Executive Summary

Heba Bou KaedBey, Min Shi, Ritesh Bachhar, Khanh Nguyen

November 7th 2025

Money laundering obscures illicit funds by blending them with legitimate transactions, on a scale estimated at 2–5% of GDP ($800B–$2T), which is a serious problem for banking institutions and anti-money laundering regulators. In this project, we evaluated in a chronological setting simulating periodic AML batch processing. Feature statistics were recomputed using data up to each evaluation window, reflecting a scenario where banks update risk profiles regularly as new data arrived. We aimed to reduce false positives while enhancing recall within anti-money laundering (AML) systems, as false positive means blocking a normal transaction while a false negative means missing an actual laundering transaction.

We used the dataset SAML-D [1], which comprised 12 characteristics and 28 typologies, selected based on existing datasets, academic literature, and interviews with AML specialists, which highlighted its broad applicability. The dataset contains 9.5 million entries, but the "isLaundering" labels are highly imbalanced, with only 0.104% representing actual money laundering activity. These properties led to the following modeling ideas.

We used the XGBoost model as the baseline model for the following reasons: it is well suited to a large dataset; its tree-based architecture is robust to outliers, and it helps prevent overfitting on noisy transactional data with its built-in regularization and shrinkage. However, the model's PR-AUC was only 0.5586 with recall 41.3%. The significant precision-recall trade off exhibited by the XGBoost model indicated the need for models capable of capturing more complex patterns — deep learning models like transformers and GNNs might be more promising avenues to reducing the precision-recall trade off.

We then tried transformer models and TGNN (temporal graphic neural network) models. We used focal loss, oversampling, and dropout to boost recall while keeping precision within a reasonable range. With these techniques, the transformer achieved a PR-AUC of 0.64 with a 59.93% recall, which was a 14.6% improvement on PR-AUC and a 45.1% improvement on recall from the baseline model. For the GNN model, Node degrees(fan-in/fan-out) were further applied, resulting in a 0.844 PR-AUC and a 81.3% recall, reflecting a further improvement of 31.9% and 35.7%, on PR-AUC and recall, respectively.

While the Transformer and TGNN ran slightly slower than the XGBoost model, their performances were much better than XGBoost model. The improved recall can help the stakeholders reduce regulatory penalties, protecting their reputation, and strengthening the integrity of their financial operations; while the improved precision will reduce false alarms, lowering operational costs and customer disruption.

In conclusion, the transformer achieved a better recall than XGBoost baseline, but at the cost of a lower precision. The TGNN model, on the other hand, delivered the strongest gains in both recall and precision, making it a robust and effective option for modern AML systems. These findings provided a solid foundation for developing scalable, graph-based approaches in real-world AML environments.

References

[1] Berkay Oztas, Dogukan Cetinkaya, Funmi Adedoyin, Marcin Budka, Hasan Dogan, and Gorkem Aksu. Enhancing anti-money laundering: Development of a synthetic transaction monitoring dataset. In *2023 IEEE International Conference on e-Business Engineering (ICEBE)*, pages 47–54, 2023.