



EBA DISCUSSION PAPER ON MACHINE LEARNING FOR IRB MODELS

EBA/DP/2021/04

11 NOVEMBER 2021



EBA

EUROPEAN
BANKING
AUTHORITY

Contents

Abbreviations	3
Responding to this Discussion Paper	4
Submission of responses	4
Publication of responses	4
Data protection	4
Disclaimer	4
Executive summary	5
1. Introduction	7
2. Machine learning: definition, learning paradigms and current use in credit risk modelling	9
2.1 Definition	9
2.2 Learning paradigms	10
2.3 Current use of ML for IRB models	10
3. Challenges and potential benefits of ML models	13
3.1 Challenges posed by ML models	14
3.2 Potential benefits from the use of ML models	20
4. How to ensure a possible prudent use of ML models going forward	22
4.1 Concerns about the use of ML	22
4.2 Expectations for a possible and prudent use of ML techniques in the context of the IRB framework	23
Annex – Summary of questions	27

Abbreviations

CRR	Capital Requirements Regulation – Regulation (EU) No 575/2013 as amended by Regulation (EU) 2019/876
CCF	Credit Conversion Factor
CRCU	Credit Risk Control Unit
CWA	Creditworthiness Assessment
EAD	Exposure At Default
ELBE	Expected Loss Based Estimate
ICAAP	Internal Capital Adequacy Assessment process
IRB	Internal Rating-Based approach
LGD	Loss Given Default
ML	Machine Learning
PIT	Point-in-Time
PD	Probability of Default
TtC	Through-the-Cycle
BD&AA	Big Data and Advanced Analytics

Responding to this Discussion Paper

The EBA invites comments on all proposals put forward in this paper and in particular on the specific questions stated in the boxes below (and in the Annex of this paper).

Comments are most helpful if they:

- respond to the question stated;
- indicate the specific point to which a comment relates;
- contain a clear rationale;
- provide evidence to support the view expressed;
- provide where possible data for a cost and benefit analysis.

Submission of responses

To submit your comments, click on the 'send your comments' button on the consultation page by 11.02.2022. Please note that comments submitted after this deadline, or submitted via other means may not be processed.

Publication of responses

Please clearly indicate in the consultation form if you wish your comments to be disclosed or to be treated as confidential. A confidential response may be requested from us in accordance with the EBA's rules on public access to documents. We may consult you if we receive such a request. Any decision we make not to disclose the response is reviewable by the EBA's Board of Appeal and the European Ombudsman.

Data protection

The protection of individuals with regard to the processing of personal data by the EBA is based on Regulation (EU) 1725/2018 of the European Parliament and of the Council of 23 October 2018. Further information on data protection can be found under the [Legal notice section](#) of the EBA website.

Disclaimer

The views expressed in this discussion paper are preliminary and will not bind in any way the EBA in the future development of any potential guidance.

Executive summary

The aim of this discussion paper is to understand the challenges and opportunities coming from the world of machine learning (ML) should they be applied in the context of internal ratings-based (IRB) models to calculate regulatory capital for credit risk.

The exponential increase in data availability and storing capacity coupled with the improvements in computing power of recent years provides an opportunity to use ML models to make sense and sort massive and unstructured data sources. Whereas standard regression models may not be able to keep track with the emergence of the so called 'Big Data', in fact, data is the fuel that powers ML models by providing the information necessary for training the model and detecting patterns and dependencies. This does not come without costs; indeed, ML models are more complex than traditional techniques such as regression analysis or simple decision trees, and often less 'transparent'. The discussion paper focuses on the more complex models, which are also the most difficult to understand and challenging should they be possibly used for regulatory capital purposes.

In the context of credit risk, ML models might be useful to improve predictive power and are not new to internal models used for credit approval processes, but they have not been incorporated into institutions' IRB models as rapidly as in other areas. The main pivotal challenge comes from their complexity which leads, at least for the more complex ones, to challenges i) in interpreting their results, ii) ensuring their adequate understanding by the management functions and iii) justifying their results to supervisors. This discussion paper is therefore a first step to engage the industry and the supervisory community to investigate the possible use of ML as IRB models and to build up a common understanding of the general aspects of ML and the related challenges in complying with the regulatory requirements.

The discussion paper ultimately aims at discussing the relevance of possible obstacles to the implementation of ML models in the IRB model space based on some practical issues. Practical issues on the use of data, explainability and other challenges are generally not new to IRB models, but may be exacerbated when using ML models and, therefore, may lead to specific challenges. These are explored in the discussion paper with the aim of seeking the industry view on those challenges.

It is clear that ML models can provide added value, but to comply with CRR requirements it should also be possible to interpret them and that the relevant stakeholders have a level of knowledge of the model's functioning that is at least proportionate to their involvement and responsibility for meeting legal requirements. Otherwise, there is a risk of developing 'black box' models. It is therefore key that institutions and all their levels of management functions and bodies have an adequate level of understanding of their IRB models, as this will be essential in order to allow ML models to be used for regulatory purposes.

Acknowledging that ML might play an important part in the way financial services will be designed and delivered in the future, the EBA is considering to provide a set of principle-based recommendations which should ensure an appropriate use of such techniques by institutions in the context of IRB models. This should ultimately ensure: i) a consistent and clear understanding of the prudential provisions, ii) how new sophisticated ML models can coexist and adhere with these requirements and so iii) that the outcome – in terms of setting capital requirements in a prudent manner – continues to be harmonised across Europe.

1. Introduction

Internal ratings-based (IRB) models currently used by institutions to calculate regulatory capital requirements for credit risk do not differ materially from the approaches used 15 to 20 years ago when the first Basel Accord was put in place. Since then, the focus of regulators and supervisors has been, in fact, more on making the estimates produced by different models comparable by improving the definition of basic concepts (as for example the definition of default), rather than on understanding the challenges coming from the world of advanced technology, i.e. machine learning (ML) and artificial intelligence (AI).

On 13 January 2020, the EBA published a report on the recent trends of big data and advanced analytics (hereinafter “Report on BD&AA”),¹ including ML, in the banking sector and on the key considerations in the development, implementation and adoption of BD&AA. The report identifies recent trends and suggests key safeguards in an effort to support ongoing technological neutrality across regulatory and supervisory approaches.

