Financial Services Advisory - Machine learning in IRB

How Grant Thornton can help you understand the challenges surrounding the implementation and validation of machine learning techniques in IRB models

June 2022

Introduction

This paper will provide an overview of the current and future uses of Machine Learning in the world of Internal Rating Based (IRB) Models which allow banks to model their own inputs for calculating Risk Weighted Assets.

What are Internal Rating Based (IRB) Models?

Risk Weighted Assets (and as a result capital requirements) for credit risk can be calculated either using the IRB or Standardised Approach. IRB models represent a more advanced and risk sensitive approach, breaking down into three component parts (PD, LGD, EAD):

1. **Probability of Default (PD):** What is the chance that the borrower will default on a 1 year time horizon?
2. **Loss Given Default (LGD):** How much will I get back (and therefore how much will I lose) in the event of default?
3. **Exposure at Default (EAD):** If the borrower defaults, what will be the size of my exposure?

A Bank's IRB approach can be classified either as Foundation or Advanced:

(1) **Foundation IRB (FIRB) Approach:** Banks estimate PD and use regulatory prescribed estimates of LGD and EAD.
(2) **Advanced IRB (AIRB) Approach:** Banks use their own estimates of PD, LGD, and EAD.

Machine Learning Techniques

Machine learning is a subset of artificial intelligence which focuses on the use of data and algorithms to automate the process of analytical model building which allows machines to learn from data, identify patterns and make predictions with minimal human intervention. The ability of machine learning algorithms to identify patterns and make predictions improves both with the use of the algorithms and with the quality of the data provided to them. The following sections give an overview of the different types of machine learning techniques. Certain methods can be used in both supervised and unsupervised learning e.g. stacking.

Supervised Learning

The algorithm learns rules for building the model from a target variable within the dataset and uses these rules to predict outcome variable values on new input data

Classification
- Neural Networks/Deep Learning
- K-Nearest Neighbours
- Decision Trees
- Support Vector Machines

Regression
- Linear Regression
- Logistic Regression

Ensemble Methods
- Bagging/Random Forests

Unsupervised Learning

The algorithm learns from a training dataset which has no target variable. The goal is to understand the distribution of the data in terms of interpretable patterns, associations and descriptive properties

Clustering
- K-Means Clustering
- Fuzzy C-Means

Pattern Search
- Apriori

Dimension Reduction
- Principal Component Analysis

Reinforcement Learning

The algorithm learns from interacting with the environment rather than from a training dataset and does not require a target variable. The algorithm learns to perform a specific task by trial and error.

- Q-Learning
- State-Action-Reward-Sate-Action (SARSA)
- Deep Q-Networks

Background

The use of machine learning techniques is an emerging trend in IRB models

In November 2021, the European Banking Authority (EBA) published a discussion paper on machine learning used in the context of IRB models to calculate regulatory capital for credit risk. The aim of the discussion paper is to set supervisory expectations on how new sophisticated machine learning models can co-exist with and adhere to the Capital Requirements Regulation (CRR) when used in the context of IRB models. The emergence of Big Data and the increase in computational power in recent years has supported the emerging trend of implementing machine learning models in the context of credit risk.

The discussion paper seeks stakeholders' feedback on many practical aspects related to the use of machine learning in the context of IRB with the aim of providing clarity on supervisory expectations regarding their use. The discussion paper is a first step of engagement between the banking industry and supervisory community to explore both the current and potential future uses of machine learning techniques in IRB models. In this publication, we define machine learning and discuss four main areas of discussion in the EBA's discussion paper: its current use in IRB modelling, challenges, potential benefits and future prudent use of machine learning models.

Our Machine Learning Approach and Services

Grant Thornton has significant experience with building and validating IRB, IFRS9 as well as Stress Testing models. In our approach we focus on the key benefits and challenges regarding the use of machine learning techniques in regulatory as well as non-regulatory modelling.

