

AML Model Validation Framework

**ThetaRay**

AML Transaction Monitoring

Model Validation Framework

REVISION A

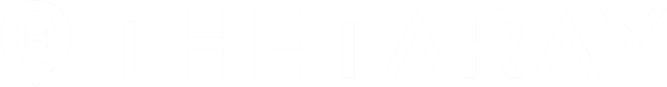


Table of Contents

**1. Legal Notice [3](#_heading=h.30j0zll)**

**2. Glossary of Terms [4](#_heading=h.3znysh7)**

**3. Document Purpose and Scope [8](#_heading=h.2et92p0)**

**4. Executive Summary [9](#_heading=h.tyjcwt)**

4.1 Model Overview [9](#_heading=h.3dy6vkm)

4.2 Appropriateness, Completeness, and Accuracy of the Input Data [27](#_heading=h.1ksv4uv)

4.3 Data Treatment [31](#_heading=h.4i7ojhp)

4.4 Data cleaning [32](#_heading=h.2xcytpi)

**5. Model Theory and Approach - TBD [34](#_heading=h.1ci93xb)**

5.1 Model Design [34](#_heading=h.3whwml4)

5.2 Parameterization [37](#_heading=h.2p2csry)

5.3 Solution Design [40](#_heading=h.41mghml)

5.4 Analysis Features Description [40](#_heading=h.2grqrue)

**6. Outcomes Analysis [41](#_heading=h.3fwokq0)**

6.1 Testing [41](#_heading=h.1v1yuxt)

**7. Model Implementation Accuracy and Ongoing Monitoring Plan [43](#_heading=h.1mrcu09)**

7.1 Model performance consistency in the implementation [43](#_heading=h.46r0co2)

7.2 Ongoing Monitoring Plan [44](#_heading=h.2lwamvv)

**8. Model Governance and Change Management [53](#_heading=h.3ygebqi)**

8.1 Model security and access control [53](#_heading=h.43ky6rz)

8.2 Model change management [54](#_heading=h.2iq8gzs)

8.3 FSD [54](#_heading=h.xvir7l)

**9. Algorithms description [55](#_heading=h.4h042r0)**

# Legal Notice

This document is proprietary and confidential and may only be used by persons who have received it directly from ThetaRay LTD. (“ThetaRay”) and may not be transferred to any other party without ThetaRay’s express written permission. The document provides preliminary and general information only and is not intended to be comprehensive or to address all the possible issues, applications, exceptions, or concerns relating to ThetaRay. All information contained in this document is confidential and shall remain at all times the sole property of ThetaRay. The recipient of this document has no right to disclose any or all of its contents or distribute, transmit, reproduce, publicize or otherwise disseminate this document or copies thereof without the prior written consent of ThetaRay, and shall keep all information contained herein strictly private and confidential. The information is intended to facilitate discussion and is not necessarily meaningful or complete without such supplemental discussion. Please note that the information procedures, practices, policies, and benefits described in this document may be modified or discontinued from time to time by ThetaRay without prior notice. As such, ThetaRay provides the document on an “As- Is” basis and makes no warranties as to the accuracy of the information contained therein. In addition, ThetaRay accepts no responsibility for any consequences whatsoever arising from the use of such information. ThetaRay shall not be required to provide any recipient with access to any additional information or to update this document or to correct any inaccuracies herein which may become apparent.

# Glossary of Terms

| **Term** | **Definition** |
| --- | --- |
| Aggregation Level | The period of analysis used by ThetaRay’s model for anomaly detection and alert processing. The period is usually Monthly, but it can be other defined time periods, such as, Daily or Weekly. |
| Alert | An alert is an alarm or a signal of the input record with suspicious, anomalous, or dangerous behaviors which will be generated for review by bank financial crimes investigations personnel. |
| AML/CFT | Anti-money laundering (AML) refers to the laws, regulations and procedures intended to prevent criminals from disguising illegally obtained funds as legitimate income. Counter-terrorist financing (CTF), or combating the financing of terrorism (CFT), seeks to stop the flow of illegal cash to terrorist organizations. |
| AML Risk Score | An optional mechanism used as part of ThetaRay’s risk-based approach in the anomaly to alert prioritization process. |
| Anomaly | An anomaly is a single record in the analyzed dataset that deviates from normality. |
| BSA | The Bank Secrecy Act (BSA) is the United States’ most important anti-money laundering regulation: banks and other financial institutions must ensure they meet the compliance obligations it involves. |
| Pattern Tagging | Pattern is based on several risk indicators each risk indicator is measured against a data point. In theory, data points that are in the same group should have similar properties and/or features. |
| Dimensionality | Dimensionality refers to how many features/columns a dataset has. |
| Dimensionality Reduction | Dimensionality reduction, or dimension reduction, is the transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data. |
| Data Engineer (DE) | A data engineer (DE) is a professional responsible for collecting, analyzing and interpreting large amounts of data |
| Data Scientist (DS) | A data scientist (DS) is a professional responsible for collecting, analyzing and interpreting large amounts of data. |
| False Negative | A false negative is an outcome where the model incorrectly predicts the negative class. |
| False Positive | A false positive is an outcome where the model incorrectly predicts the positive class. |
| FATF | The Financial Action Task Force (FATF) is an intergovernmental organization founded in 1989 on the initiative of the [G7](https://en.wikipedia.org/wiki/G7) to develop policies to combat [money laundering](https://en.wikipedia.org/wiki/Money_laundering) and terrorism financing. |
| Feature | In machine learning, a feature is an individual measurable property or characteristic of a phenomenon being observed. Features are the data columns of the dataset used in machine learning model training. |
| Fusion Score | Each unsupervised algorithm in the ThetaRay platform generates a score for each record, which is then “fused” to produce a fusion score between 0 and 1. The higher the fusion score, the more likely this record is to be an anomaly. |
| Fusion Threshold | All records for which fusion score exceeds the fusion threshold will ultimately become alerts; conversely those records where the fusion score is less than fusion threshold will not. The fusion threshold is set at 0.5 by default as this represents the natural decision boundary probability between two classes, however, it may be configured by the user. |
| Investigative Entity | The primary focal point that the algorithms focus on when detecting anomalous behavior and producing alerts. |
| KYC | Know Your Customer (KYC) refers to the process of verifying the identity of your customers, either before or during the time that they start doing business with you. The term “KYC” also references the regulated bank customer identity verification practices to assess and monitor customer risk. The KYC process is also a legal requirement intended as an anti-money laundering (AML) measure. |
| Machine Learning | Machine Learning is a form of artificial intelligence (AI) which uses computer systems that can learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data. |
| Model | “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. For the purposes of this document, the entire suite of configurable processing components of the ThetaRay platform is considered to be the model in scope. |
| Normality | Normality defines “normal” or expected behaviors based on the underlying data distribution. Please note normality is not meant to refer exclusively to the Gaussian or normal curve. |
| OFAC | The Office of Foreign Assets Control (OFAC) administers and enforces economic sanctions programs primarily against countries and groups of individuals, such as terrorists and narcotics traffickers. |
| Parameter / Hyperparameter | A model parameter/hyperparameter is a configuration variable that is internal to the model and whose value can be estimated from data. |
| Publisher | Publisher is a functionality in ThetaRay platform to post analysis, enable detected anomalies to be externalized as alerts to the Investigation Center, and/or as anomalies that can be consumed by third party applications. |
| Risk-based Approach | The risk-based approach (RBA) to AML is a regulatory expectation from banks to detect and investigate events that are more likely to be related to risks. The risks measured are linked to business activity and hence, specific to any financial institution. |
| Rule-based Approach | The approach involves human-crafted or curated rule sets; within the AML risk management space, these rules define exactly which scenarios should be alerted. |
| SAR | A Suspicious Activity Report (SAR) is a document that financial institutions, and those associated with their business, must file with the COAF whenever there is a suspected case of money laundering or fraud. |
| SME | A subject-matter expert (SME) is a person who is an authority in a particular area or topic. |
| Sonar | ThetaRay Cross-border payments detection solution |
| SWIFT | Society for Worldwide Interbank Financial Telecommunications (SWIFT) system is a vast messaging network used by banks and other financial institutions to quickly, accurately, and securely send and receive information, such as money transfer instructions. |
| TMS | Transaction Monitoring system (TMS) is an integral part of an efficient anti-money laundering solution. |
| True Negative | A true negative is an outcome where the model correctly predicts the negative class. |
| True Positive | A true positive is an outcome where the model correctly predicts the positive class. |
| Unsupervised Learning | Unsupervised learning is a type of machine learning that looks for previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision. |

# Document Purpose and Scope

The purpose of this document is to provide a framework to guide model validation testing for implementations of the ThetaRay platform using unsupervised algorithms. In this instance, the term “model” refers to the suite of configurable processing components of the ThetaRay platform, as laid out in this document. Note that while certain descriptions of the ThetaRay platform, parameter settings, best practices, etc. are provided, this document is not meant to replace model development/specification documentation, user guides, or other related materials which should be produced jointly by the model owners and developers at each implementation. This document is also not meant to be an exhaustive list of all features and settings of the ThetaRay product, as certain features are not relevant to the model validation standards outlined below.

The framework provided in this document is meant to guide model validation exercises to be compliant with relevant local regulations, including the United States’ OCC 2011-12/FRB 11-7 supervisory guidance on model risk management and the United Kingdom’s Financial Conduct Authority SS3/18 guidance on model risk management.

# Executive Summary

## Model Overview

This chapter is covering the model's purpose and the understanding of how the model operates.

The ThetaRay platform uses advanced anomaly detection techniques to identify customer or transactional behavior patterns which deviate from the expected or “normal” behaviors for a given AML use case. Such approach differs from rules-based approach by enabling a multi- dimensional view of the data, meaning, not relying on specific scenario or threshold to generate an alert but measuring several risk indicators and data points to find the unusual events.

Anomaly detection approach

The generation of anomalies is the mathematical process of identifying “outlier” activities that differ from “normality”; to define normality, the platform employs a set of unsupervised algorithms. ThetaRay transaction monitoring solution leverages unsupervised learning techniques to identify anomalous activities from an AML risk perspective. Unsupervised learning is the ability to detect mutual or common properties between records in a data set population and set a defined normality (i.e., those records which are within an expected range relative to the population). Once normality has been defined, members of the population that do not conform to it are labeled as part of the abnormal population, each with its own unique set of differences. To identify abnormality, the platform runs a series of algorithms which each identify abnormal or “anomalous” records; In particular, an “anomaly” is a single record in the analyzed dataset that deviates from normality. Each algorithm generates a score associated with each identified anomalous record which provides evidence about the probability associated with each entry. However, each algorithm analyzes the data frame in a very particular way, for example, one applies a geometrical approach whereas another uses an algebraic method, etc.; as a result, the scores generated for each anomalous record may vary significantly from one algorithm to another. Each algorithm is trained to define a “normal” space. Once trained, the algorithms are deployed in production to detect anomalies based on the normal space identified during training, which remains static until the model needs to be retrained.

When prioritizing detected anomalous activity to publish alerts for investigation, ThetaRay's transaction monitoring solution incorporates one, and at times two, mechanisms in its risk-based approach (RBA). The first mechanism, which is always used, is the "Fusion Threshold." At a high level, and explained in more detail below, the Fusion Threshold identifies anomalous behavior based on mathematical calculations produced by the algorithms. The “Fusion” is set in a strict approach to include a large population of events that is unlikely to be the portion of criminal activity within a financial data set.

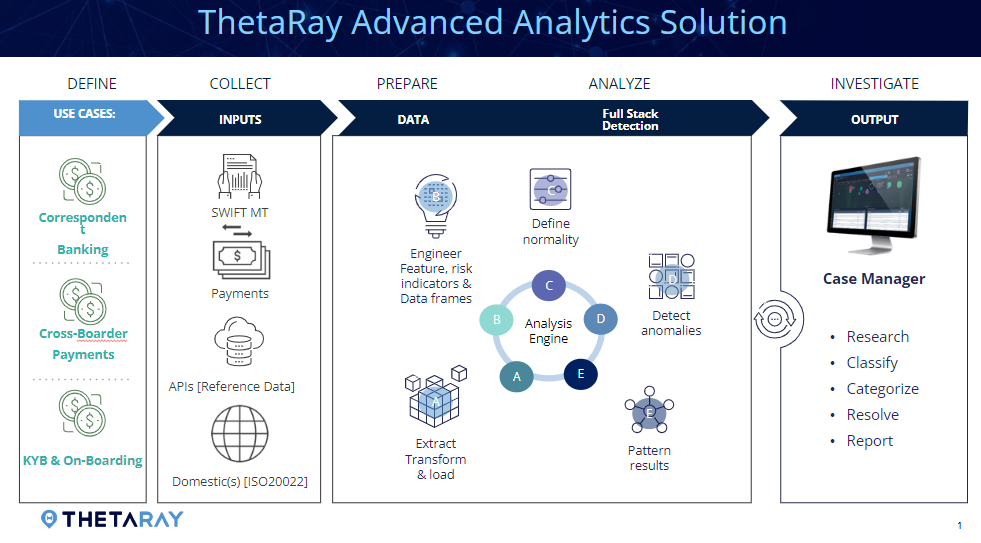
The second mechanism is the "AML Risk Score," which is an optional mechanism that is incorporated at the discretion of each Implementation. When included, the AML Risk Score is applied after the Fusion Threshold has been calculated and has produced anomalies so as to add a second layer of prioritization. Whereas the Fusion Threshold approaches the RBA from a purely mathematical basis by considering deviations from normality, as determined by comparison to the entire population of analysis, the AML Risk Score incorporates a "business aspect" by focusing its prioritization process on the weighted importance of the machine learning features included in the solution. Hence, with the AML score, the prioritization methodology is based on several steps taken at the set-up phase:

* The Financial Institution risks
* Testing of results and the impact of features by the financial institution’s AML team
* Evaluation of the alert’s quality by the financial institution’s AML team
* Updates to the model by ThetaRay team
* Finalization of the model by ThetaRay and FI’s AML teams

All the rows analyzed receive a score that enables any test, such as “below the line” testing for events that were not included as “anomalies” or such that were not converted to alerts due to the insignificant business meaning analyzed by the AML score.