In recent years, the exponential increase in data availability/storing capacity coupled with the improvements in computing power has led to the emergence of ‘Big Data’² and has offered a tremendous possibility to use new data sources in order to detect patterns and dependencies. This causes challenges when using standard regression models that may not be able to keep track with the emergence of ‘Big Data’. On the contrary, data is the fuel that powers ML models by providing the information necessary for developing and improving features and pattern recognition capabilities. Without large quantities of high-quality data it is not possible to put in place the algorithms that make ML a possible game-changing technology. Moreover, ML models might help to make sense and sort unstructured data sources. Therefore, to analyse Big Data institutions are increasingly using advanced analytics that, as clarified in the Report on BD&AA, include ‘predictive and prescriptive analytical techniques, often using AI and ML in particular, and are used to understand and recommend actions based on the analysis of high volumes of data from multiple sources, internal or external to the institution’.

In the context of credit risk, ML models might be useful and are not new to internal models used for credit approval process, but they have not been incorporated into banks’ internal models for regulatory capital requirements calculation as rapidly as in other areas. According to the Institute of International Finance (IIF) 2019 report on Machine Learning in Credit Risk (IIF 2019 Report)³ one of the main challenges mentioned by institutions for the incorporation of ML models in credit risk

1

https://eba.europa.eu/sites/default/documents/files/document_library/Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf

² There are many definitions of Big Data but, using the ESAs’ tentative definition used in the EBA Report on BD&AA, Big Data refers to large volumes of different types of data, produced at high speed from many and varied sources (e.g. the internet of things, sensors, social media and financial market data collection), which are processed, often in real time, by IT tools (powerful processors, software and algorithms).

³ https://www.iif.com/Portals/0/Files/content/Research/iif_mlcr_2nd_8_15_19.pdf

modelling was the ‘lack of understanding’ from supervisors and the uncertainty with respect to the approval of such new types of models. This is perhaps not surprising as ML models might indeed be difficult to explain and justify and difficult to understand at various level of management functions and bodies. For some types of models it could at least be challenging to demonstrate to the supervisors that the underlying economic rationale is conveyed. This discussion paper is therefore a first step to engage the industry and the supervisory community to discuss the possible use of ML for IRB models and to build up a common understanding of the general aspects of ML and the related challenges to comply with the regulatory requirements.

Acknowledging that ML might play an important part in the way financial services will be designed and delivered in the future, the EBA aims i) to identify the main challenges and possible benefits of these new models should they be possibly used in the context of IRB models as well as ii) to provide a set of principle-based recommendations which should ensure proper future use by banks for prudential purposes. This will help to provide a consistent and clear understanding of the prudential provisions, and how new sophisticated ML models might coexist with these requirements and, therefore, should ultimately ensure that the outcome - in terms of setting capital requirements in a prudent manner – continues to be harmonized across Europe.

This paper also complements the recently published report of the EBA analysing RegTech in the European (EU) financial services sector⁴. The EBA report assesses the overall benefits and challenges faced by financial institutions and RegTech providers in the use of RegTech across the EU and identifies potential risks that supervisors will need to address. The EBA Report looks at the application of technology-enabled innovation for regulatory, compliance and reporting requirements and provides a deep-dive analysis into the five most frequently used RegTech segments, including creditworthiness assessment (CWA). The chapter on CWA focuses on loan origination only and therefore does not cover prudential risk management (IRB/SA).

The discussion paper is organised as follows:

- Section 2 provides a general definition of ML models for the purpose of this discussion paper, discusses the main learning paradigms used to train ML models and, finally, discusses the current limited use of ML models in the context of IRB models;
- Section 3 analyses the challenges and the benefits related institutions may face in using ML to develop compliant IRB models;
- Finally, Section 4 provides a set of principle-based recommendations that aim at ensuring ML models adhere to the regulatory requirements set out in the CRR, should they be used in the context of the IRB framework.

⁴ <https://www.eba.europa.eu/eba-assesses-benefits-challenges-and-risks-regtech-use-eu-and-puts-forward-steps-be-taken-support>.

2. Machine learning: definition, learning paradigms and current use in credit risk modelling

2.1 Definition

The EBA has in previous communication (i.e. the report on BD&AA) used the definition of ML presented in the standard on IT governance ISO/IEC 38505-1:2017⁵ which defines ML as a ‘process using algorithms rather than procedural coding that enables learning from existing data in order to predict future outcomes. Broadly speaking, it is a field within computer science that deals with the development of models whose parameters are estimated automatically from data with limited or no human intervention.

ML covers a wide range of models with different levels of complexity. An example of simple ML models are linear regressions, which are intuitive models with a limited number of parameters, and whose study dates back to the XIX century. On the opposite end of the complexity spectrum, an example could be deep neural networks, which have been developed in the last two decades. In these models the number of parameters can rise to the millions, and their understanding and implementation represents a significant challenge.

However, the term ML, which dates from 1959, is often used by practitioners to refer only to the more complex models. This has been the case in the financial sector, where linear and logistic regressions have long been used, but the term ML is reserved for the more pioneering models. It must be noted, however, that there is no clear-cut distinction between simple and advanced models. See, for example, how a linear regression grows progressively in complexity as higher order terms are included, leading to more sophisticated, and less tractable, relations between the variables.

Some of the characteristics that are useful for evaluating the complexity of a model are:

- The number of parameters.
- The capacity to reflect highly non-linear relations between the variables accurately.
- The amount of data required to estimate the model soundly.
- The amount of data from which the model is able to extract useful information.
- Its applicability to unstructured data (reports, images, social media interactions, etc.).

⁵ <https://www.iso.org/obp/ui/#iso:std:iso-iec:38505:-1:ed-1:v1:en>.

The current discussion paper is focused on the more complex models, which are also the most difficult to understand and challenging to use for regulatory capital purposes. Therefore, for the purposes of this discussion paper, the term ML should refer to models that are characterised by a high number of parameters and, therefore, require a large volume of (potentially unstructured) data for their estimation that are able to reflect non-linear relations between the variables.