How can Grant Thornton help to design Machine Learning strategy and build/validate models

Strategy Design

Determine best use of Machine Learning

- Use of Machine Learning techniques to achieve competitive advantage for bank strategic portfolios
- Cost-benefit function maximisation

Model Build

Implement Machine Learning in Current or New Models

- Data gathering and processing
- Use of appropriate Machine Learning techniques
- Model design and implementation

Model Validation

Assess machine learning models and recommend improvements

- Validation of existing Machine Learning techniques
- Build of challenger Machine Learning models
- Added value assessment of existing Machine Learning models

In this publication, we define machine learning and discuss four main areas of discussion in the EBA's discussion paper: its current use in IRB modelling, challenges, potential benefits and future prudent use of machine learning models.

What is the current use of Machine Learning in IRB Models?

Current Use

What are the key challenges associated with the use of Machine Learning modelling techniques?

Key Challenges

What are the benefits of Machine Learning techniques use and how they can contribute to better risk capture?

Benefits

How could future model builds incorporate Machine Learning techniques in a prudent way?

Future Prudent Use

Machine Learning and Regulation

Regulation directly addressing Machine Learning

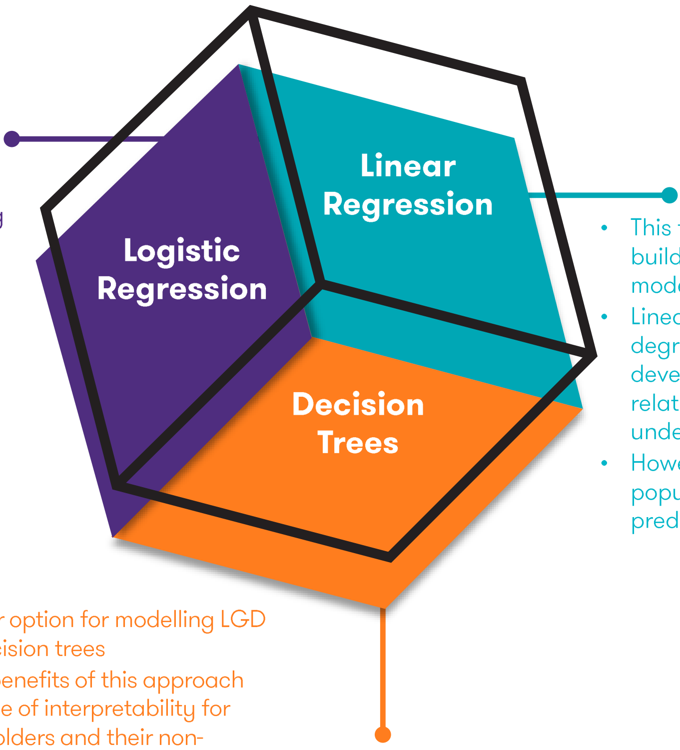
EBA Discussion Paper on Machine Learning for IRB Models - EBA/DP/2021/04
The aim of this discussion paper is to understand the challenges and opportunities coming from the world of machine learning should they be applied in the context of IRB models to calculate regulatory capital for credit risk. The discussion paper covers areas regarding machine learning definitions, current use, challenges & potential benefits of machine learning models as well as prudent future use of machine learning models going forward.