### Model Development Process Overview

The following schematic provides a high-level overview of the methodology used to configure the ThetaRay platform; the following subsections will note key touchpoints in the model development and configuration process and, where applicable, best practices for each step in the process. In this instance, the term “model” refers to the suite of configurable processing components of the ThetaRay platform. Please note that this document is focused on model development and implementation use cases which are specific to the Anti-Money Laundering (“AML”) controls developed using unlabeled data (i.e., instances in which historical SAR, case, alert, or other tagged data is not available at the time of model development and implementation). The specific AML use case is defined by the model owner and/or developer during the initial development process;



The ThetaRay solution allows for both Artificial Intelligence (“AI”)-based configurations as well as the traditional rule-enabled monitoring; the focus of this document is to outline validation procedures and model development best practices for AI-based configurations as validation standards are generally established and well understood for traditional rules-based models. Figure 1 below outlines the process flow for model development; each step in the flow will be described in further detail below.

**Step 1 - Define Business Use Case**

The first step in the model development process is to define the specific AML use case which outlines the intended purpose of the model. Examples for use cases could be AML risk detection for Correspondent Banking/ retail banking/ Payments services, Trade Finance, Capital Markets, and more. The use case must be defined for each implementation by the institution, once the use case is defined all in-scope customer information and data is collected and uploaded, For example, for retail banking- the customer use of cash and expected activity may be more relevant, while in Trade Finance the industry and geographies could provide better view of risks.

The institution must either perform a targeted risk assessment for the monitoring typologies under consideration or leverage the results of its latest risk assessment exercise. The definition of the use case is the key step in the modeling process as it will have a direct impact on the data sourcing, feature design and assessment processes.

**Step 2 - Source Data**

Source data represents a collection of raw datasets from different source systems which will be used as a foundation for the subsequent model training. During this phase, all data sources and tables which are relevant to the exercise are identified. The in-scope source data are generally determined jointly by project stakeholders from both the client’s team and the ThetaRay team. Generally, data sources for transaction monitoring can be categorized into the following types:

* Transactional data, such as ISO 20022 Messages, SWIFT Messages, retail banking transactions, etc.
* Know Your Customer (KYC) information, such as customer information and segmentation based on predefined user business logic.
* Auxiliary data, such as country risk ratings, industry risk ratings, tax haven indicators, keyword lists, etc.

**Step 3 - Data Cleansing and Preprocessing**

Data cleansing and preprocessing process enhance the quality of the model training data to avoid potential data issues flowing into the downstream model training process.

Upon data availability and load, a series of validations are performed on the ThetaRay system to assure data correctness and completeness.

**Step 4 - Data Transformation and Feature Design**

The data transformation and feature design process help transform raw data points into more business-context meaningful data points (“features”) which are used by the model.

“Features” represent the definition of the data points in the context of the analysis, as the Algo are agnostic to the meaning of the data, the features engineering serve the users in the following aspects:

1. Risk coverage- by designing specific features to focus on risks, e.g.: high risk country.
2. Explainability- by designating features to represent math calculations -provide the business language, e.g.: total value of activity.

Feature design adopts bank SME input and guidance to enhance business sense of the model inputs. SMEs from both the bank and ThetaRay review the relevant red flags and risk mapping (determined based on the specific use case and the bank’s latest risk assessment) to ensure alignment between risks, data sources, the analysis approach, and features.

The feature design process employs a risk-based approach which aims to ensure that policy requirements as derived from regulatory expectations are implemented in this system. This is achieved by holding collaborative workshops between bank SMEs and ThetaRay data scientist teams to ensure the designed features are sufficiently comprehensive of the risks identified for the use case.

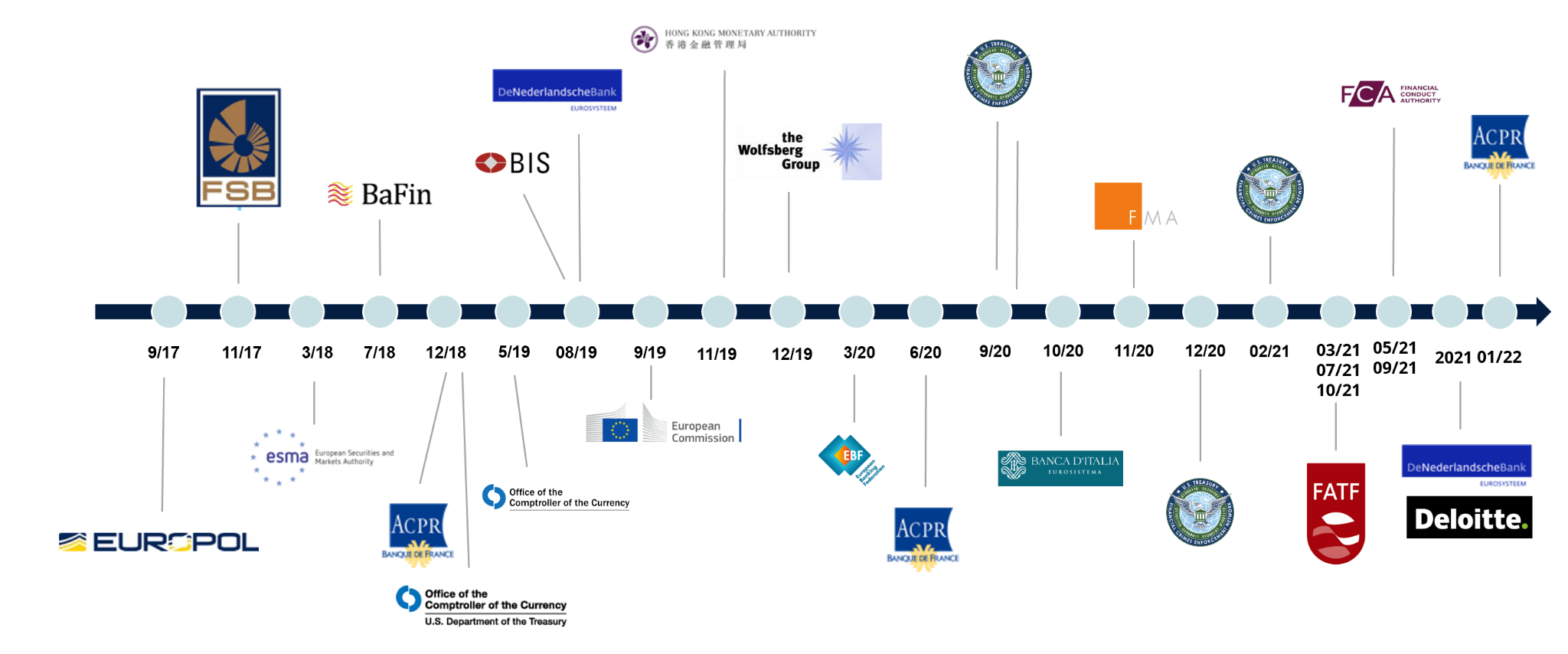
At this stage, both bank SMEs and ThetaRay data scientists and SMEs recommend qualitative “views” of the data which could improve the effectiveness of the model in addressing potentially suspicious money laundering behavior.

Based on the list of features selected, the algo in ThetaRay learns the normality in the data, based on the normality unusual events are detected and being presented in the context of the “Features”.

During the feature design process, raw data will be designed as features for processing by the platform.

The ThetaRay system provides full flexibility in the design of Features to maximize risk coverage based on available data.

The ability to focus on events that represent a combination of risk indicators is expected to be more effective (quality) and more efficient (quantity, time to resolution), In addition, such an approach is the execution of the regulatory expectation as expressed during the past 5 years. See the timeline representing these official publications:



many regulators and thought leaders have published their expectation from the Private sector to avoid “Rules Based Approach” and measure, mainly using Machine Learning techniques, unusual customer behavior in a “Risk Based Approach”.

Once features are designed, the key activity by the data science team at this stage is connecting the data sources by unique identifiers to ensure data is analyzed with the accurate context of the relevant entity. Unusual activity may be derived from four main spaces: (1) Discrepancies between the entity’s historical behavior and current; (2) Discrepancies between current activity by the entity and the current activity by the population; (3) Discrepancies between the entity in comparison to its segment or peer group in the current period and/or; (4) Discrepancies between the entity’s historical behavior and current in comparison to its segment or peer group.

To detect anomalous behavior on an aggregate level, raw input data is transformed in the following process which combines both data processing and feature design:

* Construction of historical behavior profiles for each client over specified windows of time and comparison of subsequent behavior to those historical behavior profiles
* Collect all relevant information on the transactional level.
* Aggregate transactional data by (Key, Time).
* Define time periods for analysis (AI and rules), monthly/ quarterly/ daily etc.
* Define the scope for each feature- comparison to its own history or population.
* Verify similarity between features that could duplicate the same measurement.

**Step 5 - Feature Assessment:**

Once features are designed in the prior step, they are assessed for reasonableness in both a quantitative and qualitative context. During this stage, designed features undergo common data quality tests to ensure they are within appropriate domain ranges as defined during the design process.

In addition, a qualitative review is performed by the Financial Institution’s SMEs to assess whether the designed features are adequately capturing potentially suspicious behavior. As mentioned above, before the initial model training process, SMEs from both the bank and ThetaRay side will review the designed features and assess the suitability in the current context and use case to ensure risk coverage.

Following the first analysis run, first results are reviewed by the FI’s SMEs, which assess the model output and based on the feedback, additional cycles will improve the model performance. The risk pattern identification process enriches the model with business sense and any anomalous activity that is classified into an identified pattern for future runs. This evaluation leads to the important decision concerning removal or maintenance or addition of features to achieve best results,

**Step 6 - Model Training Framework**

The model training framework is the learning phase, in which the algo learns normality in the data in order to enable anomalies detection.

The unsupervised algorithms can run partially together or altogether based on the developer selection. Selected algorithms each run independently and have unique methods for defining the activity that falls within most of the observed behavior (“normal”), during the training phase and identifying abnormal activity.

Each algorithm evaluates the historic data and leverages different combinations of mathematical techniques (e.g., algebraic, geometry, parametric, etc.) to uncover and define “normality” within the data.

Once normality is defined for each algorithm, a separate detection process is run for each algorithm to identify activity that deviates from the normality using different measurements such as density, distance, or distributions to generate anomalies.

The output by all algorithms is united (“fused”) to avoid duplications and ensure completeness. This leads to generation of the list of anomalies.

**Step 7 - Algorithm Fusion Process and Fusion Threshold Tuning**

The algorithms measure unusual events in different math calculations, to connect the dots and combine the list of potential anomalies into one, the platform employs a fusion layer which will ensemble all outputs (anomaly score) from different algorithms into one comprehensive fusion score.

This fusion score includes all findings to one range of distance from normality, the higher the score the most unusual the behavior is.

Each unsupervised algorithm generates a score for each record, which is then “fused” to produce a fusion score between 0 and 1. The higher the fusion score, the more likely this record is to be representative of anomalous behavior for the defined use case.

The Fusion score separates between the population of anomalies and the remaining data points.

Once a score is associated with each record, ThetaRay’s decision engine is used to determine which record needs to be alerted on - a basic condition that can be applied is checking whether the Fusion Score computed by ThetaRay’s algorithm exceeds a configurable threshold. It should be noted that banks may apply a risk-based approach that will be used to incorporate additional logic for prioritizing and filtering evaluated records - this mechanism enables fine tuning the alerts generated through ThetaRay machine learning algorithms, introducing rule-based alerts, as well as fine tuning the alert classification process.

Each anomaly generated from the algorithm fusion process is defined and graded by several properties:

* Trigger features: A collection of measurable features, which represent the deviation from the normal activity.
* Feature rating: For each one of the deviating features, the system rates its contribution to the deviation of this anomaly. The rating is based on “how far did this value deviate” from the rest of the population. Features are then sorted according to their rating from most significant to least significant. A trigger feature’s rating is then used to characterize the pattern of the anomaly.
* Fusion score: A value grading the overall abnormality of the record. Scores range between 0 and 1. The higher the score, the more this record deviates from normality.

For the anomalies generated from Step 7 above, there are two possible assessment methods:

1. Anomalies can be directly assessed as a part of in Step 9 (Alert Output Validation) and pushed into Step 10 (Anomaly to Alert);
2. Anomalies can be assessed in Step 8 to Step 10 on an individual (i.e., one by one) basis.

**Step 8 - Model Output Validation**

As mentioned in the executive summary above, few steps are taken to ensure model is performing within the expected quality:

* The alignment with Financial Institution risks.
* Testing of results and the impact of features by the financial institution’s AML team
* Evaluation of the alert’s quality by the financial institution’s AML team
* Updates to the model by ThetaRay team
* Finalization of the model by ThetaRay and FI’s AML teams.

Following the steps described above, a statistical sample of model outputs will be reviewed by the Financial Institution’s SMEs through an investigation process. During this process, business SMEs review and validate sample outputs to ensure compliance and intended monitoring objectives. Methodology for validation testing will be defined by each individual institution and use case.

Principals recommended by ThetaRay for the model performance testing:

Methodology for choosing alerts for investigation during pre-prod cycle testing

As stated above, one of the important steps during the development phase is a review and investigation of generated cases by the AML team. There are few considerations for selection of the sample to be tested and investigated by the team.

