dutch
banking
association

Risk-based Industry Baseline

Technical Model Documentation

strong banks
strong society

Context

Introduction

Advanced analytics, Artificial Intelligence (AI) and machine learning models for compliance with Wwft obligations are increasingly being used in the banking sector. These models can be applied extensively throughout the financial crime framework, including in client onboarding, Ongoing Due Diligence (hereafter: ODD) and transaction monitoring. Innovations in this area allow banks to effectively and efficiently improve processes to detect Money Laundering and Terrorism Financing (hereafter: ML/TF) risks. Also, providing opportunities to efficiently and adequately enhance execution of Anti Money Laundering/ Countering Financing of Terrorism (hereafter: AML/CFT) and risk management processes.

Considering that incidents as a result of the application of models could have serious reputational consequences for banks and the sector, banks are held to high integrity standards. Banks are accountable for their use of advanced analytics, AI and machine learning, as applications may not always function as intended and can result in harm, damage or loss for banks or their clients. Model applications should therefore be reliable, behave predictably, and operate within in the boundaries of rules and regulations.

This NVB Industry Baseline describes the Dutch banking practices to demonstrate that appropriate processes have been followed and ensure trustworthiness of the models. Setting-up technical model documentation helps to provide accountability and transparency, meaning that banks explain how and why they use advanced analytics, AI and machine learning models in their business processes and how these applications function.

Development and use of advanced analytics, AI and machine learning differs per bank based on their needs and business model. The development and application of such models will also depend on a bank's client portfolio, technical maturity and capabilities. The measures described in this Industry Baseline give a general overview of the required information and documentation for various types of models if and when developed and deployed in the AML/CFT framework. It describes criteria for model choices including explainability, fairness and reliability of the chosen models.

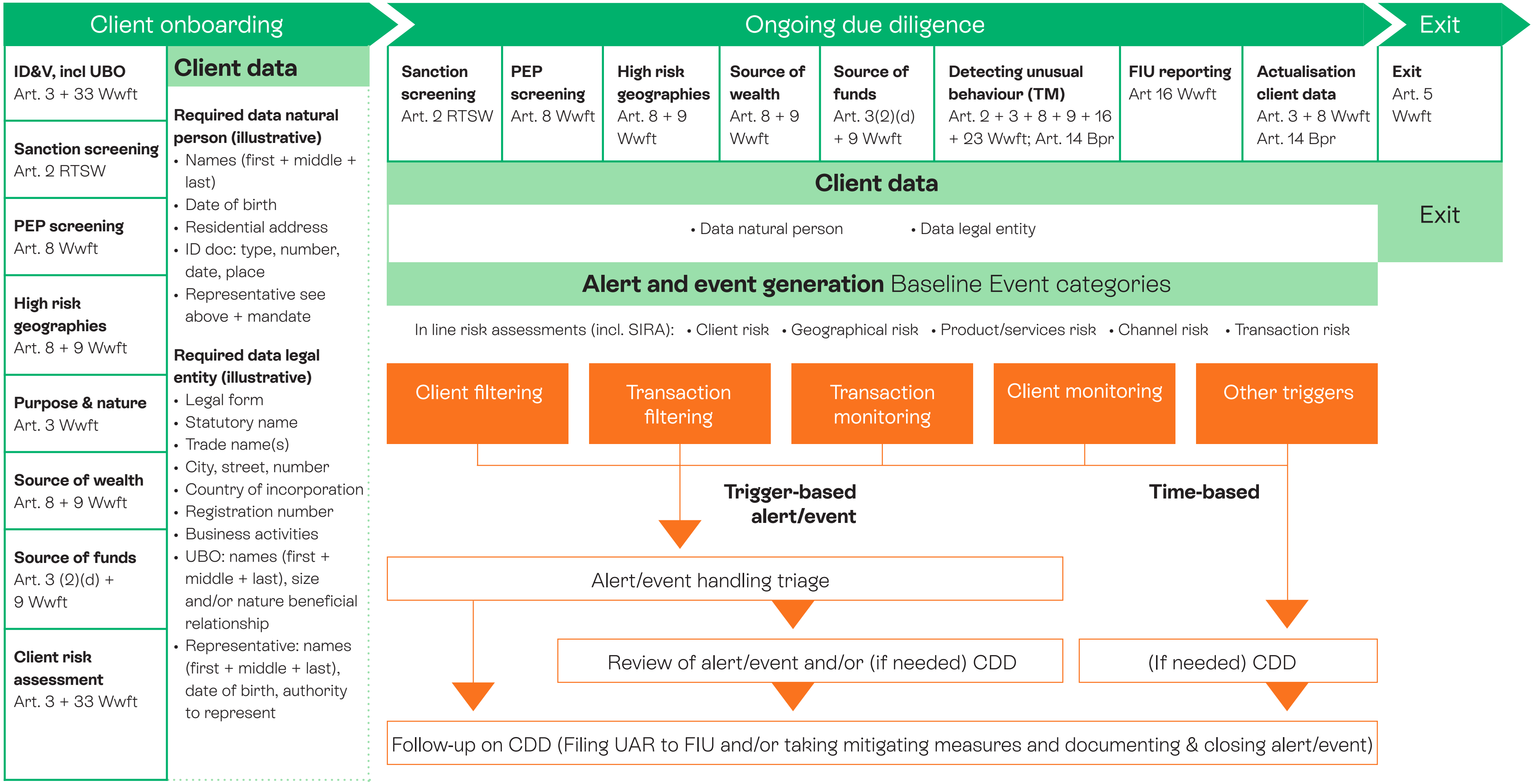
This Industry Baseline should be read in conjunction with the NVB Industry Baseline on 'Models in alert and event generation' that provides an overview of how models can be employed in a risk-based approach to detect ML/TF risks.