Beyond this general definition, several learning paradigms may be used to train the ML models, and one possible categorisation is described in the following subsection as well as in Section 1.3 of the Report on BD&AA.

2.2 Learning paradigms

There exists a range of learning paradigms that can be used to train ML models depending on the goal of the model and the type of data required. Currently the three most popular learning paradigms are:

- Supervised learning (or learning with labels): the algorithm learns rules for building the model from a labelled dataset (i.e. where the output variable such as default/non-default is known) and use these rules to predict labels on new input data.
- Unsupervised learning (or learning without labels): the algorithm learns from an input training dataset which has no labels, and the goal is to understand the distribution of the data and/or to find a more efficient representation of it.
- Reinforcement learning (or learning by feedback): the algorithm learns from interacting with the environment, rather than from a training dataset. Moreover, contrary to supervised learning, reinforcement learning does not require labelled input/output pairs. The algorithm learns to perform a specific task by trial and error.

It should be noted that there are plenty of other categorisations possible and sub-categories between the one mentioned above. For example, in the context of supervised learning, one could differentiate further between regression and classification, whether the output of the model is respectively numerical (continuous) or categorical (or discrete).

2.3 Current use of ML for IRB models

According to the IIF 2019 report, the most common use of ML within credit risk is the area of credit decisions/pricing followed by credit monitoring and collections, restructuring and recovery. In contrast, the use of ML is avoided for regulatory areas such as capital requirements for credit risk, stress testing and provisioning. In this regard, regulatory requirements are perceived as a challenge for the application of ML models, as these are more complex to interpret and explain.

For IRB models the use of ML models has been more limited where these models are used only as a complement to the standard model used for capital requirement calculation. Examples where ML techniques are currently used in the context of IRB models in compliance with CRR requirements:

- Model validation: ML models are used to develop model challengers to serve as a benchmark to the standard model used for capital requirements calculation.
- Data improvements: ML techniques can be used to improve data quality to be used for estimation, both in terms of more efficient data preparation and data exploration where ML can be used in the context of big data to analyse rich datasets.
- Variable selection: ML could be used to detect explanatory variables and combinations of them with useful predictive capacities within a large dataset.
- Risk differentiation: ML models can be used as a module for the purposes of risk differentiation of the probability of default (PD) model, where the module may allow, for example, upgrades/downgrades to the PD grade previously assigned by the 'traditional' PD model through text mining.

However, ML models might possibly be used as primary IRB models for prediction (e.g., for risk differentiation) under some challenges described in the following sections.

Comparing the 2019 IIF report with the 2018 IIF report, a clearly observed trend is that institutions have made the strategic decision to move away from using ML in regulatory areas, shifting their focus to other areas such as credit monitoring or collections and recovery. This seems to have happened in avoidance of the conflict of regulatory requirements requesting regulatory models to be intuitive, plausible and based on economic theory on the one hand, and the complex nature of ML models, on the other.

Whereas ML might help in estimating risk parameters more precisely, in fact, the increase in predictive power comes at the cost of more complexity where the relationship between input and output variables is more difficult to assess and understand. This trade-off between predictive power and interpretability – which is core to the institution's decision on whether to use ML models for prudential purposes – might have shaped the use of ML models as of today (i.e. where more complex ML models were typically used outside the regulatory remit based on the expectations that supervisors may not accept them). Given this trend, it might be important to clarify supervisors' expectations around the possible use of ML in the context of the IRB framework underlining the potential added value of ML models, provided that a safe and prudent use of ML models can be ensured. The next section clarifies the characteristics of ML models that (i) may make more challenging to fulfil the current regulatory requirements; and, in return (ii) may facilitate fulfilling some of the requirements.

Questions:

1: Do you currently use or plan to use ML models in the context of IRB in your institution? If yes, please specify and answer questions 1.1, 1.2, 1.3, 1.4; if no, are there specific reasons not to use ML models? Please specify (e.g. too costly, interpretability concerns, certain regulatory requirements, etc.)

1.1: For the estimation of which parameters does your institution currently use or plan to use ML models, i.e. PD, LGD, ELBE, EAD, CCF?

1.2: Can you specify for which specific purposes these ML models are used or planned to be used? Please specify at which stage of the estimation process they are used, i.e. data preparation, risk differentiation, risk quantification, validation.

1.3: Please also specify the type of ML models and algorithms (e.g. random forest, k-nearest neighbours, etc.) you currently use or plan to use in the IRB context?

1.4: Are you using or planning to use unstructured data for these ML models? If yes, please specify what kind of data or type of data sources you use or are planning to use. How do you ensure an adequate data quality?

2: Have you outsourced or are you planning to outsource the development and implementation of the ML models and, if yes, for which modelling phase? What are the main challenges you face in this regard?

3. Challenges and potential benefits of ML models

A consistent and clear understanding of the prudential provisions, and how new sophisticated ML models might possibly coexist with these requirements, is the first crucial step to have clear and appropriate principle-based recommendations should ML be possibly used in the context of IRB models.

The aim of this section is therefore to elaborate on the regulatory requirements for IRB models to identify those characteristics of ML which might (i) make it more challenging to comply with certain CRR requirements; and those which (ii) instead might help institutions to meet some of them.

ML models may pose challenges or provide solutions that are specific to the context in which they are used. Therefore, the next two sections analyse respectively the challenges and the possible benefits related to the use of ML in the context of IRB models based on the area of use. The identified main areas of use are:

- Risk differentiation.
- Risk quantification.
- Model validation, where ML models can be used, for example, as model challengers or for benchmarking.
- Other areas, such as, for example data preparation or use of ML models for credit risk mitigation purposes such as, for example, the valuation of collateral.

Less focus is required, when ML models are used at a lower or sub-level to improve the quantitative and qualitative part of the more traditional IRB models (e.g. where ML is used for data preparation only or where the IRB models are based on risk drivers which are the result of ML models). Therefore, as for any standard model, the regulatory requirements should be applied proportionally, considering whether ML is used as a supporting tool for predicting the parameters (e.g. in data-cleansing or variable selection) or does the prediction in full.