For each alert chosen, ThetaRay recommends a combination of the following criteria in order to include an unbiased and varied set of alerts to prioritize for investigations:

1. When alerts include multiple months for the same entity, it is preferred (but not a “must”) that the alert chosen be in the month it was first detected. This will enable the ability to understand the first unusual behavior by the entity.

2. From the list of alerts filtered in step 1, a representative sample set from all features that appear in trigger feature #1 is selected, in order to evaluate the performance of the features and the business meaning these features provide.

3. For each alert filtered by trigger feature #1, selection is determined by the following:

3a) Amount: a diverse range of amounts (i.e. low values, high values) from MT103 and MT202cov messages.

3b) Anomaly score: a diverse range of scores

3c) other criteria found in the data which can further diversify the set

The decisions for the anomaly selection are taken by the Financial Institution, in case there is an urgency to run the system in production, validation can be executed during production.

The Investigation Process

After anomalies for each use case have been processed by the ThetaRay system, alerts are published to the Investigation Center (IC). In addition to publishing the alerts, it is easy to extract files with sampled cases for review and document the evaluation by the Financial Institution teams. At the end of the evaluation process, the feedback from the AML team is embedded into the design of the analysis. In cases that the first review was concluded with unsatisfying results or in case there was a decision to change the model input, additional testing cycles are recommended to complete the development phase from a business value point of view.

Following investigation of the output sample, an assessment can be done on multiple levels:

**Feature Level**

Identify if each feature is generating valuable anomalies via business SMEs and data scientist review. If features are assessed as generating valuable anomalies (as defined by each institution and use case), In addition, potential differences between the valuable cases which the feature is signaling on vs. other types of cases are investigated. This information is further used to improve the design of the feature, as needed.

**Overall Level**

* The rates of anomalous activity per population type are calculated from the data.
* Any imbalances in the analysis and output with regards to population types are inspected.
* The overall quality of the results is assessed by bank and ThetaRay SMEs.

If any of the above are insufficient, they are addressed either via improved feature engineering or input feed segmentation. This information is assessed with regards to overall analysis design and overall approach.

Using either method, if results do not meet expectations, the process will revert to Step 5

**Step 9 - Anomaly to Alert**

The generation of anomalies, as described in the previous sections, is the mathematical process of identifying anomalous activities, and can also include AML Risk Score. Not all anomalous activities are likely to be related to Financial Crimes, hence, should be alerted on.

**Two decision points exist in the flow:**

* Anomaly externalization logic
* Alert generation logic

**Anomaly Externalization Logic**

As part of the model design, business decisions are implemented to exclude certain anomalies from being published to the consuming application, i.e. Investigation Center or a third party / existing internal investigations platform. That business logic is implemented as conditions in the publisher entity for this analysis.

Conditions are written in a common scripting language and applied to the anomaly parameters. The conditions should be tested prior to making the model operational.

**ֵAML Risk Score**

The "AML Risk Score" (Score) is a risk-based approach (RBA) mechanism that incorporates a “business aspect” to alert prioritization. Its inclusion as part of the transaction monitoring solution is discretionary. When included, the AML Risk Score is applied after the Fusion Threshold has been calculated and has produced anomalies. The effect of including the AML Risk Score is to focus the pool of anomalies identified by the Fusion Threshold to yield a high priority set of anomalies to be published as alerts. Thus, a high priority of alerts is accomplished by combining algorithmic calculations and business relevancy.

The Fusion Threshold approaches the RBA from a purely mathematical basis by calculating deviations from normality in its analysis of the entire population of data. The AML Risk Score takes a different approach by calculating a weighted importance of the machine learning features included in the solution, plus other factors, noted below. The feature importance is determined by analyzing the anomalies investigated by the client during the pre-production onboarding process. Hence, while the Fusion Threshold determines an anomaly by calculating the probability of a defined transaction set based on its distance from normality, the AML Risk Score takes the Fusion Threshold one step further by considering the relevancy of the triggered features to turn anomaly into alert for operational purposes. Based on the sampled anomalies that has been investigated by the client.

The formula used for determining the AML Risk Score considers investigated anomalies that are both detection worthy and non-detection worthy. Other factors in the calculation include a minimum value threshold, effectively not focusing on very smaller-valued anomalies, as well as consideration for multiple true positive anomalies (increases relevancy) and false positive anomalies (decreases relevancy), which have been detected in the pre-production model training phase. In this way, the statistical relevancy of the features is unbiasedly calculated using several uncorrelated and meaningful factors.

The process of applying the AML Risk Score involves detailed analysis and statistical review. Using the factors noted, the AML Risk Score is calculated for each anomaly. Since each anomaly is unique in the sense that the triggered features and their corresponding contribution to its anomalous behavior is rarely if ever the same in other anomalies, the calculated Score for each anomaly is unique. Given that, further analysis is then conducted considering all the anomalies with the Score applied to determine a threshold value of the Score. Scores above the threshold identify the statistically relevant high priority anomalies, while those below the threshold are do not convert from anomaly to alert. An above-the-line (ATL) and below-the-line (BTL) analysis is performed by the customer to validate the threshold. This ATL/BTL process involves investigations of a statistically relevant sample size performed by the client. The outcome of this review is used to confirm that the anomalies ATL are significantly more detection worthy than the anomalies BTL. If needed, fine-tuning to the Score is done to optimize the threshold. As with all RBA solutions, the objective of achieving zero detection worthy anomalies BTL is not realistic.

In the case where only the Fusion Threshold is used, all the anomalies identified become published to the Investigation Center as alerts. When the AML Risk Score is also employed, the set of anomalies produced by the Fusion Threshold are converted to alerts by the ATL AML Risk Score threshold, and that sub-set of (high priority) anomalies is then published to the Investigation Center as alerts.

In any case, anomalies with anomaly score above 80 (the exact number to be determined by the customer based on pre-production tests) should be investigated.

**Alert Generation Logic**

The Investigation Centre applies logic to optimize the generation of alerts for the purpose of increased efficiency. The two methods applied are:

* Deduplication
* Consolidation

Alert de-duplication is applied for multiple triggers (anomalies) of the same pattern (i.e., risk) arriving within a configurable time window for the same investigated entity. When this happens, only the first trigger arriving within this window will generate an alert. When this alert is generated, the timer is set until the time window elapses, at which point the time is reset and the next trigger to arrive will again generate a new alert. An example of when it may be useful to employ this feature is when the investigated entity’s activity is aggregated using a rolling window (please note the aggregation window is not to be confused with the deduplication window). That is, assume the aggregation window is of the last 30 days and that the analysis runs on a daily basis. This means that there is a potential overlap of up to 29 days between activity that’s evaluated today to the one to be evaluated tomorrow. As time elapses, the variation increases and the chance the overlap will cause the same activity to be alerted over and over is reduced. To mitigate this, duplicative alerts can be suppressed over a user-defined window, for example, to 5-10 days. Alert consolidation is a complimentary optimization option. This option applies to triggers of the same investigated entity, yet in this case it is assumed that different triggers represent different risks (patterns). Assuming an active alert, either in the queue or already assigned to an analyst user. Upon the arrival of a second trigger from the same analysis for the same investigated entity as of the active alert, and a different pattern is identified, the second trigger will not generate a new alert. Rather, it will enrich the active alert.

The user will get a more comprehensive view of the potential suspected activities by the same entity and will be able to make an informed decision about the investigated entity as a whole. In this way, efficiency is increased by the unification of information which reduces the level of effort as compared to reviewing each alert in isolation. Note that unlike for de-duplication, consolidation is not subject to a time window.

### Intended Purpose of the Model

The intended purpose of the Business Model is to monitor potential suspicious activities within the business activities in scope.

The analysis will use the relevant transactional data.

### Intended model scope of application for business need.

Regulators expect FI’s to detect unusual behavior by their customers, counterparties and other FI’s they engage with that could potentially indicate criminal activity, within the Correspondent Banking use case, for example, detecting suspicious behavior by other banks’ customers or even by banks themselves.

#### Model uses and users, and role of the model output in the business process.

The AML monitoring should enable timely detection of unusual activity through the business process, to report suspicious activity, apply enhanced Due Diligence or even terminate business relationships. Such control will increase the FI’s confidence in its AML detection and allow safer business growth.

#### Portfolio sizes/Exposure/Populations/Regions/Products

The customer of ThetaRay decides what will be the input of the data/ geographies in scope: \*\* Customer **to attach documentation detailing the procedures performed to define the scope of the ThetaRay implementation**.

This scope may be updated from time to time, for additional use cases, businesses, or territories. For any new implementations, this document remains with its basic description of the ThetaRay platform while specific chapters that are use case related will be updated.

### Applicable regulatory requirements

**Customer to add any regulatory standards that are critical for the risk assessment performed and bank’s operations (C4001?).**

### Applicable internal policies and procedures

**\*\* Travelex’s policy states:**

## Appropriateness, Completeness, and Accuracy of the Input Data

This section documents the general model development data, which is used to build, test, and calibrate models, and is distinct from data in the production.

Input data is selected and collected by ThetaRay customer, see clause 4.2.2 for the data selected and rationale.

### Design of the development data

* Data for analyzing Transactions in the context of AML requirements usually includes:
* Transactional data, such as ISO 20022 Messages, SWIFT Messages, retail banking transactions, Payment formats etc.
* Know Your Customer (KYC) information, such as customer information and segmentation based on predefined user business logic.
* Reference tables, such as country risk ratings, industry risk ratings, tax haven indicators, keyword lists, etc.

### Source of data (internal data and/or external data)

The tables below address the needs of the specific use case of Travelex. The data sources are intended to contend with the Payments challenges problem through the analyses described in the next chapter:

**Mass Payments - Remessadoras**

Source Transactions:

|  |  |
| --- | --- |
| **field** | **field\_type** |
| id\_cubo\_remessadora | string |
| numero\_remessa | string |
| operacao | string |
| transacao\_data | TIMESTAMP |
| documento\_originador | string |
| nome\_originador | string |
| tipo\_documento\_originador | string |
| cidade\_originador | string |
| estado\_originador | string |
| pais\_originador | string |
| tipo\_documento\_favorecido | string |
| documento\_favorecido | long |
| nome\_favorecido | long |
| pais\_favorecido | long |
| moeda | string |
| valor\_moeda | DOUBLE |
| valor\_reais | DOUBLE |
| remessadora\_nome | string |
| falecido | string |
| transacionou\_banco | string |
| transacionou\_corretora | string |
| aml\_pep | string |

Client KYC:

|  |  |
| --- | --- |
| **field** | **field\_type** |
| pessoa\_id | string |
| pessoa\_nome | string |
| pessoa\_data\_cadastro | string |
| pessoa\_data\_ultima\_atualizacao | string |
| pessoa\_data\_primeira\_operacao\_corretor | string |
| pessoa\_loja\_cadastro\_id | string |
| pessoa\_tipo | string |
| pessoa\_tipo\_origem | string |
| pessoa\_juridica\_capital\_aberto | string |
| pessoa\_juridica\_codigo\_cnae | string |
| pessoa\_juridica\_nome\_cnae | string |
| pessoa\_juridica\_data\_cnpg | string |
| pessoa\_juridica\_nome\_fantasia | string |
| pessoa\_juridica\_tipo\_capital | string |
| pessoa\_juridica\_tipo\_porte | string |
| pessoa\_juridica\_descricao\_ramo\_atividade | string |
| documento\_tipo | string |
| documento\_numero | string |
| documento\_numero\_sem\_formato | string |
| endereco\_principal\_cidade | string |
| endereco\_principal\_estado | string |
| endereco\_principal\_cep | string |
| enderecoprincipalpais | string |
| pessoa\_fisica\_data\_nascimento | string |
| profissao\_descricao | string |
| pessoa\_situacao\_banco | string |
| pessoa\_situacao\_corretora | string |
| cliente\_ccme | long |
| observacao\_cadastro\_banco | string |
| observacao\_cadastro\_corretora | string |
| pessoa\_cadastro\_corretorandicador\_ativo | string |
| pessoa\_juridica\_email | string |
| pessoa\_juridica\_telefone\_comercialddd | string |
| pessoa\_juridicatelefone\_comercial | string |
| pessoa\_fisica\_email | string |
| pessoa\_fisica\_telefone\_comercialddd | string |
| pessoa\_fisica\_telefone\_comercial | string |
| pessoa\_fisica\_telefone\_celularddd | string |
| pessoa\_fisica\_telefone\_celular | string |
| primeira\_operacao\_banco | string |
| ultima\_operacao\_banco | string |
| primeira\_operacao\_corretora | string |
| ultima\_operacao\_corretora | string |
| pessoa\_fisica\_sexo | string |
| pessoa\_fisica\_nacionalidade | string |
| clienteccme\_novo | string |
| risco\_ceppld | string |
| pessoa\_grau\_risco | string |

Bad Guys Info:

|  |  |
| --- | --- |
| **field** | **field\_type** |
| lista\_negativa\_id | string |
| lista\_negativa\_ativa | string |
| lista\_negativa\_data\_inclusao | string |
| lista\_negativa\_nome | string |
| lista\_negativa\_numero\_documento | string |
| lista\_negativa\_numero\_documento\_sem\_formatacao | string |
| lista\_negativa\_tipo\_documento | string |
| lista\_negativa\_pais\_id | string |
| lista\_negativa\_observacao | string |
| lista\_negativa\_usuario\_ultima\_alteracao\_id | string |
| lista\_negativa\_ultima\_alteracao | string |
| lista\_negativa\_ultima\_origem | string |
| lista\_negativa\_ultima\_tipo\_id | string |
| dtm\_last\_change | string |

Country risk:

|  |  |
| --- | --- |
| **field** | **field\_type** |
| country | string |
| risk\_classification | string |
| source\_of\_information | string |
| isso\_code\_2\_letras | string |
| isso\_code\_numerico | string |
| isso\_code\_letra | string |
| start\_date | TIMESTAMP |
| end\_date | TIMESTAMP |

