

# TRRADA: Trade Regulatory Reporting an Anomaly Detection Approach

Ramy MEZHOUD<sup>1 2</sup>, Zahia GUESSOUM<sup>1</sup>, Samer ELSAWDA<sup>2</sup>, Pavel KOCHNEV<sup>2</sup>

<sup>1</sup> LAB-I\*, Université de Reims Champagne Ardenne

<sup>2</sup> Natixis Corporate Investment Banking, Paris



## Context & Motivation

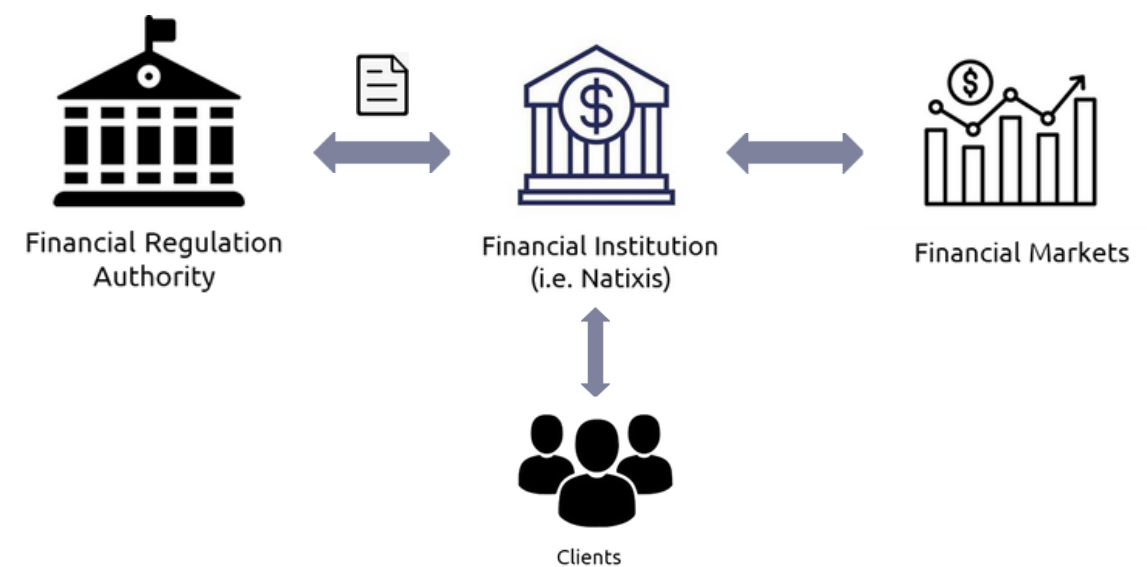
Trade Regulatory reporting is a significant and complex task that may include several anomalies, potentially leading to many consequences.

### Source of anomalies:

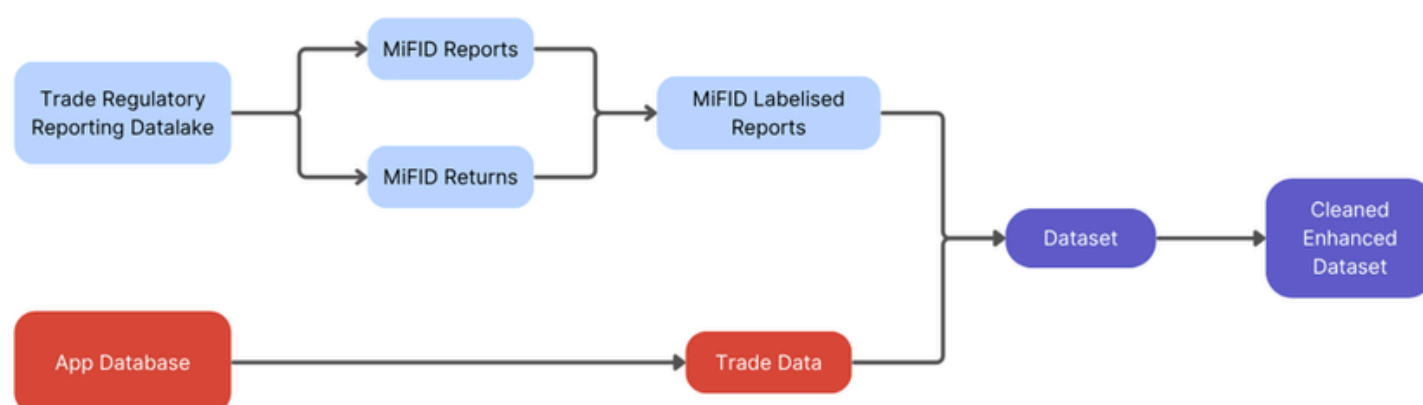
- Volume and complexity of the data.
- Fragmentation of information systems.
- Regulatory updates and changes.
- Human errors and manual corrections.