Finally, the possible decision of going forward with ML in credit risk models should not be based exclusively on prudential terms, but also on other relevant aspects such as ethical and legal aspects as well as consumer and data protection. However, to keep this paper focused, we concentrate on the CRR requirements solely.

3.1 Challenges posed by ML models

Depending on the context of their use, the complexity and interpretability of some ML models might pose additional challenges for the institutions to develop compliant IRB models. Nevertheless, this might not necessarily rule out the use of these models.

Among the specific challenges of using ML models for the purpose of **risk differentiation** there are those related to the following CRR requirements:

- The definition and assignment criteria to grades or pools (Article 171(1)(a) and (b) of the CRR⁶) may be difficult to analyse should sophisticated ML models be used as the main models for risk differentiation. This may constrain the use of models where you do not have a clear economic link between the input and the output variables. This does not mean that ML techniques are not compatible with this requirement, but rather that finding the required clear economic theory and assumptions behind the model may be a challenge. In order to avoid these issues, institutions should look for suitable tools to interpret these complex ML models.
- Complementing human judgement requires understanding of the model and this aspect could hinder the use of difficult to interpret ML models. When an institution uses statistical models for the assignment process to grades or pools, this should be complemented by human judgement. In particular, the complexity of ML may create specific challenges related to human judgement which depend on whether this is applied in the model development and/or in the application of the estimates. Concerning human judgement applied in the model development, the complexity of ML models may make the assessment of the modelling assumptions and whether the selected risk drivers contribute to the risk assessment in line with their economic meaning (as required by Article 174(e) of the CRR⁷) more challenging. Expert judgment may also be used when setting the hyperparameters (as also explained in the technical box on hyperparameters) that are required by specific ML models. Concerning human judgement in the application of the estimates, the complexity of ML models may make it more difficult to take into account aspects which are not, but should be, embedded in the specific predictions (Article 172(3) of the CRR).⁸
- In terms of documentation requirements, Articles 175(1), 175(2) and 175(4)(a) of the CRR⁹ require that if the institution uses a statistical model in the rating assignment process it should document the modelling assumptions and theory behind the model. However, and not limited to the IRB approach, the complexity of some ML models can make it challenging to provide a clear outline of the theory, assumptions and mathematical basis of the final

⁶ Article. 24(1) of the [RTS on the specification of the assessment methodology for competent authorities regarding compliance of an institution with the requirements to use the IRB Approach](#) (RTS on AM).

⁷ Article 42(a)(b) of the RTS on AM and paragraphs 35(a)(b)(c) and 58 of the [Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures](#) (GL on PD and LGD).

⁸ Article 24(2) and 42(c) of the RTS on AM and Section 8.2 of the GL on PD and LGD.

⁹ Articles 3(3) and 32(5)(b), 32(6) of the RTS on AM.

assignment of estimates to grades, individual obligors, exposures or pools. This is in particular true where ML models are used for risk differentiation. Also, the documentation of the model's weaknesses¹⁰ requires that the institution's relevant staff fully understands the model's capabilities and limitations.

Among the specific challenges of using the ML model for the purpose of **risk quantification** there are those related to the following CRR requirements:

- Concerning the estimation process, the plausibility and intuitiveness of the estimates is also required in Article 179(1)(a) of the CRR, but ML models can result in non-intuitive estimates, particularly when the structure of the model is not easily interpretable. Additionally, it can be difficult to correctly make judgmental considerations, as requested by Article 180(1)(d) of the CRR, when combining results of techniques and when making adjustments for different kinds of limitations.
- Article 180(1)(a) in combination with Article 180(1)(h) CRR and Article 180(2)(a) in combination with Article 180(2)(e) CRR requires institutions to estimate PDs by obligor grades or pools from long-run averages of one-year default rates and, in particular, that the length of the underlying historical observation period used shall be at least five years for at least one data source. Similarly, according to Articles 181(1)(j) and 181(2) the estimates of LGD shall be based on data over a minimum of five years. This might be a problem for the use of big data or unstructured data, which might not be available for a sufficiently long-time horizon. Moreover, data retention rules related to the General Data Protection Regulation (GDPR)¹¹ may create further challenges in meeting the minimum five years length of the underlying historical observation in case of natural persons.

Questions:

3: Do you see or expect any challenges regarding the internal user acceptance of ML models (e.g. by credit officers responsible for credit approval)? What are the measures taken to ensure good knowledge of the ML models by their users (e.g. staff training, adapting required documentation to these new models)?

4: If you use or plan to use ML models in the context of IRB, can you please describe if and where (i.e. in which phase of the estimation process, e.g. development, application or both) human intervention is allowed and how it depends on the specific use of the ML model?

5. Do you see any issues in the interaction between data retention requirements of GDPR and the CRR requirements on the length of the historical observation period?

Due to the nature of ML, some specific challenges may arise in performing the **validation** of ML models. These challenges are of two types:

¹⁰ Article 41(d) of the RTS AM.

¹¹ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN>.

- 1) Difficulties in interpreting/resolving the findings of the validation.
- 2) Difficulties related to the validation tasks themselves.

Some of the features of ML lead to both potential challenges, for example, a more complex model may lead to a more complex documentation, which in turn will make its validation harder for the validation function.

In the first category, ML models may make the resolution of identified deficiencies more complex: it may, for example, not be straightforward to understand a decrease in the core model performance (as required by Article 174(d) of the CRR) if the link between input data and risk parameters is not properly understood. Also, the validation of internal estimates may be harder and, in particular as required by Article 185(b) of the CRR, institutions may find it challenging to explain any material difference between the realised default rates and the expected range of variability of the PD estimates for each grade. The comparison between the predictions of the models and the observed default rates can in fact be more challenging due to difficulties in assessing the effect of the economic cycle on the logic of the model.

