

Risk-based Industry Baseline

Models in alert and event generation

Context

Introduction

The use of models enables banks to analyse their client data more effectively and efficiently to detect money laundering and terrorist financing (hereafter: ML/TF) risks. It allows banks to work in a risk-based manner and allocate resources to the most prominent ML/TF risks and priorities. Models can improve the detection of ML/TF risks and unusual transactions, support the automation of processes and reduce false positive alerts and events, strengthening the fight against ML/TF and resulting in fewer unnecessary actions towards clients. Prudent application of models in the Ongoing Due Diligence (hereafter: ODD) framework ^[1] requires proper management and understanding of risks related to the use of models as well as safeguards against these risks.

This NVB Industry Baseline aims to provide an overview of how models can be employed in a risk-based approach to detect ML/TF risks. Its goals are:

- to create a common understanding of the way banks can use models to detect ML/TF and comply with laws and regulations;
- to provide guidance for banks in their transition towards using more advanced models in their risk detection processes. Specifically for smaller banks in the Netherlands, this provides guidance they need to prepare for further innovation and use of more sophisticated models;
- conditions to consider to ensure reliable and responsible application of models.

Models related to alert and event generation processes applied by banks is the focus, as detailed below in section 1.2.

This NVB Industry Baseline outlines the following important principles.

- Banks are allowed to determine where and how models are used in their ODD processes in a risk-based manner.
- Banks may use their models in the execution of ODD processes provided that they demonstrate that the models sufficiently cover ML/TF risks.
- Reliance on these models is supported by model risk management to identify and mitigate relevant technical, compliance, and ethical risks.
- By definition, all models have some degree of uncertainty and inaccuracy. Therefore, banks accept that models used in their ODD processes may occasionally fail to detect certain cases of

ML/TF in accordance with their risk appetite.

- It is essential to maintain and update models through a learning loop to cover new and changing ML/TF risks.
- The conditions included in this NVB Industry Baseline are meant as considerations for banks to take into account when using (advanced) models. They are not minimal requirements that must be met before banks can start using advanced models in their ODD framework. It is within the banks' discretion and risk appetite to determine whether the conditions are sufficient to (partly) replace other ML/TF risk detection mechanisms.

¹ The ODD framework consists of fundamental components required for an effective Ongoing Due Diligence. It contains a set of ongoing screening and monitoring processes and controls to be performed by banks after a client has been onboarded. However, the considerations and conditions listed in this document apply for models used in client onboarding as well.

