INTESA SANPAOLO
Transparency Register 24037141789-48

EBA DISCUSSION PAPER ON MACHINE LEARNING FOR IRB MODELS (EBA/DP/2021/04)

Foreword

Intesa Sanpaolo welcomes the discussion paper on machine learning used in the context of internal ratings-based (IRB) models to calculate regulatory capital for credit risk and supports the EBA's goals to build up a common understanding of their general aspects, and the related challenges and opportunities in complying with the prudential requirements.

Intesa Sanpaolo deems the subject under discussion to be of great relevance and has attentively followed the developments regarding the correct application of the machine learning approach for the IRB models. The introduction of some components developed with ML techniques into the SME Retail rating model has already been validated by the Regulator in 2021, while the validation of the Retail rating model is awaited for 2022.

As a general introduction, Intesa Sanpaolo would like to anticipate the following most important and impacting issues of the new principle-based recommendations regarding the use of machine learning models in the context of the IRB framework:

- the necessity to avoid eventual overlaps with the existing legislative framework;
- the importance to define a commonly used taxonomy regarding the machine learning approach, as the notions such as "artificial intelligence" and "machine learning" cover a lot of different techniques and estimation methods varying in complexity and interpretability;
- the regulatory approval of the models should be technology-neutral, and it is important to take into account not only the drawbacks of the machine learning approach, but also the improvement of model performance and a more extensive coverage of potential discriminating risk drivers;
- the approval process of the machine learning approach should not be specific, based on excessively sever rules, thus potentially precluding the possibility of its application;
- during the regulatory validation of the innovative algorithms, the improvement of the performance with respect to the results produced by traditional regression models should be given due consideration, without necessarily requiring additional redundant tests that increase the supervisory costs and do not add any value.

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.).

Yes, we currently use a machine learning approach for the regulatory rating models dedicated to Retail and SME Retail customers, with regards to which the sample size proved to be adequate. The SME Retail rating model has been validated by the Regulator in 2021 and we are waiting for the validation of the Retail model in 2022. Furthermore, Internal Validation developed in parallel a challenger model for Retail portfolio for monitoring and benchmarking purposes of the results obtained in the IRB context.

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?

We use machine learning for the estimation of PD models.

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.

The machine learning approach is adopted in the risk differentiation phase of the model development process (estimation of the scoring functions based on new data sources) for some components of the PD risk parameter . The ML algorithms have been tested and compared with the performance of the traditional Logistic Regression in the risk differentiation phase, obtaining quite satisfactory results. This has been done with the objective to improve the discriminatory power of the rating models and to include the largest information set possible to be used in order to obtain a comprehensive risk assessment of the Retail and SME Retail clients.

The machine learning approach is also adopted in the validation, in both initial (model change request) and in the ongoing validation phase in order to fulfill the monitoring of the results obtained by the same technique in the IRB context.

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?

The algorithm currently used in our rating models is the Extreme Gradient Boosting technique available in the Python library “XGBoost”. The Gradient boosting is an approach where new models are created with the objective to predict the residuals or errors of prior models and then added together to make the final prediction. The approach is called gradient boosting because it uses a gradient descent algorithm to minimize the loss when adding new models. This approach supports both regression and classification predictive modeling problems.

In addition, the Random Forests approach has been tested as well.

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?

For the time being, we don’t use unstructured data for ML models, given the general challenges associated with the necessity to ensure an adequate level of data quality and representativeness of the application portfolio.

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?

No, in line with the internal development and implementation of the traditional IRB models, we have not outsourced or are planning to outsource the development and implementation of the ML models.

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)?

With regards to the human judgement applied in the model development phase (e.g. model assumptions or economic meaning of the risk drivers), while it is true that there could exist certain difficulties related to the verification of model assumptions or economic meaning, it is also true that there is a rather sufficient number of modelling and validation techniques that could be used in order to mitigate or overcome these difficulties. For example, if an initial attentive data preparation and selection of risk drivers, that have a clear economic meaning for a certain portfolio of clients, has been conducted, the redundancy, excessive correlation, sample-dependence and poor understanding of the final model can be avoided. In addition, the feature importance and interpretability analysis, out-of-sample and out-of-time validation tests (including the annual back-testing analysis and feedback collection from model users), as well as benchmarking analysis (e.g. with respect to traditional logistic regression models) can further help to avoid the biases due to overfitting or lack of representativeness and "black box" models.

As for the human judgment in the application of the ML models, the rating analysts using the model should be adequately trained with regards to the general model structure and model design, as well as in relation to the economic and credit meaning of the risk drivers contributing to the model output. That is why, regardless of the model estimation techniques, the final model should have a clear and intuitive meaning. For those portfolios of clients where an override application is possible, the rating analyst should be able to assess if all the relevant information contributing to the correct credit risk assessment is already embedded in the model or should be evaluated by applying the human judgment, as it cannot be taken into account in an automatic way. For this purpose, there should be defined clear guidelines describing the possible override motivations based on specific cases to be analyzed in the rating attribution process. Finally, we are going to apply the ML approach to the segments covered by the models with a higher level of automatization, where the human judgement and overrides are less frequent (e.g. Retail and SME Retail).

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?

The machine learning approach is currently used for the Retail and SME Retail rating models. In the model development phase, the human judgment is present when selecting the initial set of risk drivers, as well as when assessing the appropriateness of the final model. In addition, all the analyses and statistical tests aimed at assessing the model performance and interpretability are naturally

accompanied by human judgment. As for the model application phase, the human judgment is currently present only for the SME Retail rating model that includes a specific override framework.

---

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?

With regards to the Retail rating model, validated by the competent authorities, we don't use sensitive personal data. Other personal data is used in compliance with the GDPR requirements and retained for as long as its need is justified.

