics from Chapter II. The present chapter is structured according to the risks identified for ML and integrates these into the cycle of model development and maintenance.

1. Supervisory practice endures

Pillar 1 includes extensive regulations for reviewing and approving internal models formulated in a technology-neutral manner and therefore also addresses the risks of ML

methods. Principle-based requirements for risk management and IT provide a sound footing in Pillar 2.[17,18]

Supervisors are systematically pursuing an inspection approach geared towards banking processes that establishes general, overarching inspection areas for each risk type to be inspected (e.g. credit or market risk) and is continuously adjusted to current circumstances.

Supervisory practice for ML methods can therefore also be derived from the existing framework. At the same time, an outlier analysis, also supported by this consultation, is currently surveying the areas in which the supervisory inspection approach needs to be fleshed out in order to cater to the peculiarities of using ML methods.

This not only takes account of their mathematical/methodological aspects, but also of how these ML methods are integrated into processes, which is just as important for their controlled and thus successful and efficient use.

Supervisors are focusing on any new or much more pronounced risks that arise from ML methods. These are revealed in the data basis, validation (from model development to the test procedure to operation), model changes and management.

At a more abstract level, new draft legislation such as the AI Regulation drafted by the European Commission has placed special emphasis on aspects of consumer protection relating to ML. Supervisors are facing the task of inserting new requirements into the existing requirements of Pillar 1 and 2 models in a consistent manner.

[17] The CRR and MaRisk primarily form the legal basis for Pillar 1 and 2 inspections, supported by EBA and SSM standards. In this case, Commission Delegated Regulation (EU) Nos 2014/529 and 2015/942 are to be applied for Pillar 1 models. The EBA/GL/2020/06 (Guidelines on loan origination and monitoring) states the requirements for the use of automated models in creditworthiness assessments and credit decision-making covered by the German supervisory approach.
[18] BaFin, 2021, "Big data and artificial intelligence – principles for the use of algorithms in decision-making processes", available online (but in German only) at:
https://www.bafin.de/SharedDocs/Downloads/DE/Aufsichtsrecht/dl_Prinzipienpapier_BDAI.html

Questions on Chapter III.1:

e) In your opinion, do existing regulations already contain prudential requirements that appear to hinder the use of ML methods? Do you believe that contradictions will arise between prudential regulations for Pillar 1 and 2 models and the draft AI Regulation? Please state any relevant references to the corresponding regulations and explain the challenges.

f) To what extent do you believe the requirements laid out in EBA/GL/2020/06 with reference to the use of automated models in creditworthiness assessments and credit decision-making are also suitable for other ML methods in Pillar 2 (MaRisk) and should be taken on?

g) Are there any other points where you believe current supervisory practice requires adjustment in order to appropriately acknowledge ML procedures and their associated risks?

h) Do ML methods entail specific risks for IT implementation and outsourcing management? Are "adversarial attacks" conceivable in the financial sector and should ML methods be given particular protection against such attacks?

2. Methods invite to „believe in data"

Data quality is already an issue of key importance in supervisory action. Here, however, the characteristics of ML methods make clear that the data basis should be viewed particularly as a starting point and as a success factor. Unstructured data can now be exploited by and for ML methods. Furthermore, ML methods allow for calculations that factor in a large number of determinants. This makes it easy for modellers to quickly scale ML to large datasets.

ML methods learn what they find in the data provided and replicate any patterns contained therein. The black box characteristic may result in problems being obscured by seemingly good performance. For example, it is possible that models may learn correlations between input data that do not actually represent real relationships but are only based on coincidental characteristics of the learning dataset (model overfitting).

ML methods can utilise large volumes of data; the quality of these data must be continually ensured. This applies not only when developing and validating models, but also in their application.

Supervisors expect banks to undertake additional efforts to guarantee the quality of the underlying data. In particular, this involves ensuring that the training data are free from any systematic bias regarding the functional relationships to be learned by the model.

Questions on Chapter III.2:

i) What challenges do you see when selecting data and when ensuring data quality with regard to ML methods?

j) In your opinion, what aspects of data quality are made easier through the application of ML methods?