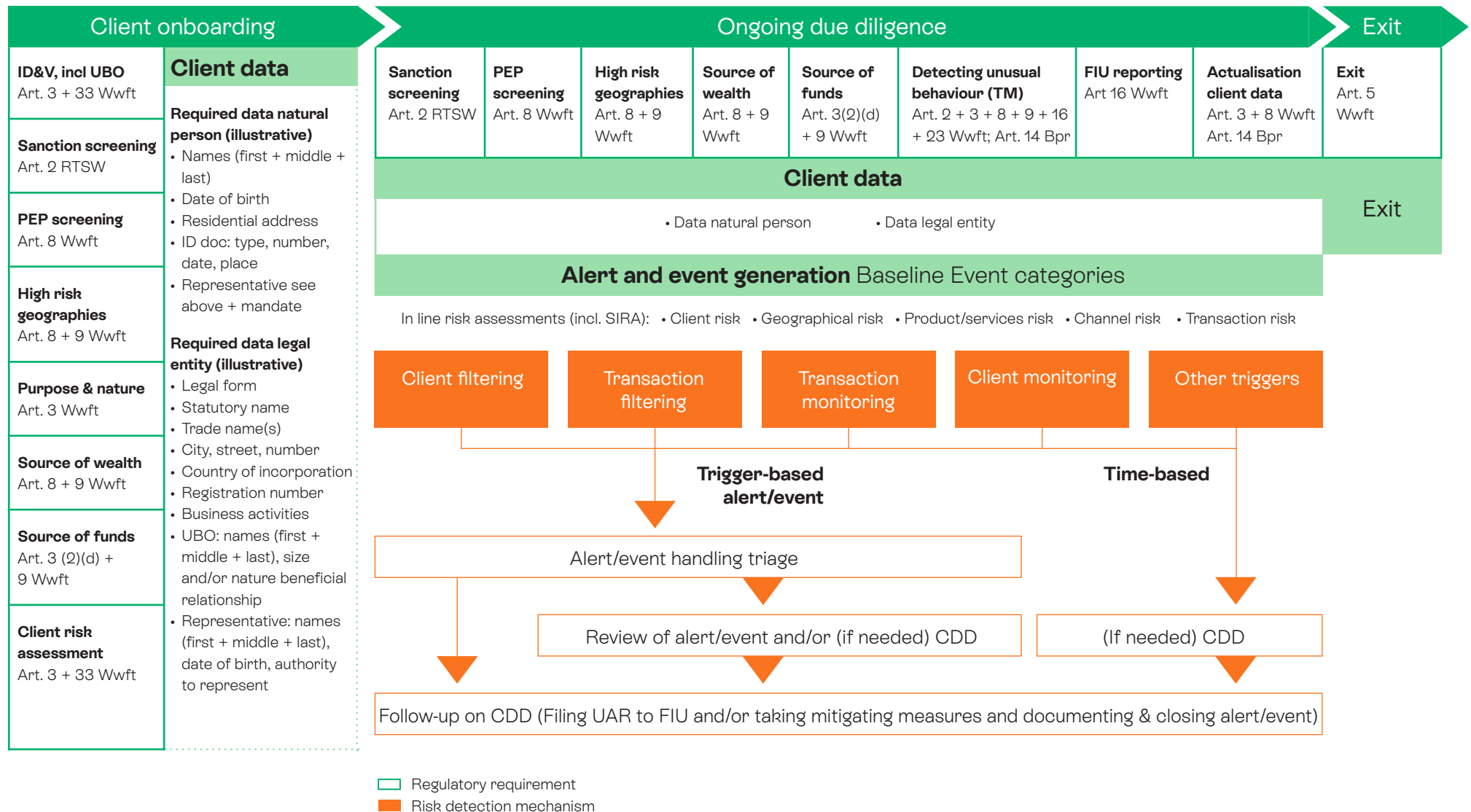
Positioning within the Financial Crime Framework

Models can be applied by banks to generate alerts and events in any risk detection mechanism of their Financial Crime Framework. This means models can be used in ODD processes such as client filtering, transaction filtering, transaction monitoring and client monitoring. The models use client and transaction data as input and generate outcomes that form the basis for subsequent alert and event handling.

The diagram below represents the Financial Crime Framework; alert and event generating models that are in scope for this Baseline are positioned in the risk detection mechanisms outlined in purple. Triage models (also 'noise reduction' or 'prioritisation' models) can be said not to generate alerts or events as such. They are nevertheless in scope of this Baseline because they provide an intermediate step towards incorporating more advanced models for alert and event generation.

Financial Crime Framework

Risk-based



Risk-based Industry Baseline

1 Industry Baseline

This NVB Industry Baseline describes possible applications of models in the alert and event generation processes of banks and under what conditions banks can make use of models. It provides details regarding:

- definition of a model;
- use of models in alert and event generation;
- guidance to transition towards using advanced models in alert and event generation;
- conditions when using models in alert and event generation;
- criteria to demonstrate effective implementation.

1.1 Definition of a model

This NVB Industry Baseline adopts the widely established model definition as stated in the Federal Reserve's guidance on model risk management SR 11-7 [2]: "The term model refers to a quantitative method, system, or approach that applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to process input data into quantitative estimates". This Baseline covers models that are used to assess ML/TF risk levels, (deviations from) expected client behaviour or otherwise find indications of possible ML/TF risks. The SR 11-7 definition does not specify by which method the model should be developed.

Consequently, it encompasses models based on advanced technologies such as machine learning as well as those based on more traditional methods using business rules. Thus, the conditions and requirements outlined below apply to models based on business rules just as they do to models based on advanced techniques.

Examples of techniques that can be used to create models in an ODD framework are (list not exhaustive):

- business rules;
- machine learning (e.g. anomaly detection, supervised learning methods);
- statistical analyses (e.g. linear or logistic regression);
- network analysis (e.g. community detection, path analysis);
- process analyses (e.g. process discovery algorithms);

1.2 Use of models in alert and event generation

Models can play various roles in alert and event generation. Models can, for example, be used to estimate the probability of possible ML/TF risks based on client, transaction and other data.