In the second category, specific challenges related to the validation of the core model performance are related to the following CRR requirements:

- With respect to the assessment of the inputs to the models, where all relevant information is considered when assigning obligors and facilities to grades or pools in accordance with Article 172(1) CRR, it may be more difficult to assess the representativeness and to fulfil the more operational data requirements (e.g. data quality or data storage and maintenance). Moreover, extra care should be taken in evaluating the quality of the data input to avoid cases where the score obtained by a ML model is used as an explanatory variable for another model which could lead to feedback loops.
- In the assessment of the model outcomes, special attention must be placed on the use of out-of-sample and out-of-time samples (as already requested in Article 175(4)(b) CRR) due to the high risk of overfitting.
- In addition, the validation function is expected to analyse and challenge the model design, assumptions and methodology (Article 185 CRR).¹² As such, a more complex model will be harder to challenge efficiently. For instance, the validation of the hyperparameters (more details on hyperparameters are given in the following technical box) may require additional statistical knowledge, and therefore institutions should ensure that the staff in the validation function is appropriately trained.

¹² Article 11(2)(a) of the RTS on AM.

Technical box: Hyperparameters

In order to estimate the parameters of a ML model from data, it is often required to specify a set of **hyperparameters**, which are used to describe the structure of the model and to customise the learning algorithm. For example, in generalised regression models, the highest order terms used is a hyperparameter that determines the structure of the model, and the loss function to minimise is a hyperparameter that characterises the learning process.

Some examples of parameters and hyperparameters that determine the structure of a ML model are given in the following table:

	Tree	Generalised regression	Neural network	Boosting
Hyperparameter	Depth of the tree. Min. observations per leaf.	Polynomial order of the terms. Transformations on the variables.	Number of nodes. Activation functions.	Number of weak prediction models. Hyperparameters of the weak models.
Parameter	Variable and threshold in each split.	Weight of each term.	Weight of each connection.	The parameters of the weak models.

It is important to note that the structure and capacities of a model are greatly determined by the values of the hyperparameters. For example, a binary tree with depth two is a very simple model that can predict at most four different values and can use the information of only three different variables, but if the depth is set to ten the complexity of the model becomes significant, as the number of different possible predictions climbs to 1024 and the number of variables used to 1023. Further, in linear regressions the relationship between an explanatory variable and the objective variable is always monotonic and proportional, but if higher order terms are included in a regression, the relationship can become arbitrarily involved. Hyperparameters can be set by expert judgement (using default values or values which have been proven adequate in a similar problem) and, in the context of supervised learning, can also be obtained by minimising the error of the predictions. This error minimisation cannot be performed on the dataset used to determine the parameters as, in many cases, it would lead to an overly complex model prone to overfitting. Therefore, three different samples are needed to develop a model if hyperparameters are determined from the data:

- Training sample: used to determine the values of the parameters.
- Validation sample: used to determine the values of the hyperparameters.
- Test sample: used to measure the performance of the model.

Questions:

6: Do you have any experience in ML models used for estimating credit risk (if possible, please differentiate between models where ML is used only for risk differentiation, only for risk quantification or used for both)? If so, what are the main challenges you face especially in the areas of:

- a) Methodology (e.g. which tests to use/validation activities to perform).
- b) Traceability (e.g. how to identify the root cause for an identified issue).
- c) Knowledge needed by the validation function (e.g. specialised training sessions on ML techniques by an independent party).
- d) Resources needed to perform the validation (e.g. more time needed for validation)?

7: Can you please elaborate on your strategy to overcome the overfitting issues related to ML models (e.g. cross-validation, regularisation)?

Other challenges related to the use of a ML models are:

- Corporate Governance is also a requirement related to interpretability. The institution's management body remains responsible for meeting legal requirements. The use of ML models makes this more challenging but does not alleviate these parties from their responsibilities. As required by Article 189 of the CRR, all material aspects of the rating and estimation processes shall be approved by the institution's management body or a designated committee thereof and senior management. These parties shall possess a general understanding of the rating systems of the institution and detailed comprehension of its associated management reports.
- The soundness and integrity of the implementation process (Article 144 of the CRR) can be jeopardised by the complex specifications of ML models. Additionally, requirements from Article 171 of the CRR also affect the processes of the rating system, among which implementation processes are included. In particular, the complexity of ML models may make it more difficult to verify the correct implementation of internal ratings and risk parameters in IT systems.¹³
- Categorisation of model changes (as required by Article 143(3) of the CRR¹⁴) may be challenging for models updated at a high frequency with time-varying weights associated to variables (sometimes with change in the sign of the effect). These kinds of models might moreover be difficult to be validated as different 'interim model versions' might exist for one approved model and, therefore, it may be more challenging for biases to be detected

¹³ Article 11(2)(b) of the RTS AM. Another challenge is the setup of a robust and safe infrastructure (i.e. protected from cyberattacks) as required by Article 78 of the RTS on AM.

¹⁴ Article 2 of the RTS for assessing the materiality of extensions and changes of the Internal Ratings-Based Approach and the Advanced Measurement Approach ([Regulation \(EU\) No 529/2014](#)).

(e.g. due to permanent overfitting). A recalibration is generally required, where there is a break in the economic conditions, institutions' processes or in the underlying data. If a potential model change only has a minor impact, the question to be analysed is whether an adaptation of the model in the course of retraining is in fact needed.

- The use of big and unstructured data may pose challenges to institutions related to:
 - Putting in place a process for vetting data inputs into the model which ensures the accuracy, completeness and appropriateness of the data as required by Article 174(b) CRR.
 - Ensuring that the data used to build the model is representative of the application portfolio as requested by Articles 174(c) and 179(1)(d) CRR.

Questions:

8: What are the specific challenges you see regarding the development, maintenance and control of ML models in the IRB context, e.g., when verifying the correct implementation of internal rating and risk parameters in IT systems, when monitoring the correct functioning of the models or when integrating control models for identifying possible incidences?

9: How often do you plan to update your ML models (e.g., by re estimating parameters of the model and/or its hyperparameters) Please explain any related challenges with particular reference to those related to ensuring compliance with [Regulation \(EU\) No 529/2014](#) (i.e. materiality assessment of IRB model changes).

