Artificial Intelligence in Regulatory Transaction Anomalies Detection

(An Experimentation)

Date: 08-May-2026

CRÉDIT AGRICOLE
CORPORATE & INVESTMENT BANK

Abstract

Financial institutions (FI) operating in foreign exchange (FX) markets are facing increasing challenges in detecting abnormal market behavior within fast-moving and complex trading environments. In addition, FI's must ensure that derivative transaction data submitted to Monetary Authority of Singapore (MAS) regulators are complete, timely and accurate to support market transparency and effective regulatory oversight. Accurate and timely reporting enables regulators to monitor systemic risks and ensures FI's meet their regulatory obligations.

The Securities and Futures (Reporting of Derivatives Contracts) Regulations 2013 sets out the reporting requirements for Over-the-counter (OTC) counterparties for OTC derivatives. The Guidelines to the Securities and Futures (Reporting of Derivatives Contracts) Regulations 2013 [SFA 06A-G01] set out MAS' expectations for reporting entities complying with the OTC derivatives reporting requirements.

guidelines-to-the-securities-and-futures-reporting-of-derivatives-contracts-regulations_31-may-24_3.pdf

To comply with the OTC derivatives reporting requirements, Crédit Agricole CIB currently operates rule-based monitoring tools that are effective in validating data completeness, conformity and consistency and enforcing predefined reporting checks. Building on these controls, this experiment explores the application of Artificial Intelligence deep learning techniques-specifically an autoencoder model-to complement existing capabilities by improving data quality monitoring and anomaly detection. Unlike rule-based checks that focus primarily on individual fields, the autoencoder model analyses patterns across multiple attributes simultaneously, allowing detection of outliers and unusual behavior through cross feature relationships and pattern deviations.

The pilot implementation focuses on FX derivatives, where data volume and transaction complexity present immediate monitoring challenges. The approach will be extended to other derivative products to further enhance reporting data quality across asset classes.

The study demonstrates how Artificial Intelligence (AI)-driven methods can complement existing reporting frameworks by enhancing data quality assurance while reducing operational burden. These findings provide a foundation for developing scalable monitoring capabilities that support accurate regulatory submissions and strengthen risk management and compliance effectiveness with production trading environments.

Table of Contents

1. Challenges for FI in FX market anomalies

Singapore is the third largest FX centre globally after London and New York and is the largest in Asia Pacific (MAS, 2025). In October 2025, outright forwards and FX swaps in Singapore accounted for US$3.7 trillion and US$13.8 trillion in total monthly volume respectively (SFEMC, 2025). This immense scale and complexity present significant challenges for Financial institutions (FI) in detecting and ascertaining market anomalies that could impact trading decisions, risk management, and regulatory compliance. The inherent characteristics of FX markets create a unique environment where traditional anomaly detection methodologies face substantial limitations, requiring sophisticated approaches to distinguish between legitimate market behaviour and genuine anomalies.

Unlike centralized markets, FX trading occurs across decentralized venues, instruments and counterparties. As a result, distinguishing legitimate market movements from genuine irregularities is complex, and traditional rule-based detection methods often struggle to adapt to rapidly changing market conditions.

1.1. Market Microstructure Complexity

Singapore is a major global FX trading centre, with transactions executed across multiple venues, counterparties, and products rather than a single exchange. This structure results in differences in pricing, liquidity and execution patterns, making it challenging to establish consistent baselines for anomaly detection.

Market activity also varies across trading sessions and reacts quickly to global economic and geopolitical developments. As a result, legitimate market movements may appear anomalous when assessed using static rules, requiring detection approaches that can adapt to changing and interconnected market behavior.

1.2. Data Quality and Baseline Establishments

The effectiveness of anomaly detection systems is fundamentally dependent on data quality, yet FX markets present unique data integrity challenges that undermine detection capabilities. Financial institutions frequently encounter inconsistent data formats across different liquidity providers and trading venues, creating standardization obstacles that compromise analytical consistency. The decentralized nature of FX markets means that trade reporting standards vary significantly across jurisdictions and counterparties, resulting in incomplete or delayed data transmission that can mask or create false anomalies.

These challenges are further exacerbated by fragmented legacy infrastructures, where disparate technology platforms create significant barriers to unified data analysis and real-time surveillance capabilities, that were not designed for comprehensive data aggregation and real-time analysis. The resulting data quality issues including missing timestamps, inconsistent trade identification protocols, and varying precision in price reporting create noise that significantly increases false positive rates in anomaly detection systems.

Even when data quality reaches acceptable standards, establishing reliable baselines for normal market behaviour remains problematic due to the inherent volatility and regime-dependent nature of FX markets. Macroeconomic factors such as central bank interventions create legitimate but sudden market movements that appear anomalous when measured against historical patterns. Economic event-driven volatility, including monetary policy announcements and geopolitical developments, can cause rapid shifts in exchange rates and trading volumes that exceed traditional statistical thresholds without representing genuine market anomalies.

Cross-currency correlations introduce additional complexity, as anomalies in one currency pair can propagate through related pairs, creating cascading effects that complicate root cause analysis. The interconnected nature of FX markets means that what appears to be an anomaly in a specific currency pair may reflect legitimate arbitrage activities or risk management strategies across multiple instruments. Moreover, structural regime changes in market behavior, driven by evolving monetary policies or shifts in global economic conditions, cause baseline decay that renders historical patterns inadequate for current anomaly detection.

These challenges collectively demonstrate that effective FX market anomaly detection requires sophisticated methodologies capable of adapting to dynamic market conditions while maintaining sensitivity to genuine irregularities that could indicate operational failures, market manipulation, or systemic risks.

2. Overview of The Experiment

The objective of this experiment was to assess whether AI techniques such as deep learning could improve the detection of anomalies and data quality issues in FX derivatives reporting by learning patterns of normal trading behavior across multiple transaction attributes. Unlike conventional technology projects where requirements are fully defined upfront by business teams, this initiative followed an experimental approach. Given the emerging nature of AI solutions, requirements evolved iteratively as models were developed and tested.