Models can also calculate relevant deviations in client behaviour, for example, compared to a peer group. In this case, a model outcome typically

indicates a degree of deviation relevant to possible ML/TF risks, which is used to select an appropriate operational response. Banks can then use the estimated risk level to decide on a response, such as risk-differentiated review, comprehensive review, or pre-defined automated handling and closing of alert or event (see NVB Industry Baseline 'Ongoing Due Diligence', section 1.3 'Handling methods').

The following paragraphs list some examples out of many possible uses of models in alert and event generation. These examples aim to highlight use cases, not to provide an exhaustive list of model use for alert and event generation.

- In **client filtering**, the use of models can strengthen banks' ability to identify potential risks and compliance issues. Models such as fuzzy matching algorithms can identify potential matches to Politically Exposed Persons, sanctions and watchlists. Named entity recognition and named entity resolution models can support adverse media retrieval by identifying and resolving named entities in news articles or other sources, aiding in the identification of possible ML/TF risks related to a client.

2 Supervisory Guidance on Model Risk Management, Board of Governors of the Federal Reserve System Office of the Comptroller of the Currency, 4 April 2011.

- In **transaction filtering**, models support compliance with sanctions regulations. Also in this process, fuzzy matching algorithms can help to identify individuals, organisations or entities that are subject to sanctions or trade restrictions.
- Models are commonly used in **transaction monitoring** in several ways. Rule-based and supervised machine learning models can identify transactions that exhibit known, specific ML/TF risk patterns. Anomaly detection models can help detect new and less defined ML/TF risks by highlighting transactions that deviate from expected behaviour. Network analysis models analyse transactional networks to detect unusual transactions and thus can highlight ML/TF risks involving multiple parties.
- In **client monitoring**, models can help identify (changes in) client attributes and activities that indicate potential ML/TF risks or imply a need to reassess the client risk classification. Rule-based and supervised machine learning models can identify specific ML/TF risk patterns and changes in client risk profiles. Anomaly detection models can identify unexpected behaviour to help identify new ML/TF risks. Network analysis models can identify hidden relationships and potential links with illicit activities.
- Advanced **alert and event handling triage** models (also called ‘noise reduction’ or ‘prioritisation’ models) can supplement traditional rule-based models that have high false positive rates. These supervised machine learning models estimate the probability that an

alert or event is a true positive (e.g. leading to an FIU report or an adjustment of the risk-classification). This estimate then enables a differentiation of the outcome response as described below.

Note that the automation of processes (robotics) is not a model in the context of this Baseline and therefore out of scope. Similarly, single-use models only used for ad hoc analyses (e.g. in response to current events) are out of scope.

Outcome response

Models can directly lead to an alert and event, or they can be combined with other models in an ensemble that then triggers an alert and event. In either case, an outcome response must be specified as interface between models and subsequent operational review of the alert and event.

Most advanced models deliver a granular estimate, for example, of the probability or severity of an ML/TF risk. A common industry practice is to bucket these estimates in intervals. Each interval could then be translated into a specific response based on its risk level. This defines a risk-based approach that differentiates the response based on the quantitative estimate of the risk level that the model delivers. The diagram below shows the relation between models for alert and event generation as a risk detection mechanism, the alert and event handling triage and the outcome response (i.e. the risk-based handling method).

