



Comments

EBA discussion Paper on Machine Learning for IRB Models
(EBA/DP/2021/04)

Our ref

Ref. DK: EBA-DP

Ref. DSGV: 7715/10

Contact: Mr Silvio Andrae
Telephone: +49 30 20225-5437
Telefax: +49 30 20225-5404
E-Mail: silvio.andrae@dsgv.de

Berlin, February 11, 2022

The **German Banking Industry Committee** is the joint committee operated by the central associations of the German banking industry. These associations are the Bundesverband der Deutschen Volksbanken und Raiffeisenbanken (BVR), for the cooperative banks, the Bundesverband deutscher Banken (BdB), for the private commercial banks, the Bundesverband Öffentlicher Banken Deutschlands (VÖB), for the public-sector banks, the Deutscher Sparkassen- und Giroverband (DSGV), for the savings banks finance group, and the Verband deutscher Pfandbriefbanken (vdp), for the Pfandbrief banks. Collectively, they represent more than 1,700 banks.

Coordinator:
German Savings Banks Association
Charlottenstraße 47 | 10117 Berlin | Germany
Telephone: +49 30 20225-0
Telefax: +49 30 20225-250
www.die-deutsche-kreditwirtschaft.de

On November 11, 2021, the European Banking Authority (EBA) issued the "Discussion Paper on Machine Learning for IRB Models" for consultation. We are pleased to take the opportunity to comment.

1. General comments

The EBA is considering formulating principles-based recommendations. These rules aim is to ensure the appropriate use of ML techniques by institutions in the context of IRB models. It shall provide a coherent and clear understanding of the supervisory rules, and in this way, harmonization in the EU is achieved. DK supports the distinction made by the EBA that the term machine learning should mainly refer to the more complex models and that statistical methods of regression do not fall under this term. However, since even more complex ML methods for assessing credit risk have so far only been used sporadically, the EBA should not publish its recommendations too early in order not to block further developments from statistical methods to ML methods.

In general, we would like to note that the algorithms underlying ML models and the data used are not per se more complex than in traditional methods. The decisive factor is always the respective application case.

2. Specific comments

Q1. 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.).

At present, complex ML models are still rarely used to assess credit risk. This is often because sufficient data are available for existing portfolios (e.g., high default portfolios). The portfolios are exceptionally homogeneous, and the predictive power of the probability of default is based on statistical regression methods (e.g., discriminant analysis, logistic regression) is outstanding. However, ML models use in IRBA to some extent, and new applications for various purposes are being considered.

Q1.1. For the estimation of which parameters does your institution currently use or plan to use ML models, i.e. PD, LGD, ELBE, EAD, CCF?

ML models are already used for PD and LGD estimation. In the future, they will also use them for EAD estimation (see Q1.3).

Q1.2. 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.

PD: Bucketing in order to estimate alpha and beta for PD calibration

LGD: Several ML models were used in the risk differentiation phase

We see an advantage in using a lot of data (including metadata) for ML methods, as this can strengthen the robustness of ML models. ML methods can intelligently replace missing data sets and thus increase model quality. Also, ML can detect anomalies (e.g., systematic biases in the data) that are not detected by the standardized DQ checks.

We want to point out that the "sufficiency" of the data size depends very much on the approach used. Indeed, the additional benefit of using enormous data sets in classical statistical modeling may be

statistically insignificant, so the cost of further data processing is not justified. On the other hand, high-dimensional big data is required to build nonparametric predictive models. Therefore, to make sense of Big Data in the context of risk analysis, discussions about data properties and methods should go hand in hand.

Q1.3 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?

PD: K-means clustering algorithm is used to compute the score buckets for the calibration. The use of other machine learning algorithms (in particular, Artificial Neural Networks, Random Forest algorithm) should be evaluated e.g. for the estimation of different scores for PD modules. Particularly, the performance obtained by using such models could be compared with the performance of the currently used models.

LGD: knn-method was used for bucketing purposes. Several regression models were also used in the risk differentiation to estimate the relationship between input variables and their associated features. Decision trees were tested but not finally selected for risk differentiation.

EAD: It is planned to use Decision Trees.

Q1.4. 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?

Not at the moment since the use of unstructured data seems very complicated, both because of the memory space required and the challenges with regard to interpretability.

Q2. 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?

N/A

Q3. 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)?

The low explanatory power of ML-driven models for credit risk remains perhaps their biggest drawback. Visual inspection of, say, a random forest is not possible, and although there are some tools (such as feature importance) that provide insight into the inner workings of this type of model, the logic of ML models is much more complicated than that of a traditional logistic regression approach. In particular, in most cases, it is impossible to see a final formula of the model (as in the case of logistic regression).

Acceptance problems would probably arise especially in the cases where non-linear effects are present which credit officers may not be aware of. Therefore, the use of such models would certainly require more documentation and explanation of the reasons for their use. Nevertheless, also some currently used models are already difficult to interpret for users with no statistical knowledge.

Due to the typical "black-box" characteristic of ML models and the fact that "new" computational tools, whose idiosyncratic feature is to be non-deterministic, would be included in models' methodologies,

teams that are entitled to supervise the efficacy of the models will apply even more prudential approaches in order to avoid unexpected results, leading to a loss in efficiency in the model process lifecycle. Moreover, the computational time could undermine the acceptance of ML models.

Q4. 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?

They are only used in model development but are not applied in production. The results of the ML methods are assessed by developers together with business who decide whether the method should be applied for development instead of the classic approach.

Q5. 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?

We do not see any additional problems compared to the existing ones.

Q6. 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)?