The technology team worked closely with the business stakeholders to demonstrate the model output progressively. Workshops were conducted iteratively with business stakeholders to gather use cases and real scenarios causing anomalies. Insights from these sessions were continuously incorporated to refine model logic and improve detection accuracy. This collaborative process allowed practical operational insights to shape model development while enabling business users to better understand AI-driven monitoring capabilities.

This approach differs from traditional project delivery by emphasizing rapid experimentation, learning cycles, and feedback incorporation rather than fixed requirement delivery. It enabled faster identification of feasible use cases and ensured that solutions developed addressed real operational challenges rather than theoretical requirements.

2.1. Functional Perspective-Business Objectives

From a functional standpoint, the experiment aimed to enhance existing monitoring capabilities supporting regulatory reporting obligations. Current systems are effective at enforcing rule-based checks and ensuring transaction data completeness; however, complex cross-field inconsistencies or emerging abnormal patterns may not always be detected through predefined rules. The experiment therefore sought to determine whether machine learning models could complement existing controls by:

- Improving detection accuracy for unusual transaction patterns
- Identifying inconsistencies across multiple transaction attributes simultaneously
- Supporting timely and accurate regulatory submissions

Business users evaluated model outputs during pilot testing and provided feedback to refine alert relevance and usability.

2.2 Technical Perspective-Business Objectives

From a technical perspective, a typical Machine Learning (ML) experiment cycle requires historical data to be partitioned into training, validation, and testing sets. The training and validation sets serve the purpose of tuning model parameters and optimizing performance, while the testing set provides an ultimate, unseen evaluation of model effectiveness. This partitioning strategy ensures that we avoid look-ahead bias and data-snooping issues[1] that could compromise the integrity of performance assessments.

[1] Look-ahead bias refers to the mistake of using future information to make a prediction or assessment on current time. A common example would be to use this year's data to determine last year's anomalies. Data-snooping bias refers to overfitting parameters to random ethereal market patterns.

Beyond standard model parameters, the experiment itself requires careful parameterization. One critical experimental parameter involves empirical validation against known anomalies. Specifically, we must determine whether the model can successfully identify anomalies that have already been labelled as such through domain expertise or prior analysis. Furthermore, we need to assess whether we can identify and understand the structural characteristics of these anomalies to enable the construction of additional test cases that enhance model robustness.

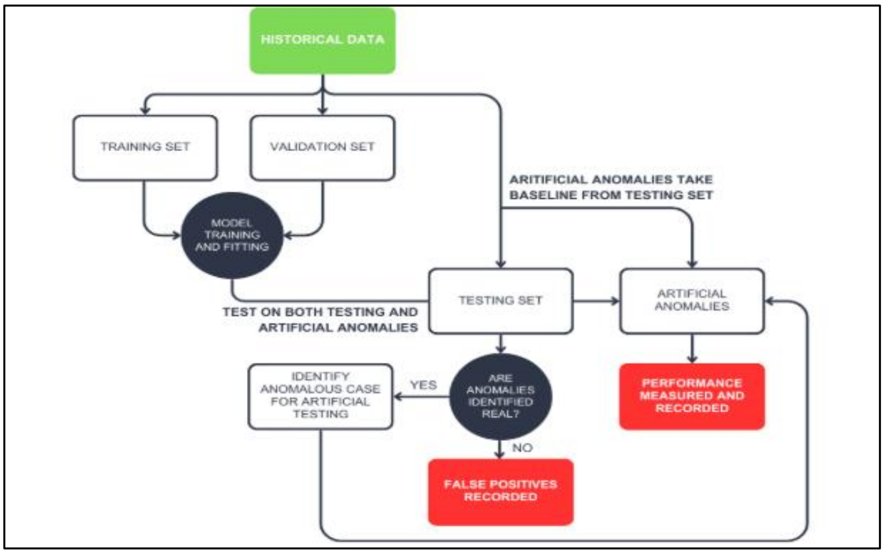


Figure 1: Experimentation Flow

The end-to-end experimentation framework, built around the traditional ML experiment cycle, creates a systematic approach to model development and validation. The key principle underlying this framework is that benchmarks for detection rates on artificial anomaly cases must be established from baselines derived from the testing set to avoid circular bias during iterative training cycles. It is imperative to ensure that model parameters are tuned specifically for fitting normality patterns rather than being optimized toward flagging these artificial test cases. This approach prevents the model from overfitting known anomalies while maintaining its ability to generalize to previously unseen irregularities. The framework enables iterative refinement of both feature engineering approaches and training methodologies while maintaining rigorous evaluation standards that preserve the independence between model optimization and anomaly detection validation.

2.3 Iterative Anomaly Discovery and Case Development

As we iterate through feature engineering phases and successive training cycles, the framework facilitates the identification of additional anomalous cases through both automated detection and collaboration with domain experts within the financial institution. This iterative process begins with relatively straightforward anomaly types that can be identified through basic data inspection. For example, simple data entry errors such as incorrect currency pair orientations, where USD/EUR trades are erroneously recorded as EUR/USD, can be detected through elementary validation checks on traded exchange rates that fall outside reasonable market ranges.

However, as the experimentation progressed and our understanding of the data deepened, we uncovered more sophisticated patterns involving relationships between counterparties, clearing houses, market and transaction characteristics. This enhanced pattern recognition capability allowed for the development of artificial test cases that are both more complex and more representative of realistic market scenarios. These advanced test cases provide a more comprehensive evaluation of model performance across diverse anomaly types that may occur in production environments.