3. Focus on explainability

As the hypothesis space that can be reflected by the model becomes more complex and its dimension inceases, it also becomes more difficult to describe the functional relationship between input and output (i.e. the hypothesis specified in the training) verbally or using mathematical formulae, and the details of the calculations are less transparent for modellers, users, validators and supervisors. As a result, it is more difficult to comprehend the modelling and, if applicable, to check the validity of the model output as well. User acceptance may also suffer.

This black box characteristic can be considered the price of better model performance. However, it can be entirely justified, for example by higher predictive ability. Yet this may have to be weighed up against potentially greater model risk, depending on the significance of the model within the context of the associated banking processes.

Modellers must justify why the benefits gained are worth the trade-offs in the comprehensibility of the model. The extent to which a black box could be acceptable in supervisory terms is also dependent on how the model concerned is treated in the bank's risk management.

The black box may hide the fact that, amongst other things, ML methods learn relationships in the data that lack any real basis and do not allow for any general conclusions to be drawn.

As a result, it is the explainability and plausibility of the model behaviour, rather than its comprehensibility in detail, that gain in significance overall.

The term "explainability" is multifaceted, as modellers, validators, supervisors and users have different specialist backgrounds and require different information. In order to take account of these user-specific requirements, "explainable AI" (XAI) methods were developed.

From a supervisory perspective, XAI methods are highly promising with regard to mitigating the black box characteristic. However, XAI methods themselves represent models with

assumptions and weaknesses, and, in many cases, are still in the testing phase. As a result, it is a challenge to integrate these methods into structured processes. For example, it should be determined when which methods can be used, whether global or local approaches should be employed, which selection criteria and sample sizes are needed for local explanations, and which user groups should be targeted.

Banks must employ validation methods that are tailored to the model structure and that adequately cover the model risks. In the case of machine learning procedures, there is a particular danger of overfitting, for example due to the large number of parameters. Alongside the usual approaches, such as out-of-sample validation and backtesting, XAI approaches can help to identify overfitting. In addition, synthetic or stress/extreme scenarios as well as tests against traditional procedures could support the plausibility and explainability of ML methods.

Questions on Chapter III.3:

k) In your opinion, what impact does the black box characteristic have on the validation of the procedure?

l) How important is the trade-off between performance and explainability for you?

m) Do you believe XAI methods (always) offer a way out of the black box dilemma? Which processes are most promising and for which ML methods?

n) How do you think XAI should be incorporated into the validation process?

4. Adaptivity: model changes are more difficult to identify

Institutions and enterprise are obligated to inform supervisors of any changes to Pillar 1 models and, if applicable, only implement these changes after they have been approved. There is no clear-cut distinction between regular model maintenance and model change, which continually leads to discussions with supervisors, especially as the term "model change" is also dependent on the prevailing supervisory context.[19]

The flexibility and, in some cases, high-frequency adaptivity – i.e. frequent adjustment due to new data, for example – of ML procedures make it more difficult to draw a clear line between

[19] For IMM models, see the table in Section 3.10 of the ECB guide on materiality assessment (EGMA), which presents illustrative examples of model maintenance and model change. For market risk, see also Commission Delegated Regulation (EU) No 2015/942, Annex III, Section 2, which lists examples of model changes.

adjustments and changes; this is, however, indispensable for supervisors. As a general rule, the need for high-frequency adaptivity should be thoroughly justified.

While this document cannot provide any ML-specific definition of the term "model change", it uses examples to illustrate how supervisors can classify changes.

- Procedural criteria, such as the first-time use of an ML method for a given task or changes to processes in conjunction with application of the method, can, as is the case for traditional methods, be used to identify and classify model changes.

- The question of whether retraining – i.e. training following an update to the underlying data – an ML method should be considered a change cannot be generally answered on the basis of comparison with traditional approaches.

  For example, in the case of counterparty credit risk, recalibrating the probability distributions of risk factors is generally classified as model maintenance and not model change.[20] However, the hypothesis space that can be depicted by these models is significantly smaller than those of modern ML methods, and the functional relationship between input and output is not fundamentally changed by the recalibration of such a model.

