fintegral

11th August 2023

AI and Machine Learning for Credit Rating Models – Part IV

A brief overview of the EBA's follow-up report on machine learning for IRB models

fintegral

AI and Machine Learning for Credit Rating Models

Principle-based recommendations are welcomed by the industry

The EBA recently released its follow-up report[1] to its 2021 discussion paper[2] (DP) on the use of machine learning (ML) for internal ratings-based (IRB) models. The principle-based recommendations, largely unchanged from the 2021 DP, are welcomed by the industry providing much-needed clarity and a sound basis for the development of ML models.

Summary of EBA's principle-based recommendations

The following principle-based recommendations by the EBA are designed to help firms apply ML in their IRB model landscape:

- Ensure models are properly understood by their users including the model development unit, validation team and credit risk management functions to enable stakeholders to assess the relevance and appropriateness of the risk drivers, and economic rationale of the models.
- Avoid unnecessary complexity if not justified by a significant improvement in predictive power.
- Ensure models can be interpreted and documented clearly. This includes providing details on the rules and controls used in the preparation of data.
- Enhance the understanding of assumptions and behaviour of models on specific predictions, when using human judgement in model development and application respectively.
- Justify and monitor frequent updates of a model.
- Validate complex ML models with limited explainability or frequently updated models, which may require increased depth/frequency. This includes assessing model overfitting issues, model design, data representativeness and data quality issues, and stability of estimates.

The aim of the recommendations will ensure:

- The capital requirements are set in a prudent manner which continues to be harmonised across Europe.
- The development of sophisticated ML models can coexist and adhere to the Capital Requirements Regulation (CRR).
- A consistent and clearer understanding of regulations.

Potential benefits

There are several potential benefits from the use of ML as primary IRB models. These include:

- Improved data quality in terms of more efficient data preparation and mining.
- Superior risk quantification and model discriminatory power for example by detecting useful predictive explanatory variables in large datasets or make use of non-linear relationships.
- Robust model validation and monitoring techniques for example by developing ML based model challengers to serve as benchmark to the standard models.

Use case scenarios

The consultation findings show firms are using, or intending to use, ML for some steps of the IRB (and also ECL) approach including:

- Development of PD models.
- Validation of PD segmentation.
- Ranking exposures including modelling non-linear relationships.
- Identifying risk drivers.
- Resolving data quality issues.
- Estimating PD/LGD score ranges.
- Model validation including developing challenger models and benchmarking.
- Collateral valuation.

[1] EBA, 2023, "Machine Learning for IRB Models" – EBA/REP/2023/28
[2] EBA, 2021, "Discussion Paper on Machine Learning for IRB Models" – EBA/DP/2021/04

AI and Machine Learning for Credit Rating Models

Consideration of other legal frameworks when using ML

The EBA illustrates firms need to acknowledge more than the existing prudential requirements on IRB models (CRR/CRD)[1] but also other legal frameworks namely the GDPR[2], AI Act[3], and CCD[4] which are designed to ensure ethical standards are upheld and consumers are protected when using ML models.

CRR/CRD

- The general governance and guidance on areas including data quality, model development and calibration, model validation, and model documentation are covered within CRR/CRD. Therefore, it's essential the introduction of other legal standards like the proposed AI Act do not create legal uncertainty.
- Majority of the prudential requirements align with the proposed AI Act, with the remaining requirements of the Act largely relating to administrative and procedural obligations.

GDPR

- Firms must ensure data is fit for purpose, complete, and meets data quality standards.
- Stringent controls on data are required to ensure adherence to GDPR rules, with firms subject to additional supervisory reviews.
- The collection and storing of personal client information may be argued as valid for capital calculation purposes.

Proposed AI Act

- The use of AI for creditworthiness assessments (CWA) is considered high-risk, potentially discriminating and adversely affecting consumers' access to financial and essential resources.
- The EBA requests clarification on the scope of the AI Act - confirm if the requirements apply to the CWA at the point of loan origination only, and not extended to IRB models used for capital calculations.
- The EBA calls for a framework to ensure ML models used for CWAs are designed according to their intended purpose.
- Although, the AI Act requirements may indirectly apply to IRB models through the prudential use-test requirement.

Proposed CCD

- Referencing Art. 9 of the GDPR, the assessment of creditworthiness must be based on relevant and accurate data which excludes the use of specific personal data.
- Firms not allowed to use data collected from social networks.