2.4 Model Selection and Architecture

2.4.1 Selection Phase

The experimentation phase evaluated multiple unsupervised and semi-supervised learning approaches to identify the most suitable methodology for FX market anomaly detection. The candidate models included Autoencoders, One-Class Support Vector Machines (OCSVM), and Isolation Forest algorithms, each offering distinct advantages for different aspects of the anomaly detection problem. Through systematic evaluation across various data characteristics and anomaly types, the Autoencoder architecture demonstrated superior performance when handling sparse data matrices and proved effective across both simple and complex anomaly patterns. The Autoencoder also handles multicollinearity well as it compresses redundant information in lower-dimensional representations compared to Isolation Forest that relies on random selection of features that may potentially reduce learning diversity of the ensemble, and the OCSVM that particularly faces distortion in decision boundary in high-dimensional collinear space. This performance advantage outweighed the inherent risks typically associated with deep learning structures, such as interpretability challenges and computational complexity.

2.4.2 Autoencoder Structure

The selected Autoencoder implementation follows a Multi-Layer Perceptron architecture incorporating latent dimensional representations designed to capture underlying patterns of normal market behavior. The fundamental principle underlying this approach relies on reconstruction error analysis, where the discrepancy between input transaction features and their reconstructed outputs serves as the primary mechanism for anomaly scoring. Transactions that exhibit high reconstruction errors indicate deviations from learned normal patterns, suggesting potential anomalous behavior. The architecture of the Autoencoder was built with PyTorch.

FX transaction data encompasses both categorical variables such as currency pairs and counterparty identifiers, and numerical features including notional amounts and pricing data, a specialized dual-head Autoencoder architecture was implemented. This design incorporates two distinct reconstruction heads: one optimized for categorical feature reconstruction utilizing Binary Cross-Entropy with Logits Loss, and another designed for numerical feature reconstruction employing Huber Loss to provide robustness against outliers. Despite this dual-head structure, both reconstruction pathways share common input layers and latent dimensional representations, ensuring that the model learns unified representations of transaction normality while accommodating the distinct statistical properties of different feature types.

This architectural approach enables the model to simultaneously process mixed data types while maintaining the coherence of learned normal patterns across all transaction attributes, providing a comprehensive foundation for detecting anomalies that may manifest across either categorical or numerical dimensions, or combinations thereof.

2.4.3 Key Model Parameters

Hyperparameters were optimized using Bayesian Optimization via Optuna for efficient convergence. While many hyperparameters were tuned, the optimization process found encoding dimensions, activation layer, and batch size to be the most important model hyperparameters. It should be noted that encoding dimensions and batch sizes are highly dependent on input data dimensionality, while the Exponential Linear Unit activation function was identified as the most efficient activation layer.

2.5 Feature Selection and Engineering

The experimental approach to feature selection was an idea-first stepwise approach where new features were engineered (or not) and added to the list of input features to the model. Eventually, the selection boiled down to 3 primary categories:

1. Numerical features focused on price-related variables including notional amounts, exchange rates, and tenor calculations.
2. Categorical features that define trade structure and execution context such as event types, delivery mechanisms, locations, and directional indicators.
3. Regulatory/operational indicators include, but not limited to clearing status, central counterparty involvement, and collateral portfolio classifications.

The experiments also deliberately adopted a minimal feature engineering strategy, guided by both practical deployment considerations and theoretical ML principles. The decision was motivated by key factors that align with the objectives of creating a product-ready system suitable for FI environments.

A primary consideration involved maintaining feature interpretability for business end users who must understand and validate model outputs in production environments. The introduction of engineered features that do not correspond to original trade attributes creates additional complexity for business stakeholders who need to interpret anomaly alerts and take appropriate action. To ensure transparency with regulators, preserving original feature dimensions was strongly considered.

Another strong case was made here for the Autoencoder. By preserving original feature dimensions, the selected model was required to learn key representations in the FX space without additional features. One of which is the linearity in:

$$Leg\ 2\ Amount = FX\ Rate * Leg\ 1\ Amount$$

In our testing, we were unable to get the One Class SVM and Isolation Forest model to consistently learn this representation.

Scaling numerical features, however, proved to be challenging and key in each experimented model's ability in discerning between the vast number of currencies and pairing. We further explore this challenge and solution in 3.4 Standardizing values across currency pairs and currencies.

3. Challenges Faced

3.1 Technical Challenges

Pang, Shen et al. (2020) outlines 6 primary challenges for Deep Anomaly Detection tasks:

1. Low anomaly detection recall rate: facing false positives
2. Anomaly detection in high-dimensional and/or not-independent data
3. Data-efficient learning of normality/abnormality
4. Noise-resilient anomaly detection
5. Detection of complex anomalies
6. Anomaly explanation

The Autoencoder covers challenges 1, 2, 4, 5, lacking primarily in 3 and 6.

Challenge 3 arises with the lack of data, specifically labelled data. Without labels on what is anomalous and what is not, the risk that a small but arbitrary amount of anomalous data may or may not be ingested in training data is accepted. The risk is also managed by avoiding a Variational Autoencoder that imposes probabilistic structure and priors; a regular Autoencoder will not fit the potential anomalous data into latent representations.

Challenge 6 was addressed by ranking models (Pang, Shen, Cao, & Hengel, 2020). The assumption is that there exists an observable ordinal variable that captures some data abnormality in a function τ (x; ϴ). In our implementation, this ordinal variable is derived from the reconstruction errors, where each feature's "wrongness score" is ranked against its training distribution, enabling both anomaly detection through percentile thresholds and explainability through feature-level rankings.

3.2 Exchange rates as a model feature

Exchange rates are inherently non-stationary because they exhibit trends, cycles and random walks. This characteristic presents a fundamental challenge for anomaly detection models: using raw price levels as features may not effectively distinguish between normal market movements and genuine anomalies, as both can appear as unpredictable deviations from historical patterns.

Given this nature, we made the decision to measure the spread between the transacted rate and t-1 market spot close rate. Empirically, our hypothesis is that the spread should be an indicator of price slippage and off-market executions; and that there exists some pattern in this spread given some profiling features such as tenor, counterparty and currency pair. As such, this spread serves alongside the transacted exchange rate to warn of anomalies not just to the transaction but to irregularities against the market as a whole.

