OVERVIEW

No. RADAR
4956

Responsible level
Germany

Competent authority
BaFin – Federal Financial Supervisory Authority

Standard designation
Machine learning in risk models - characteristics and supervisory priorities

Title of Standard
Machine learning in risk models - characteristics and supervisory priorities

Abbreviation (standard)

Short Title

Abbreviation (standard)

Abbreviation

Implementation status of the standard
published

Industry relevance
Banking, insurance

category
01. Banking and banking supervisory law

Document type
Miscellaneous

Management Summary
This joint paper by BaFin and the Deutsche Bundesbank addresses the application of artificial intelligence and machine learning (ML methods) in risk management in financial institutions. The discussion builds on previous national and international publications and links them to the prudential risks of ML and existing supervisory practice, with a focus on solvency supervision. The aim of the document is to identify the characteristics of ML methods that may be of supervisory relevance and for which supervisory practice may need to be further developed.

Management Summary

This present joint document of BaFin and Deutsche Bundesbank focuses on the use of artificial intelligence and machine learning (ML methods) in the risk management of financial firms. The paper builds on existing national and international publications and focuses on solvency supervision connect the prudential risks of ML with current supervisory practices. It intends to identify characteristic traits of ML methods, which may have relevant supervisory implications for supervisors and to develop supervisory practices when necessary.

CONTENTS

Main content

B. Essential content

I. Characteristics of ML (Chapter II)

This paper identifies characteristics of artificial intelligence and machine learning (ML methods) for prudential purposes; a general definition of ML methods is not provided.

1. Dimensions and examples (Section 1)

The characteristics of ML are grouped into the three dimensions of the AI/ML scenario: "Methodology and Data Basis," "Use of Output," and "Outsourcing and IT Infrastructure," and are clearly presented in Table 1. Using two examples from the area of IRBA (Internal Ratings-Based Approach) rating methods (classic use of logistic regression vs. ML-supported use of neural networks), the different ranges of the model characteristics of these two ML methods are visualized. The individual model characteristics are discussed below.

2. Methodology (Section 2)

a) Regarding the complexity and dimension of the hypothesis space, it is explained that the hypothesis represents the causal relationship between input (e.g., market and portfolio data) and output (e.g., prices or risk measures). The hypothesis space contains all hypotheses specified by the modeler that are relevant to the problem description. The self-optimizing algorithm used to describe the problem learns and describes the given problem structure based on available data.

The possible hypotheses are determined by the choice of ML method (e.g., neural network, random forest, k-nearest neighbors) and its specification (model design, hyperparameters). Multilayer neural networks (deep neural networks, DNNs) allow hypotheses to be linked at multiple levels, resulting in a significantly higher dimension of the hypothesis space and a loss of traceability of the description of the relationship between input and output (black box property). This limited traceability has consequences for model development, validation, inferences about the data basis, the explainability of the model results, and thus for the internal and external justification of the chosen ML method.

b) The complexity of training refers to the learning process for establishing a concrete hypothesis, also called calibration or training. The number of specified calculation operations and the sequence of nested calculation rules may affect the availability of hardware resources or the numerical stability of the calculation method. The more complex the

The more sophisticated the training procedure is, the more it becomes the focus of supervisory attention.

c) The adaptability of some ML methods carries the risk of no longer being able to distinguish between model development and operation, as well as model maintenance and modification, due to the continuous inclusion of new data. This may be relevant for the supervisory approval of model changes under Pillar 1, the validation and reproducibility of model results, and the assurance of data quality.

3. Data basis (Section 3)

By using a large number of data sources, linking them, and using synthetic, even unstructured, data, ML methods are based on a multitude of input parameters and thus on a large amount of data. The amount of data used in training, its relevance, and quality determine the performance of ML methods and are therefore subject to supervisory oversight.

4. Use of the output (Section 4)

a) Depending on the area of application, the ML method plays a different role in the model. Applications can be, for example, supporting (e.g., data preparation), as a subcomponent (e.g., in the rating process), as a central component of the model, or outside the model (e.g., as a validation procedure or as a proxy for the "real" model). The increasing importance of the ML method leads to increasing intensity of supervisory oversight.
b) The results incorporated into the business process and their relevance are described by the scope of application (e.g., early risk detection, rating procedures). Supervisory interest is based on the impact of the scope of application of the ML method on the risk situation.

c) The degree of automation is divided into algorithm-determined processes, with extensive automation of the exploitation of results, but with a higher risk due to inadequate monitoring of the ML method, and algorithm-based processes, which rely more heavily on human control. Depending on the degree of automation, the resulting operational risks must be assessed differently.

5. Outsourcing and IT infrastructure (Section 5)

Possible outsourcing of ML methods to specialized service providers or the use of a specific IT infrastructure are possible and subject to the relevant regulations such as those of the *Circular 10/2017 (BA) - Banking Supervisory Requirements for IT (BAIT, Dataset 3275)*. Integration problems of outsourced models or purchased software into existing systems (legacy IT) are possible.