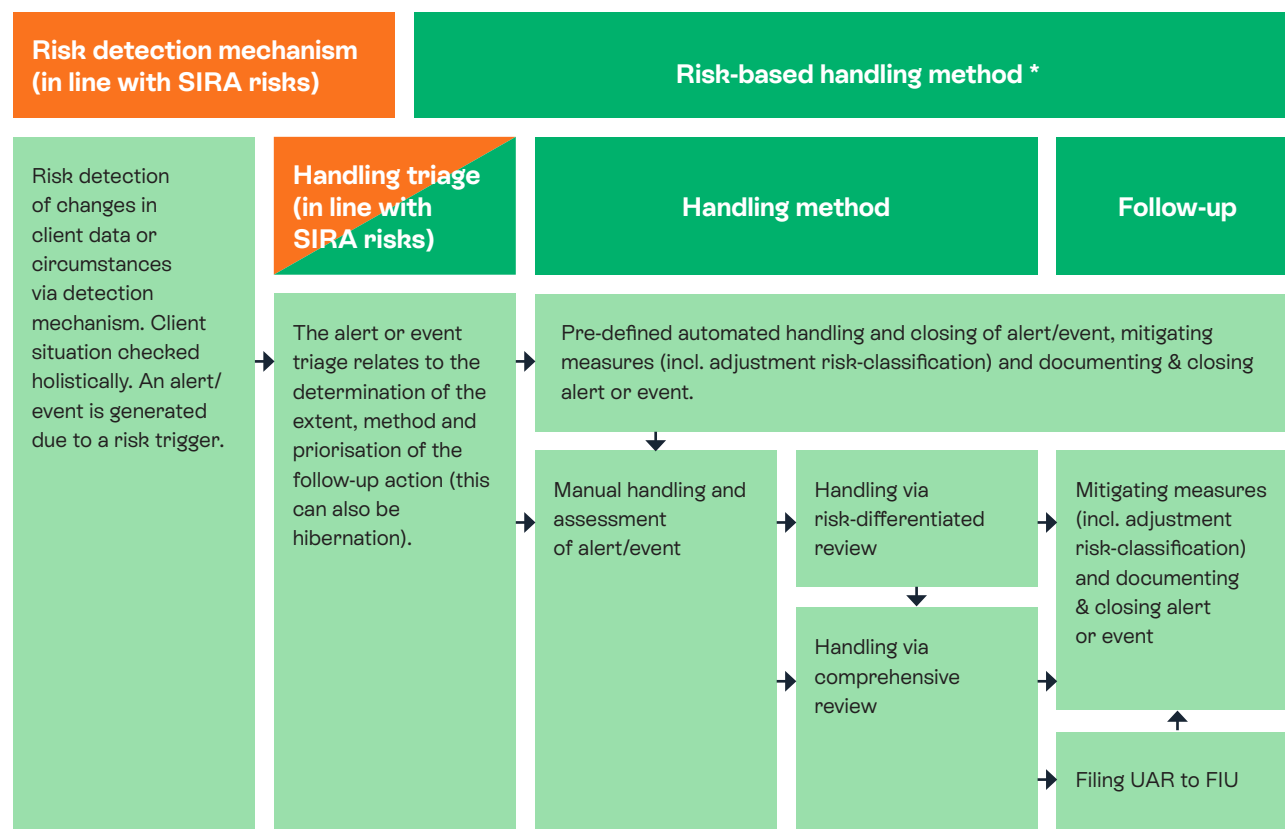
An example of a model that delivers alerts bucketed into low-, medium-, and high-risk intervals, with subsequent handling assigned as follows: low-risk alerts or events go through a predefined automated handling and closing process, alerts and events at medium risk level are processed manually in a risk-differentiated review, and high-risk alerts and events lead to a manual comprehensive review (see NVB Industry Baseline ‘Ongoing Due Diligence’, section 1.3 ‘Handling methods’). More detailed bucketing schemes than the one described here are also possible.

1.3 Transitioning from rule-based to advanced models

Banks currently strive for a transition from rule-based to advanced models in the context of ODD. A common way to start this transition is to add an alert and event handling triage model to an existing set of business rules. These additional advanced models enable better risk-based tailoring of the response to the alert or event. This increases efficiency by focussing the manual alert and event handling efforts where they are risk relevant.

Further steps in the transition to more advanced models for alert and event generation could include the following.

- The advanced models are monitored and updated in a continuous learning loop to reflect new ML/TF risks and changes in circumstances. Identification of new risks can lead to the development of new advanced models that



* This flow is a simplified version for illustrative purpose. Please note that for different risk detection processes (TF, TM, CF, CM) different and more specific process flows would apply.

● Risk detection mechanism ● Risk-based handling method

3 Precision and recall are evaluation metrics to assess the effectiveness of a model in identifying relevant cases. Precision measures the accuracy of the alerts, focussing on the ratio of correctly generated alerts to the total number of alerts. A higher precision value indicates fewer false positives. Recall, on the other hand, measures the effectiveness of the alerts and emphasises how many relevant alerts were successfully identified and aims to minimize false negatives.

specifically detect these risks.

- For transaction and client monitoring, advanced models such as anomaly detection or alert-generating versions of a triage model can be added to detect previously undetected and more complex ML/TF risks, thus increasing the effectiveness of risk detection.
- Rule-based models can be replaced by advanced models with better precision and/or recall. [3]

This increases both efficiency and effectiveness because advanced models can detect more complex combinations of risk factors than a rule-based model. Note that it may not always be useful to replace rule-based models with advanced models, for example, in the case of objective indicators.

