Discussion Paper on machine learning for IRB models

European Banking Authority

www.managementsolutions.com

List of abbreviations

Abbreviations	Meaning
CRR	Capital Requirements Regulation
IRB	Internal Rating-Based approach
LGD	Loss Given Default
ML	Machine Learning
PD	Probability of Default

1 Context

The Discussion Paper aims to understand the challenges and opportunities which financial institutions may face when using machine learning to develop CRR-compliant IRB models, and identify some possible obstacles and practical issues

- Since the first Basel Accord was put in place almost two decades ago, **the approaches for developing IRB models by institutions to calculate regulatory capital requirements for credit risk have not materially developed**. Moreover, **regulators and supervisors have focused more on making the estimates produced by different models comparable**, by improving definitions of basic concepts (i.e. default), rather tan focusing on the increasing challenges that rise from the world of advanced technology (i.e. machine learning (ML) and artificial intelligence (AI)).
- In the recent years, **Big Data has emerged as a result of the increase in data availability and storing capacity**, coupled with the improvements in computing power. This reality has created challenges for standard regression models, due to their inability to keep track with the emergence of Big Data.
- As a response, **ML models use this data as fuel that provides the necessary information for developing and improving features and pattern recognition capabilities**. However, this data should have high quality and be available in large quantities for these models to properly work.

In this context, the EBA aims to identify the main challenges and possible benefits of these new models in the IRB context, as well as to provide a set of principle-based recommendations which should ensure proper future use by banks for prudential purposes. The main objective is to provide a consistent and clear understanding of the prudential provisions, and how new sophisticated ML models might coexist with these requirements. Therefore, ultimately the aim is to ensure that the outcome (in terms of setting capital requirements in a prudent manner) continues to be harmonized across Europe.

Following this objective, the EBA published:

- In **December 2020**, a report on the recent trends of **big data and advanced analytics**, including ML, in the banking sector and on the key considerations in the development, implementation and adoption of big data and advanced analytics.
- In **June 2021**, a report on the **current RegTech landscape in the EU**. This report assesses the benefits and challenges faced by financial institutions and RegTech providers in the use of RegTech.

In addition, the EBA has published **a Discussion Paper on Machine Learning for Internal Ratings-based (IRB) models**, which aims at discussing the relevance of possible obstacles to the implementation of ML models in the IRB model space based on some practical issues. The use of data, explainability and other challenges are generally not new practical issues to IRB models, but may be exacerbated when using ML models and, therefore, may lead to specific challenges.

Access the Discussion Paper

2 Executive summary

The main objective is to provide a consistent and clear understanding of how new sophisticated ML models might coexist with prudential requirements, identify the main challenges and possible benefits of ML models as well as providing a set of principle-based recommendations in the context of IRB models

1. Learning paradigms

Learning paradigms can be used to **train ML models depending on the goal of the model and the data** required. Some popular learning paradigms are:

- **Supervised learning**
- **Unsupervised learning**
- **Reinforcement learning**

There are plenty of other categorizations possible.

2. Current use in Credit Risk Modelling

For **IRB models the use of ML has been limited, and these models are used only as a complement to a standard model** used for regulatory purposes (CRR). Examples where ML techniques are currently used in the context of IRB models in compliance with CRR requirements are:

- **Model validation (e.g. challenger models)**
- **Data improvement (e.g. filling in missing values)**
- **Variable selection (e.g. optimizing the selection of variables)**
- **Risk differentiation (e.g. in changes of the risk grade)**

3. Challenges and potential benefits

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.

- **Risk differentiation and quantification challenges**
- **Model validation challenges**

The use of ML models might be beneficial in terms of **risk differentiation, risk quantification, data collection, credit risk mitigation techniques, validation and performing stress testing.**

4. Development and recommendations

The following **four pillars for development** are necessary to support the rollout of advanced analytics, along with a set of trust elements[1] that should be properly and sufficiently addressed:

- **Data management**
- **Technological infrastructure**
- **Organization and governance**
- **Analytics methodology**

Specific recommendations for ML models include: appropriate **knowledge of the models**, **model interpretability**, **low complexity for model use**, and **adequate model validation** techniques.