  By contrast, retraining an ML method can also involve structural changes to the function reflected in the method, such as the number of layers in a DNN. Even though the functional relationship between input and output to be reflected by the model may have only marginally changed (e.g. due to market movements), the realised "hypothesis" may differ to the previous one to such a significant degree that it would have to be considered a model change in supervisory terms.

  Regardless of their classification, the high flexibility and adaptivity of ML methods mean that, following successive instances of training or adjustment, they can become far removed from the original models – in the sense that the inputs flow into the outputs with entirely different weights – within just a short period of time. For this reason, this adaptability – more so than in the case of traditional methods – must be subject to checks to prevent the model structures from fundamentally changing, which may nullify their explainability and validation, without the models having to undergo renewed supervisory evaluation.

- An extension of the risk factors used in the model would still be considered a model change. However, this distinction can be difficult to make in the case of complex datasets, especially of unstructured data.

These examples also illustrate the necessary shift in supervisory focus and the approach towards model adjustments:

[20] EGMA, p. 18. For market risk, see also the simplifications for smaller model changes in rapidly changing markets listed on p. 4 of EBA/RTS/2014/10.

- Model approval: The possible methods of model maintenance (including ongoing internal model validation and model changes) must already be assessed within the scope of model approval. The model change policy to be drawn up by the bank must take account of the particular features of the utilised ML methods when classifying adjustments. XAI processes could potentially also be employed here.

- Communication: Supervisors must also be able to trace activities related to regular model maintenance in order to identify in good time whether the original model is being significantly changed through a number of incremental model changes.

- Internal validation: Compared with traditional methods, the internal validation of ML methods has a greater focus on monitoring and ensuring the suitability of the methods employed on a continuous (and not just annual) basis. Model validation must make clear which change to the model structure will have what effect on the model output. This places greater demands on validators. This is the only way to ensure that unintended effects arising from model changes can be identified and avoided in good time. High-frequency retraining raises the question of which version of the model should be validated (which is critical in the case of ongoing retraining, for example).

Many ML methods are used in the supervisory area of Pillar 2 where approval is not required. This results in greater flexibility for model retraining and changes, yet existing requirements in this regard (e.g. from MaRisk) remain in effect here, too. From a supervisory perspective, it is nevertheless crucial, despite this flexibility, to adapt the training cycle to the specific use case, as well as provide the necessary justification, in order to create a balance between the explainability and validation of the model and ensuring that the data are up-to-date.

Questions on Chapter III.4:

o) What questions on supervisory practice do you see arising with regard to model adjustments for ML methods?

p) Do you believe it is necessary for certain ML methods to be retrained at very high frequencies?

q) Do you see ML methods necessitating changes in model governance? How do traditional modelling units, validators and new “data science” units work together?

IV. Outlook

With regard to the choice of supervisory focus, in addition to the general BDAI principles, it is essential to achieve clarity in the development and application of ML methods in the context of models relevant for supervision in Pillars 1 and 2. This will create an environment which enables enterprises to invest in ML and address the risks of these methods as early as possible. At the same time, the impetus to adopt ML must come from the enterprises themselves. Supervisors do not insist on the use of ML methods as long as "traditional" methods of meeting regulatory requirements are still seen as suitable.

The next step will be to use this consultation paper to enter into dialogue with enterprises. Supervisors also consider it necessary to harmonise international approaches as far as possible and set identical cross-sector criteria for the use of ML methods. The European Commission's digital finance strategy[21] plays a role in European standardisation.

V. Consultation

BaFin and the Bundesbank invite the industry to comment on this consultation paper and answer the questions posed within. Please respond by email to Konsultation-11-21@bafin.de and ai-b3@bundesbank.de by 30.09.2021.

Comments submitted will be consolidated and published in anonymised fashion.

[21] European Commission, 2020, "Digital Finance Strategy for the EU", available online at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0591

Annex

Frequently asked questions

Do procedures which explicitly reflect presumed functional relationships also count as ML?