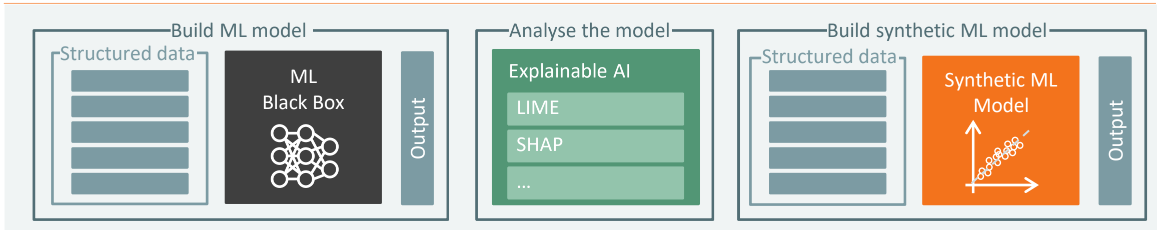
[1] Capital Requirements Regulation (CRR) / Capital Requirements Directive (CRD)
[2] General Data Protection Regulation
[3] The Artificial Intelligence (AI) Act (under proposal) was drafted to fast-track the adoption of AI by the EU in a safe manner whilst addressing the risks of using the technology and safeguarding users.
[4] The new proposed Consumer Credit Directive (CCD) to replace the existing 2008/48/EC directive was originally established to promote high levels of consumer protection and install consumer confidence within the credit market.

AI and Machine Learning for Credit Rating Models

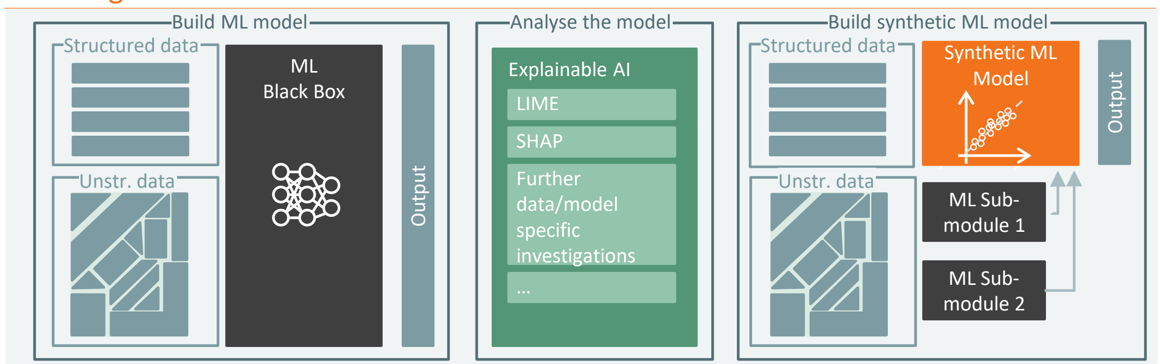
Fintegral's approach to developing compliant models: synthetic ML models

Fintegral holds a wealth of experience with "synthetic ML Models", understanding complex statistical models which are informed by ML models. Based on our experience, the advantages of enhanced explainability, improved stakeholder acceptance, and streamlined model validation outweigh the relatively higher time investment required for constructing synthetic ML models.

Structured data only



Including unstructured data



High level explanation

In the construction of synthetic ML models, we follow these outlined steps:

1) Building of one or more ML models. This step includes data cleansing, building and comparing multiple models.
2) The best performing model(s) are analysed using methods of explainable AI. The goal is to understand the main features and non-linearities.
3) A statistical model is built based on the knowledge gained in the analysis of the ML model(s). These models may include decision trees, and particularly when involving unstructured data, incorporate submodules which leverage Natural Language Processing (NLP) or other machine learning techniques to generate features integrated within the statistical model.

In practice, the three outlined steps are interconnected rather than being strictly independent, potentially giving rise to valuable feedback loops. This is particularly evident between the analysis of the model and the construction of the ML model, as well as between the creation of the synthetic ML model and the subsequent analysis of the ML model.

Comparison of the different approaches

	Statistical Model	ML Model	Synthetic ML Model
Expected workload for model development	🟠🟠🟠	🟠⚪⚪	🟠🟠⚪
Expected workload for model validation	🟠⚪⚪	🟠🟠🟠	🟠⚪⚪
Model performance (In-Time)	🟠⚪⚪	🟠🟠🟠	🟠⚪⚪
Danger of Overfitting	🟠🟠🟠	🟠⚪⚪	🟠⚪⚪
Model explainability	🟠🟠🟠	🟠⚪⚪	🟠🟠🟠

Key:
High 🟠🟠🟠
Low ⚪⚪⚪

Key points of synthetic ML models

- Easier to comply with laws and regulations.
- Easier to explain and validate than traditional ML Models.
- The additional time and resources required by the development of synthetic ML models is typically offset by *"simpler"* model validation.
- Synthetic ML models typically suffer less from overfitting, which outweighs the lower in-time-sample performance.

Contact

Fintegral

London ¦ Frankfurt ¦ Zurich

www.fintegral.com

Dr. Andreas Peter
Managing Partner
Fintegral Group

+49 160 583 40 66
andreas.peter@fintegral.com

Fintegral Deutschland AG
Steinweg 5
60313 Frankfurt am Main
Germany

Tobias Kesselring
Senior Manager
Fintegral Schweiz AG

+41 79 271 19 00
tobias.kesselring@fintegral.com

Fintegral Schweiz AG
Claridenstrasse 35
8002 Zürich

Abdul Qaiyum
Senior Consultant
Fintegral UK Ltd.

+44 7496 363 298
abdul.qaiyum@fintegral.com

Fintegral UK Ltd.
City Tower, 40 Basinghall St.
London EC2V 5DE
United Kingdom