Key Capital Requirements Regulation (CRR) relevant in Machine Learning context

Article 171 – Assignments to grades or pools
There must be consistent assignment of similarly risky borrowers to the same grade or pool. Adding complex modelling techniques may make it difficult to precisely detail the different grade definitions. This can make verifying the correct implementation of the internal ratings challenging. *(a) the grade or pool definitions and criteria shall be sufficiently detailed to allow those charged with assigning ratings to consistently assign obligors or facilities posing similar risk to the same grade or pool. This consistency shall exist across lines of business, departments and geographic locations;*
Article 174 – Use of Models
There must be a clear understanding of how the model functions by both the model developers and model users. Machine Learning techniques may add an increased level of complexity to the model comprehension. *(e) the institution shall complement the statistical model by human judgement and human oversight to review model based assignments and to ensure that the models are used appropriately. Review procedures shall aim at finding and limiting errors associated with model weaknesses. Human judgements shall take into account all relevant information not considered by the model. The institution shall document how human judgement and model results are to be combined.*
Article 175 – Documentation of Rating Systems
An increased level of complexity will require more extensive model documentation. *4. Where the institution employs statistical models in the rating process, the institution shall document their methodologies. This material shall:* *(a) provide a detailed outline of the theory, assumptions and mathematical and empirical basis of the assignment of estimates to grades, individual obligors, exposures, or pools, and the data source(s) used to estimate the model;* *(b) establish a rigorous statistical process including out-of-time and out-of-sample performance tests for validating the model;* *(c) indicate any circumstances under which the model does not work effectively.*
Article 179 – Overall Requirements for Estimation
There must be an obvious relationship between risk drivers and default. Machine Learning techniques may make it difficult to establish this link. *(a) an institution's own estimates of the risk parameters PD, LGD, conversion factor and EL shall incorporate all relevant data, information and methods. The estimates shall be derived using both historical experience and empirical evidence, and not based purely on judgmental considerations. The estimates shall be plausible and intuitive and shall be based on the material drivers of the respective risk parameters.*
Article 185 – Validation of Internal Estimates
Machine Learning models can be used in validation to establish a ceiling for IRB model performance. *(a) institutions shall have robust systems in place to validate the accuracy and consistency of rating systems, processes, and the estimation of all relevant risk parameters. The internal validation process shall enable the institution to assess the performance of internal rating and risk estimation systems consistently and meaningfully.*

Current Use in IRB modelling

Up to now the use of advanced machine learning techniques in IRB modelling has been limited due to regulatory requirements and in particular, the difficulty in interpreting and explaining the functionality of more complex models. The standard suite of modelling techniques in IRB today mainly consists of logistic regressions, linear regressions and decisions trees based on our industry experience. These more basic machine learning models are used currently in the following ways in IRB modelling.



- Logistic regression is typically used in the context of building probability of default (PD) models within credit risk
- This machine learning technique is commonly used for binary classification problems such as predicting whether default will occur

- This technique can be used when building loss given default (LGD) models for credit risk
- Linear regression allows for a high degree of automation in the development process and are relatively quick to build and easy to understand
- However, these are becoming less popular due to their often-low predictability

- Another option for modelling LGD are decision trees
- Some benefits of this approach are ease of interpretability for stakeholders and their non-sensitivity to distributions or statistical instabilities
- However, it requires a lot of manual work compared to other more advanced machine learning techniques

While these techniques are considered supervised machine learning algorithms, these models tend to be more simplistic making it difficult to obtain complex relationships. More powerful and compact algorithms such as Neural Networks can easily outperform these algorithms.The most common use of advanced machine learning techniques within credit risk can be seen outside of IRB in the areas of decisions/pricing, credit monitoring and collections, restructuring and recovery.

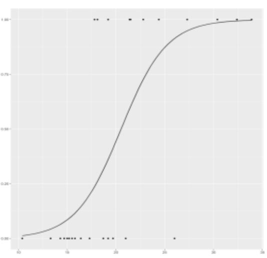
Simple vs Complex Models

Currently there are a few simpler and more traditional machine learning techniques being used within IRB, namely, linear and logistic regression and decision trees. While technically these techniques are machine learning algorithms, these models would lie on the simpler and more intuitive side of the machine learning spectrum. Below we discuss the more complex machine learning models.