There are other aspects (not necessarily challenges) related to the use of ML models for the purposes of own funds requirements that are also relevant to discuss:

- Use test: Article 144(1)(b) CRR prescribes that internal ratings and default and loss estimates used in the calculation of own funds requirements play an essential role for internal purposes like risk management, credit approval and decision-making processes. The rationale for the use test is to avoid banks using internal models only to reduce on capital requirements, but rather that they believe in them and use them also for internal purposes. This 'use test' requirement may hamper the introduction of the ML models for internal purposes due to the challenges ML may encounter in complying with strict CRR requirements. In this context, this discussion paper seeks to discuss the supervisory expectations around the use of ML in the context of IRB models, in order to clarify a possible proper future use of such techniques.
- The EU-wide legislative proposal on artificial intelligence (AI act)¹⁵ includes among the high-risk use cases the use of AI for evaluating the creditworthiness of natural persons or for establishing their credit scores. Whereas the focus of the AI legislative proposal is on credit

¹⁵<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>.

granting, considering that the decision-making process is covered by the use-test criteria, the requirements of AI should be taken into consideration, where relevant, also in the context of IRB models.

Questions:

10: Are you using or planning to use ML for credit risk apart from regulatory capital purposes? Please specify (i.e. loan origination, loan acquisition, provisioning, ICAAP).

11. Do you see any challenges in using ML in the context of IRB models stemming from the AI act?

12. Do you see any additional challenge or issue that is relevant for discussion related to the use of ML models in the IRB context?

3.2 Potential benefits from the use of ML models

Having analysed in detail the challenges that institutions might encounter when using ML models for prudential purposes, it is fair to say that ML models might prove to be useful in improving IRB models, even helping them to meet some prudential requirements.

In particular, the following are areas where the use of ML models might be beneficial:

- Improving **risk differentiation**, both by improving the model discriminatory power (Article 170(1)(f) and (3)(c) of the CRR)¹⁶ and by providing useful tools for the identification of all the relevant risk drivers or even relations among them (Articles 170(3)(a) and (4) and 171(2) of the CRR).¹⁷ ML models might be used to optimise the portfolio segmentation, to build robust models across geographical and industry sectors/products and take data-driven decisions that balance data availability against the required model granularity. Moreover, ML models might help to confirm data features selected by expert judgement used for ‘traditional’ model development giving a data-driven perspective to the feature selection process (Articles 172(3) and 174(e) of the CRR).
- Improving **risk quantification**, by improving the model predictive ability and detecting material biases (Article 174(a) of the CRR)¹⁸ and also by providing useful tools for the identification of recovery patterns in LGD models.¹⁹ ML models might also help in the calculation of the necessary appropriate adjustments (Article 180(1)(d) and 181(1)(b) of the CRR²⁰).

¹⁶ And Articles 36(1)(a) and 37(1)(c) of the RTS on AM.

¹⁷ Articles 35(2) and 45(1)(a) of the RTS on AM and paragraphs 21, 25, 121 of the GL on PD and LGD.

¹⁸ Articles 36(1)(a) and 37(1)(c) of the RTS on AM.

¹⁹ Paragraph 159 of the GL on PD and LGD.

²⁰ Section 4.4.1 and 4.4.2 of the GL on PD and LGD.

- Improving **data** collection and **preparation** processes including, for example, cleaning of input data or by providing a tool for data treatment and data quality checks (as requested by Article 174(b) of the CRR).²¹ ML models might be useful tools for assessing representativeness (e.g. through unsupervised learning techniques) as requested by Article 174(c). Moreover, ML models might be used for performing outlier detection and for error correction. ML models might allow institutions to use unstructured data (e.g. qualitative data such as business reports), which would expand the data sets that can be used for parameter estimation.
- Improving **credit risk mitigation techniques** where ML models might be used for collateral valuation (e.g. through haircut models).
- Providing robust systems for **validation** and monitoring of the models. ML models might be used to generate model challengers or as a supporting analysis for alternative assumptions or approaches (Article 190(2) of the CRR).²²
- Performing **stress testing**, by assessing the effect of certain specific conditions on the total capital requirements for credit risk and by identifying adverse scenarios (Article 177(1) and (2) of the CRR).

Questions:

13: Are you using or planning to use ML for collateral valuation? Please specify.

14. Do you see any other area where the use of ML models might be beneficial?

²¹ Articles 32(3)(b), 76(1) and (2)(a)(g) of the RTS on AM and paragraph 72 of the GL on PD and LGD.

²² Article 41(b) of the RTS on AM and paragraph 220(b) of the GL on PD and LGD.

4. How to ensure a possible prudent use of ML models going forward

All internal models used for capital purposes require supervisory approval and will need to comply with the requirements set out in the CRR. It may nonetheless be helpful to stress the economic principles underlying requirements for IRB models in the CRR. The use of principles allows the focus to be on the economic and supervisory elements, which banks should consider when seeking approval to apply ML models for regulatory purposes. This section follows this approach and seeks to provide useful guidance to institutions in the context of assessing whether ML models can be ultimately approved by supervisors.

4.1 Concerns about the use of ML

ML models are more complex than traditional techniques such as regression analysis or simple decision trees, and sometimes less 'transparent', therefore the existing risk management approaches and governance framework used for traditional model types may require further enhancements.

The main concerns stemming from the analysis of the CRR requirements relate to the complexity and reliability of the ML models where the main pivotal challenges seem to be the interpretability of the results, the governance with a special reference to increased needs of training for staff and the difficulty in evaluating the generalisation capacity of a model (i.e. avoiding overfitting). To understand the underlying relations between the variables exploited by the model, several interpretability techniques have been developed by practitioners. As highlighted in more detail in the technical box below, the choice of which of these techniques to use can pose a challenge by itself, while these techniques often only allow a limited understanding of the logic of the model.

These concerns form the basis for the principle-based guidance on the minimum expectations from ML used in the context of the IRB framework provided in the following section.

Technical box: Interpretability techniques

When dealing with complex ML models, one of the most significant challenges is to explain why a model produces some given outcomes. To address this difficulty, a number of techniques have been developed that allow some insight into the internal logic of a model to be obtained. Some of the most widely used techniques are:

1. Graphical tools showing the effect of an explanatory variable on the model. Partial dependence plots analyse the effect on the average prediction, while individual conditional expectations show the effect on a specific prediction.
2. Feature importance measures reveal the relevance of each explanatory variable in the overall model.