Ideally, the spread should exhibit stationarity such that the Autoencoder learns the habitual spread of each profile. To test this, we've taken transactions with USD/HKD, USD/TRY, and USD/SGD as representatives for liquid, exotic, and domestically relevant pairs respectively, each with a sufficient sample size for asymptotic test validity ($n \geq 50$). All pairs were tested under the Augmented Dickey-Fuller (ADF) test of unit root and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test of level stationarity. Results are broadly supportive of stationarity, with all pairs rejecting the unit root under ADF. Two pairs exhibit structural breaks under KPSS, which we attribute to evolving market regimes rather than fundamental non-stationarity of the spread. Furthermore, given that the conditioning on profiling features such as currency pair, counterparty, and tenor effectively segments the data into sub-populations, within which the spread is expected to be more uniformly stationary.

Pair	ADF Test	KPSS Test	Conclusion
USD/HKD	Test Statistic = -14.697 p-Value = 0.000	Test Statistic = 0.062 p-Value = 0.01	Stationary with structural breaks.
USD/TRY	Test Statistic = -9.366 p-Value = 0.000	Test Statistic = -14.697 p-Value = 0.1	Stationary.
USD/SGD	Test Statistic = -18.698 p-Value = 0.000	Test Statistic = 15.481 p-Value = 0.01	Stationary with structural breaks.

3.3 High Cardinality of Currency Pairs and Counterparties

The extensive universe of currency pairs and counterparties presents significant challenges for anomaly detection systems. While not all currency combinations are actively traded, the theoretical cardinality of currency pair permutations remains substantial, following the mathematical relationship:

$$^{n}P_{2} = \frac{n!}{(n-2)!} = n(n-1)$$

This quadratic growth pattern means that introducing a single additional currency to the trading universe increases complexity by 2n pairs. Consequently, as cardinality expands, many trade archetypes become increasingly sparse, hindering the Autoencoder model's ability to learn robust underlying patterns. To address these cardinality challenges, we adopt differentiated risk assessment strategies for currency pairs and counterparties.

We accept the elevated risk of false anomaly detection for rare currency pairs. This approach is justified by the premise that exotic pairs inherently represent pattern-breaking behavior. Given that the majority of institutional trading activity centers on commonly traded pairs, unusual exotic pair transactions warrant heightened scrutiny as potentially anomalous events.

Conversely, we actively mitigate false positive risks for rare or new counterparties. Since infrequent counterparties and newly onboarded, entities should not trigger immediate anomaly alerts (assuming proper due diligence has been conducted), we implement a categorization approach. Counterparties with fewer than 100 historical transactions are consolidated into an "others" category. This methodology enables the Autoencoder to:

- Learn more robust patterns for established counterparties with sufficient transaction history and;
- Generate a distinct latent representation for rare counterparties without penalizing legitimate but infrequent trading relationships.

3.4 Currency-Specific Standardization Framework

The diverse nature of FX markets presents challenges when making cross-currency comparisons and detections – for instance, exchange rates between different pairs operate on vastly different numerical scales. This is depicted in figure 2 where the log frequency distribution of both USD/SGD and USD/INR rates floats around 1.16 to 1.30 and 86 to 98 respectively. This disparity creates numerical instability within machine learning models, as features with larger absolute values can disproportionately influence model weights.

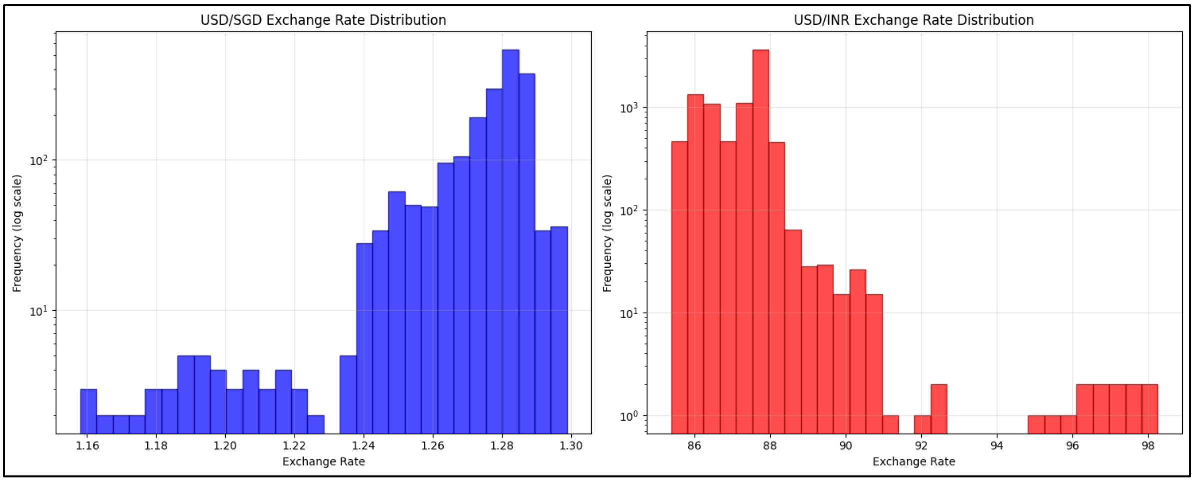


Figure 2: Distribution of Exchange Rates between USD/SGD (left) and USD/INR (right)

To address this challenge, we implemented a currency-pair-specific standardization approach that ensures numerical consistency across all products. All (except notional values) numerical features undergo standardization to achieve a normal distribution ~N(0,1) within their respective currency pair groupings. This methodology preserves the relative magnitude relationships within each currency pair while eliminating scale disparities that would distort learning process.

Notional values require different treatment due to their currency-denominated nature. Rather than applying cross-currency standardization, notional amounts are normalized within their respective currency groups to maintain business significance of transaction sizes relative to typical market activity for each currency.