### Consequences:

- Financial and legal impacts.
- Loss of reputation.
- Loss of activity license.

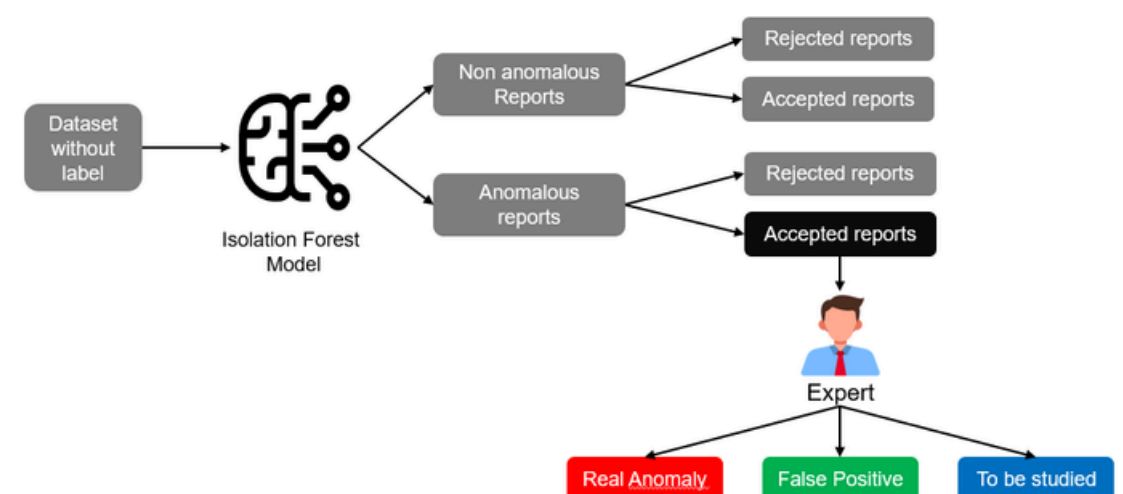


## Dataset Construction



Use case: Jurisdiction: ESMA / Directive: MiFID II / Trade type: FX\_SWAP, FX\_FWD