Tax Havens:

|  |  |
| --- | --- |
| **field** | **field\_type** |
| country\_code | string |
| country\_name | string |
| start\_date | TIMESTAMP |
| end\_date | TIMESTAMP |

Currency Exchange Rate

|  |  |
| --- | --- |
| **field** | **field\_type** |
| source\_currency | string |
| target\_currency | string |
| date | TIMESTAMP |
| exchange\_rate | DOUBLE |

**Confidence - Forex + Remesas**

Source Transactions

built from 4 different datasets with shared schema:

* trx\_ppc
* trx\_especie
* trx\_corretora\_remmitance
* trx\_forex

|  |  |
| --- | --- |
| **field** | **field\_type** |
| operacao\_id | string |
| numero\_proposta | string |
| proposta\_id | string |
| numero\_operacao | string |
| data\_boleto | string |
| proposta\_data\_hora\_criacao | TIMESTAMP |
| ano\_operacao | string |
| mes\_operacao | string |
| dia\_operacao | string |
| data\_mes\_operacao | string |
| data\_operacao | TIMESTAMP |
| tipo\_parceria | string |
| tipo\_operacao | string |
| remessadora | string |
| produto | string |
| moeda | string |
| codigo\_promocao | string |
| descricao\_promocao | string |
| status\_proposta | string |
| operacao\_cancelada\_apos\_bacen | string |
| reportado\_bacen | string |
| nome\_loja\_previsao\_entrega | string |
| fato | string |
| fato\_codigo | string |
| flag\_loja\_diferente | string |
| flag\_canal\_diferente | string |
| receita\_bruta | string |
| volume | DOUBLE |
| volume\_usd | DOUBLE |
| lucro\_partner | string |
| taxa\_minima | DOUBLE |
| volume\_me | DOUBLE |
| taxa\_sugerida | DOUBLE |
| tarifa\_operacao | string |
| tarifa\_admcartao\_credito\_debito | string |
| taxa\_negociada | DOUBLE |
| custo\_oportunidade | DOUBLE |
| comissao\_operacao | string |
| parametro\_comissao | string |
| percentual\_comissao\_operacao | string |
| grupo\_produto | string |
| cliente\_id | string |
| operador\_proposta\_id | string |
| operador\_boleto\_id | string |
| loja\_criacao\_id | string |
| loja\_finalizada\_id | string |
| indicador\_operacao\_cliente\_id | string |
| comercial\_responsavel\_cadastro\_cliente\_id | string |
| indicador\_cadastro\_cliente\_id | string |
| comercial\_responsavel\_indicador\_cadastro\_cliente\_id | string |
| comercial\_responsavel\_operacao\_cliente\_id | string |
| comercial\_responsavel\_indicador\_operacao\_cliente\_id | string |
| proposta\_data\_criacao | string |
| proposta\_boletada | string |
| operacao\_cancelada | string |
| status\_seguro | string |
| receita\_liquida | string |
| rentabilidade\_bruta | string |
| rentabilidade\_liquida | string |
| segmento | string |
| despesa\_parceiro | string |
| percentual\_parceiro | string |
| percentual\_matriz | string |
| comissao\_parceiro | string |
| comissao\_matriz | string |
| tipo\_recebimento | string |
| tipo\_pagamento | string |
| receita\_bpp | string |
| numero\_referencia\_moneygram | string |
| tarifa\_moneygrammn | string |
| comissao\_moneygrammn | string |
| comissao\_conf\_sobre\_tarifa\_moneygrammn | string |
| pais\_saida\_entrada | string |
| originador\_nome | string |
| originador\_pais | string |
| favorecido\_nome | string |
| favorecido\_pais | string |
| loja\_taxa\_id | string |
| loja\_cadastro\_id | string |
| loja\_boleto\_id | string |
| loja\_previsao\_entrega\_id | string |
| margem | string |
| operacao\_iof | string |
| receita\_controladoria | string |
| socio\_receita\_parcial | string |
| socio\_tarifa\_operacao | string |
| receita\_confidence | string |
| receita\_confidence\_remessadoras | string |
| receita\_smallworld | string |
| numero\_cartao | string |
| cartao\_data\_validade | string |
| cartao\_data\_recebimento | string |
| cartao\_migrado | string |
| cartao\_status | string |
| consignado\_para | string |
| tipo\_documento\_consignatario | string |
| numero\_documento\_consignatario | string |
| tipo\_movimentacao\_cartao | string |
| proposta\_link\_parceiro | string |
| valor\_comercial | string |
| valor\_publicado\_matriz | string |
| iof | string |
| seguradora | string |
| proposta\_data\_finalizada | string |
| proposta\_data\_hora\_finalizada | string |
| supervisor\_loja\_criacao | string |
| supervisor\_loja\_boleto | string |
| data\_emissao\_voucher | string |
| comissionado\_fx | string |
| tipo\_entrega | string |
| fatlastchange | string |
| proposta\_codigo\_campanha\_id | string |
| proposta\_codigo\_promocional\_id | string |

Area Risco

|  |  |
| --- | --- |
| **field** | **field\_type** |
| id\_area\_risco | string |
| cep | string |
| cidade\_municipio | string |
| bairro | string |
| lougradouro | string |
| uf | string |
| area\_restrita\_correio | string |
| area\_restrita\_fronteira | string |
| area\_restrita\_mineracao | string |
| area\_restrita\_portuaria | string |

CNAE Sensival

|  |  |
| --- | --- |
| **field** | **field\_type** |
| cnae\_id | string |
| cnae\_codigo | string |
| cnae\_nome | string |
| cnae\_status | string |
| cnae\_grupo\_risco | string |
| sensivel | string |
| flg\_ativo | string |

Corretora Transacoes

|  |  |
| --- | --- |
| **field** | **field\_type** |
| proposta\_id | string |
| details | string |

Loja

|  |  |
| --- | --- |
| **field** | **field\_type** |
| loja\_id | string |
| loja\_nome | string |
| loja\_nome\_exibicao | string |
| loja\_numero | string |
| loja\_associada\_id | string |
| loja\_regional\_id | string |
| loja\_status | string |
| loja\_canal\_digital | string |
| loja\_socio | string |
| loja\_canal | string |
| loja\_pessoa\_juridica\_id | string |
| loja\_inauguracao\_data\_abertura | string |
| loja\_inauguracao\_data\_encerramento | string |
| loja\_inauguracao\_status | string |
| dtmlastchange | string |
| loja\_rank | string |
| loja\_data\_cadastro | string |
| loja\_correspondente\_cambial | string |
| lojacr\_estrutura | string |
| lojacr\_nome | string |
| lojacl\_estrutura | string |
| lojacl\_nome | string |
| loja\_regional | string |
| loja\_diretor | string |
| loja\_gerente | string |
| loja\_cnpj | string |
| loja\_email | string |
| loja\_telefone | string |
| loja\_logradouro | string |
| loja\_numero\_endereco | string |
| loja\_complemento | string |
| loja\_bairro | string |
| loja\_cidade | string |
| loja\_uf | string |
| loja\_cep | string |

Natureza Tipica

|  |  |
| --- | --- |
| **field** | **field\_type** |
| id | string |
| tipo | string |
| ramo\_atividade | string |
| codigo\_natureza\_operacao\_codigo | string |
| codigo\_natureza\_operaco\_descricao | string |
| flg\_ativo | string |

Naturezas Sensiveis

|  |  |
| --- | --- |
| **field** | **field\_type** |
| categoria | string |
| codigo | string |
| produto\_transacao | string |
| anr | string |
| possibilidade\_de\_operar\_em\_especie | string |
| rastreabilidade\_de\_contrapartes | string |
| possibilidade\_de\_terceiros\_atuarem | string |
| amparo\_documental\_padronizado | string |
| burla\_or\_fracionamento | string |
| canais\_de\_alto\_risco | string |
| limite | string |
| pontos | string |
| risco | string |

Profissao Sensivel

|  |  |
| --- | --- |
| **field** | **field\_type** |
| profissao\_id | string |
| profissao\_codigo | string |
| profissao\_area | string |
| profissao\_descricao | string |
| profissao\_grupo\_risco | string |
| sensivel | string |
| flg\_ativo | string |

Client KYC:

|  |  |
| --- | --- |
| **field** | **field\_type** |
| pessoa\_id | string |
| pessoa\_nome | string |
| pessoa\_data\_cadastro | string |
| pessoa\_data\_ultima\_atualizacao | string |
| pessoa\_data\_primeira\_operacao\_corretor | string |
| pessoa\_loja\_cadastro\_id | string |
| pessoa\_tipo | string |
| pessoa\_tipo\_origem | string |
| pessoa\_juridica\_capital\_aberto | string |
| pessoa\_juridica\_codigo\_cnae | string |
| pessoa\_juridica\_nome\_cnae | string |
| pessoa\_juridica\_data\_cnpg | string |
| pessoa\_juridica\_nome\_fantasia | string |
| pessoa\_juridica\_tipo\_capital | string |
| pessoa\_juridica\_tipo\_porte | string |
| pessoa\_juridica\_descricao\_ramo\_atividade | string |
| documento\_tipo | string |
| documento\_numero | string |
| documento\_numero\_sem\_formato | string |
| endereco\_principal\_cidade | string |
| endereco\_principal\_estado | string |
| endereco\_principal\_cep | string |
| enderecoprincipalpais | string |
| pessoa\_fisica\_data\_nascimento | string |
| profissao\_descricao | string |
| pessoa\_situacao\_banco | string |
| pessoa\_situacao\_corretora | string |
| cliente\_ccme | long |
| observacao\_cadastro\_banco | string |
| observacao\_cadastro\_corretora | string |
| pessoa\_cadastro\_corretorandicador\_ativo | string |
| pessoa\_juridica\_email | string |
| pessoa\_juridica\_telefone\_comercialddd | string |
| pessoa\_juridicatelefone\_comercial | string |
| pessoa\_fisica\_email | string |
| pessoa\_fisica\_telefone\_comercialddd | string |
| pessoa\_fisica\_telefone\_comercial | string |
| pessoa\_fisica\_telefone\_celularddd | string |
| pessoa\_fisica\_telefone\_celular | string |
| primeira\_operacao\_banco | string |
| ultima\_operacao\_banco | string |
| primeira\_operacao\_corretora | string |
| ultima\_operacao\_corretora | string |
| pessoa\_fisica\_sexo | string |
| pessoa\_fisica\_nacionalidade | string |
| clienteccme\_novo | string |
| risco\_ceppld | string |
| pessoa\_grau\_risco | string |

Bad Guys Info:

|  |  |
| --- | --- |
| **field** | **field\_type** |
| lista\_negativa\_id | string |
| lista\_negativa\_ativa | string |
| lista\_negativa\_data\_inclusao | string |
| lista\_negativa\_nome | string |
| lista\_negativa\_numero\_documento | string |
| lista\_negativa\_numero\_documento\_sem\_formatacao | string |
| lista\_negativa\_tipo\_documento | string |
| lista\_negativa\_pais\_id | string |
| lista\_negativa\_observacao | string |
| lista\_negativa\_usuario\_ultima\_alteracao\_id | string |
| lista\_negativa\_ultima\_alteracao | string |
| lista\_negativa\_ultima\_origem | string |
| lista\_negativa\_ultima\_tipo\_id | string |
| dtm\_last\_change | string |

Country risk:

|  |  |
| --- | --- |
| **field** | **field\_type** |
| country | string |
| risk\_classification | string |
| source\_of\_information | string |
| isso\_code\_2\_letras | string |
| isso\_code\_numerico | string |
| isso\_code\_letra | string |
| start\_date | TIMESTAMP |
| end\_date | TIMESTAMP |

**Data Model**

\*\* Project is still in the model design phase. Once the model design is completed this section will be updated.

**Diagram

Description automatically generated**

### Data appropriateness

The specific data sources used and their appropriateness will be specific to each institution and their use case. However, certain best practices should be applied to ensure that the input data sources are complete from a business and risk perspective. A recommended process is that a risk assessment is performed (or the latest BSA/AML risk assessment is leveraged) to identify risks. ThetaRay’s customer is responsible for the risk mapping and coverage needed, ThetaRay’s Business SMEs and Data Science then map a data source to each area of risk to ensure coverage during collaborative workshops.

### Data quality assessment, data limitations & weakness

Upon data availability and load, a series of validations should be performed to assure data correctness and completeness. Certain validations can be performed automatically while others require manual inspection or review; the set of data validations used should be customized for each use case and feature set. These validations entail review of each data field in each data source supplied and include the following:

* Data is of the expected structure.
* Data is of the expected type.
* Data is of the expected time span.

**The validations below are applied to the data described in the previous section. They are an expansion of the 3 types of validations described.**

* Data validation: Check that each data file contains relevant data in the expected format.
* Data Integrity - Check that rows and columns in the input file match loaded data.
* Compare column order and names with reference schema.
* Compare null ratios.
* Perform payment or transaction type specific logic validations, for example: no missing BIC information when it is mandatory in a SWIFT transaction.
* Validate field data types according to reference schema.
* Compare field lengths to reference data (if available) to detect invalid values or encrypted fields.
* Check for zero\negative amounts.
* Check for duplicate transactions.
* Check distributions of fields, especially monthly distribution of transactions.
* Check for missing dates.
* Check the number of distinct values for specific fields.
* Check joins can be run across various data sources.
* Generate & inspect summary statistics for each field where applicable.

## Data Treatment

Invalid Row Treatment

Rows will be rejected upon data source upload if they have a number of columns not matching the number of columns in the data source schema. In addition, rows with invalid data will not be passed into the TR algorithms unless the invalidity is restricted to forensic features that do not represent the primary key (== investigated entity) or the event time (== occurred on).