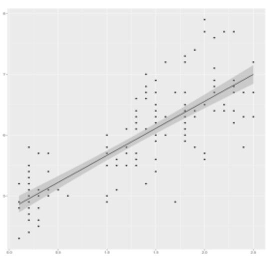
Generally, when machine learning models are discussed (e.g., in the EBA discussion paper), references are made to more complex models such as neural networks, deep learning, reinforcement learning and ensemble methods. These machine learning models can significantly improve model performance but are generally less transparent, more difficult to explain and require more computational power.

The complexity of these models leads to pivotal challenges in interpreting model results. IRB model regulators will be concerned about the use of "black box models" where the process used to generate the output can be confusing. Despite the above challenges, institutions are interested in using these complex machine learning methods due to the benefits they possess. In the later sections, we go into more detail about the current challenges as well as the potential benefits/future prudent uses of these advanced machine learning techniques.

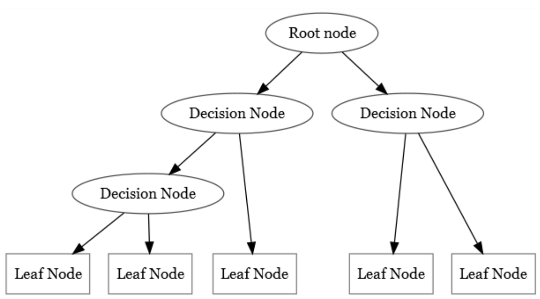
Traditional Approach



Logistic Regression

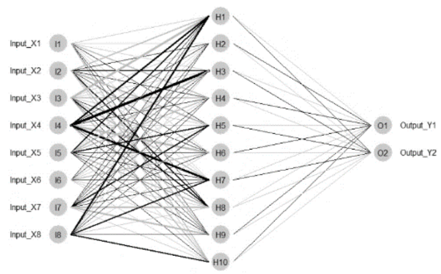


Linear Regression

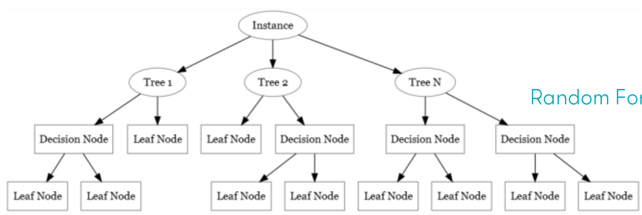


Decision Trees

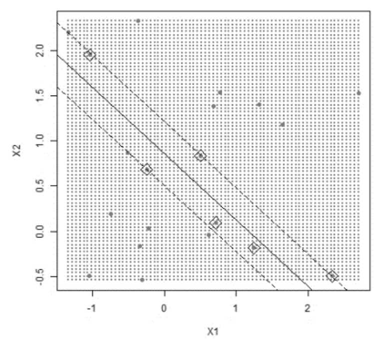
Modern Approach



Neural Networks



Random Forests



Support Vector Machines

Challenges & Potential Benefits with Using Machine Learning Techniques in Recent Periods

Challenges

Listed below are some challenges related to the implementation of machine learning in IRB models.

Complexity/ Interpretability

- The primary challenges posed using advanced machine learning models is their complexity and interpretability in the highly regulated space of regulatory capital modelling.
- Enhancements to the current risk management approaches are required to ensure that more advanced techniques can be correctly understood and implemented.
- It is noted that major improvements have been made in recent times in terms of interpretability and documentation of these complex machine learning techniques.

Computing Power

- Major improvements can be seen in computing power and the use of Big Data with regards to machine learning in recent years.
- The computer power requirements needed for some of the more advanced machine learning techniques still remains a major challenge e.g. deep learning.

Potential Benefits

Listed below are some potential benefits of implementing machine learning techniques for the purposes of IRB modelling.

Enhanced risk differentiation

- Improvements in the model's discriminatory power and a greater ability to identify the optimal risk drivers and portfolio segmentation.

Improved risk quantification

- Detection of material biases and identification of recovery patterns in LGD models.

Compatibility with large, unstructured datasets

- Enhancements in data quality assessments and representativeness of the model development sample. Possibility to use data from additional sources.