Positioning within the Financial Crime Framework

Models can be applied by banks in various stages of the Financial Crime Framework, such as client onboarding and for ODD processes as well as transaction monitoring. In this risk management framework models typically use client and transaction data as input and generate outcomes that form the basis for client risk assessments, alert and event generation. The coverage of ML/ TF risks, as described in banks' risk assessments (such as the Systematic Integrity Risk Analysis), is an important consideration.

Financial Crime Framework

Risk-based



Client onboarding

ID&V, incl UBO
Art. 3 + 33 Wwft

Sanction screening
Art. 2 RTSW

PEP screening
Art. 8 Wwft

High risk geographies
Art. 8 + 9 Wwft

Purpose & nature
Art. 3 Wwft

Source of wealth
Art. 8 + 9 Wwft

Source of funds
Art. 3 (2)(d) + 9 Wwft

Client risk assessment
Art. 3 + 33 Wwft

Client data

Required data natural person (illustrative)

- Names (first + middle + last)
- Date of birth
- Residential address
- ID doc: type, number, date, place
- Representative see above + mandate

Required data legal entity (illustrative)

- Legal form
- Statutory name
- Trade name(s)
- City, street, number
- Country of incorporation
- Registration number
- Business activities
- UBO: names (first + middle + last), size and/or nature beneficial relationship
- Representative: names (first + middle + last), date of birth, authority to represent

Ongoing due diligence

Sanction screening
Art. 2 RTSW

PEP screening
Art. 8 Wwft

High risk geographies
Art. 8 + 9 Wwft

Source of wealth
Art. 8 + 9 Wwft

Source of funds
Art. 3(2)(d) + 9 Wwft

Detecting unusual behaviour (TM)
Art. 2 + 3 + 8 + 9 + 16 + 23 Wwft; Art. 14 Bpr

FIU reporting
Art 16 Wwft

Actualisation client data
Art. 3 + 8 Wwft
Art. 14 Bpr

Exit

Exit
Art. 5 Wwft

Client data

- Data natural person
- Data legal entity

Exit

Alert and event generation Baseline Event categories

In line risk assessments (incl. SIRA): • Client risk • Geographical risk • Product/services risk • Channel risk • Transaction risk

Client filtering

Transaction filtering

Transaction monitoring

Client monitoring

Other triggers

Trigger-based alert/event

Time-based

Alert/event handling triage

Review of alert/event and/or (if needed) CDD

(If needed) CDD

Follow-up on CDD (Filing UAR to FIU and/or taking mitigating measures and documenting & closing alert/event)

□ Regulatory requirement
🟧 Risk detection mechanism

Risk-based Industry Baseline

1 Industry Baseline

This NVB Industry Baseline describes the types of information and documentation to be detailed when developing and using advanced analytics, AI and machine learning models in the context of the AML/CFT framework. With this technical model documentation banks can demonstrate that diligent processes have been followed and ensure trustworthiness of the advanced analytics, AI and machine learning models. The documentation also allows for transparency and can serve as future reference when updating and validating models.