## Data cleaning

In general - whatever way data is being ingested into the system, it is validated against a predefined schema of the tables into which data is uploaded (known as data sources). The data upload may result in rows being either rejected or corrupted. Rows are rejected when they don’t pass one of the validations (field type mismatch, missing primary key, main date field is missing or has different format, etc.). Rows are considered corrupted when the upload process fails to read the rows properly from the file due to line breaks, wrong number of delimiters, etc. In both cases the system generates a report for those invalid rows, for the inspection of governing systems or personnel. When the discrepancy is manageable (e.g., missing or type mismatched data), user-defined code may be put in place to work around the gap, so that effect on downstream processes is controlled.

Schema validation also applies to tables populated with transformed data, aka data frames. Similar principles to the ones applied for data sources also apply for data frames, considering, of course, the nature of the transformation: wrangling, join, union or copy.

In addition, each deployment should articulate its risk-based approach to dealing with missing and/or invalid values (e.g. setting tolerance thresholds for percentage of invalid rows).

Development data is a “point in time” snapshot of production data. That is, generally the entire set of production data is used over a relevant historical time period. The relevant time window and data points selected are reliant on the particular use case and conditions provided via workshops with business SMEs and will need to be customized by each individual institution.

In addition, in the model output validation step, a sample of anomalous records will be taken for SME review. Some statistical sampling methods allow reviewers to use a sample’s results to make inferences about the entire population under review. Items for a statistical sample must be selected randomly from the population, but can be stratified. The two statistical sampling methodologies included in the OCC Sampling Guidance booklet [3] (numerical and proportional) are the methods widely used in the industry and supported by statistical theory.

# Model Theory and Approach - TBD

This chapter documents the general areas of model risk related to the model development methodology and approach.

## Model Design

### Alternative modeling approach assessment

ThetaRay’s algorithms are more accurate, faster and provide a better unsupervised approach than known alternatives in the market.

For a discussion of the selection of the algorithms used by the ThetaRay platform, their appropriateness and comparison to leading industry and academic practices, please refer to the Algo chapter.

Unsupervised Machine Learning enables several KPIs that are unlikely to be achieved using old methods:

* Effectiveness -
  + The ability to find suspicious events
  + The ability to find patterns with no predefined characteristics- “Unknown Unknowns”
* Efficiency -
  + The ability to focus on the relevant cases, usually with lower number of alerts and anyway with low portion of “False Positives”
  + Investigation process - usually the investigation time in ThetaRay is faster due to the availability and structure of Forensics data in the Investigation Center.

These characteristics maximize the ability to find anomalies and operationalize the methodology for AML use cases.

### 

### Model assumptions

Example assumptions are provided below. This is not an exhaustive list as additional assumptions may apply at each institution based on the specific use case, data sources, etc. Model validators should carefully consider all assumptions.

| **Assumption/Limitation** | **Associated Model Risk(s)** | **Model Risk Mitigants/Remediation** |
| --- | --- | --- |
| All of the algorithms in ThetaRay assume explicitly and implicitly processing imbalanced data which means that it is assumed that the majority is normal without financial crimes and the abnormalities (anomalies) are minority. | The model may not be able to detect money laundering behaviors which do not deviate from normal activity. | The ThetaRay solution is not meant to detect “normal” behaviors which may be explained via static processes and tools (e.g., currency transaction reporting). For use cases which require the detection of known but normal behaviors, the rule-based capabilities of the ThetaRay platform can be leveraged. |

| The features used in the model are sufficiently comprehensive to capture the AML / CTF risks posed to the institution. | If features are not well-designed, the model may not be able to detect some money laundering behaviors associated with the monitoring objective. | A risk assessment exercise took place prior to the definition of each use case. The feature design and assessment process closely maps back to the identified risks to provide comfort that the system is sufficiently comprehensive for the monitoring objective. In addition, bank SMEs perform model output validation on a sample of results. |
| --- | --- | --- |
| Unsupervised anomaly detection closely depends on well-engineered features, which greatly relies on highly skilled domain experts on data integrity concern, null completion, codes vs. measurements handling, etc. | Failure to correctly address these issues could result in high false positive alarm rates. | The recommended ThetaRay model development methodology includes several steps to mitigate this risk, including ThetaRay domain area experts assisting with the design and configuration of the system at each implementation and performance testing to support key parameter settings. |
| Dependency on historical data:  Normality is usually measured against entity’s activity in previous months, | Lack of historical data may impact the effectiveness of the analysis. | In cases that there is no historical data at all (new business) - the FI will need to apply rules for a short period.  In cases that the FI is exposed to activity by an entity from time to time, such as the “Pseudo” customer analysis, features with comparison to the normality in the population will apply. |

### Development data suitability assessment

Development data selected includes:

* Transactional data
* Risk tables
* KYC

## Parameterization

### Feature Design

As described above, features were selected in accordance with: data availability, risks and potential valuable views of the data, followed by testing and improvements cycles.

### Feature transformation

To detect anomalous behavior on an aggregate level (as opposed to the transaction level), raw input data is transformed in the following process which combines both data processing and feature engineering.

The analysis in ThetaRay is based on the historical data that was uploaded to the MiniO. This data could be further manipulated by introducing new features from time to time in case of specific regulatory requirement, for example.

The data uploaded to Thetaray platform is going through transformation that enables feature normalization or scaling, as well as construction of a historical profile against every new event is measured. These aggregations by time-window or event type give a more relevant view of the data and behavior.

Normalization / scaling is also performed to efficiently organize data with the same units of measurement and minimize the columns that are repetitive and add noise to the model.

### Model parameter estimation

Please see the assumptions section for how the dimensionality reduction process is conceptually supported by the ThetaRay solution. Further details on other relevant customizations is provided below.

**1. Model training framework parameter optimization**

Certain parameters within the unsupervised algorithms can be tuned manually by model developers. These parameters are set as default values which were determined by ThetaRay data scientists based on past experience with implementing the algorithms.

**2. Fusion threshold optimization**

Fusion Score Generation: The fusion scores generated by each algorithm may vary significantly between algorithms, since some of the algorithms apply a geometrical approach while others use algebraic methods. Consequently, before fusing all the scores they need to be rescaled to generate a posterior probability ranging from 0 to 1. Low and high scores are transformed to probabilities close to 0 and close to 1, respectively. The final fused score of each entry is then calculated as the joint probability of all the probabilities calculated by each algorithm.

Fusion Threshold Definition: The fusion threshold is 0.5 by default and represents the decision boundary probability between two classes. All records where the score exceeds the fusion threshold are identified as abnormal which will ultimately become anomalies and those where score is less than fusion threshold will not. Such an approach guarantees a wide population of potential anomalies which may be fine-tuned per the use case and the FI’s risk appetite.

In the following cases, the fusion threshold may need to be tuned:

* Ensure valuable cases are not being missed
* Reduce the number of false positive alerts
* Apply different risk-based approaches, for example - for specific use cases - detect only the extreme cases.

Tuning Methodology: The fusion threshold will be optimized to ensure that the number of relevant or potentially suspicious alerts not flagged by the system are within an acceptable range or risk tolerance. The tuning of this parameter can be like testing practices for traditional rules-based monitoring, but in a simplified manner as there is only a single threshold or “line” to evaluate, i.e., the fusion threshold.

Two fusion threshold testing methods are provided to optimize the fusion threshold.

* BTL (Below the Line) testing method:
  + A lower fusion threshold will be determined to perform the testing.
  + A random sample will be selected from the records between the lower fusion threshold and current fusion threshold, and a risk-based approach will be applied to investigate them.
  + If the number of potentially suspicious anomalies in the sample exceeds the risk tolerance, the fusion threshold will be lowered again, and the investigation process will proceed until the fusion threshold is set so that the number of potentially suspicious anomalies in the selected sample is within the established risk tolerance.
* ATL (Above the Line) testing method:
  + A higher fusion threshold will be determined to perform the testing.
  + A random sample will be selected from the records between the current fusion threshold and higher fusion threshold, and a risk-based approach will be applied to investigate them.
  + If the number of potentially suspicious anomalies in the sample does not exceed the risk tolerance, the fusion threshold will be increased again, and the investigation process will proceed until the fusion threshold is set so that the number of potentially suspicious anomalies in the selected sample is above the established risk tolerance.
  + As mentioned above, setting the Fusion threshold to 0.5 is expected to provide relatively large population of anomalies, in some cases there will be assigned also AML risk score to separate between alerted cases and other cases. In such cases the BTL methodology is expected to apply considering the last criteria used- the AML score.

## Solution Design

Work in progress. To be completed by Travelex and ThetaRay

The following snapshot describes the flow the data goes through in the alert generation process:



1. data\_prep: In this step both transaction tables (outgoing & incoming) get consolidated into one, also flags and auxiliary columns which will be used later on get added to the table.
2. data\_prep\_check: In this step the process checks the count of transactions for the current month lies within the expected distribution based on the historical data.
3. compute\_features: In this step the computation of the analysis and forensic features occurred.
4. detect: In this step the previously train anomaly detection algorithms evaluate the current month’s activity of account holders and detect the anomalies.
5. drift: In this step the drift detection process takes place, certain parameters of the features and the anomalies get checked and if any is out of bounds and exceptions gets raised and the process is stopped.
6. identify: In this step the anomalies get tagged to risk indicators using the most important trigger features.
7. distribute: In this step the anomalies get published as alerts in the Investigation center.

## Analysis Features Description

Work in progress. To be completed by Travelex and ThetaRay

### 5.4.1 Mass Payments Remessadoras

The following sections details the analysis features used in the solution:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Feature Name** | **Display Name** | **Business Description** | **Investigated Entity** |
| 1 | sum\_risky\_receiver | Valor Saida Favorecido Risco | The total outgoing transaction value to PEPs and/or risky receivers in current period | Sender |
| 2 | sum\_bad\_guy\_receiver | Valor Saida Favorecido Bad Guy | The total outgoing transaction value to “Bad Guys” receivers in current period | Sender |
| 3 | sum\_trx | Valor Saida/Entrada Total | The total outgoing/incoming transaction value in current period | Sender/Receiver |
| 4 | cnt\_trx\_sc | Volume Saida/Entrada Total | The total outgoing/incoming transaction volume in current perior | Sender/Receiver |
| 5 | z\_score\_cnt\_trx | Volume Entrada Historico | The number of all incoming transactions in the current month compared with the customer average for the past 6 months. | Receiver |
| 6 | z\_score\_cnt\_trx\_seg\_country | Volume Saida/Entrada Segmentado Pais | The volume of outgoing/incoming transactions in current period compared to the volume of transactions of all entities in the same segment (using sender country) in the same period | Sender/Receiver |
| 7 | z\_score\_sum\_trx\_seg\_country | Valor Saida/Entrada Segmentado Pais | The value of outgoing/incoming transactions in current period vs the value of transactions of all entities in the same segment (using sender country) in the same period | Sender/Receiver |
| 8 | z\_score\_cnt\_dstnct\_rec | Um para Muitos Favorecidos | The count of distinct counterparties (using "receiver\_id") transacted with in current month compared with the customer average for the past 6 months | Sender |
| 9 | z\_score\_max\_sum\_trx\_sender | Valor Maximo Unico Originador | The max value received from an unique sender compared with the customer average for the past 6 months | Receiver |
| 10 | z\_score\_max\_cnt\_trx\_sender | Volume Maximo Unico Originador | The max volume received from an unique sender compared with the customer average for the past 6 months | Receiver |
| 11 | z\_score\_sum\_hgh\_rsk\_cntry | Valor Historico Entrada Pais Risco | The sum of all outgoing transactions to a high risk country compared with the customer average for the past 6 months. | Receiver |
| 12 | z\_score\_sum\_trx | Valor Transacao Historico | The Sum of all outgoing/incoming transactions in the current month compared with the customer average for the past 6 months. | Sender/Receiver |

#### 

The following sections details the pseudocode of each the features:

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Feature Name** | **Display Name** | **Pseudocode** |
| 1 | sum\_risky\_receiver | Valor Saida Favorecido Risco | 1. Sum the value of transactions to receivers whose risk is high. |
| 2 | sum\_bad\_guy\_receiver | Valor Saida Favorecido Bad Guy | 1. Sum the value of transactions to receivers who are tagged as BadGuy |
| 3 | sum\_trx | Valor Saida/Entrada Total | 1. Sum the value of outgoing/incoming transactions |
| 4 | cnt\_trx\_sc | Volume Saida/Entrada Total | 1 . Count the number of outgoing/incoming transactions  2. Normalization of the feature between 0 and 1 using percent rank among the population |
| 5 | z\_score\_cnt\_trx | Volume Entrada Historico | 1. Count the number of outgoing/incoming transactions 2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using points 1 and 2. |
| 6 | z\_score\_cnt\_trx\_seg\_country | Volume Saida/Entrada Segmentado Pais | 1. Count the number of outgoing/incoming transactions 2. Compute the average and the standard deviation segmented per sender country 3. Compute ZScore using the points 1 and 2. |
| 7 | z\_score\_sum\_trx\_seg\_country | Valor Saida/Entrada Segmentado Pais | 1.Sum the value of outgoing/incoming transactions 2. Compute the average and the standard deviation segmented per sender country 3. Compute ZScore using the points 1 and 2. |
| 8 | z\_score\_cnt\_dstnct\_rec | Um para Muitos Favorecidos | 1. Count the distinct number of receivers 2. Compute the average and the standard deviation segmented per sender country 3. Compute ZScore using the points 1 and 2. |
| 9 | z\_score\_max\_sum\_trx\_sender | Valor Maximo Unico Originador | 1. Sum the value of incoming transactions  2. Find the max value of point 1  3. Compute the average and the standard deviation for the past 6 months.  4. Compute ZScore using the points 2 and 3. |
| 10 | z\_score\_max\_cnt\_trx\_sender | Volume Maximo Unico Originador | 1. Count the number of incoming transactions  2. Find the max value of point 1  3. Compute the average and the standard deviation for the past 6 months.  4. Compute ZScore using the points 2 and 3. |
| 11 | z\_score\_sum\_hgh\_rsk\_cntry | Valor Saida (/Entrada) Especie Pais Alto Risco | 1. Sum the value of transactions where the sender country s risky.  2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using points 1 and 2. |
| 12 | z\_score\_sum\_trx | Valor Transacao Historico | 1. .Sum the value of outgoing/incoming transactions 2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using points 1 and 2. |