The coverage of ML/TF risks (as described in a bank's risk assessment such as the Systematic Integrity Risk Analysis – SIRA) with models is an important consideration in this transition. When introducing advanced models, banks can opt for large models that cover all or multiple ML/TF risks or they can choose to develop multiple models that each detect specific ML/TF risks. It depends on each bank's specific circumstances which approach is to be preferred.

1.4 Conditions when using models

In any transition path to more advanced models, banks will need to manage risks associated with the use of models. To ensure the reliable, ethical, and responsible use of models, a number of conditions must be met. This Industry Baseline describes how banks can adequately manage those conditions.

To a large extent, these conditions are applicable to the use of models beyond the scope of alert and event generation. For this generic purpose, banks should have a sufficiently mature model risk management framework in place. Model risk management would be expected at least to establish and define:

- standards for ownership of and responsibilities and accountability for models (e.g. an overview of competent bodies or functions and their decision-making processes and procedures);
- requirements and standards for (metrics of) model performance, fairness, and explainability;
- requirements and standards for model monitoring;
- requirements and standards for data quality management and data governance;
- requirements and standards for the technical infrastructure in which models operate;
- requirements for training and awareness for stakeholders;
- a process for managing model development and modification, including testing, validation, and approval before implementation;

- model documentation standards (see the NVB Industry Baseline on ‘Technical Model Documentation’ [4]);
- policies that describe the model approval and risk acceptance procedures.

The details of a model risk management framework are not in scope of this Baseline. See SR 11-7 for more comprehensive information.

In general, advanced models do not imply other or more stringent requirements and controls than traditional rule-based models. If the general model risk management conditions outlined below are met, valid and responsible use of the model should be ensured. Comparative performance analysis between old and new models can facilitate model validation. Banks can then decide which model most adequately helps to mitigate ML/TF risks, without having to prove that the new model delivers the exact same outcomes (and specifically the same true positive results) as the old model.

Specific conditions when using models in the ODD context

In addition to generic model risk management, the ODD context poses several specific challenges to use models. The paragraphs below outline a set of considerations when using models in event and alert generation, rather than the minimal requirements that must be met before banks would be able to start using models. It is within the banks’ discretion to determine whether these are sufficient to transition towards the use of more sophisticated models. It is recognized that the

processes and their maturity will evolve and improve over time as banks gain insights from incrementally building more and better models.

Governance

Models used in ODD are subject to specific regulations such as the Wwft. They are typically being developed by teams located in the first line of defence for this purpose and validated by the banks’ AML/CFT compliance function. Additionally, banks’ model risk management function also has a mandate to validate models and will consider them from a more technical perspective. These overlapping responsibilities create uncertainty about the demarcation between the various competencies involved. To address this, banks should have the following condition in place.

- Documented and clear operationalisation of the roles and responsibilities of the model owners, users, and validators of ODD models to avoid uncertainty and conflict.

ML/TF coverage

Banks want to ensure that the introduction of (advanced) models maintains or increases the quality, effectiveness and efficiency of their ODD framework. This means that all alert and event generating mechanisms, including models, together sufficiently cover the ML/TF risks as identified in relevant risk assessments (including SIRA).

4 The ‘Technical Model Documentation Baseline’ is the output of the DNB innovation roundtables.

To this end,

- Banks can determine and document the positioning of the model in their ODD framework. This includes defining and documenting the model's:
 - purpose and use within the framework also related to the other ML/TF risk detection mechanisms;
 - relation to other processes or controls (e.g. relation of alert and event generating model with the risks as identified in relevant risk assessments, data actualisation processes, threshold setting, outcome responses, etc.).
- Banks have procedures in place to document and validate the risk coverage [5]. Outcome analyses (i.e., back-testing, portfolio management, audits) provide banks with insights into the coverage of the relevant risks and the ability to improve this. Banks should perform such coverage testing prior to implementation and during the use of the model. Specifically, banks should monitor if the risk coverage of the models is adequate and have procedures in place to follow-up on the discovery of new or missed ML/TF risks to enhance or adjust their control framework and safeguard that learnings are fed back to optimise or (re)develop existing models.
- Banks can define model performance metrics that enable the objective comparison of models, e.g. for model selection or to detect deterioration of performance. Model performance must be expressed in reliable, quantitative, objective and measurable metrics (see the NVB Industry

Baseline on 'Technical Model Documentation', section 5.2 'Validation of results'). The metrics can eliminate the need for additional tests such as 'shadowing' new models against old systems in parallel runs when introducing or updating advanced models.