Analytics models come in various sizes and shapes. Sufficient capabilities regarding advanced analytics, AI and machine learning methods and techniques are important to ensure that the developed and applied advanced analytics, AI and machine learning models are adequate and well understood. Accountability and oversight are important elements in advanced analytics, AI and machine learning. The responsible use of these models should be supported by banks' model governance and risk management framework for which the technical model documentation can serve to identify and mitigate relevant technical, compliance, and ethical risks.

The technical model documentation presented in this Industry Baseline provides a framework to support banks in demonstrating how their application of advanced analytics, AI and machine learning models is developed and assessed. The technical model documentation can serve to report model performance in reliable, quantitative, objective and measurable metrics, which enables reliable, transparent, ethical and responsible use of these models. Furthermore, it allows for demonstrating fairness, explainability and interpretability of the model.

The technical model documentation provides banks with a framework that ensures that advanced analytics models, AI and machine learning are used according to their intended purpose.

The correct working of advanced analytics, AI and machine learning applications should be regularly assessed. Especially, for scenarios where the model behaves unexpectedly, an incident response plan that includes steps for rapid detection, response, and recovery should be in place. Issues with data integrity and bias, both in development and production, should be evaluated and documented in a structural manner for future reference. Similarly, feedback loops can be applied to ensure that changing patterns in input or output are flagged at an early stage so models can be adapted accordingly and remain effective over time.

The technical model documentation also fits within the overall governance framework that oversees the entire lifecycle of advanced analytics, AI and machine learning models. This includes the roles and responsibilities of different internal and external stakeholders involved in the development, deployment, and maintenance of the models. The documentation will also enable accountability, clear communication with various stakeholders regarding the capabilities and limitations of the models and can guide training of relevant staff to understand and correctly interpret model outputs.

The technical model documentation is expected to establish and define:

- **model governance**: all governance roles and processes;
- **conceptual model design**: all major decision points made together with business parties which impacted the design;
- **technical model design and data**: description of data sources, data quality, data processing and pre-processing steps;
- **model development**: methodology used, validation of results, and stability tests used;
- **explainability**: ability to explain what happens in a model from input to output;
- **risk considerations**: overview of any risks and limitations of the model(s) and technologies used;

- **model maintenance in production**: plan to monitor and retrain the model's performance in production;
- **implementation of the model**: how will the model be deployed in production.

In the annex a more detailed description of the technical model documentation is provided.

2 Impact

Advanced analytics, AI and machine learning models can significantly contribute to an effective and efficient AML/CFT framework by improving banks' information position on clients and transactions and by accurately detecting ML/TF risks. These systems can limit unnecessary burdens on clients and improve risk relevant AML/CFT controls.

However, advanced analytics, AI and machine learning and innovation come with additional responsibilities and accountability for banks. Specifically, sufficient human agency and oversight is necessary before any decisions are taken regarding a client. Impact on society and clients can be major if advanced analytics, AI and machine learning are not used responsibly. The safety and fundamental rights of people and businesses should be guaranteed by preventing biases and avoiding risks of violations of privacy rights.

The technical model documentation for these systems as set out in this Industry Baseline will contribute to a consistent approach to developing, assessing and using advanced analytics, AI and machine learning models.

3 Use cases

The use case below is an example to illustrate a practical application of this Industry Baseline and not intended to be exhaustive.

Combining a rule-based model and machine learning

Example

A bank needs to enhance its rule-based model for transaction monitoring to a risk-based model that combines its rule-based system with machine learning.

Industry Baseline

- During the risk assessment a bank has identified 20 red flags that might indicate ML risks. These risks have been addressed to date by a rule-based transaction monitoring set-up, which resulted in imprecise outcomes (i.e., a large number of false positives). The bank decides to analyse the possibility of tackling the problem with a machine learning model.
- The bank forms a group of business, compliance and modelling experts, that agrees on the objective and design of the model.
- As part of the implementation testing, the bank makes a comparison between the initial outcomes of the model and the outcomes of the existing business rules over the last 2 years. The model can identify 93 of the 96 cases flagged earlier by the rule-based system that got escalated and reported. Only 3 of the SARs are missed. However the model also finds an additional 16 SAR-worthy cases.
- The compliance expert in the group performs an analysis on the 3 missed SARs and concludes that 2 of the 3 SARs were bycatch, purely based on non-transactional information. The third missed SAR however is a complex case involving clear ML risk factors that require attention. The group decides to analyse the SAR together, but is incapable of finding a targeted way of detecting the SAR with the new model.
- The group documents their effort and proposes that the benefit of the model outweighs the missing single SAR.
- After approval of Compliance, the model is implemented in parallel to the rules, within the production pipeline, to ensure technical and performance stability.
- After a month of running the model in parallel with the current set-up, the team is certain the model runs without issues and is ready for deployment.
- The proposal is tabled at the Compliance committee, outlining the pros and cons and receives final approval.
- The model is implemented and the model monitoring efforts start in parallel.