3. Shapley values quantify the impact of each explanatory variable on a specific prediction of the model.
4. Local explanations, such as LIME and anchors, provide simple approximations of the model on a vicinity of an observation.
5. Counterfactual explanations indicate how a specific prediction of the model could be modified by altering the values of the explanatory variables as little as possible.

The use of these techniques can also pose a challenge by itself in relation to the mathematical hypotheses on which they rely, the difficulty to implement them or due to the computational capacities required. Also, it must be noted that each technique provides only some **partial understanding** of a model, and that their usefulness can greatly vary depending on the case.

Questions:

15: What does your institution do to ensure explainability of the ML models, i.e. the use of *ex post* tools to describe the contribution of individual variables or the introduction of constraints in the algorithm to reduce complexity?

16. Are you concerned about how to share the information gathered on the interpretability with the different stakeholders (e.g. senior management)? What approaches do you think could be useful to address these issues?

4.2 Expectations for a possible and prudent use of ML techniques in the context of the IRB framework

The EBA has already identified the four pillars for the development, implementation and adoption of BD&AA in its report on BD&AA – namely data management, technological infrastructure, organisation and governance and analytics methodology, which are necessary to support the rollout of advanced analytics, along with a set of trust elements that should be properly and sufficiently addressed (namely, ethics, explainability and interpretability, traceability and auditability, fairness and bias prevention/detection, data protection and quality, and consumer protection aspects and security).

Along these lines, instead of concluding which specific ML model might be accepted for which specific prudential function of IRB modelling, this section seeks to discuss a set of recommendations in the form of a principle-based approach to which IRB models should adhere. These principles are intended to make clearer how to adhere to the regulatory requirements set out in the CRR for IRB models.

ML models might add value, provided they ensure acceptable monitoring, validation and explainability of the methodology and of the model outcomes. A good level of institutional understanding about their IRB models is a key element, and with even more relevance when ML models are used for regulatory purposes.

ML might be used for different purposes and at various levels: data preparation, risk differentiation, risk quantification and internal validation purposes. All of the following recommendations apply where ML models are used for risk differentiation²³ and risk quantification purposes, except for explicit divergent indications.

If institutions want to use ML models for regulatory capital purposes, all the relevant stakeholders should have an **appropriate level of knowledge of the model's functioning**. In particular, the EBA recommends that institutions ensure that:

- a. The staff working in the model development unit, credit risk control unit (CRCU) and the validation unit is sufficiently skilled to develop and validate ML models and, therefore, to assess the relevance and appropriateness of the risk drivers used, as well as the soundness of the underlying economic rationale in the overall model. For these purposes, appropriate actions are taken, such as, for example, on the organisation of technical in-depth training sessions.
- b. The management body and senior management are in the position to have a good understanding of the model, by providing them with appropriate high-level documentation. That documentation should at least clarify which indicators or variables are the key drivers for the assignment of exposures to grade or pools, as well as – if relevant – how ML models impact the risk quantification.

It is recommended that institutions find an appropriate balance between model performance and explainability of the results. Indeed, a higher level of complexity may lead to better model performances, but at the cost of lower explainability and comprehension of the model's functioning. Therefore, institutions are recommended to **avoid unnecessary complexity** in the modelling approach if it is not justified by a significant improvement in the predictive capacities. Institutions should avoid:

- a. including an excessive number of explanatory drivers or drivers with no significant predictive information;
- b. using unstructured data if more conventional data is available that provides similar predictive capacities; and
- c. overly complex modelling choices if simpler approaches yielding similar results are available.

In addition, to ensuring that the **model is correctly interpreted and understood**, institutions are recommended to:

- a. Analyse in a statistical manner: i) the relationship of each single risk driver with the output variable, *ceteris paribus*; ii) the overall weight of each risk driver in determining the output variable, in order to detect which risk drivers influence model prediction the most. These

²³ The ML model may be the main model used for risk differentiation purposes or may be used in modules or sub-modules that are combined with other modules, potentially estimated with other simpler techniques. The recommendations are valid for both situations and should be applied at a level consistent with the application of the techniques (for instance, if ML is used only for one module in the PD model, the recommendation should be applied for that module).

- analyses are particularly relevant where a close and punctual representation of the relationship between model output and input variables is not determinable due to the complexity of the model.
- b. Assess the economic relationship of each risk driver with the output variable to ensure that the model estimates are plausible and intuitive.
 - c. Provide a summary document in which the model is explained in an easy manner based on the outcomes of the analyses described in point a. The document should at least describe:
 - i. The key drivers of the model.
 - ii. The main relationships between the risk drivers and the model predictions.
The addressees of the document are all the relevant stakeholders, including the staff which uses the model for internal purposes.
 - d. Ensure that potential biases in the model (e.g. overfitting to the training sample) are detected.

Where ML techniques are used, a good level of understanding of the model is required especially where **human judgment** is applied, even though the exact concerns can be different depending on where human judgement is applied. In particular, when human judgment is used in the development of the model, staff in charge should be in the position to assess the modelling assumptions and whether the selected risk drivers contribute to the risk assessment in line with their economic meaning. If human judgment is used in the application, on the other hand, staff in charge of performing overrides need to be able to consider the behaviour of the model on a specific prediction, taking into account aspects which are not already embedded in it, or identify the cases where the model's logic could be misleading. Finally, institutions are recommended to grant the performance of the override process of the automated model outputs, considering only those aspects which are insufficiently embedded in the automatic rating.

Where ML models are frequently updated, the reason for such **regular updates** needs to be analysed in detail and monitored by the institution. Generally, a break in the economic conditions or in the institutions' processes or in the underlying data might justify a model update. As credit risk is, however, not supposed to change frequently (in contrast to e.g. the market risk), such updates should, in general, not occur frequently. Therefore, the parameters of the model should generally be stable. Moreover, institutions should always compare the changes to the last approved model in order to make sure that many insignificant changes at a high frequency do not lead to an unnoticed material change within a certain time period. Therefore, institutions need to weigh the benefits against the potential risks of automatic and frequent updates of its IRB model.