(1) Stated in the report on the recent trends of big data and advanced analytics.

3 Machine Learning (1/2)

Definition and learning paradigms

Complex ML models are characterized by the use of a high number of parameters and, therefore, require a large volume of 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

Machine Learning (ML): Process using algorithms rather than procedural coding that enables learning from existing data in order to predict future outcomes.

- 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**.

Uses

- The term ML is often used by practitioners to refer only to the more complex models. In the financial sector, **the term ML is reserved for the more pioneering models**[1].
- 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.).

Different **learning paradigms** can be used to train ML models, depending on the goal of the model and the data required. The most popular learning paradigms are:

- **Supervised learning**: the algorithm learns rules for building the model from a labelled dataset and use these rules to predict labels on new input data.
- **Unsupervised learning**: 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**: the algorithm learns from interacting with the environment, rather than from a training dataset. Moreover, reinforcement learning does not require labelled input/output pairs. The algorithm learns to perform a specific task by trial and error.

(1) Although there is no clear-cut distinction between simple and advanced models, the discussion paper focuses on complex ML models.

3 Machine Learning (2/2)

Current use in Credit Risk Modelling

Within credit risk, decision, monitoring and collections are areas where most commonly ML is used. In this regard, regulatory requirements are perceived as a challenge for the application of ML models, since these are more complex to interpret and explain

✓ Within Credit risk, ML is mostly used in credit decisions/pricing followed by credit monitoring and collections, restructuring and recovery.
✗ In contrast, the use of ML is more limited for regulatory areas such as capital requirements for credit risk, stress testing and provisioning.

Current use of ML in IRB Models

- For IRB models, the use of ML has been more limited and 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**: 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 analyze 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 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.
- ML models might possibly be **used as primary IRB models for prediction** (e.g., for risk differentiation), but there exist some challenges. Institutions tend to make strategic decisions focusing on areas such as credit monitoring or collections and recovery.
- 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** might have shaped the use of ML models as of today, 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.

4 Challenges and potential benefits of ML models (1/5)

Risk differentiation challenges

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 for the purpose of risk differentiation

		Related CRR Arts.
Definition and assignment criteria	• According to CRR, an institution shall have specific definitions, processes and criteria for assigning exposures to grades or pools within a rating system that comply with several requirements[1]. The **definition and assignment criteria to grades or pools may be difficult to analyze should sophisticated ML models** be used as the main models for risk differentiation. This may constraint the use of these models where there is not a clear economic link between input and output variables. • In order to avoid these issues, institutions **should look for suitable tools to interpret these complex ML models.**	1. Article 171(1)(a) and (b) CRR
Complementing human judgement	• When an institution uses statistical models for the assignment process to grades or pools, this should be complemented by human judgement. • 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 (required in CRR[2]) more challenging.	2. Article 174(e) CRR
Documentation requirements	• CRR requires that if the institution uses a statistical model in the rating assignment process it should document the modelling assumptions and theory behind the model[3]. 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 assignment of estimates to grades, individual obligors, exposures or pools[3]. • **This happens specially when ML models are used for risk differentiation.** • The documentation of the model's weaknesses requires that the institution's relevant staff fully understands the model's capabilities and limitations.	3. Articles 175(1), 175(2) and 175(4)(a) CRR

4 Challenges and potential benefits of ML models (2/5)

Risk quantification challenges

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 for the purpose of risk quantification

		Related CRR Arts.
Estimation process	• An institution's own estimates of the risk parameters PD, LGD, CCF must be plausible and intuitive according to CRR[1]. However, **ML models can result in non-intuitive estimates**, particularly when the structure of the model is not easily interpretable. • Institutions shall recognize the importance of judgmental considerations in combining results of techniques and in making adjustments for limitations of techniques and information[2]. It can be difficult to correctly make judgmental considerations	1. Article 179(1)(a) of the CRR 2. Article 180(1)(d) of the CRR
PDs by obligor grades or pools	• The 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[3]. • This might be a **problem when using 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) may create further challenges** in meeting the minimum five years length of the underlying historical observation in case of natural persons.	3. Article 180(1)(a) and (h) and Article 180(2)(a) and (e) CRR