### 5.4.2 Confidence Remittance & Forex

The following sections details the analysis features used in the solution:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Feature Name** | **Display Name** | **Business Description** | **Investigated Entity** | **Comments** |
| 1 | z\_score\_sum\_trx | Valor Saida(/Entrada) Historico | Sum of all outgoing(/incoming) transactions in the current month compared with the customer average for the past 6 months. | sender(/receiver) | only applies for finalized transactions |
| 2 | z\_score\_cnt\_trx | Volume Saida (/Entrada) Historico | Count of all outgoing(/incoming) transactions in the current month compared with the customer average for the past 6 months. | sender(/receiver) | only applies for finalized transactions |
| 3 | z\_score\_sum\_trx\_cash\_seg | Valor Saida (/Entrada) Cash Segmentado | Sum of all outgoing(/incoming) cash transactions in the current month, compared with the average sum of cash transactions of all clients from the same type (Pessoa Fisica/Juridica) for the past 6 months. | sender(/receiver) | only applies for Travelex clients and finalized transactions |
| 4 | z\_score\_sum\_trx\_remittance\_seg | Valor Saida (/Entrada) Rem Segmentado | Sum of all outgoing(/incoming) remittance type transactions in the current month, compared with the average sum of remittance transactions of all clients from the same type (Pessoa Fisica/Juridica) for the past 6 months. | sender(/receiver) | only applies for Travelex clients and finalized transactions |
| 5 | z\_score\_sum\_trx\_ppc\_seg | Valor Saida (/Entrada) PPC Segmentado | Sum of all outgoing(/incoming) PPC type transactions in the current month, compared with the average sum of PPC transactions of all clients from the same type (Pessoa Fisica/Juridica) for the past 6 months. | sender(/receiver) | only applies for Travelex clients and finalized transactions |
| 6 | z\_score\_sum\_hgh\_rsk\_cntry | Valor Saida (/Entrada) Pais Risco | Sum of all customer’s outgoing(/incoming) not cash transactions to or from a high risk country compared with the customer average for the past 6 months. | sender(/receiver) | only applies for finalized transactions |
| 7 | z\_score\_cnt\_dstnct\_rec\_id/z\_score\_cnt\_dstnct\_send\_id | Um para Muitos Favorecidos(/Muitos Para Um) | Unusual number of different counterparties (using "receiver\_id"/”sender\_id”) transacted with in current month compared with the average number of different counterparties for the customer in the last 6 months | sender(/receiver) | only applies for finalized transactions |
| 8 | z\_score\_cnt\_dstnct\_crrncy | Volume Moedas Distintas | Unusual total number of currencies transacted in outgoing(/incoming) transactions to or from risky countries in current month compared with the average for the past 6 months. | sender(/receiver) | only applies for finalized transactions |
| 9 | z\_score\_sum\_rsk\_cstmr | Valor Saida (/Entrada) Cliente Risco | Sum of outgoing(/incoming) transactions to(/from) high risk customer in current month compared with the customer average for the past 6 months. | sender(/receiver) | only applies when counterparty is Travelex client and for finalized transactions. |
| 10 | z\_score\_cnt\_trx\_n\_day | Volume Saida(/Entrada) N Dias | Number of outgoing(/incoming) transactions in current N=5 day period compared with the customer average for the past 6 months. | sender(/receiver) | only applies for finalized transactions |
| 11 | me\_to\_me\_ratio | Me to Me | Unusual ratio of "me-to-me" transactions out of customer total transactions in current month. | sender(/receiver) | only applies for finalized transactions |
| 12 | many\_small\_trx | Estruturacao | A break of one big transaction (amount higher than 1000) to many outgoing(/incoming) small (more than 5) amount transactions in current month. | sender(/receiver) | only applies for finalized transactions |
| 13 | z\_score\_sum\_cash\_hgh\_rsk\_cntry | Valor Saida (/Entrada) Especie Pais Alto Risco | Sum of outgoing(/incoming) cash transactions to or from a high risk country for a customer in current month. | sender(/receiver) | only applies for finalized transactions |
| 14 | z\_score\_cnt\_dstnct\_rsk\_cntry | Volume Paises Risco Distintos | A change in number of high risk countries transacted with in current month compared with the customer average for the past 6 months. | sender(/receiver) | only applies for finalized transactions |
| 15 | new\_customer | Novas Contrapartes | Count of different counterparties that appear this month and did not appear in the preceding 6 months. | sender(/receiver) | only applies for finalized transactions |
| 16 | round\_amounts | Valores Redondos | High volume of outgoing(/incoming) transactions above $100 and with round amounts that is an increment of 1000 in original currency. | sender(/receiver) | only applies for finalized transactions |
| 17 | z\_score\_sum\_cash\_ppc | Valor Saida(/Entrada) Especie para PPC | Sum amount of outgoing(/incoming) transactions in cash to a pre paid card in current month compared with the customer average for the past 6 months. | sender(/receiver) | only applies for finalized transactions |
| 18 | z\_score\_cnt\_cash\_ppc | Volume Saida(/Entrada) Especie para PPC | Number of outgoing(/incoming) transactions in cash to a pre paid card in current month compared with the customer average for the past 6 months. | sender(/receiver) | only applies for finalized transactions |
| 19 | sensitive\_nature | Natureza Sensivel | Unusual ratio of sensitive nature outgoing(/incoming) transactions out of customer total transactions in current month. | sender(/receiver) | only applies for finalized transactions |
| 20 | high\_risk\_areas | Area de Risco | Indicator if the client area is risky | sender(/receiver) | only applies for Travelex client and for finalized transactions. |
| 21 | sensitive\_cnae\_profession | Ocupacao Sensivel | Indicator if the cnae profession is sensitive | sender(/receiver) | only applies for Travelex client and for finalized transactions. |
| 22 | z\_score\_cnt\_cancelled\_trx | Volume Total Cancelado Incomum | Number of all canceled outgoing(/incoming) transactions in the current month compared with the customer average for the past 6 months. | sender(/receiver) | only applies for canceled transactions. |
| 23 | z\_score\_sum\_cancelled\_trx | Valor Total Cancelado Incomum | Sum of all canceled outgoing(/incoming) transactions in the current month compared with the customer average for the past 6 months. | sender(/receiver) | only applies for canceled transactions. |

The following sections details the pseudocode of each the features:

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Feature Name** | **Display Name** | **Pseudocode** |
| 1 | z\_score\_sum\_trx | Valor Saida(/Entrada) Historico | 1. Sum the value of transactions with Finalized status for the current month. 2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using points 1 and 2. |
| 2 | z\_score\_cnt\_trx | Volume Saida (/Entrada) Historico | 1. Count the number of transactions with Finalized status for the current month. 2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using points 1 and 2. |
| 3 | z\_score\_sum\_trx\_cash\_seg | Valor Saida (/Entrada) Cash Segmentado | 1. Sum the value of transactions with Finalized status where the client is Travelex client and product group segment is cash.  2. Compute the average and the standard deviation segmented by client type, for the past 6 months.  3. Compute ZScore using points 1 and 2. |
| 4 | z\_score\_sum\_trx\_remittance\_seg | Valor Saida (/Entrada) Rem Segmentado | 1. Sum the value of transactions with Finalized status where the client is Travelex client and product group segment is remittance.  2. Compute the average and the standard deviation segmented by client type, for the past 6 months.  3. Compute ZScore using points 1 and 2. |
| 5 | z\_score\_sum\_trx\_ppc\_seg | Valor Saida (/Entrada) PPC Segmentado | 1. Sum the value of transactions with Finalized status where the client is Travelex client and product group segment is PPC.  2. Compute the average and the standard deviation segmented by client type, for the past 6 months.  3. Compute ZScore using points 1 and 2. |
| 6 | z\_score\_sum\_hgh\_rsk\_cntry | Valor Saida (/Entrada) Pais Risco | 1. Sum the value of transactions with Finalized status where payment type is not cash and sender country or receiver country is risky.  2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using points 1 and 2. |
| 7 | z\_score\_cnt\_dstnct\_rec\_id | Um para Muitos Favorecidos(/Muitos Para Um) | 1. Count the number of distinct receiver ids transacted with for the current month. 2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using points 1 and 2. |
| 8 | z\_score\_cnt\_dstnct\_crrncy | Volume Moedas Distintas | 1. Count the number of distinct currencies used in transactions with Finalized status to or from a risky country for the current month. 2. Compute the average and the standard deviation for the past 6 months. 3. Compute ZScore using the points 1 and 2. |
| 9 | z\_score\_sum\_rsk\_cstmr | Valor Saida (/Entrada) Cliente Risco | 1. Sum the value of transactions with Finalized status where the counterparty is high risk customer for the current month.  2. Compute the average and the standard deviation segmented by client type, for the past 6 months.  3. Compute ZScore using points 1 and 2. |
| 10 | z\_score\_cnt\_trx\_n\_day | Volume Saida(/Entrada) N Dias | 1. Count the number of transactions with finalized status in the past n days.  2. Compute the average and the standard deviation for the past 6 months. 3. Compute ZScore using the points 1 and 2. |
| 11 | me\_to\_me\_ratio | Me to Me | 1. Create column is\_me\_to\_me created with fuzzy matching of ‘sender\_name', 'receiver\_name’, that indicates whether names in those columns are matching.  2. Count the number of transactions with finalized status defined as me\_to\_me for the current month.  3. calculate the ratio of the number of me\_to\_me transactions out of total transactions for the current month. |
| 12 | many\_small\_trx | Estruturacao | 1. Calculate number of transactions and sum of transactions for current month.  2. Calculate percentile rank over count of transactions for all customers with more than 5 transactions.  3. Calculate percentile rank over sum of transactions for all customers with sum higher than 1000.  4. For all customers with point 3 higher than 0.75 and number of transactions higher than 5, calculate weighted average for points 2 and 3, otherwise 0. |
| 13 | z\_score\_sum\_cash\_hgh\_rsk\_cntry | Valor Saida (/Entrada) Especie Pais Alto Risco | 1. Sum the value of transactions with Finalized status where the sender or receiver country is risky and the transaction type is cash for the current month.  2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using the points 1 and 2. |
| 14 | z\_score\_cnt\_dstnct\_rsk\_cntry | Volume Paises Risco Distintos | 1. Count the number of distinct counterparty countries, where the counterparty country is risky.  2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using the points 1 and 2. |
| 15 | new\_customer | Novas Contrapartes | 1. Define new customer equals true if counterparty has not been transacted with in the past 6 months.  2. Count number of counterparties defined as new customers. |
| 16 | round\_amounts | Valores Redondos | 1. Count the number of transactions with finalized status where the amount above $100 with round amounts in original currency for the current month.  2. calculate the ratio of point 1 out of all transactions in the current month. |
| 17 | z\_score\_sum\_cash\_ppc | Valor Saida(/Entrada) Especie para PPC | 1. Sum the value of transactions with Finalized status where the group segment is PPC and the transactions was made in cash.  2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using the points 1 and 2. |
| 18 | z\_score\_cnt\_cash\_ppc | Volume Saida(/Entrada) Especie para PPC | 1. Sum the value of transactions with Finalized status where the group segment is PPC and the transactions was made in cash.  2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using the points 1 and 2. |
| 19 | sensitive\_nature | Natureza Sensivel | 1. Count the number of transactions with finalized status where the trx\_sensitive\_natures\_risk is ALTO  2. calculate the ratio of point 1 out of all transactions in the current month. |
| 20 | high\_risk\_areas | Area de Risco | Indicator - 1 if client is Travelex client AND ('client\_is\_area\_restricted\_post\_office' is True OR 'client\_is\_area\_restricted\_border’ is True OR 'client\_is\_area\_restricted\_mining’ is TRUE OR client\_is\_area\_restricted\_port is True) , otherwise 0. |
| 21 | sensitive\_cnae\_profession | Ocupacao Sensivel | Indicator - 1 if client is Travelex client AND client\_is\_cnae\_prof\_sensitive is True, otherwise 0. |
| 22 | z\_score\_cnt\_cancelled\_trx | Volume Total Cancelado Incomum | 1. Count the number of transactions where status is canceled for the current month.  2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using the points 1 and 2. |
| 23 | z\_score\_sum\_cancelled\_trx | Valor Total Cancelado Incomum | 1. Sum the value of transactions where status is canceled for the current month.  2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using the points 1 and 2. |