4. Outcome

4.1 False Positives

The operational deployment of anomaly detection systems in financial institutions requires careful consideration of false positive rates, as these represent a significant operational burden that can undermine system effectiveness and user adoption. False positives impose what can be characterized as an "operational tax" on business users who must investigate each alert, leading to resource misallocation and potential alert fatigue that may compromise the detection of genuine anomalies. While FX market anomaly identification may not demand immediate real-time response capabilities, delayed detection and reporting of legitimate irregularities can significantly impact regulatory compliance and institutional credibility with supervisory authorities.

With an unsupervised approach, it is difficult and/or extremely costly to label our test data to identify proper false positives *a priori*.

The experimental evaluation demonstrates robust performance characteristics under rigorous testing conditions. Statistical analysis of model performance yields a **false positive rate of 11.22% with 95% confidence intervals**, derived from a comprehensive evaluation dataset comprising approximately 21,000 historical transaction records. This sample size provides sufficient statistical power to generate reliable performance estimates while maintaining confidence in the generalizability of results to production environments.

The testing protocol was designed to ensure the validity of performance metrics through careful selection of evaluation data. All testing was conducted exclusively on historical transaction data that had previously undergone regulatory reporting processes, thereby establishing a verified baseline of legitimate transactions. This methodological approach ensures that all records included in the evaluation dataset represent confirmed true negatives, eliminating potential contamination from unidentified anomalies that could artificially inflate performance metrics. The use of historically validated data provides confidence that the measured false positive rate reflects realistic operational conditions rather than artificially optimistic laboratory results.

This performance level represents a substantial improvement over traditional rule-based systems while maintaining the sensitivity required for effective regulatory compliance and risk management in FX trading operations.

4.2 Detection Test Cases

Why is Autoencoder more than just a rule-based method?

A rule can flag a single condition in isolation, but the autoencoder assesses the full trade pattern against historical behavior. It identifies whether the trade looks unusual not only because one value breaches a threshold, but because the combination of attributes is inconsistent with what has been seen historically. While some scenarios may also be captured by rules, the autoencoder adds value by analyzing the trade holistically. For FX forwards, it does not look at one field alone. It compares the full set of trade attributes against historical patterns and highlights when overall trade behavior is unusual. Refer to Appendices A1-A4 (where actual data is redacted for confidentiality purposes) for examples.

To ensure that test cases (TC) were as realistic as possible and without creating dummy trades that were obviously anomalous, each case has 1,000 data points taken from the unseen data set and was synthetically adjusted accordingly. The following true positive results (TPR) from training with data from Q1-Q3 2025 were the following:

1. TC1 (100% TPR) – Delayed reporting cases where reporting date exceeds the threshold of 2 working days between reporting and execution
2. TC2 (80% TPR) – USD trades with Notional > 13 billion
3. TC3 (100% TPR) – Flipped exchange rates (e.g. USD/INR trades had rates recorded as INR/USD)
4. TC4 (78% TPR) – Counterparty unusual patterns where counterparties would trade beyond their usual currencies and values
5. TC5 (50% TPR) – Cases where exchange rates would severely differ in historic $\pm 3\sigma$

The True Positive Rate (TPR) per use case is expected through evolve as model is further refined and exposed to broader business scenarios. Potential improvements include increasing the number of hidden layers and dimensions. Prior to any production deployment, the acceptable performance threshold will be reviewed and agreed with the business.

4.3 Overall Performance & Scalability

The current implementation demonstrates promising overall performance, with primary optimization efforts focused on enhancing detection sensitivity. This approach represents a significant improvement over alternative strategies, which would necessitate either comprehensive manual review of all transactions or statistical sampling methodologies that inherently risk allowing anomalous activities to bypass detection systems.

Autoencoder architecture also offers substantial scalability advantages over traditional ML methodologies. Unlike the OCSVM[2], which requires the construction of high-dimensional decision boundaries that become computationally prohibitive as feature dimensionality increases, or Isolation Forest algorithms that rely on recursive feature partitioning that may not capture complex inter-feature relationships, the Autoencoder approach maintains consistent computational complexity regardless of transaction complexity.

Further adding to the scalability of our experiments, the model's design minimizes specialized feature engineering requirements, reducing dependency on extensive data science expertise for system maintenance and expansion. Business domain experts can readily understand and configure input features based on standard trade characteristics.

Despite these advantages, numerical feature standardization remains instrument-specific and still requires domain expertise. Features such as exchange rates and tenor calculations must continue to be normalized within their contexts[3], necessitating ongoing research and calibration to ensure appropriate and numerically stable scaling methodologies.

[2] One-Class Support Vector Machines (OCSVM)

[3] As mentioned in 3.4 Currency-Specific Standardization Framework, all, except notional, numerical features are scaled by their respective currency pairs.

5. Expected Benefit and Key Considerations

5.1 Simplified Threshold Management

The Autoencoder-based approach significantly reduces the complexity inherent in traditional rule-based anomaly detection systems by eliminating the need for extensive threshold calibration and rule maintenance. For instance, conventional systems typically require exchange rates to be benchmarked against historical market rates (often with simple lagged comparisons such as t-1) or against static reference values that require periodic manual updates every few days or weeks.

5.2 Recognition of Multivariate Interactions

Traditional rule-based frameworks require establishment and maintenance of (_^n)P_2 relationships corresponding to all possible currency pair permutations. Complexity only increases exponentially with interactions with other features within the trade. The Autoencoder can detect intricate multivariate interactions, say, patterns of a said counterparty while benchmarking the patterns of the currency pair traded. While both rule-based and ML approaches face challenges adapting to market regime changes, the Autoencoder offers advantage in pattern complexity. However, both approaches require regular maintenance – rule-based systems through threshold recalibration and Autoencoders through periodic retraining to ensure continued effectiveness in evolving market conditions.

5.3 Architectural Efficiency

The rule-based paradigm creates high-dimensional threshold space that becomes increasingly unwieldly as new currency pairs or market conditions are introduced. In contrast, the Autoencoder learns adaptive representations of normal market behavior that automatically adjust to evolving market conditions without requiring manual recalibration. This approach transforms a rigid, multi-dimensional rule matrix into a flexible learning system that maintains detection sensitivity while reducing operational overhead and improving responsiveness to legitimate market evolution.