For complex ML models with limited explainability or for frequently updated models, a reliable **validation** is particularly important and might require increased depth and/or frequency. The institutions are recommended to take care of:

- i. Overfitting issues: ML models are very prone to suffer from overfitting, i.e. performance optimisation of the development sample, which leads to very high performance of the development sample that may not be confirmed on the current and foreseeable application portfolio. Therefore, they should put particular attention on the comparison of the model

performances measured within the development sample with those obtained using out-of-sample and out-of-time samples.

- ii. Challenging the model design: the hyperparameters used to describe the structure of the model and to customise the learning algorithm are often based on human judgement. The validation unit should therefore place particular attention on verifying the rationale behind the choice of these hyperparameters. This check may prove to be particularly challenging for complex models considering that a deep knowledge of the methodology is required to understand all implications of hyperparameters. If, on the contrary, hyperparameters are selected by minimising the error of the model, it should be ensured that this process does not introduce an undesired bias.
- iii. Representativeness and data quality issues: If ML techniques used for risk differentiation and risk quantification purposes are fed with a large amount of data, sufficient data quality needs to be ensured. Where these data are external data, institutions are recommended to place particular care on the assessment of the representativeness of the external data with respect to the application portfolio. In particular, institutions are recommended to verify whether a diminished representativeness leads to a reduction in the performance of the model measured strictly on the internal customers. Institutions should also be particularly careful when using unstructured data in ensuring accuracy, completeness and appropriateness of the data.
- iv. Analysis of the stability of the estimates, also in light of the institution's rating philosophy. It is useful to analyse both the stability:
 - In the assignment process of each debtor/exposure to grades or pools. Indeed, ML algorithms may introduce point-in-time (PiT) elements in the models that may hamper the stability of the rating assignment process compared to more through-the-cycle (TtC) models leading to potential rapid changes in capital requirements;
 - Of the relationship between the output variable and the drivers in subsequent releases of the model based on ML techniques especially in light of the model change policy, to provide an assessment of whether changes between inputs and outputs require regulatory approval, *ex ante* or *ex post* notification.

Where ML techniques are used for **data preparation** purposes, institutions are recommended to ensure that there are clear rules and documentation. Institutions should ensure the appropriateness of the methodology applied to data by means of the application of a proper set of checks and controls.

Question:

17: Do you have any concern related to the principle-based recommendations?

Annex – Summary of questions

1: Do you currently use or plan to use ML models in the context of IRB in your institution? If yes, please specify and answer questions 1.1, 1.2, 1.3. 1.4; if no, are there specific reasons not to use ML models? Please specify (e.g. too costly, interpretability concerns, certain regulatory requirements, etc.)

1.1: For the estimation of which parameters does your institution currently use or plan to use ML models, i.e. PD, LGD, ELBE, EAD, CCF?

1.2: Can you specify for which specific purposes these ML models are used or planned to be used? Please specify at which stage of the estimation process they are used, i.e. data preparation, risk differentiation, risk quantification, validation.

1.3: Please also specify the type of ML models and algorithms (e.g. random forest, k-nearest neighbours, etc.) you currently use or plan to use in the IRB context?

1.4: Are you using or planning to use unstructured data for these ML models? If yes, please specify what kind of data or type of data sources you use or are planning to use. How do you ensure an adequate data quality?

2: Have you outsourced or are you planning to outsource the development and implementation of the ML models and, if yes, for which modelling phase? What are the main challenges you face in this regard?

3: Do you see or expect any challenges regarding the internal user acceptance of ML models (e.g. by credit officers responsible for credit approval)? What are the measures taken to ensure good knowledge of the ML models by their users (e.g. staff training, adapting required documentation to these new models)?

4: If you use or plan to use ML models in the context of IRB, can you please describe if and where (i.e. in which phase of the estimation process, e.g. development, application or both) human intervention is allowed and how it depends on the specific use of the ML model?

5: Do you see any issues in the interaction between data retention requirements of GDPR and the CRR requirements on the length of the historical observation period?

6: Do you have any experience in ML models used for estimating credit risk (if possible, please differentiate between models where ML is used only for risk differentiation, only for risk quantification or used for both)? If so, what are the main challenges you face especially in the areas of:

a) Methodology (e.g. which tests to use/validation activities to perform).

- b) Traceability (e.g. how to identify the root cause for an identified issue).
- c) Knowledge needed by the validation function (e.g. specialised training sessions on ML techniques by an independent party).
- d) Resources needed to perform the validation (e.g. more time needed for validation)?

7: Can you please elaborate on your strategy to overcome the overfitting issues related to ML models (e.g. cross-validation, regularisation)?

8: What are the specific challenges you see regarding the development, maintenance and control of ML models in the IRB context, e.g., when verifying the correct implementation of internal rating and risk parameters in IT systems, when monitoring the correct functioning of the models or when integrating control models for identifying possible incidences?

9: How often do you plan to update your ML models (e.g., by re estimating parameters of the model and/or its hyperparameters) Please explain any related challenges with particular reference to those related to ensuring compliance with [Regulation \(EU\) No 529/2014](#) (i.e. materiality assessment of IRB model changes).

10: Are you using or planning to use ML for credit risk apart from regulatory capital purposes? Please specify (i.e. loan origination, loan acquisition, provisioning, ICAAP).

11. Do you see any challenges in using ML in the context of IRB models stemming from the AI act?

12. Do you see any additional challenge or issue that is relevant for discussion related to the use of ML models in the IRB context?

13: Are you using or planning to use ML for collateral valuation? Please specify.

14. Do you see any other area where the use of ML models might be beneficial?

15: What does your institution do to ensure explainability of the ML models, i.e. the use of ex post tools to describe the contribution of individual variables or the introduction of constraints in the algorithm to reduce complexity?

16. Are you concerned about how to share the information gathered on the interpretability with the different stakeholders (e.g. senior management)? What approaches do you think could be useful to address these issues?

17: Do you have any concern related to the principle-based recommendations?



EUROPEAN BANKING AUTHORITY

Tour Europalaza, 20 avenue André Prothin CS 30154
92927 Paris La Défense CEDEX, FRANCE

Tel. +33 1 86 52 70 00

E-mail: info@eba.europa.eu

<https://eba.europa.eu>