- Banks decide, based on their SIRA and risk appetite, the minimum performance they require of each model.
- Banks define requirements and standards for data governance that apply to models in the context of this Baseline. The AML setting imposes additional requirements on data governance because data used in ODD should be updated regularly (Wwft, art 3.11).

Risks with the use of advanced models

As techniques in the domain of AI are developing rapidly, opportunities and risks develop just as quickly, including for the use in ODD. Advanced models can be hard to interpret, and their use in alert and event generation imposes requirements to clarify their operation. In order to explain what drives each individual alert and to ensure fairness:

- Transparency is required to be able to ascertain which ML/TF risks the model covers and thus enable model validation to assess its fitness for purpose. Explainability is required for informed (risk-based) manual handling of alerts and events.
- It is vital that models do not inadvertently disadvantage certain client groups. Banks should be able to define their concept of fairness and demonstrate how they ensure that their (advanced) models operate accordingly.

1.5 Client types

Banks can make use of models in their ODD framework in a risk-based manner to generate alerts and events for all client types. Depending on the model's objective and performance, they can differentiate its use and operation for different client groups.

1.6 Criteria to demonstrate effective implementation

Banks demonstrate their controlled and effective use of models by documenting and substantiating how they meet the conditions covered in section 1.4 'Conditions when using models.'

Furthermore, banks should have documentation in place to demonstrate effective implementation of the risk-based approach within their respective ODD controls. An overview of this documentation can be found in the overarching NVB Industry Baseline on 'Ongoing Due Diligence', section 1.6. The section covers the following categories:

- I Risk & control documentation;
- II Effectiveness testing documentation;
- III Process validation documentation;
- IV Documentation of relevant decision making;
- V Client data documentation.

5 See NVB Industry Baseline 'Ongoing Due Diligence', section 1.6 Criteria to demonstrate effective implementation, I. Risk & control documentation.

2 Impact

Adequate use of models is paramount for an effective and efficient risk-based ODD framework. It enables banks to accurately analyse ML/TF risks while balancing potential adverse impact on clients and operations. At the same time, the use of models comes with risks and implies additional obligations for banks so that they can adopt models in a reliable and responsible way. The banking industry experiences the need for standards on how to achieve this. Conditions set out in this Baseline offer a common language and understanding of ODD controls based on models.

Society as a whole will benefit from the ensuing improved effectiveness and efficiency of ODD controls as it delivers a safe and trustworthy financial system while limiting unnecessary burden for well-intended citizens.

3 Use cases

Alert triage model

Example

A bank's rule-based transaction monitoring system leads to a large proportion of the analysts spending their time manually sifting through obvious false positive alerts. The output of the rules does not provide information to allow prioritisation in the handling of alerts that are more likely to be true positive.

Industry Baseline

- The bank develops a model for alert handling triage. The model is developed using supervised learning on representative and recent data with known outcomes.
- An alert is labelled as positive if it leads to a reassessment of the client's risk classification. Other alerts are labelled as negative.
- For each alert, the model delivers an estimate of the probability that it is a positive alert. Based on this estimate, the model assigns alerts to one of three buckets that determine the operational follow-up.

Testing and monitoring

- Prior to model implementation, model validation is conducted involving all aspects of the bank's model risk management framework (e.g. including model objective and soundness, design, technical implementation, bias testing and governance).
- Model acceptance includes a comprehensive testing phase with out-of-sample data that is more recent than the data used to develop the model.
- During the use of the model, the performance is regularly monitored to confirm that the outcomes are still valid and that the process still sufficiently mitigates the targeted AML/CFT risks.

Model performance metric

Example

A bank wants to define a quantitative metric to determine whether an alert generating model performs sufficiently adequate and to allow for objective comparisons between different models and model versions.

Industry Baseline

- The bank defines a performance metric that combines the precision in the generated alerts and the recall in risk-based transaction samples.
- The target outcome for the model is defined as positive for
 - a alerts that lead to an enhanced due diligence review of the client situation; and for
 - b alerts that lead to an unusual activity report to the FIU.
 Alerts that do not meet either of these two criteria are labelled negative.
- The bank defines a minimum required performance level that is within their risk appetite and compliant with relevant legislation.