This paradigm shift from static rule enforcement to dynamic pattern learning represents a fundamental improvement in both system maintainability and detection accuracy across diverse market conditions.

5.4 Implementation Challenges and Considerations

5.4.1 Computational Resource Requirements

The Autoencoder implementation introduces significant computational overhead compared to traditional rule-based systems. Training requires substantial historical data spanning multiple market cycles to establish robust baseline patterns, with initial model training taking several hours to days depending on data volume and architectural complexity. Additionally, the production pipeline necessitates regular retraining schedules (weekly to monthly) to maintain model effectiveness, requiring dedicated computational resources and automated orchestration systems.

In our experimentation, end-to-end training from data preparation to model deployment with 100k data points over 100 iterations and 100 epochs per iteration using a g4dn.xlarge instance in Amazon SageMaker averaged ~120 minutes. Given that computational optimization was not the primary objective of this experiment, we expect significant improvements in training duration through infrastructure scaling and algorithmic optimizations in production environments.

5.4.2 Model Performance Monitoring

Continuous monitoring infrastructure is essential to detect model degradation over time. Key monitoring requirements include:

- Reconstruction error drift tracking to identify when model performance deteriorates
- Feature distribution monitoring to detect changes in market behavior patterns
- False positive rate analysis to ensure operational efficiency
- Automated alerting systems for model performance the market thresholds

5.4.3 Operational Integration Complexity

Deployment requires significant integration with existing FI infrastructure:

- Legacy system integration with established trading and surveillance platforms
- Real-time processing requirements to meet regulatory reporting timelines
- Workflow integration with existing investigation and case management systems
- Staff training for interpreting ML-generated anomaly scores versus traditional rule alerts

6. Expected Benefit and Key Considerations

The FX market's decentralized structure, high velocity, and sensitivity to macroeconomic and geopolitical events create inherent challenges for traditional rule-based anomaly detection approaches. As outlined in this paper, these limitations are further compounded by fragmented data sources, dynamic reporting standards, and evolving market regimes.

This experiment demonstrates that machine learning techniques (e, g. Autoencoder)—particularly those capable of learning patterns across multiple transaction attributes—offer a meaningful complement to existing controls. By moving beyond static rules and leveraging data-driven pattern recognition, the model was able to surface anomalies that are difficult to detect through conventional methods. Notably, the model successfully identified flipped exchange rates, deviations from counterparties' typical currency trading behavior, delayed trade reporting beyond T+2 timelines, and unusual notional sizes relative to historical patterns. These findings validate the ability of ML models to detect both data quality issues and subtle behavioral irregularities across interconnected FX transactions.

If deployed into production, such capabilities are expected to deliver several tangible benefits. First, enhanced detection accuracy would reduce false positives while improving the identification of genuinely suspicious or erroneous transactions. Second, the ability to analyze multiple attributes simultaneously enables earlier detection of cross-field inconsistencies that may otherwise go unnoticed. Third, improved monitoring of reporting timeliness and trade completeness would strengthen regulatory compliance and data integrity. Collectively, these improvements can lead to more efficient surveillance processes, reduced manual investigation effort, and better risk management outcomes.

Beyond detection performance, the experiment also highlights important implementation learnings. The iterative, workshop-driven approach with business stakeholders proved critical in refining model logic and ensuring alignment with real operational scenarios. Continuous feedback loops allowed domain expertise to be embedded into the model, improving both accuracy and usability of outputs. In addition, the exercise underscored the importance of data quality readiness, robust feature engineering, and clear governance frameworks to support sustainable deployment of AI-driven solutions.

In conclusion, while machine learning is not a replacement for existing rule-based controls, it provides a powerful augmentation that enhances the ability of financial institutions to navigate the complexities of FX markets. By combining adaptive AI techniques with domain expertise and strong data foundations, institutions can build more resilient, scalable, and forward-looking anomaly detection capabilities that are better suited to today's dynamic market environment.

7. Appendix

Interpreting the Results

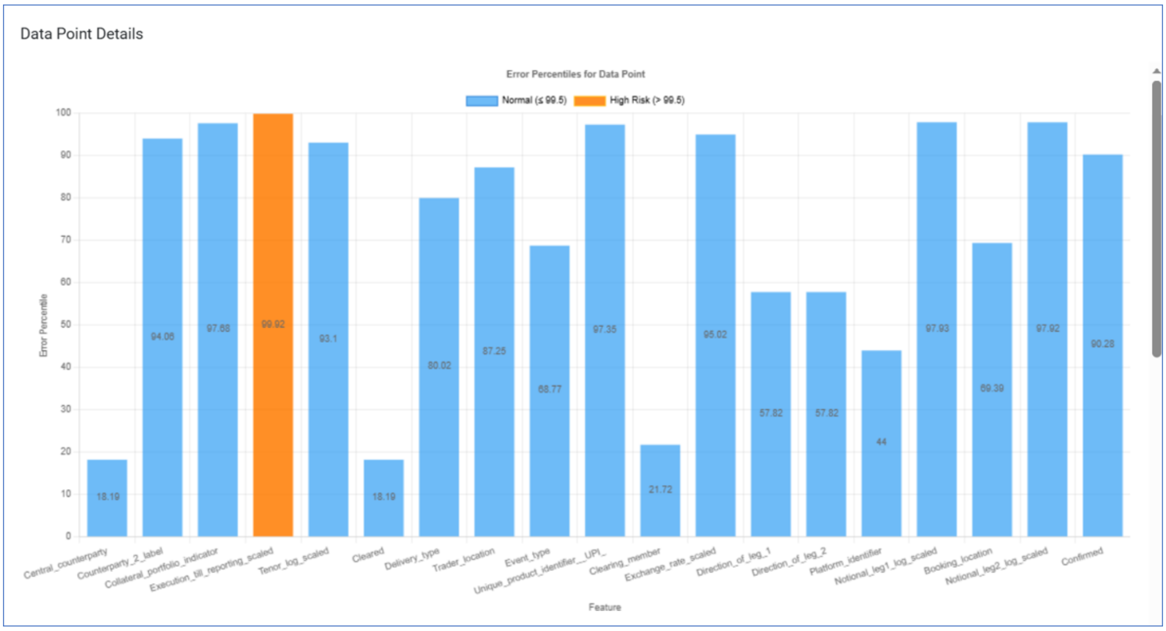
- The model evaluates multiple trade attributes and assigns anomaly scores based on deviation from learned historical patterns.
- Attributes highlighted in **orange** indicate pattern breaking behavior thus scoring higher.
- Attributes shown in blue fall within expected behavioral ranges and are not flagged.