The gain in the predictive power of pure ML models (such as neural networks, support vector machines, and random forests) has been limited for institutions. For this reason, many institutions have been hesitant to adopt them. Which model is best suited for a given situation usually depends on data availability and the predictive outcome, which must be relevant for a bank as a whole.

However, there are also institutions that do have experience for certain predefined steps that include both risk differentiation and risk quantification. The experience arises from development activities irrespectively on whether the ML method was chosen as the final approach. In general, from a local perspective, the answers to the single points below (a. – d.) would be pre-specified by Group Risk Governance.

Q7. Can you please elaborate on your strategy to overcome the overfitting issues related to ML models (e.g. cross-validation, regularisation)?

For the above reasons, some banks are experimenting with so-called hybrid methods. Such methods allow for more advanced feature engineering than conventional models while retaining the setting of a logit model so that the explainability factor is preserved (as much as possible).

Some institutions make use of training, test, and out-of-sample validation samples as well as k-fold cross validation, feature selection (already widely used for the purposes mentioned above - even with current techniques), reduction of layers and size of the model (ANNs case).

Sequential Forward Selection (SFS) is usually used to select features. Only those features that perform optimally are to be chosen. For SFS stability, cross-validation (fitting to multiple subsamples) is applied. These hybrid procedures have not yet been used in the context of IRBA. It is often noted that although the hybrid approach is more parsimonious (at least in some applications), it achieves essentially the same area under the AUC curve as the longstanding approaches.

Q8. 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?

Build up expertise in the latest ML methods and techniques to ensure that the tools used to develop, maintain, and control the resulting model are adequate and well understood. Other than that, the models need to be monitored even more closely than traditional models as the behavior of those models have a higher uncertainty than classic models. To build up knowledge and a better expectational wisdom, it would be beneficial to work with ML models firstly in a shadow environment.

A challenge is posed by the shorter re-calibration cycles and extended testing activities to monitor the functionality of the ML models. Both mean a higher use of working time.

Q9. 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).

In uncertain times, however, the effectiveness of a synthetic hyperparameter with high predictive power is unclear. Models built before the pandemic may be driven by feature-engineering techniques that rely on pre-2020 data, which of course, do not include pandemic risk factors. Experts can identify such risk factors more quickly than complex ML models with data-poor hyperparameters. Indeed, a hyperparameter carries more model risk than the original features.

The benefit of ML techniques is that they can quite easily adapt to new data / extended time series, so a re-estimation of the parameters can be done with new incoming data. In case this update of parameters within the scope of existing models would be seen as a model change, this would lead to a non-applicability of dynamic modeling.

Q10. 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).

There are activities in the area of credit risk monitoring.

Q11. Do you see any challenges in using ML in the context of IRB models stemming from the AI act?

Under the current reading of the Commission's draft AI regulation, ML procedures would be categorized as high-risk applications in the context of the IRBA. The regulatory proposal would not, at this stage, promote the introduction of an ML model for regulatory capital adequacy purposes.

Apart from that, as the application of ML models in banking is restricted to bank related topics which are quite strictly regulated anyway (with respect to ethics and applicability), we do not expect additional challenges.

Q12. Do you see any additional challenge or issue that is relevant for discussion related to the use of ML models in the IRB context?

N/A

Q13. Are you using or planning to use ML for collateral valuation? Please specify.

N/A

Q14. Do you see any other area where the use of ML models might be beneficial?

Besides the mentioned areas ML models might of course be beneficial in the whole process of banking, i.e., customer acquisition (identification techniques), segmentation (clustering techniques), CRM, extended pre-warning and monitoring systems, etc. Generally, the use of machine learning techniques should be evaluated in all those areas of the bank where a large amount of data is available.

Q15. 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?

Indeed, ML methods for IRBA focus on both the explainability and the model's performance. The real challenge, in our view, is to match the ML model to the complexity of the problem. The key is whether a result is plausible using ML. Explainability should not be equated with causality. Explainability is achieved, for example, through a detailed understanding of primary model drivers and their sensitivities.

Accordingly, explainability can also serve to improve the model. However, ease of explanation does not necessarily mean that the model is per se better.

We consider several XAI methods to be applicable in practice. They work on the principle that a simplified model is trained on a complex model to reduce its complexity and make it tangible to humans. In the sense of necessary conditions and transparency, they are well suited to provide plausibility approaches and thus falsify a model. They are not suitable to provide sufficient "explanations" because the models are usually not causal. There is always a decision as to how much remaining complexity in the model is acceptable. There is a trade-off.

Feature-Importances, Shapley Values, and the LIME method provide hints and should also be made available to the users in the application itself. The deeper analysis can be clustered and used to find explanatory patterns and approximate causal relationships.

Q16. 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?

N/A

Q17. Do you have any concern related to the principle-based recommendations?

As already noted in the general comments, it seems to us to be too early to formulate principles at this stage. The critical point is where these principles would be enshrined in law (e.g., in the CRR concerning Pillar 1 or the form of guidelines for Pillar 2).

In principle, attention should be paid to the explainability of the risk metrics (PDs, LGDs) using the ML model during development, production, and validation.

Model governance can remain unchanged. It should maintain the proven division of labor between modeling and validation. We do not see a need for adjustments. As is already the case with existing statistical methods, it should meet the requirements of model risk management accordingly. Model adaptations should be limited to the central computational process of the model. Changes to the IT architecture and non-ML functions should not be classified as such. This fact would be significant, especially for hybrid approaches.

The institute has to ensure the ongoing monitoring of (ML) models. The existing regulatory framework and model approval are sufficient for this. In our opinion, an extension is not necessary. The banks address specific aspects through their governance.