Greater efficiency in remediation exercises

- Machine Learning Models can be used to detect cohorts of interest for business sampling. For example time series of loan repayment can be analysed to detect potential modifications that were not historically recorded.

Future Prudent use of Machine Learning Models

Given the challenges laid out above it is unlikely that we will see a full-scale machine learning IRB model soon but there is scope to incorporate aspects of machine learning into the model development process.

Variable Pre-Processing

Machine learning can be used to detect the best explanatory variables or combination of variables from a large dataset

Segmentation

Unsupervised machine learning techniques can be used to classify segments of a portfolio with similar risk characteristics which can form potential segmentation variables

Risk Differentiation

Machine Learning models can be used as a module for the purposes of risk differentiation for probability of default

Model Validation

Machine learning models can be used in model validation as a challenger model to establish an upper bound for model performance

It is essential that there is sufficient monitoring, validation and explain-ability of the methodologies and the model outcomes from any machine learning techniques used. All relevant stakeholders, including both model developers and senior management will require an appropriate understanding of the techniques used and how the models function.

Contact

Grant Thornton's Financial Services Risk, Consulting and Advisory teams are comprised of dedicated experts who are experienced in supporting banks and investment firms with a variety of regulatory challenges, including those arising from the development of IRB Models. Our highly qualified Quantitative Risk team provides support to financial institutions across the full spectrum of risk measurement and modelling strategies, including the development, deployment and validation of key models and risk measurement methodologies in regulatory capital, stress testing and IRB, IFRS9 and bank risk modelling. They have experience implementing machine learning techniques in the context of credit risk modelling as well as a keen interest in emerging trends within the machine learning space.

Partner group

Dwayne Price
Partner
Financial Services Advisory
T +353 (0)1 436 6494
E dwayne.price@ie.gt.com

Brian O'Dwyer
Partner
Financial Services Advisory
T +353 (0)1 433 2538
E brian.odwyer@ie.gt.com

Frankie Cronin
Partner
Financial Services Advisory
T +353 (0)1 646 9044
E frankie.cronin@ie.gt.com

Nuala Crimmins
Partner
Financial Services Advisory
T +353 (0)1 483 8577
E nuala.crimmins@ie.gt.com

Amanda Ward
Partner
Financial Services Advisory
T +353 (0)1 433 2440
E amanda.ward@ie.gt.com

Machine Learning & IRB group

Lukas Majer
Director
Quantitative Risk
T +353 (0)1 646 9006
E lukas.majer@ie.gt.com

Mark Perry
Director
Quantitative Risk
T +353 (0)1 408 6909
E mark.perry@ie.gt.com

Andreas Spyrides
Director
Quantitative Risk
T +357 (0)2 260 0270
E andreas.spyrides@cy.gt.com

Stephen Speight
Associate Director
Quantitative Risk
T +353 (0)1 408 6909
E stephen.speight@ie.gt.com

Killian Buckley
Manager
Quantitative Risk
T +353 (0)1 408 6909
E killian.buckley@ie.gt.com

Phanis Ioannou
Manager
Quantitative Risk
T +357 22 600 296
E phanis.ioannou@cy.gt.com

Rachel Bernard
Assistant Manager
Financial Services Advisory
T +353 (0) 8 948 35851
E Rachel.Bernard@ie.gt.com

Cian Mellett
Assistant Manager
Quantitative Risk
T +353 (0)1 680 5805
E cian.mellett@ie.gt.com

Joseph Glynn
Consultant
Quantitative Risk
T +353 (0)1 408 6988
E joseph.glynn@ie.gt.com

Conor Ryan
Consultant
Quantitative Risk
T +353 (0)1 680 5805
E conor.ryan@ie.gt.com

Christina Savva
Consultant
Quantitative Risk
T +357 22600271
E Christina.savva@cy.gt.com

Grant Thornton

grantthornton.ie