Not in the narrower sense of ML – in a statistical context, learning instead focuses on recognising functional relationships and regularities on the basis of past observations in order to make better predictions in the future.[22] Machine learning hinges on one specific learning problem: how to apply statistical methods to a collection of data (input-output) to learn a function which can be used to predict the output for new input.[23] Machine learning is used because modellers can neither keep track of the technical relationship between input and output, nor find if-then conditions for it.[24]

Do "traditional", long-standing models already known to supervisors (e.g. logistic regression) fall under the characteristics of ML?

Even existing internal models use self-learning procedures which, to varying degrees, display the characteristics described above. In addition to simple regression approaches for individual risk factors (e.g. short-rate models), there are also sophisticated approaches to recreating changes in own funds under combined risk factor shifts (market, credit, underwriting and operational risks) whose specification requires a similarly high amount of processing power as training a neural network, which can be found in, among others, internal models used by insurance groups. Said specification can also include "training cycles", i.e. the iterative addition or removal of polynomial building blocks using an information criterion in combination with subsequent retraining of the expanded/reduced function via regression.

The hypothesis space these models have is generally more limited than that of the aforementioned ML methods, which is why certain problems such as the black box characteristic occur less often or not at all. However, similar limitations can arise in relation to validation concerning the progressive plausibility of calculations, making one reliant even here on checking the plausibility of results, e.g. via out-of-sample comparisons.

Are supervisors more sceptical of unsupervised learning than supervised learning?

No. The use case and type of ML are directly linked, as every type solves different problems and is therefore suited to different use cases. It is not possible to decide which type should

[22] See S. Russel, P. Norvig, 2016, Artificial Intelligence: A Modern Approach, third edition, p. 693.
[23] The focus here is initially on supervised learning. With unsupervised learning, the learning task consists of pattern recognition via systematising input data (without output or labels).
[24] See S. Russel, P. Norvig, 2016, Artificial Intelligence: A Modern Approach, third edition, p. 693 ff.; Bank of England, 2019, Machine learning in UK financial services.

be used based on the level of risk involved. Each individual use case should be categorised and appraised according to its characteristics. Fundamentally, banking and financial supervisors in Germany use the same evaluation criterion for all types of ML. Even so, use cases with supervised learning currently outnumber all others.

Table 2: Types of ML

Type
Supervised learning: The algorithm is trained to map input data to given output, also known as “label”. The trained model can then be used on new data.
Unsupervised learning: Finding hidden patterns or structures (clustering, association, dimension reduction).
Reinforcement learning: Learning tasks during which the model’s own solution strategies are repeatedly re-used and evaluated as input parameters.

Selected examples of definitions of the term „Machine Learning“

Table 3: Selected definitions of ML

Academia	Regulators	IT companies
▪ „Set of methods that can automatically detect patterns in data, then use the un-covered patterns to predict future data, or to perform other kinds of decision making under uncertainty.“ Murphy, S. 1 ▪ „Machine Learning is the science (and art) of programming computers so they can learn from data.“ Geron, S. 2 ▪ „Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.“ Samuel in Geron, S. 2 ▪ „A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E“ Mitchell in Geron, S.2 ▪ „The objective of machine learning is the generation of ‘knowledge‘ out of ‘experience‘ through learning algorithms using examples to develop a complex model. This model, and thus the automatically obtained knowledge representation, can in turn be used on	▪ „The standard on IT governance ISO/IEC 38505-1:2017 defines ML as a ‘process using algorithms rather than procedural coding that enables learning from existing data in order to predict future outcomes‘.“ EBA, S. 14 ▪ „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.“ Bank of England, S. 6 ▪ „Machine learning may be defined as a method of designing a sequence of actions to solve a problem, known as algorithms, which optimise automatically through experience and with limited or no human intervention.“ FSB, S. 4	▪ „ML, a subset of AI, focuses on the ability of machines to receive data and learn for themselves without being programmed with rules. ML differs from traditional programming by allowing you to teach your program with examples rather than a list of instructions. [It] enables you to "train" an algorithm so that it can learn on its own, and then adjust and improve as it learns more about the information it is processing.“ Google ▪ „ML is a form of AI that enables a system to learn from data rather than through explicit programming. […] ML enables models to train on data sets before being deployed. Some ML models are online and continuous [leading] to an improvement in the types of associations made between data elements. Due to their complexity and size, these patterns and associations could have easily been overlooked by human observation.“ IBM ▪ „ML is the process of using mathematical models of data to help a computer learn without direct instruction. It’s considered a subset of AI. ML uses algorithms to identify patterns within data, and those patterns are then used to create a data model that can make predictions. With increased data and experience, the results of machine learning are more accurate – much like how humans improve with more practice.“ Microsoft