4 Challenges and potential benefits of ML models (3/5)

Model validation challenges

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 for the purpose of model validation

		Related CRR Arts.
Interpreting and resolving the findings of the validation	• According to CRR, the institution shall have a regular cycle of model validation that includes monitoring of model performance and stability[1]. ML models may make the **resolution of identified deficiencies more complex** for example. • It may not be straightforward to **understand a decrease in the core model performance** if the link between input data and risk parameters is not properly understood. • Furthermore, CRR requires institutions to **regularly compare realized default rates with estimated PDs for each grade** and, where realized default rates are outside the expected range for that grade, institutions shall specifically analyze the reasons for the deviation. • The validation of internal estimates may be harder, and institutions may find it challenging to explain any material **difference between the realized default rates and the expected range of variability of the PD estimates** for each grade[2].	1. Article 174(d) CRR 2. Article 185(b) CRR
Validation of the core model performance	• 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[3] , it may be more difficult to assess the representativeness and to fulfil the more operational data requirements. 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. • The validation function is expected to **analyze and challenge the model design**, assumptions and methodology[4] .As such, a more complex model will be harder to challenge efficiently (i.e. evaluation of the hyperparameters may require additional statistical knowledge).	3. Article 172(1) CRR 4. Article 185 CRR

4 Challenges and potential benefits of ML models (4/5)

Other challenges

The use of ML models may pose other challenges with regards to corporate governance, implementation process, categorization of model changes and the use of big and unstructured data

Other challenges related to the use of ML models are:

		Related CRR Arts.
Corporate Governance	• 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[1]. Therefore, corporate governance is a requirement related to interpretability.	1. Art.189 CRR
Implementation process	• The requirements contained in CRR in relation to the assignment to grades or pools[2] 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**.	2. Art. 171 CRR
Categorization of model changes	• CRR requires institutions to obtain the prior permission of the competent authorities for material changes to the range of application of a rating system or material changes to a rating system[3]. This **categorization of model changes may be challenging for models updated at a high frequency with time-varying weights** associated to variables. 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 analyzed is whether an adaption of the model in the course of retraining is in fact needed.	3. Art. 143(3) CRR
Use of big and unstructured data	This may pose challenges to institutions related to: ○ Putting in place a **process for vetting data inputs** into the model which ensures the accuracy of the data[4]. ○ Ensuring that the **data** used to build the model is **representative of the application portfolio[5]**.	4. Art. 174(b) CRR 5. Art. 174 (c) and 179(1)(d) CRR

Aspects related to the use of ML models for the purposes of own funds:

- 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. This 'use test' requirement may hamper the introduction of the ML models for internal purposes.
- The EU-wide legislative proposal on artificial intelligence includes 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 granting, the requirements of AI should be taken into consideration, in the context of IRB models.

4 Challenges and potential benefits of ML models (5/5)

Potential benefits

ML models might prove to be useful in improving IRB models, even helping them to meet some prudential requirements. In fact the use of ML models might be beneficial in terms of risk differentiation, risk quantification, data collection, credit risk mitigation techniques, validation and performing stress testing

		Related CRR Arts.
Improving risk differentiation	In relation to the structure of rating systems contained in CRR[1], ML models can improve the model **discriminatory power** and by provide useful tools for the i**dentification of all the relevant risk** drivers or even relations among them. ML models might be used to **optimize the portfolio segmentation**, **take data-driven decisions that balance data availability** against the required model granularity.	1. Art.170(1)(f) and (3)(c) CRR and Art.170(3)(a) and (4) and 171(2) CRR
Improving risk quantification	In relation to the use of models, CRR requires that **the model shall have good predictive power and capital requirements shall not be distorted** as a result of its use[2]. In this sense, ML improves risk quantification, by improving the model predictive ability and detecting material biases. Furthermore, ML models might also help in the **calculation of the necessary appropriate adjustments that institutions shall use in the PD estimation techniques[3]**.	2. Article 174(a) CRR 3. Art. 180(1)(d) and 181(1)(b) CRR
Improving data collection and preparation	CRR requires institutions to establish a **rigorous statistical process including out-of-time and out-of-sample performance tests for** validating the model[4]. ML models can improve data collection and preparation processes including, for example, cleaning of input data or by providing a tool for data treatment and data quality checks. Furthermore, CRR requires institutions to indicate any circumstances under which the model does not work effectively[5]. ML models might be used for performing outlier detection and for error correction.	4. Article 174(b) of the CRR 5. Art. 174(c) CRR.
Improving credit risk mitigation techniques	The ML models might be used for **collateral valuation** (e.g. through haircut models).	N/A
Providing robust systems for validation	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[6]**.	6. Art. 190(2) CRR
Performing stress testing	According to CRR, an institution shall have in place sound stress testing processes for use in the assessment of its capital adequacy[7].ML can assess the effect of certain specific conditions on the total capital requirements for credit risk and by **identifying adverse scenarios**.	7. Art. 177(1) and (2) of the CRR