II. Supervisory approach (Chapter III)

The use of ML methods does not require a fundamentally new supervisory practice, which is essentially based on the *Regulation (EU) No. 575/2013 (CRR, Dataset 558, Dataset 559 and Dataset 560)*, the *Delegated Regulation (EU) No 529/2014 (dataset 918)* and the requirements of the *Circular 09/2017 (BA) (MARisk, dataset 1925)*. However, adjustments are necessary in certain places, following the principle of proportionality.

1. Supervisory practice is maintained (Section 1)

The supervisory practice for ML methods will be derived from the existing Pillar 1 framework

(Review and approval of internal models) and Pillar 2 (principle-based requirements for risk management and IT). However, more specific approaches may arise from the review of mathematical/methodological aspects and the procedural embedding of ML methods for controlled and thus successful and efficient use, as well as with regard to new and/or more pronounced risks in the data basis, validation, model changes, and control.

2. Methods invite data belief (Section 2)

Institutes should make additional efforts to ensure the quality of the data basis in model development, validation, and application. Training data must be free of systematic biases with regard to the causal relationships to be learned by the model. When using a large data basis, the learning of correlations between input data that do not represent a real relationship but are based on random properties of the training data set (model overfitting) must be excluded.

3. Explainability comes into focus (Section 3)

The complexity and dimensionality of the hypothesis space influences the traceability of the model and complicates the verification of model results. The use of a black box property may be justified by a potentially higher predictive quality, but depending on the importance of the model in the banking process, it entails a higher model risk. Losses in traceability must be weighed against the achieved advantages and justified; supervisory acceptance depends on the treatment of the model in risk management. In addition to the explainability of the results, their plausibility or the use of XAI (Explainable Artificial Intelligence) techniques are increasingly becoming important. However, these themselves represent models with assumptions and weaknesses and are in the testing phase.

Institutions carry out appropriate validation activities according to the model; in addition to the commonly used out-of-sample validation and/or backtesting, the use of XAI techniques, synthetic or stress/extreme scenarios, and tests against traditional methods may be considered.

4. Adaptivity: Model changes are harder to detect (Section 4)

Distinguishing between regular model maintenance and model changes requiring approval under Pillar 1 is difficult due to the flexibility and high-frequency adaptability of ML processes, but is essential for supervisory purposes. The need for frequent model adjustments must therefore always be well-founded.
The source uses examples to illustrate the classification of the term "model change." This is based on (see source for details):

- procedural criteria (including first use of an LM method for a task);
- Training gaps (comparison of a training session before and after an update of the data basis (re-training));
- Extensions of the risk factors used.

With regard to model adjustments, the supervisory focus is on (see source for details):

- the model approval on the possible techniques of model maintenance (model change policy);

- communication on regular model maintenance activities to detect significant model changes through "creeping" adjustments;

- internal validation on the frequency and appropriateness of the methods used.

For many ML methods that do not fall within the regulatory scope of Pillar 2, which requires approval, the existing requirements already apply (e.g., MARisk). Adapting the training cycle to the use case, with appropriate justification, is expected to ensure a balance between data timeliness and its explainability and verifiability.

CATEGORIZATION

Keywords

Adaptivity, algorithm, supervisory practice, outsourcing, degree of automation, backtesting, black box property, data preparation, data basis, data quality, DNN, deep neural network, explainability, explainable artificial intelligence, hypothesis, hypothesis space, input, input parameters, IT infrastructure, knearest neighbors, legacy IT, learning methods, logistic regression, MARisk, machine learning, methodology, ML method, model change policy, model change, model adaptation, model operation, model design, model development, model result, model maintenance, model overfitting, traceability, neural network, operational risk, output, random forest, reproducibility, risk situation, systematic bias, training, training cycle, validation, XAI technology

Legal and information bases

- Circular 10/2017 (BA) (BAIT) (Data set 3275)
- Circular 09/2017 (BA) (MARisk) (Dataset 1925)
- Delegated Regulation (EU) No 529/2014 (Dataset 918)
- Regulation (EU) No. 575/2013 (CRR) (Dataset 558, Dataset 559 and Dataset 560)

Related Standards

Target group – credit institutions

Yes

Target group – financial services institutions

Yes

Target group – Other companies in the financial sector

Yes

Target group – payment institutions

Yes

Target group – insurance companies

Yes

Target group – supplement

Comments

Statement by (date)

Implementation status Explanation

Responses to the consultation paper

Status – Further Details

Responses to the consultation paper

Date of entry into force/publication

February 18, 2022

Entry into force estimated?

No

Date of first application

February 18, 2022

Application appreciated?

No

Date Standard repealed

Remark (Entry into force/Publication)

Comments

Sources

The sources are not shown in this working paper.