## Dataset Enhancement



## Implementation & Results

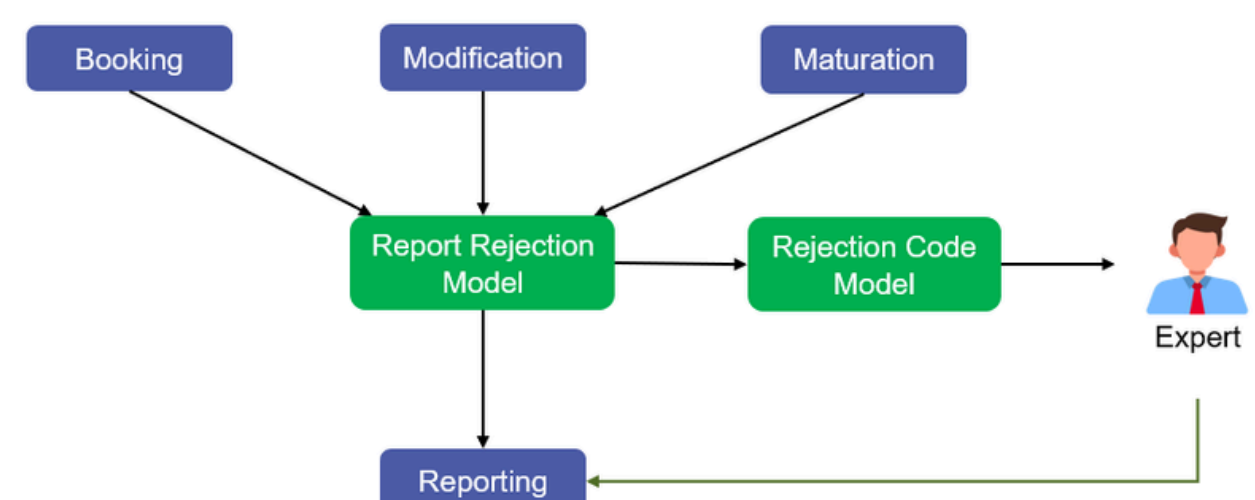
Classifier	Accuracy	Macro-Precision	Macro-Recall	Macro-F1 Score	FPR	FNR
XGBoost	<b>0.991</b>	0.907	<b>0.870</b>	<b>0.888</b>	<b>0.257</b>	0.003
Decision Tree	0.988	0.929	0.756	0.820	0.486	0.002
Random Forest	0.989	0.883	0.813	0.844	0.369	0.004
SVM	0.990	<b>0.983</b>	0.769	0.844	0.462	<b>0.001</b>
LightGBM	<b>0.991</b>	0.919	0.857	0.885	0.283	0.003

Table 1. PERFORMANCE EVALUATION OF BINARY CLASSIFICATION

Classifier	wA-Accuracy	wA-Precision	wA-Recall	wA-F1-score
Decision Tree	0.991	0.610	0.854	0.679
Random Forest	0.989	0.500	0.834	0.591
SVM	0.984	0.508	<b>0.893</b>	0.598
XGBoost	0.998	<b>0.875</b>	0.707	<b>0.770</b>
LightGBM	<b>0.994</b>	0.614	0.865	0.695

Table 2. PERFORMANCE EVALUATION OF MULTILABEL CLASSIFICATION

Rejection Code	Count	Accuracy	Precision	Recall	F1 Score
Rejection Code 1	17794	99.26%	0.06%	0.01%	0.06%
Rejection Code 2	17833	99.48%	0.00%	0.00%	0.00%
Rejection Code 3	17896	99.83%	0.03%	0.12%	0.02%
Rejection Code 4	17851	99.58%	0.04%	0.23%	0.15%
Rejection Code 5	17874	99.71%	0.01%	0.03%	0.01%
Rejection Code 6	17893	99.82%	0.00%	0.00%	0.00%
Rejection Code 7	17843	99.54%	0.02%	0.05%	0.02%
Rejection Code 8	17892	99.81%	0.04%	0.11%	0.04%



## Conclusion & Future Work

Due to confidentiality issues, there's a lack of online data for trade regulatory reports. No current method forecasts report rejections or detects anomalies. We proposed a dataset construction approach and forecasting methods to predict report rejections before submission. Our results were presented to experts for production deployment. Future work should aim to expand this approach to cover a broader range of instruments and jurisdictions. We also suggest using Large Language Models (LLMs) to offer expert-based error remediation, enhancing accuracy and efficiency in reporting.

## Related Literature

- A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine learning for anomaly detection: A systematic review," IEEE Access, vol. 9, pp. 78 658–78 700, 2021.
- V. Jain, A. Balakrishnan, D. Beeram, M. Najana, and P. Chintale, "Leveraging Artificial Intelligence for Enhancing Regulatory Compliance in the Financial Sector," International Journal of Computer Trends and Technology, vol. 72, no. 5, pp. 124–140, May 2024.
- F. Moshirian, "The global financial crisis and the evolution of markets, institutions and regulation," Journal of banking & Finance, vol. 35, no. 3, pp. 502–511, 2011