ManagementSolutions Making things happen

 |

5 Prudent use of ML models (1/2)

Challenges and development for ML

ML models are more complex than traditional techniques and sometimes less ‘transparent’ . The main concerns stemming from the analysis of the CRR requirements relate to the complexity and reliability.

Challenges

- The **main challenges** highligted are:
  - The **interpretability** of the results[1]
  - The **governance,** with a special reference to increased need for training the staff
  - The difficulty in **evaluating the generalization capacity of a model** (i.e. avoiding overfitting).

EBA Development for ML tecniques

- There are **four pillars** for the development:
  - **Data management**
  - **Technological infrastructure**
  - **Organization and governance**
  - **Analytics methodology**

These pillars are necessary to support the rollout of advanced analytics, along with a set of trust elements that should be properly and sufficiently addressed (i.e. ethics, explainability and interpretability, traceability and auditability, fairness and bias prevention/detection, data protection and data quality, and consumer protection and security).

(1) One of the most significant challenges dealing with complex ML models is to explain why a model produces some given outcomes. There are techniques that provide only some partial understanding of a model, and that their usefulness can greatly vary depending on the case (i.e. Graphical tools, Shapley values, LIME, differential alteration of explanatory variables…).

5 Prudent use of ML models (2/2)

Specific recommendations for ML

ML models are more complex than traditional techniques and sometimes less 'transparent'. The main concerns stemming from the analysis of the CRR requirements relate to the complexity and reliability

Model knowledge

- Analysts should have sufficient knowledge to **develop and validate** the ML model.
- **The relevance** and appropriateness of the **risk drivers** used should be assessed.
- The **underlying economic rationale should be clear.**
- The documentation should clarify **which indicators or variables are the key drivers for the assignment of exposures to grade or pools**.

Easy to understand

Institutions should:

- Detect which **risk drivers influence** model prediction the most
- Assess the **economic relationship** of each risk driver
- Ensure that **potential biases** in the model **are detected** (i.e. overfitting)
- Ensure that the documentation adequately describes the model, the risk drivers and their relation with the model predictions
- **Analyze and monitor regular updates** in detail

No complexity

Institutions should avoid:

- Including an **excessive number** of explanatory **drivers**
- Using **unstructured data**
- Overly **complex modelling choices**
- **Unnecessary complexity** in the modelling approach if it is not justified by a significant improvement

Validation

The validation should pay particular attention to:

- **Preventing overfitting issues** (i.e. performance optimization of the development sample)
- **Challenging** the model design
- Ensuring representativeness and **data quality issues**
- **Analyzing the stability** of the estimates

6 Why Management Solutions

Areas of collaborations

Management Solutions has an expert working group that supports its clients in the development and implementation of their Machine Learning projects with focus on interpretability of outcomes and business integration - in each of the 6 defined lines of activity, and bringing expertise in each area

MS capabilities regarding Machine Learning models

1 Data & IT Infrastructure

- New data as input to traditional models.
- ML for data sourcing/ capturing
- ML and Deep learning for Recovery, Fraud Detection, etc.
- ML for data lake information (Personal Data, Data relationship, etc.)
- Best Practices for Big Data

2 Modelling

- New algorithms used as challenger models for traditional modelling techniques.
- Development of tools for comparison: traditional algorithms Vs. ML/ Deep Learning
- Discretization techniques (e.g. Optimal binning with gradient boosting, random forest)
- Feature Engineering and Selection for Machine Learning modelling.