Testing and monitoring

- Prior to model implementation, the bank reviews the performance of new models by means of performance scores that follow from model testing procedures.
- During the model lifecycle, the bank has monitoring and reporting procedures in place for the defined performance metric to keep track of model performance.

- Based on this monitoring, the bank regularly evaluates if the model still meets the intended objectives, conditions and updates and replaces or withdraws it when needed.

Anomaly detection model

Example

A bank wants to augment their rule-based transaction monitoring models to detect more complex ML/TF schemes for which no concrete red flags have been identified.

Industry Baseline

- The bank develops an anomaly detection model using the Isolation Forest algorithm.
- The model's inputs consist of features that are relevant to known ML/TF risks. These features relate to a client's behaviour over a four-week period: they do not describe individual transactions, but rather describe the behaviour over the period. Typical examples of features are ratio of international transactions, ratio of cash transactions, etc.
- The model's relevance is validated on historical data by measuring if cases identified as an outlier by the model have a higher portion of true positive alerts.
- The model also generates information for each alerted case listing the most relevant risk features to support the alert handling. These details describe which features contributed most to the detection of the anomaly.
- The model outcomes are bucketed into extreme, high, and other anomalies. Extreme anomalies always lead to an alert, cases in the 'high' bucket are randomly sampled as a continuous below-the-line test.
- The model is validated and monitored according to the bank's model risk management framework.

Context

Regulatory framework

Anti-Money Laundering and Countering the Financing of Terrorism (hereafter: AML/CFT) legislation requires that banks perform ongoing monitoring of their business relationships, including scrutiny of transactions undertaken throughout the course of the relationship. In addition, it requires reporting of unusual (as defined in the Wet ter voorkoming van witwassen en financieren van terrorisme, hereafter: Wwft) or suspicious (as defined in the Anti-Money Laundering Directive, hereafter: AMLD) transactions and attempted transactions to the Financial Intelligence Unit (hereafter: FIU).

There are no legal requirements that specify what method or technology banks must apply to generate ML/TF alerts and events. However, DNB and the Wolfsberg Group provide guiding principles for the use of models in banks. Moreover, the proposed EU AI Act [6] will have an impact on the requirements for banks to make responsible use of models and therefore also on this NVB Industry Baseline.

6 Proposal for a 'Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending certain Union Legislative Acts', 21 April 2021. The industry expects the Act to be formally adopted by Q1 2024.

- **General principles for the use of Artificial Intelligence in the financial sector, DNB, 2019**

“The principles in this chapter should be seen in the context of controlled and sound business operations. Proportionality applies to these principles, and their applicability should be considered in light of the scale, complexity and materiality of an organisation’s AI applications. The applicability of these principles is also determined by the role of an AI application in the organisation’s decision-making process; this means whether the AI application serves a descriptive, diagnostic, predictive, prescriptive, or automation purpose.”

“The principles are divided over six key aspects of responsible use of AI, namely (i) soundness, (ii) accountability, (iii) fairness, (iv) ethics, (v) skills, (vi) and transparency (or ‘SAFEST’). For each principle suggestions are provided on how the principle can be operationalised in an organisation.”

- **The Wolfsberg Principles for Responsible AI and Machine Learning, The Wolfsberg Group, 2022**

The Wolfsberg Group identified principles that support banks’ responsible use of Artificial Intelligence (hereafter: AI) and machine learning in their financial crime compliance applications.

These principles consist of the following five elements:

- 1 legitimate purpose;
- 2 proportionate use;
- 3 design and technical expertise;
- 4 accountability and oversight;
- 5 openness and transparency.

Relationship between ‘DNB Good Practices’ and ‘NVB Industry Baseline’

DNB aims to illustrate its supervisory practices to the benefit of supervised entities by, for example, providing an interpretation of regulatory requirements (Q&As) and examples on how regulatory requirements can be met (Good Practices). It is important to note that neither the DNB Q&As nor Good Practices are legally binding.

The NVB Industry Baseline describes the application and execution of the risk-based approach, supported by models, in more detail.



© July 2023

Dutch Banking Association
Gustav Mahlerplein 29-35
1082 MS Amsterdam
www.nvb.nl