Context

Regulatory framework

The regulatory context for this topic is described in relevant parts of applicable laws, regulations and guidelines from various authorities. The Wwft requires banks to assess client integrity risks at onboarding and continuously during the business relationship, and to perform ODD, including scrutiny of transactions undertaken throughout the course of the business relationship.

There are no legal requirements that specify if and how banks should use models when implementing the requirements of the Wwft. However, DNB and the Wolfsberg Group provide general guiding principles for the use of advanced analytics, AI and machine learning models in banks. Moreover, the EU Artificial Intelligence Act (AI Act) will have an impact on requirements for banks to make responsible use of models and therefore also on this NVB Industry Baseline.

The use of advanced analytics, AI and machine learning models may create issues beyond the scope of AML/CFT. Specifically, the interaction with the General Data Protection Regulation (GDPR) and the AI Act should be adhered to by banks.

Below an overview of relevant laws, regulations and guidance.

- **General Principles for the use of AI in the Financial Sector, DNB, 2019**
  “AI applications in the financial sector should be reliable and accurate, behave predictably, and operate within the boundaries of applicable rules and regulations. Firms should also be accountable for their use of AI, as AI applications may not always function as intended and can result in damages for the firm itself, its clients and/or other relevant stakeholders. it is vital for society’s trust in the financial sector that financial firms’ AI applications – individually or collectively – do not inadvertently disadvantage certain groups of clients.”
- **Wolfsberg Principles for using AI and Machine Learning in Financial Crime Compliance, The Wolfsberg Group, 2022**
  “By leveraging the advances in data science underpinning AI/ML, FIs can holistically analyse the client and transactional data created by their products and services more effectively and efficiently to detect, investigate, and manage the risk of financial crime, and satisfy regulatory requirements.”
  “Design of AI/ML systems should be driven by a clear definition of the intended outcomes and ensure that results can be adequately explained or proven given the data inputs.”
- **EBA Machine Learning for IRB Models, Follow-up Report from the Consultation on the Discussion Paper on Machine Learning for IRB Models, EBA/REP/2023/28, August 2023**
  “Another challenge for financial institutions relates to the explainability and interpretability of the results of ML models. A higher level of complexity may lead to better model performance but at the cost of lower explainability and comprehension of the model’s functioning.”
  In addition, to ensuring that the model is correctly interpreted and understood, institutions are recommended to (...) provide a summary document in which the model is explained in an easy manner based on the outcomes of the analyses described in point a. The document is recommended to describe:
  i. The key drivers of the model.
  ii. The main relationships between the risk drivers and the model predictions. The addressees of the document are all the relevant stakeholders, including the staff which uses the model for internal purposes.”
- **Charter of fundamental rights of the European Union, 2012**
  “The peoples of Europe, in creating an ever closer union among them, are resolved to share a peaceful future based on common values. Conscious of its spiritual and moral heritage, the Union is founded on the indivisible, universal values of human dignity, freedom, equality and solidarity; it is based on the principles of

democracy and the rule of law. It places the individual at the heart of its activities, by establishing the citizenship of the Union and by creating an area of freedom, security and justice."

- **AI Act (final text to be published) – Seven principles**
  "The seven non-binding ethical principles for AI should help ensure that AI is trustworthy and ethically sound. The seven principles include: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being and accountability."
- **AI Act (final text to be published) Article 17**
  "1. Providers of high-risk AI systems shall have a quality management system in place that ensures compliance with this Regulation. It shall be documented in a systematic and orderly manner in the form of written policies, procedures or instructions, and can be incorporated into an existing quality management system under Union sectoral legislative acts. It shall include at least the following aspects: [...]
  (b) techniques, procedures and systematic actions to be used for the design, design control and design verification of the high-risk AI system;
  (c) techniques, procedures and systematic actions to be used for the development, quality control and quality assurance of the high-risk AI system;
  (d) examination, test and validation procedures to be carried out before, during and after the development of the high-risk AI system, and the frequency with which they have to be carried out;
  (e) technical specifications, including standards, to be applied and, where the relevant harmonised standards are not applied in full, or do not cover all of the relevant requirements, the means to be used to ensure that the high-risk AI system complies with the requirements set out in Chapter 2 of this Title;
  (f) systems and procedures for data management, including data acquisition, data collection, data analysis, data labelling, data storage, data filtration, data mining, data aggregation, data retention and any other operation regarding the data that is performed before and for the purposes of the placing on the market or putting into service of high-risk AI systems;
  (g) the risk management system referred to in Article 9;
  (h) the setting-up, implementation and maintenance of a post-market monitoring system, in accordance with Article 61;
  (i) procedures related to the reporting of serious incidents and of malfunctioning in accordance with Article 62;
  (j) the handling of communication with relevant competent authorities, including sectoral ones;
  (k) systems and procedures for record keeping of all relevant documentation and information;
  (l) resource management, including security of supply related measures;
  (m) an accountability framework setting out the responsibilities of the management and other staff with regard to all aspects listed in this paragraph."