3 Validation

- Adjustment and adaptation of Model validation procedures and techniques for new algorithms.

4 Interpretability

- Interpretability module development for Machine learning models via surrogate models.

5 Regulation & Benchmarking

- Interpretation of Supervisory opinion and benchmarking capabilities in Europe and the Americas.

6 Other

- Documentation and data quality auditing.
- Training on the use, importance and integration of ML techniques.,
- Integration in decision making processes and interpretability for model users.

Business

- *Retail*
- *Consumer*
- *Insurance*
- *Corporate and Markets*
- *Asset Management*

6 Why Management Solutions

Experience and Capabilities

Management Solutions has proven experience in supporting and developing Machine Learning Models within the Financial Industry

Management Solutions' differential knowledge and experience (illustrative)

1	▪ **Admission Models**: By analysing Social Networks and digital footprint data, the admission models can be improved in terms of predictive power..
2	▪ **Behavioural Models and early warnings**: Probabilistic Graph Models based on commercial interactions between clients and providers to determine Probability of Defaults and early warnings.
3	▪ **Segmentation and strategy for Recovery administration**: Trough a dynamic segmentation on client profiles using ML techniques, the recovery management can be improved by taking into account multiple effects at the same time.
4	▪ **Reputational risk assessment through Sentiment Analysis** using social media and news articles from several external digital sources in combination with Machine Learning techniques.
5	▪ **Fraud detection models and Anti-Money Laundering warnings**: Through knowledge discovery techniques, it is possible to find anomalous behaviours in clients or pool of clients.
6	▪ **Prospective model for CRM**: ML techniques are used to extract valuable information from new data sources to enrich Databases and Data Lakes used in CRM and marketing campaigns.
7	▪ **Interpretability**: Model unboxing to rationalize the behavior of models in specific ranges of application.
8	▪ **Network Geolocalizer (ATMs and Bank offices)**: Application of ML to geodemographic information, to optimize the distribution of ATMs and bank offices.
9	▪ **Data Lakes**: Improved traceability of personal data for GDPR compliance.
10	▪ **Extensive experience in information systems and system architecture** in the field of risk management and modelling, ample knowledge on Big Data, with certified professionals in Amazon Web Services and Microsoft Azure

A| Annex: a practical use case

A Annex: a practical use case

Credit Risk Models (1/3)

The performance of several machine learning (ML) models for credit default prediction is compared with the statistical performance of a simple and traditionally used model like Logistic Regression (Logit)

Description of the use case[1]

- **Use of ML models by the Spanish regulator (BoE) to study their advantages over traditional models** such as logistic regression.
- The use of advanced analytical models such as Lasso regression, CART, Random Forest, XGBoost and neural networks for credit default prediction (i.e. rating models and PD calibration) is proposed.
- For this purpose, an anonymized database of consumer loans (3.95% average DR), from a Spanish global systemic bank, with >75k transactions and 370 variables, is used.

Objectives of the study

- Determine the **feasibility and advantages** of using ML models.
- To study the **trade-off between the risk added by the use of ML models, and the improvement in the predictive power** of the models.
- Analyze whether the **improvement in predictive power is due to the greater use of data or intrinsic to the model**.
- Determine the **economic impact on capital** due to the use of ML models.

Tasks performed

- Development of **rating models for different sample sizes (S) and different number of variables available (N)** for training.
- Comparison between models and **measurement of discriminatory power** using AUC-ROC & Brier's Score and comparison of TPR/FPR under different thresholds.
- Comparison of whether the **improvement in AUC-ROC depends on S/N through simulations** with a random number of variables and sample size.
- Comparison between models and **measurement of the predictive power through the Brier's score**, measuring its sensitivity to variations in S and N.
- Comparison of the **economic impact on capital by the use of ML models**.