The accompanying summary table provides traceability by listing the raw features and their connection to engineered features, supporting auditability and downstream review workflows.

Appendix A1: For Illustration Purposes Only

The primary struggle of having datetime formats in input features is the complexity in encoding. While cyclical encoding or sinusoidal transformation could have encoded for raw datetime features to be interpretable, we found that creating a feature for business context such as “*Tenor*” or “*Execution_till_reporting*” (A.K.A. timeliness) allowed for us to:

- Compress information from 2 features into a single feature
- Give business context and explainability to what the anomalies are from raw datetimes

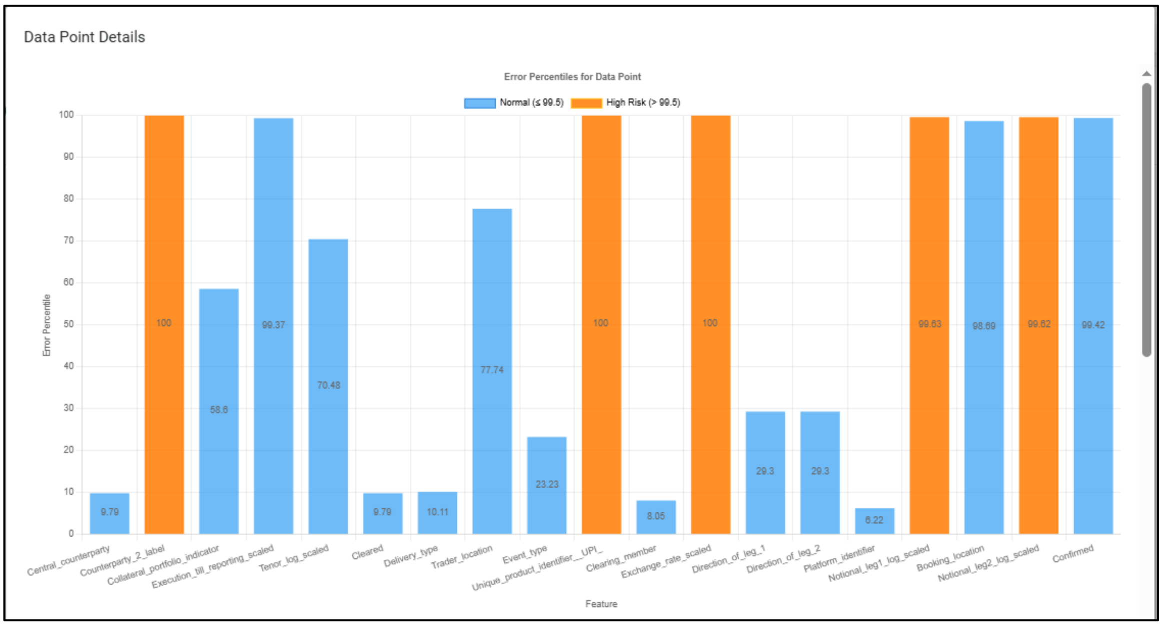


Field	Value
Execution Timestamp	2025-08-19 03:03:06+00:00
Reporting Timestamp	2025-08-27 03:03:06+00:00

Appendix A2: For Illustration Purposes Only

In this sample, the model does not only flag that the amount is above a threshold (TC2). It can also indicate:

- Which leg carries the abnormal notional
- That the FX rate is also an outlier
- That the counterparty is unusual for this trade size and rate pattern based on historical behavior



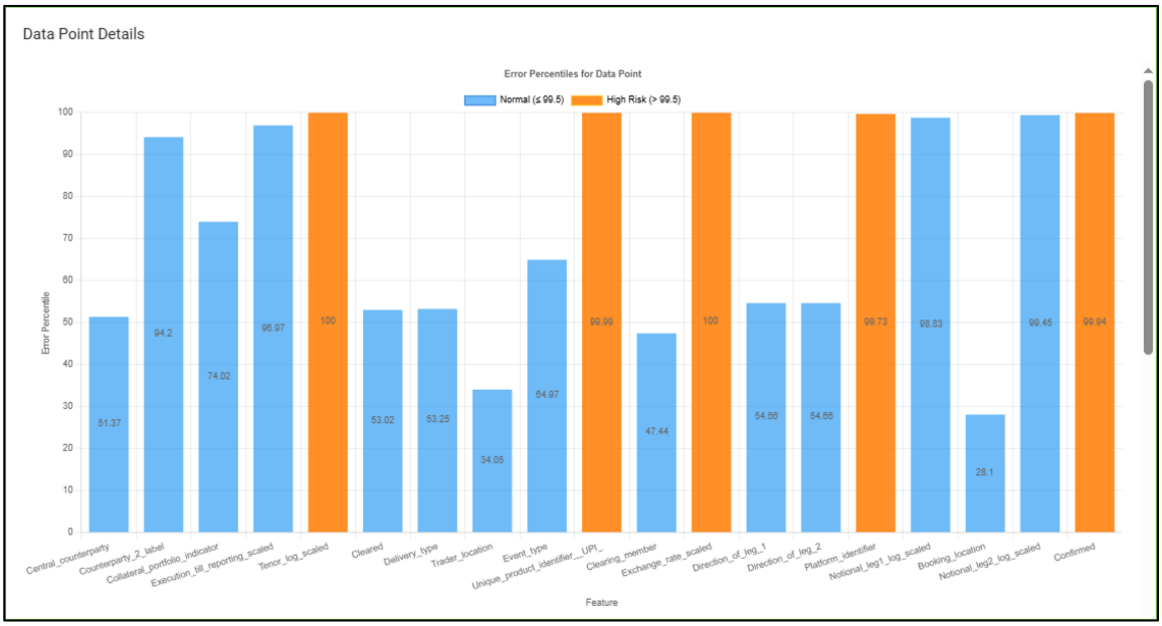
Field	**Value**
Exchange Rate	1.27
Notional Amount of leg 1	67,000,000,000
Notional amount of leg 2	85,251,805,000
Counterparty 2	ABC
Currency 1	USD
Currency 2	SGD