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 technical model documentation banks can use when applying advanced analytics, AI and machine learning models. Banks decide how to use the NVB Industry Baselines and each bank individually determines its own policy and risk appetite.

This NVB Industry Baselines was designed in consultation with the DNB, the Dutch Authority for the Financial Markets (AFM) and the Dutch Data Protection Authority.

Annex

1 Model Process Control and Model Governance items

In this section, all items are to be explained from the process perspective only. All technical details belong to later sections.

1.1 Model Governance roles
1.2 Authors of TMD
1.3 Document history log
1.4 Approvals and signoffs (including use of data)
1.5 Ongoing monitoring plan
1.6 Model review requirements

2 Executive summary

In this context, a model can refer to a single machine learning or analytics model, rule-based model or a combination of machine learning models and rules.

2.1 Business problem statement

Describe the composition of the model (machine learning, rule-based or combination), and elaborate on the purpose of the model i.e., which business problem is it addressing, what it is expected to do, how does it fit into the existing business context and processes, and what is the customer scope.

2.2 Intended use of the model

Describe in full detail the scope of the model and its use in production. For instance, how frequently is output produced and how is it applied in the business processes. Describe all intended uses of the model. Also include situations when the model may not be used.

2.3 Data and Modelling

Describe the applicability of the data used and the actual data circumstances. Outline the final model selected, main modelling decisions, and why these choices were made.

2.4 Performance

Describe the expected impact of the model.

3 Conceptual model design

List or refer to all major decisions made together with business parties which impacted the design.

3.1 Target variable definition

Describe the target variable (the variable that we are trying to measure/predict), mention which business processes it relates to and the link between the target variable and the business problem.

3.2 Definition of an observation in the model

Describe in business terms a generic observation (data points / features) used in the model. This description should provide clarity for both subject matter experts, developers, and validators.

3.3 Scope of the model

Describe the observations that are in scope when the model is (going to be) used in production, using relevant business terminology. Include all relevant aspects such as time, definition of e.g., client, transaction and/or events that may impact the model in this description.

4 Technical model design and data

4.1 Data sources

Describe all data sources used, e.g., the database, table/file names and external sources (if relevant). For each table, include a list of the extracted variables and their definitions. Also describe the relationship between this data, the business process, the problem statement, and the solution. It should be clear from this list how data usage is minimized, considering the GDPR.

4.2 Data retrieval

Detail from a technical perspective how data flows between IT systems and is ingested to be used for the model purpose.

4.3 Data quality and representativeness assessment

Describe data quality checks that have been executed and the resulting outcomes in a precise manner. Conclude with whether these assessments showed that the data used is sufficiently representative for the defined business problem.

4.4 Data processing

4.4.1 Data pre-processing
Describe the main selections and pre-processing steps that have been undertaken, such as selections of rows in tables or otherwise specified selections of raw data, treatment of missing values, aggregations, joins, text pre-processing etc. This does not include feature and/or target values, design and/or definition.

4.4.2 Feature variables design
Describe how the preprocessed data is transformed when deriving the values for each feature. Provide the business expert input that led to these features. The descriptions of the features should be able to be understood with limited reference required to existing code. Also, describe features that have been selected directly from a database (e.g., feature store). For the final feature set (analytical base table), provide the completed quality checks including tests for e.g., multi-collinearity.

4.4.3 Target variable processing design
Describe how you processed data to target values, in accordance with the target variable definition. Your descriptions should be clear as to be understood by non-technical stakeholders. Ensure that proper documentation on the exact target variable implementation choices is either included here, or in the code base. Where applicable, state the temporal relationship between business terms and target definition.

5 Model development

5.1 Methodology

5.1.1 Generic description of machine learning methods applied here
Describe the different machine learning algorithms that were used during experimentation. It is important to find balance between how much detail is used in this description versus what details can be referenced to in existing well-established publications.