(1) *Alonso, A., & Carbó, J. M. (2021). Understanding the performance of machine learning models to predict credit default: a novel approach for supervisory evaluation*.

A Annex: a practical use case

Credit Risk Models (2/3)

The performance of several machine learning (ML) models for credit default prediction is compared with the statistical performance of a simple and traditionally used model like Logistic Regression (Logit)

Benefits obtained / Impact

Classification (Risk Differentiation)
-
Ability of the model to discriminate defaulted loans from those who have been repaid

- **Approach:** As shown in **Figure 1**, the AUC-ROC plots the true positive rate (TPR) vs the false positive rate (FPR) to study the discriminatory or classification power of the selected models.
- **ML models outperform traditional models:**
  - The further up from the red dotted area, the more classification power the model would have. Logit achieves the smallest area (0.78), while XGBoost the largest (0.84).
  - As shown in **Figure 2,** Random Forest and XGBoost outperform the rest of the models even when a small amount of data (5,000 observations) is used
  - On the other hand, Logit and Lasso can outperform the rest of the models when the sample is less than 5,000 loans (e.g., in LDPs).

Figure 1

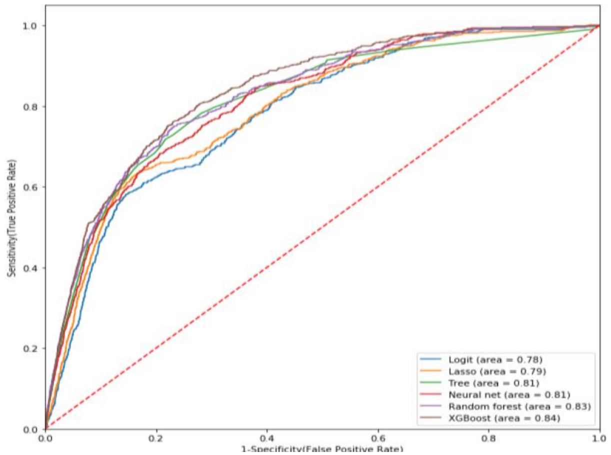
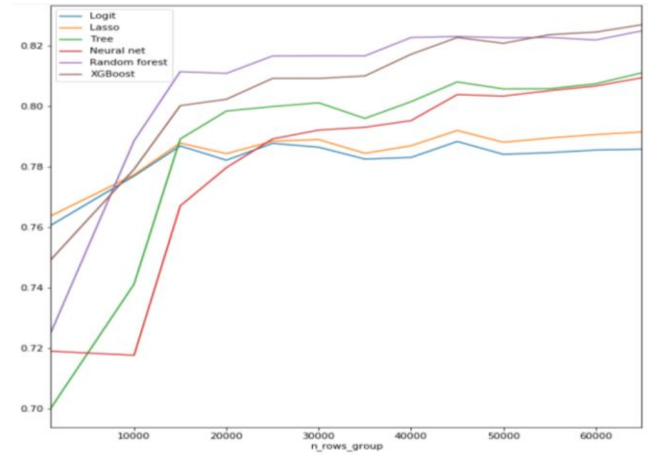


Figure 2



Calibration (Risk Quantification)
-
The quality of the estimation of the probability (how good the estimated probability fits the observed default rate)

- **Approach:** The Brier score is a key measure to quantify the accuracy of a probability forecast. The Brier score is used to study the calibration power of the six ML models.
- **ML models present the lowest error:**
  - **Figure 3** shows for each model the resulting box plots from different simulations with dissimilar sample sizes. The models with the lowest average Brier score are Random Forest and XGBoost.
  - **Figure 4** shows for each model the resulting box plots from different simulations with dissimilar number of features available. XGBoost and Random Forest also achieve the lowest Brier scores, reinforcing the existence of a model advantage from a calibration point of view.