This gives more context than a simple rule. It suggests that the anomaly is not just “notional > threshold,” but that this counterparty typically does not trade this size of notional, with this type of FX rate, historically.

Appendix A3: For Illustration Purposes Only

While a flipped FX rate (TC3 & by extension TC5) can sometimes be captured by rules, the model highlights that the issue is not isolated, but part of a broader abnormal trade pattern. In this case, the model flags multiple attributes (in light orange) as outliers:

- Exchange rate – significantly deviates from historical patterns, indicating a likely flipped value
- Platform identifier – unusual compared to typical trades for this currency pair
- Effective date & expiration date – gap exceeds expected norms (e.g., more than 2 business days)
- Currency pair direction (USD/KRW) – inconsistent with typical rate ranges when compared historically

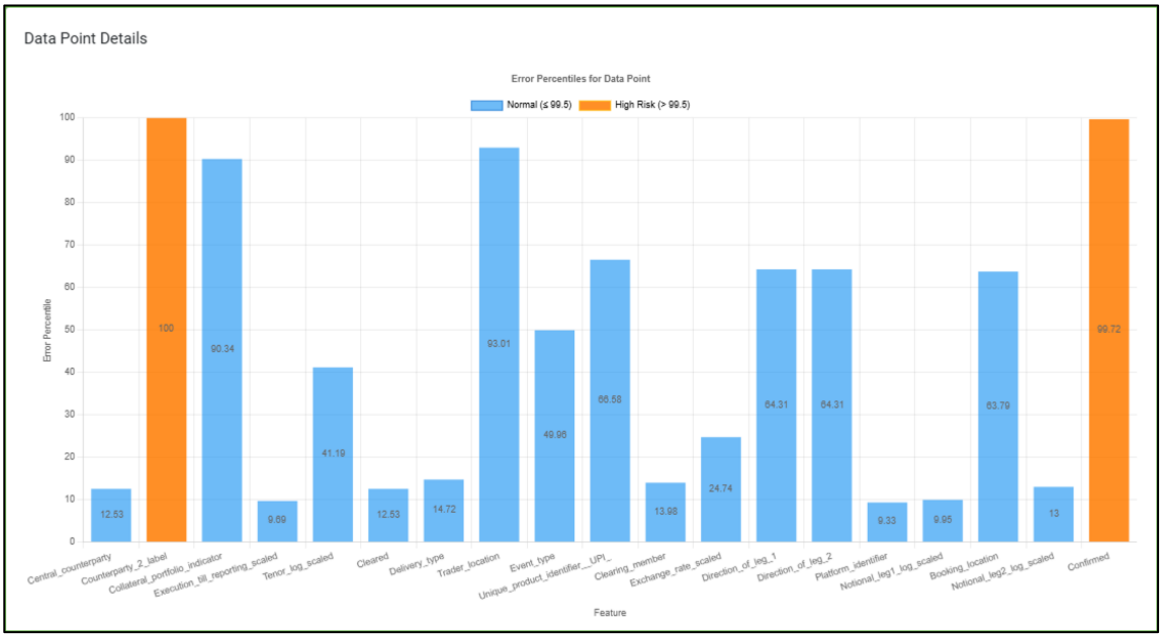


Field	Value
Type of Confirmation	Not Confirmed
Effective date	2025-12-05
Expiration date	2025-12-06
Exchange rate	0.00068
Notional Amount of leg 1	14,772,500
Notional amount of leg 2	10,000
Platform identifier	AACA
Currency 1	USD
Currency 2	KRW

Appendix A4: For Illustration Purposes Only

The model detects not just unusual values, but unusual behavior — identifying when a counterparty trades something it typically does not (TC4).

- Counterparty has no historical activity for this currency pair
- Trade deviates from established behavioral patterns
- May indicate data issue or booking error
- In some cases, a genuine business activity that may require validation



Field	Value
Type of Confirmation (Electronic, Physical, Not Confirmed)	Electronic
Counterparty 2	ABC
Currency 1	USD
Currency 2	SGD

8. Reference List in Alphabetical Order

1. Goetz von Peter-Bank for International Settlements (2025). Triennial Central Bank Survey of foreign exchange and Over-the-counter (OTC) derivatives markets in 2025.

2. Guansong Pang, Chunhua Shen, Longbing Cao, and Anton van den Hengel (2020). Deep Learning for Anomaly Detection.

3. Monetary Authority of Singapore (2025). Foreign Exchange & Derivatives.

4. Monetary Authority of Singapore (2024). guidelines-to-the-securities-and-futures-reporting-of-derivatives-contracts-regulations_31-may-24_3.pdf

5. Richard K. Lyons (2001). https://www.chesler.us/resources/academia/lyons_microstructure.pdf. The Microstructure Approach to Exchange Rates.

6. Singapore Foreign Exchange Market Committee (2025). October 2025 Survey of Singapore FX Volume.

CRÉDIT AGRICOLE
CORPORATE & INVESTMENT BANK