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. specialized training sessions on ML techniques by an independent party).

d) Resources needed to perform the validation (e.g. more time needed for validation)?

We used the Machine learning approach only for the risk differentiation phase, giving preference to the Extreme Gradient Boosting (XGB) algorithm due to its best results in terms of accuracy and stability, as well as based on the consolidated experience in its application for the rating models development.

The approach adopted for the long list definition is the same used for other modules estimated with traditional methods.

After that, the long list is analyzed with both machine learning techniques and traditional logistic regression in order to be able to comparing the performances. The development sample has been split in 2 sub-samples, following a specific stratification for significant variables:

- A training database consisting of 80% of the entire population
- A test database consisting of the remaining 20% of the entire population

In order to fit the model, an optimization with a grid search hyperparameters was firstly conducted, to explore a wide combination of parameters. This approach has been done with a Cross Validation methodology that works by splitting the dataset into k-parts (e.g. k=5 or k=10). Each split of the data is called a fold. The algorithm is trained on k-1 folds with one held back and tested on the held back fold. This is repeated so that each fold of the dataset is given a chance to be the held back test set. The result is a more reliable estimate of the performance of the algorithm on new data given your test data. It is more accurate because the algorithm is trained and evaluated multiple times on different data.

After the abovementioned step, the performance has been analyzed and compared both on training and test datasets. In order to reduce the number of selected risk drivers, at the same time maintaining

the model performance sufficiently invariant, features importance of variables has been calculated. Then only the most important ones have been selected and used in the chosen model.

Finally, interpretability techniques have been performed (as pointed out in reply to question 15).

In our experience, the main challenges, other than those presented by the compliance with the existing regulatory framework, regard the acquisition of adequate-level technical skills needed for model development and validation, as well as the annual model review management once the model is approved and used for regulatory purposes.

With reference to the specific challenges in validating ML models, in our internal practice, the most relevant aspects regarding the model design, assumptions and methodology were specifically challenged by the Internal Validation Function in the initial model validation stage. More recently, in relation to the Retail model change waiting for the supervisory authorization in 2022, the IVF has also developed a dedicated challenger model to be used as a benchmark within ongoing validation activities, once deployed. In this context, further enhancement of the internal validation framework might however be employed to address the inherent specificities of ML models.

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

We used Cross Validation methodology and optimized the hyperparameters, with particular attention on regularization ones (lambda, alpha, etc.), to make the model more conservative and avoid overfitting. The XGBoost has shown good performance also out-of-sample and out-of-time. Moreover, the rating models are annually assessed by the internal validation function within the review of estimates and back-testing framework.

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?

Similar to the rating models developed by applying traditional regression analysis, the ML models should be accompanied by clear and objective frameworks setting the general rules for the model development and validation, as well as for the governance, data quality and periodic monitoring by relevant internal bodies in order to be able to promptly identify eventual deficiencies both in terms of correct functioning and robust performance in the application portfolio. In addition, in our view, a set of "core" model elements (in terms of relative weights) should be identified and monitored with more attention based on the materiality principle. For example, it should be avoided that the risk drivers that have a relatively small weight in the model output determine the necessity of model redevelopment thus rendering the rating system less stable over time.

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).

The current rating models developed by applying the machine learning approach do not have self-learning solutions, in compliance with the Regulation (EU) No 529/2014 and internal review of estimates and materiality assessment frameworks, the model performance is analyzed annually by integrating the time series underlying the model development with the most recent data. This assessment can result in a confirmation of the existing model, as well as in the necessity of model re-calibration or re-estimation, based on the materiality principle. In the meanwhile, the model parameters remain stable over time.

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).

At the moment, the machine learning approach is introduced for the IRB Retail and SME Retail models, which are also used for managerial purposes.

---

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

Not at the moment, unless further complexity regarding the regulatory framework of IRB models is introduced.

---

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?

In our view, apart from taking correctly into account the peculiarities of the ML approach discussed in the previous points, a possible additional challenge could be presented by eventual excessive regulation aimed specifically at the models developed by applying the machine learning approach. The existing general regulatory framework is rather extensive and already covers sufficiently all the relevant aspects regarding the development, application and monitoring of the IRB models. The eventual adjustments should be made instead at the internal framework level, giving major details whenever deemed necessary. On the other hand, a principle-based set of recommendations, taking into account the experience of all the market participants, could be very useful in order to adequately keep track of all the important aspects to be considered when developing the models based on the machine learning approach. In addition, it could help to exploit its vast opportunities and maximize the potential positive effects that should not be disregarded (e.g. use of new data sources covering the risk drivers relevant for a correct credit risk level assessment, as well as improvement of the discriminatory power and predictive ability of the model).

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

Not at the moment.

---

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

The bank is using the Machine Learning approach for Fraud Risk, for the second level Credit Controls and Reporting, as well as in some other areas. The application of ML techniques is deemed

useful to identify unexpected data patterns and thus to also enhance the data quality framework.

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?

In order to ensure explainability of the results produced by the models developed with ML approach, we have used several commonly used methodologies present in the literature, aimed at interpreting the variables. The SHAP methodology based on the Shapley values has shown the best results, in our experience. With this approach, we can analyze a marginal contribution of each variable considering every possible combination with the remaining ones. In particular, we can observe the change in the score for each combination, assigning a weight to the single variable. In this way, we can assess which features are more important in the model.

---

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?

The feature importance analysis renders the information regarding the interpretability of risk drivers and final model similar to the one obtained with traditional regression analysis approach. The results can be adequately explained and demonstrated to the relevant stakeholders.

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

Apart from the suggestions highlighted in the previous questions and pointed out in the foreword, no other concern at the moment.