Figure 3

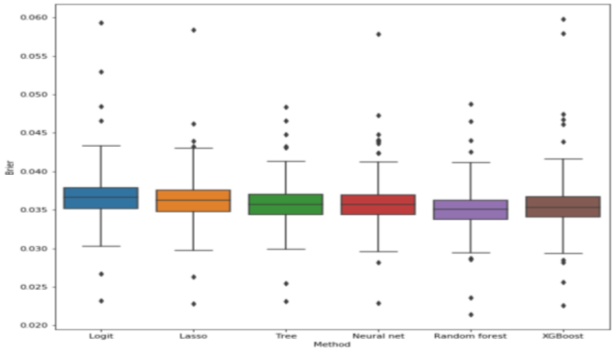
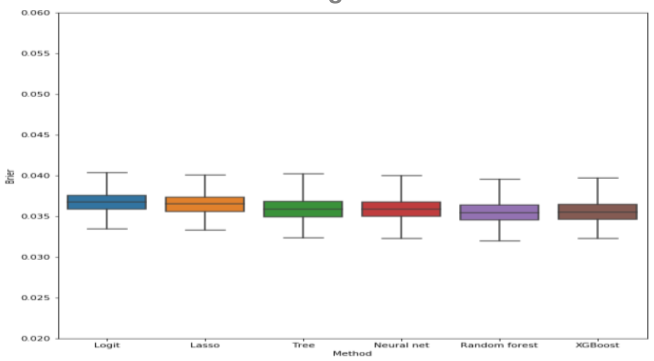


Figure 4



Annex: a practical use case

Credit Risk Models (3/3)

The performance of several machine learning (ML) models for credit default prediction is compared with the statistical performance of a simple and traditionally used model like Logistic Regression (Logit)

Benefits obtained / Impact

Economic impact

- Assuming that the institution follows an IRB approach, **the benefit of using ML models can be quantified in terms of regulatory capital** between using a commonly use model nowadays like Lasso compared to using **XGBoost**, which is **the model found to be the most efficient** in terms of predictive performance in the dataset.
- **Figure 5** shows the distribution of loans per final rating bucket. The distribution of loans per bucket differ between each model: **XGBoost has a more granular and smooth distribution over buckets whereas Lasso accumulates more loans in buckets 2 and 5**.
- In **Figure 6,** on the left hand side the average PD is plotted in each bucket for both XGBoost and Lasso, and on the right, the corresponding capital requirement. First of all, **it can be seen that the higher the PD of a bucket, the higher would be the regulatory capital.**
- Taking the average capital requirement for each bucket (Figure 6 right), and weighting it by the amount of loans that are in the bucket (see Figure 5), **capital requirements are 12.4% lower under the XGBoost rating scale than under the Lasso one**.
- Summarizing, the fact that **XGBoost is able to deliver a more granular distribution of loans and a smoother classification of loans per buckets of PD, ensures higher capital savings**.

Figure 5

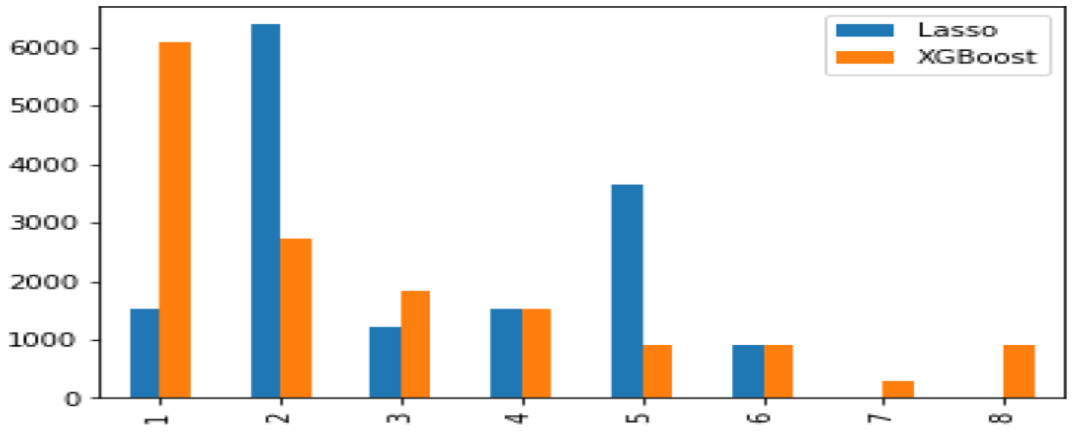


Figure 6