The following sections details the analysis features used in the solution:

|  |  |  |
| --- | --- | --- |
| **No** | **Feature Name** | **Business Description** |
| 1 | **one\_to\_many\_history\_spike** | Unusual number of different outgoing counterparties in the current month compared with the customer average for the past 6 months. |
| 2 | **many\_to\_one\_history\_spike** | Unusual number of different incoming counterparties in the current month compared with the customer average for the past 6 months. |
| 3 | **new\_account\_high\_value** | High value of transactions for new accounts when compared to other new accounts in the history. |
| 4 | **possible\_lob\_changes** | Unusual change in the distribution of source of payments compared to the customer historical profile. |
| 5 | **high\_volume\_new\_loaders** | Unusual total count of new loaders in the current month. |
| 6 | **high\_volume\_new\_beneficiaries** | Unusual total count of new beneficiaries in the current month. |
| 7 | **new\_currency\_activity** | New currency used in the current month with counterpary in high risk country. |
| 8 | **unique\_currency\_history\_spike** | Unusual number of currencies for the current month compared with the average number of currencies for the customer in the last 6 months |
| 9 | **unique\_countries\_history\_spike** | Unusual of number of countries for the current month compared with the average number of countries for the customer in the last 6 months |
| 10 | **tax\_haven\_country\_activity** | Total sum of transactions from/to tax havens in the current month |
| 11 | **high\_risk\_country\_activity** | Total sum of transactions from/to risky countries for transactions in the current month |
| 12 | **new\_country\_activity** | Number of new countries of activity when considering the last 6 months and compared against the population. |
| 13 | **charity\_keyword\_match** | Count of incoming transaction where receiver name or free text matchs with charity related keywords list. |
| 14 | **cryptocurrency\_activity** | Sum of transactions in the current month where bank name or free text matchs with crypto related keywords list |
| 15 | **sof\_value\_spike\_not\_trusted** | Unusual sum of transactions from not trusted loaders for the current month compared with the average sum of transactions from not trusted loaders for the customer in the last 6 months |
| 16 | **sof\_volume\_spike\_not\_trusted** | Unusual count of transactions from not trusted loaders for the current month compared with the average count of transactions from not trusted loaders for the customer in the last 6 months |
| 17 | **unusual\_value\_atm\_withdrawal** | Unusual sum of atm transactions for the current month compared with the average sum of atm transactions for the customer in the last 6 months |
| 18 | **unusual\_volume\_round\_amounts** | High volume of transactions above $1000 and with round amounts. |
| 19 | **risky\_outgoing\_activity\_ratio** | Ratio between the sum of transactions with method POS/PON/ATM and the total outgoing value of transactions for the current month |
| 20 | **same\_day\_distinct\_atm\_terminal** | Maximum number of same day distinct atm terminals for the current month. |
| 21 | **high\_value\_atm\_withdrawals** | High value of atm withdrawals for the country and high ratio of atm to total outgoing value. |
| 22 | **high\_value\_ipsp\_ewallet** | Sum of transactions where initiator vertical is IPSPs or eWallets. |
| 23 | **high\_volume\_receiver\_mismatch** | Count of transactions where IPSPs mismatches receiver names. |
| 24 | **high\_value\_risky\_mcc** | Unusual total sum of transactions in the current month where mcc code defined as risky. |

The following sections details the pseudocode of each the features:

|  |  |  |
| --- | --- | --- |
| **No** | **Feature Name** | **Pseudocode** |
| 1 | **one\_to\_many\_history\_spike** | 1. Create column aux\_beneficiary\_id as (if transaction\_method equals PON assing beneficiary\_bank\_account\_number, if transaction\_method equals MAP assing beneficiary\_id\_map, else NULL) 2. Count the distinct number of aux\_beneficiary\_id in the current month. 3. Compute average and standard deviation of point 2 for the ah's past 6 months. 4. Compute ZScore using the points 2 and 3. |
| 2 | **many\_to\_one\_history\_spike** | 1. Count the distinct number of loader\_id in the current month when trusted\_loader equals false. 2. Compute average and standard deviation of point 2 for the ah's past 6 months. 3. Compute ZScore using the points 2 and 3. |
| 3 | **new\_account\_high\_value** | 1. Define new customer equals true if ah is in the first three months of activity (if has at least one transaction in that month). 2. Define total turnover as (sum of not trusted incoming activity plus sum of outgoing activity minus sum of outgoing activity where method is partner charge) 3. Compute average and standard deviation for all new customers in the current month. 4. Compute ZScore using the points 2 and 3. |
| 4 | **possible\_lob\_changes** | 1. Compute the sum of map, the sum of payment services, the sum of mass payout and the total sum of payments. 2. Compute the ratio of each type of payments as (sum of map / sum of total). 3. Compute the averages of each ratio defined in point 2 for the ah's past 6 months. 4. Compute the absolute difference between the current month ratios and the past months averages 5. Compute the maximum between the differences in point 4. |
| 5 | **high\_volume\_new\_loaders** | 1. Create list of distinct loaders for the current month for the ah. 2. Create list of distinct loaders for the past 6 months for the ah. 3. Create list of loaders which are in point 1 and not in point 2. 4. Count the number of loaders in point 3. |
| 6 | **high\_volume\_new\_beneficiaries** | 1. Create column aux\_beneficiary\_id as (if transaction\_method equals PON assing beneficiary\_bank\_account\_number, if transaction\_method equals MAP assing beneficiary\_id\_map, else NULL) 2. Create list of distinct beneficiaries for the current month for the ah. 3. Create list of distinct beneficiaries for the past 6 months for the ah. 4. Create list of beneficiaries which are in point 2 and not in point 3. 5. Count the number of beneficiaries in point 4. |
| 7 | **new\_currency\_activity** | 1. Create list of distinct currencies from payments when loader billing country is high risk for the current month. 2. Create list of distinct currencies from transactions when transaction destination is high risk for the current month. 3. Create list of distinct currencies from the points 1 and 2. 4. Create list of distinct currencies in point 3 for the past 6 months. 5. Create list of currencies which are in point 3 and not in point 4. 6. Count the number of currencies in point 5. |
| 8 | **unique\_currency\_history\_spike** | 1. Count the number of distinct currencies for the current month of the ah. 2. Compute the average and the standard deviation for the past 6 months of the ah. 3. Compute ZScore using the points 1 and 2. |
| 9 | **unique\_countries\_history\_spike** | 1. Count the number of distinct countries for the current month of the ah. 2. Compute the average and the standard deviation for the past 6 months of the ah. 3. Compute ZScore using the points 1 and 2. |
| 10 | **tax\_haven\_country\_activity** | 1. Sum the amount of usd transacted to/from countries considered tax havens for the current month of the ah. |
| 11 | **high\_risk\_country\_activity** | 1. Sum the amount of usd transacted to/from countries considered high risk for the current month of the ah. |
| 12 | **new\_country\_activity** | 1. Create list of distinct countries transacted with in the current month. 2. Create list of distinct countries transactied with in the past 6 months. 3. Create list of countries which are in point 1 and not in point 2. 4. Count the number of countries in point 3. |
| 13 | **charity\_keyword\_match** | 1. Count the number of transactions that match charity keywords using the field glps\_receiver\_name. 2. Count the number of transactions that match charity keywords using the field payment\_description. 3. Sum the values in point 1 and 2. |
| 14 | **cryptocurrency\_activity** | 1. Sum the usd value of transactions that match crypto keywords using the field beneficiary\_bank. 2. Sum the usd value of transactions that match crypto keywords using the field outgoing\_payment\_description. 3. Sum the values in point 1 and 2. |
| 15 | **sof\_value\_spike\_not\_trusted** | 1. Sum the usd value of transactions where the loader is not in trusted loader list or the payment method is not mass payout for the current month. 2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using points 1 and 2. |
| 16 | **sof\_volume\_spike\_not\_trusted** | 1. Count the number of transactions where the loader is not in trusted loader list or the payment method is not mass payout for the current month. 2. Compute the average and the standard deviation for the past 6 months.  3. Compute ZScore using points 1 and 2. |
| 17 | **unusual\_value\_atm\_withdrawal** | 1. Sum the usd value of transaction where transaction method is ATM for the current month. 2. Compute the average and the standard deviation for the past 6 months. 3. Compute ZScore using points 1 and 2. |
| 18 | **unusual\_volume\_round\_amounts** | 1. Count the number of transaction above $1000 with round amounts for the current month for incoming transactions where not trusted loader. 2. Count the number of transaction above $1000 with round amounts for the current month for outgoing transactions. 3. Sum the values in point 1 and 2. |
| 19 | **risky\_outgoing\_activity\_ratio** | 1. Sum the usd value of the following transaction methods for the current month: POS, PON, ATM. 2. Sum the usd value of the outgoing transactions. 3. Compute the ratio point 1 / point 2. |
| 20 | **same\_day\_distinct\_atm\_terminal** | 1. For each day in the current month compute the distinct number of atm terminals used. 2. Compute the maximum value of the values in point 1. |
| 21 | **high\_value\_atm\_withdrawals** | 1. Sum the usd value of transactions where transaction method equals ATM; for the current month. 2. Compute the average and standard deviation of ah for the same billing country in the current month. 3. Compute ZScore using points 1 and 2. 4. Compute the following ratio as (point 1 / sum of outgoing usd value for the current month). 5. Multiply the values of points 3 and 4. |
| 22 | **high\_value\_ipsp\_ewallet** | 1. Sum the usd value of payments where initiator vertical equals IPSPs of eWallets. |
| 23 | **high\_volume\_receiver\_mismatch** | 1. Count the number of transactions where initiator vertical equals IPSPs and the glps receiver name mismatches the ah's name. |
| 24 | **high\_value\_risky\_mcc** | 1. Sum the usd value of transactions where mcc code is in risky mcc list. |

# Outcomes Analysis

This chapter confirms that the model is functioning and performing appropriately.

## Testing

In order to approve the model, several analysis runs were conducted and tested both by the customer and ThetaRay team, once achieved sufficient results both from effectiveness and efficiency perspectives, the model was approved to be executed in production.

### Back testing

Machine Learning techniques and especially Unsupervised Learning are new by nature and represent a new approach, for most use cases, the old methods were concluded as inappropriate, inefficient, ineffective or were not specific for the use case.

Considering the above, in most implementations of ThetaRay, and specifically for Correspondent Banking, Trade Finance in particular or cross border payments in general the comparison to previous methods is not always applicable.,

In most cases, the quality was measured without comparison to old methods. (The customer tested the results against previously known cases, see documentation for the results analysis done: Sensitivity Testing.

### Benchmarking

ThetaRay solution provides eight (8) different algorithms for benchmarking. Each algorithm can generate separate outputs which can be used to form the model performance process. FIs can also use either internal developed models or external vendor models to benchmark with ThetaRay solution, based on their own model risk management policies and standards.

### Establish preliminary expectation of the testing output

Customer Compliance and business teams measured the performance of the ThetaRay solutions.

**Summary of the tests performed and the results:**

**\*\* Travelex’s team to include the summary of the sample testing/ results/ conclusions**

# Model Implementation Accuracy and Ongoing Monitoring Plan

The objective of this section is to ensure the transfer of the developed model into a production environment is robust and the required ongoing monitoring activities are well defined and completely understood from both qualitative and quantitative perspectives.

## Model performance consistency in the implementation

**\*\* As referenced in section 5, the results during the development phase were as follows:**

**\*\* Travelex and ThetaRay to document the results based on investigation of sampled alerts and conclusions.**

## Ongoing Monitoring Plan

This chapter is designed to document the ongoing testing for the model ongoing performance assessment, both from technical as well as operational aspects.

The testing should be comprehensive and test a variety of scenarios (e.g., model log files should be reviewed to ensure the model consistency).

ThetaRay analysis workflows are automated through a built-in orchestration component which is based on Apache Airflow. Each workflow consists of multiple tasks which cover activities that include - data upload, data pre-processing and feature engineering, algorithms execution and alert generation. The state of each execution, start and completion times are recorded and made available through a web-based user interface to the administrator of the system. Moreover, the logs of each step in the workflow are recorded and are made available through the same user interface.

To enable proactive notifications in case of execution failures, ThetaRay bundles a monitoring component based on OpenSearch which is pre-integrated to the workflow automation component - in case of execution failures (either technical or functional) an alert is triggered and can be delivered to destinations such as Slack or e-Mail if needed.

Operations such as data upload, derived datasets write operations record their execution metrics into a ‘runs tracking component’ that is based on MLFlow. Metrics are recorded as a result of API calls performing the above operations and can be back referenced to the associated workflow step generating them.

Technical information and metrics are the following:

|  |  |
| --- | --- |
| **Recorded info** | **Applicable to** |
| Total records written | Upload |
| Corrupted records (failed parsing) | Upload |
| Rejected records (failed validation) | Upload |
| Total records written | Dataset write |

### Baseline Settings

Drift is a phenomenon that may occur when the fundamentals on which a model is built have changed. In the case of ThetaRay’s analytics, profiling of normality lies at the base of the model. These changes from the baseline (the time period on which the model was trained on) could be caused by a change of operational activity (expanding the activity in risky countries, general growth in customers and transactions) or behavioral (customer's financial activity profile). When normality changes in the real world, it is imperative that this is identified to identify preface drift, which is a phenomenon that may occur when the fundamentals on which a model is built have changed. ThetaRay offers a recommended ongoing monitoring methodology to identify drift, as described below as a template. The final drift monitoring will be decided with the customer: what tests to perform, what output to log, what results should trigger investigation or stop analysis from publish new alerts to the IC, etc.

Assuming normality has not changed, it is expected that over time the percentage of anomalous observations is steady. Hence, if that percentage is significantly bigger or smaller than in the past, it is potentially considered as drift in normality. To enable drift identification, a baseline is set for the expected anomaly percentage and for a sensitivity threshold. The anomaly percentage is calculated as the average percentage of anomalies in a batch. The batches partaking in this calculation are those used in the training of the model. The sensitivity threshold is calculated as StdDev \* N, where StdDev is the standard deviation of (anomaly count/batch) per all batches used in training/testing the model, with the default N=2 (this parameter may change per customer request.

### Drift Identification

Once the baseline is established, it is used to evaluate potential drift in new batches. As a deviation from the baseline may be anecdotal, drift is declared if X of Y (which is by default three of five) consequent batches are identified as deviating from the baseline. As optimization, as soon as three deviated metrics are logged, drift is declared without waiting for the completion of all five batches. Once drift is declared, the batch count is reset.

### Visual Indication of Drift

An identified drift is visually available through the dedicated Jupyter Notebooks.