new, potentially unknown
data of the same kind."
Fraunhofer IAIS, S. 8

Sources of the ML definitions

1. A. Geron, 2019, "Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow", 2. ed. O'Reilly.

2. T. Hastie, R. Tibshirani und J. Friedman (2009), "The Elements of Statistical Learning: Prediction, Inference and Data Mining", 2. ed. Springer-Verlag.

3. G. James, D. Witten, T. Hastie und R. Tibshirani (2013), "An Introduction to Statistical Learning", Springer-Verlag

4. K.P. Murphy, 2012, "Machine Learning: A Probabilistic Perspective", MIT Press.

5. EBA, 2020, "Report on Big Data and Advanced Analytics", available online: https://eba.europa.eu/sites/default/documents/files/document_library/Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf

6. Bank of England, 2019, "Machine learning in UK financial services".

7. Bundesbank, 2020, "Policy Discussion Paper, The Use of Artificial Intelligence and Machine Learning in the Financial Sector", available online: https://www.bundesbank.de/resource/blob/598256/d7d26167bceb18ee7c0c296902e42162/mL/2020-11-policy-dp-aiml-data.pdf

8. Financial Stability Board (FSB), 2017,"Artificial intelligence and machine learning in financial services, Market developments and financial stability implications", available online https://www.fsb.org/wp-content/uploads/P011117.pdf

9. Fraunhofer IAIS, 2018, "Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung", available online: https://www.bigdata-ai.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/Fraunhofer_Studie_ML_201809.pdf

10. Google, "Using AI for a social good", available online: https://ai.google/education/social-good-guide/?category=introduction

11. IBM, 2020, "Data science and machine learning", available online: https://www.ibm.com/hk-en/analytics/machine-learning#:~:text=Machine%20learning%20enables%20models%20to,associations%20made%20between%20data%20elements

12. Microsoft Azure, “What is Machine Learning?” available online: https://azure.microsoft.com/en-us/overview/what-is-machine-learning-platform/

Impressum

Herausgeber
Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin)
Gruppe Kommunikation
Graurheindorfer Straße 108, 53117 Bonn
Marie-Curie-Straße 24 – 28, 60439 Frankfurt am Main
www.bafin.de

Redaktion und Layout
BaFin, Interne Kommunikation und Internet
Redaktion: Rebecca Frener
Tel.: +49 0 228 41 08 22 13
Kathrin Jung
Tel.: +49 0 228 41 08 16 28
Layout: Christina Eschweiler,
Tel.: +49 0 228 41 08 38 71
E-Mail: journal@bafin.de

Designkonzept
werksfarbe.com | concept + design
Humboldtstraße 18, 60318 Frankfurt
www.werksfarbe.com

Bezug
Das BaFinJournal* erscheint jeweils zur Monatsmitte auf der Internetseite der BaFin. Mit dem Abonnement des Newsletters der BaFin werden Sie über das Erscheinen einer neuen Ausgabe per E-Mail informiert.
Den BaFin-Newsletter finden Sie unter: www.bafin.de/Newsletter.

Disclaimer
Bitte beachten Sie, dass alle Angaben sorgfältig zusammengestellt worden sind, jedoch eine Haftung der BaFin für die Vollständigkeit und Richtigkeit der Angaben ausgeschlossen ist.

Ausschließlich zum Zweck der besseren Lesbarkeit wird im BaFin Journal auf die geschlechtsspezifische Schreibweise verzichtet. Alle personenbezogenen Bezeichnungen sind somit geschlechtsneutral zu verstehen.