5.1.2 Train/validate/test (for supervised learning models)
Describe how the train/validate/test sets were constructed and if any sampling techniques were used.

5.1.3 Hyperparameter tuning
Describe the hyperparameter tuning process used for the final model that was selected.

5.1.4 Domain expertise
State here or refer to another document that details all major decision points made together with business parties that impacted your model design.

5.2 Validation of results

5.2.1 Performance metrics and constraints
Provide definitions of the performance metrics that were selected to validate the suitability of your model, describe why they were relevant.

5.2.2 Results achieved
Provide actual estimates of the metrics from the previous section for your final model as well as an overview of the results achieved for models that have been rejected. Also, describe limitations of the model that result directly from the metrics.

5.3 Stability tests

Stability tests are used to determine how a model changes over time.

5.3.1 Description of procedures and statistics used
Describe all procedures used to test stability of your model.

5.3.2 Stability estimates and justification rationale
Provide actual estimates of all considered stability statistics for your final model. Also provide rationale for accepting the obtained results. A key aspect to consider is to obtain satisfactory results that sufficiently guarantee model’s performance in production.

5.3.3 Sensitivity Analysis

Sensitivity analysis determines how changes in the feature set may affect the target variable. Describe all procedures used in testing the sensitivity of the model.

6 Explainability

Refers to the ability to explain what happens in your model from input to output.

6.1 Describe the required degree of explainability

Describe the applicability of explainability with respect to how the model is used (e.g., whether the model is integral to the risk mitigation process, hence the extent to which interpretability is required). Explainability describes the functioning and results of the model sufficiently as to be understood by stakeholders.

Two aspects of explainability include transparency and interpretability. Transparency refers to describing the internal workings of a model, which this documentation describes. For interpretability, this describes how a human can interpret the model outputs.

6.2 Describe how interpretability is ensured (based on the requirements of 6.1)

For interpretability (which refers to understanding the reasoning behind a model's predictions), in cases where models are embedded in existing processes with interpretability requirements, describe the list of features and their relations toward the integrity risk framework where applicable. Describe the feature importance across all samples (global explainability) and describe the relations of features to the model output for specific samples (local explainability).

7 Risk considerations

Provide a detailed overview of any risks and limitations of the model(s) and technologies used. The list below is not meant to be exhaustive.

7.1 Regulatory assessment

Describe the results of a regulatory self-assessment to show that the model is compliant with relevant regulations and policies.

7.2 Preconditions and restrictions

Describe the preconditions and restrictions for the model, e.g., 'Data and labelling quality', 'Limitations in scoping' and 'Model complexity'. Include a description and the potential impact.

7.3 Assumptions

Describe the underlying assumptions made during development of the model, e.g: 'We assume that entities that engage in money laundering behave differently or can be profiled by certain characteristics.'

7.4 Ethical considerations

Describe the ethical considerations associated with this model. This can include, for example, considerations related to specific ethical topics (fairness, explainability, equity, human judgement, and bias) based on internal and external regulation, policies, and/or guidelines. Considerations should include a brief description of the issues, cover both data and modeling, assess potential impact, cover business considerations, and document technical choices and analyses performed (possibly by including references to dedicated documents when in place).

7.5 Technology landscape

Impact of (expected) changes in the technology landscape which could impact the validity of the model over time.

7.6 Conclusion

Provide a conclusion on the risk considerations, balancing out the impact on the financial institutes' clients and the model's purpose. The explanation should warrant the conclusion that any residual risks have been reduced to an acceptable level when compared to the necessity to adequately prevent money laundering for the financial institutes as well as for society.

8 Model maintenance in production

8.1 Monitoring

Describe in detail the plan to monitor the model's performance in production. This should include all relevant model performance metrics as well metrics associated with model and features stability assessment.

8.2 Retraining (if applicable)

Describe how the model will be retrained once in production in full details including what criteria would trigger a model retrain such as a specific timeline or degradation of the model performance outside a set range.

9 Implementation of the model

Describe how the model will be deployed in production. This should entail all relevant details, such as referral to platform, but also tests (unit/integration/functional) etc. Also describe limitations in implementing the model (especially in the context of decision making).
If relevant, give special attention to the (scheduled) re-estimation of the model at regular intervals, i.e., when the model (parameters/coefficients) is dynamically updated, given new data. If an algorithm is used to do this, clearly describe this algorithm and the tests performed on this algorithm.

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