### Drift Identification Test Example

The following tests are examples of ongoing monitoring of feature calculation, algorithm performance and their suitability to the ongoing data feed. The developers (of the customers as TR data scientists) can decide in which deviation from baseline the periodic run will stop until further investigation.

For each metric in them, the following process is applied:

* Calculation of Metric;
* Assessment of Metric (Pass/Fail);
* Following failed validation, an investigation is conducted into the root cause – e.g., data drift, concept drift, or other underlying reasons;
* Following the investigation, any issues identified are addressed according to their domain, for example:
  + Data drift may require retraining the model.
  + Concept drift may require a re-design of the feature set.
* Type 1 Tests - Data Frame Distributions
  + Summary Statistics of Analysis Features across Data frame

This test calculates summary statistics of values for each analysis feature for the entire data frame compared to past periods. It allows us to identify any large and sudden change in analysis feature distributions which may indicate a data drift.

In this example, the average of each feature is calculated.

The test may also run on other summary statistics, including:

* Percentiles
  + Minimum - 0%
  + 25%
  + Median - 50%
  + 75%
  + Maximum - 100%
* Standard Deviation
* Interquartile Range
* Range



**Example Metric:**

* Compute the Z score of the test statistic compared to the past 12 months.
* If the absolute value of the Z score exceeds a benchmark value (such as two – matching 95% confidence) then the metric is triggered and considered as failed validation
* Characteristics Stability Index of Analysis Features across Data frame

Characteristics Stability Index (CSI) is similar to Population stability Index (PSI). It is applied on each analysis feature to quantify how much has the distribution of the analysis feature changed compared to the data used in the training.

Unlike PSI where there is a rule of thumb on the interpretation of the index, the thresholds to trigger further investigation on each analysis feature need to be established for each feature using the historical record. This is due to the wide range of behavior for each feature. Some are more volatile whereas some are more stable.

**Example:**

CSI Results for features A,B,C,D,E,F,G,H,I:

A: 0.007228

B: 0.008873

C: 0.009127

D: 0.002664

E: 0.007325

F: 0.009035

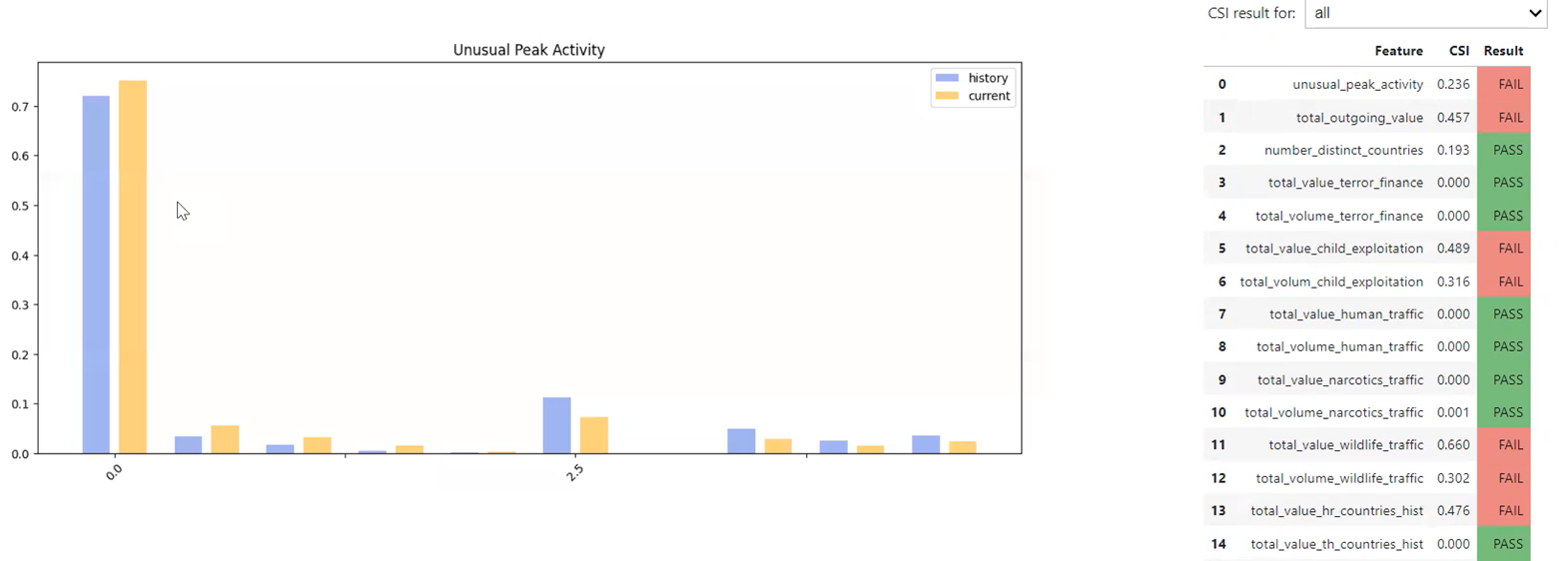
G: 0.006423

H: 0.006295

I: 0.242430

Across these features, feature I had a significant change whereas all of the others had very little change. This change could be due to differences in data or behavior.

**Example Metric:**

* If the value of the CSI score exceeds a benchmark value (which can be determined per feature), then the metric is triggered and considered as failed validation.

2. Type 3 Tests - Anomaly Composition

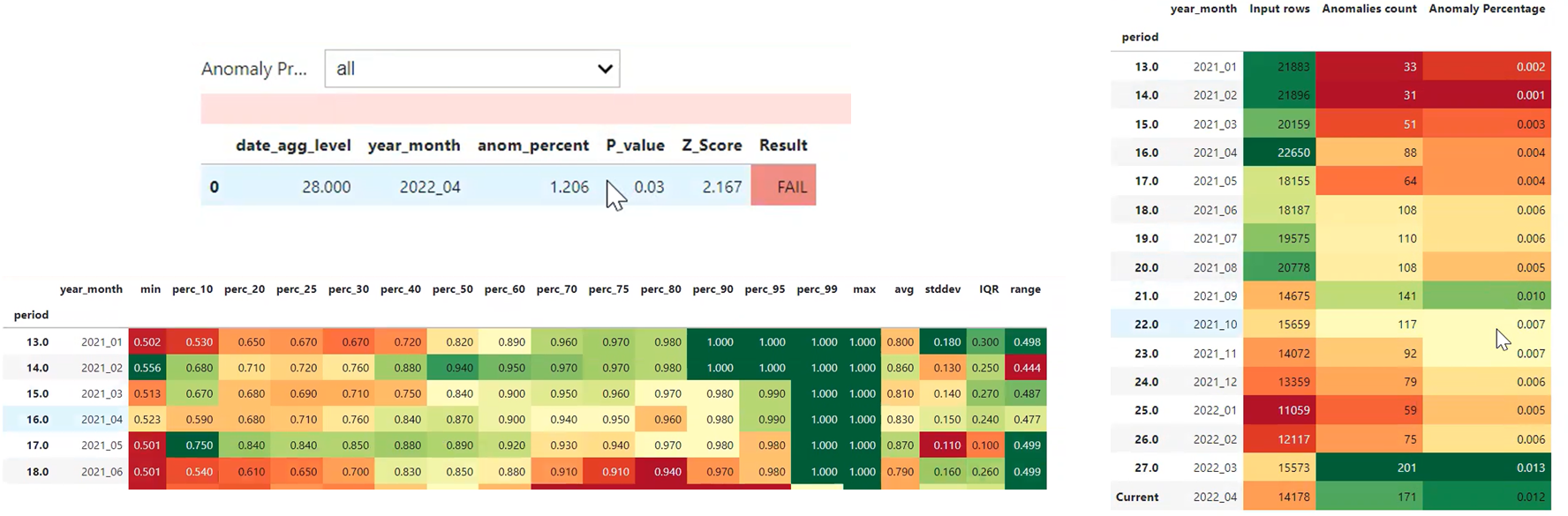
* + Number and Percentage of Anomalies per month

This test calculates the percentage of rows each month which are identified as anomalous and compared to previous months without taking into account its population segment.

Every anomaly identified by the ThetaRay algorithms is assigned an abnormality score in the range of 0-100. Higher scores indicate more abnormal behavior.

Population stability Index (PSI) is a measure of how much the measure of a population has changed over a period of time. PSI can be applied at an anomaly score level, by binning the scores.

**Example:**



**Formula:**



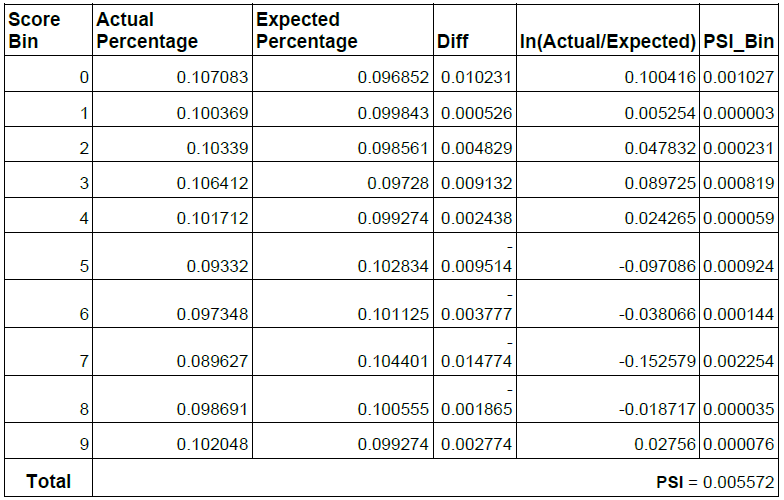
* “Actual” – Measured Values
* “Expected” – Projected Values based on past observations
* Sum over bins of feature values, for example [0, 0.2, 0.4, … , 4.6, 4.8, 5.0]

Suggested interpretation of results:

|  |  |
| --- | --- |
| **PSI Range** | **Interpretation** |
| PSI <= 0.1 | Little change (no action required) |
| 0.1 < PSI <= 0.25 | Moderate change |
| 0.25 < PSI | Significant change |

This test computes the PSI for anomaly scores on the new month/batch compared to a previous time period.

**Example:**



The final PSI value for anomaly score is 0.005572 which indicates high stability due to its low value.

**Example Metric:**

* If the value of the CSI score exceeds a benchmark value (such as 0.25), then the metric is triggered and considered as failed validation.

### Model ongoing performance tests

This chapter should document thresholds or triggers in drift analysis/performance tests that would require additional actions to ensure the model fundamentals remain unchanged with statistical and business rationale. And the model output is continuing to support the business decisions.

Drift analysis or drift identification test includes several analyses or tests such as data frame distribution test, anomaly distribution test, anomaly composition test to make sure that raw data quality/normality or anomaly distribution is consistent with that in model training data. The thresholds or triggers in these tests should be supported from both a statistical standpoint as well as from the line of business that uses the output. Thresholds should exist for each test performed in section 5.2.1 or rationale should be provided for why a threshold is not necessary.

# Model Governance and Change Management

This section is to provide information on the governance and model control to ensure the overall soundness of the proper supervision of the model performance and the accuracy/consistency/integrity of the model output.

During the ongoing Governance procedures, committees and Management meetings, The Travelex’s Compliance will have clear view on the performance of the ThetaRay solution for the specific use case, recommended KPIs to track:

* SARs identified and their severity
* Stats on detection worthy versus false positive cases
* Number of new and unknown patterns identified per period
* The risk profile of the cross-border activity

These stats will enable minimum understanding to re model performance and effectiveness.

Additional measurements can and will serve Travelex’s management with better business & operational decisions:

* Efficiency- time to resolution of cases
* Efficiency- number of alerts per period
* Risk management/ business relationship- portion of alerts per respondent bank, for example

The above-mentioned measurements will enable a transparent view of model output, timely identification of potential places for improvement and validation of risk coverage.

## Model security and access control

The ThetaRay platform allows for the creation of user permissions which can be configured and customized by each institution based on the specific roles defined pursuant with internal policies and procedures. Please request a copy of the latest user guide for more information.

**\*\* Within Travelex implementation, define the user roles (YH: preferable to detail by functions/ roles and keep a detailed list with names in operational processes and not update this document upon each change in roles, addition of analysts to the team etc.)**

## Model change management

Each institution should be responsible for defining it’s own model change management procedures for updates to configurable parameters, including roles, responsibilities and review and approvals before model changes are released into production. ThetaRay is responsible to update the customer for any modifications/ configurations and changes to the software that impact the performance of the model, upgrades and new functionalities included. Table updates, such as adding a country to a risk list or a specific keyword, does not mean a model change. Travelex will manage versions of the model, as model may be subject to changes mainly due to:

* Regulatory updates
* Business activity update
* Timely evaluation of features performance.

# Algorithms description

\*\* See separate documentation: “A Survey of ThetaRay Algorithms,” March 31, 2022.

**About ThetaRay**

Developed by two of the world’s preeminent mathematicians, ThetaRay brings a transformational approach to big data analytics, isolating problems buried beneath big data and pinpointing opportunities concealed within traditional systems.   
Drawing inspiration from the manner in which complex systems operate in nature, ThetaRay’s patented algorithms grow more intelligent over time, yielding objective results in seconds.

Revolutionary in its approach, ThetaRay lets math discover meaning in the data without making any pre-assumptions. Its rule-free approach pinpoints items that matter now and, in the future, mapping them in context, time, and location.

With ThetaRay, Clients obtain measurable value in days and achieve full deployment in a matter of weeks. From that point on, ThetaRay’s system runs automatically, without intervention from Client personnel and with the continuous support of an adaptable platform.

ThetaRay helps financial institutions, cybersecurity divisions, and critical infrastructure businesses make giant strides in managing risk and generating new growth. With ThetaRay’s solution, organizations become as resilient as these times require: safeguarding assets, recovering from setbacks, and capturing future growth amid continuous change.