Nicolás Robayo Pardo

AML Monitoring System for Bitcoin Transactions

This report presents the development of an Anti-Money Laundering (AML) monitoring system designed to detect illicit Bitcoin transactions. The business objective is to identify transactions associated with money laundering, darknet marketplaces, and other criminal activities. Each transaction is linked to either licit or illicit entities, and our goal was to create a robust predictive model to classify new transactions.

The dataset presents two significant challenges. First, we face severe class imbalance with only 2% of transactions being illicit, yet these carry substantial financial and legal consequences for legitimate payment systems. Second, only 21% of observations are labeled, creating a semi-supervised learning environment. The dataset consists of Bitcoin transactions connected through money inflows and outflows, divided into 49 timesteps representing networks of transactions, which allows us to leverage graph structures for enhanced detection capabilities.

Since the original variables are anonymized, we augmented the dataset with two types of derived features. We implemented the GuiltyWalker algorithm from Oliveira et al. [1], which creates random walks that attempt to reach known illicit transactions by following money flows backward in time. The algorithm was run looking to achieve 500 successful walks (finding an illicit node is success) with a maximum length of 50. With the results of the algorithm some features were computed that summarized the random walks. We also trained a Graph Convolutional Network (GCNs) as defined by Weber et al. [2] with two layers to generate embeddings that capture the structural properties of the Bitcoin transaction network, implementing careful controls to prevent information leakage during model training. This was done to extract more of the network features inside the data and a network was trained for every time step.

We then constructed a combined approach model where an Isolation Forest was deployed as a feature engine to generate anomaly scores. This model learns patterns of "normal" transactions, providing a strong signal for detecting outliers that might represent illicit activity. An XGBoost ensemble model was then implemented for final classification, incorporating both the original features and the engineered features. To address class imbalance, we employed SMOTE (Synthetic Minority Over-sampling Technique) to generate synthetic examples of illicit transactions, balancing the training data. Both

1. https://doi.org/10.48550/arXiv.2102.05373
2. https://arxiv.org/pdf/1908.02591

models were calibrated using stratified cross-validation with asymmetric thresholds, and we optimized threshold selection to maximize the F1 score, balancing precision and recall.

The model achieved an F1 score of 0.8725 on the test dataset, with 92.17% precision (percentage of flagged transactions that are truly illicit) and 82.84% recall (percentage of illicit transactions successfully caught). When applied to unlabeled data, our model flagged 6.48% of previously unknown transactions as potentially illicit. The algorithm processes 95,339 transactions per second, making it suitable for production environments. We can compare the F1 score to that obtained by [1] who used only an expanded dataset with the Guiltywalkers variables in a random forest model and only obtained a 0.796 F1 score. We have gained 0.0765 in F1 Score.

The most influential features for classification were two graph embedding features and three transaction features (specifically features 4, 57, and 60 using 0-based indexing). The importance of the graph embeddings show that the leveraging the structure of the net is primordial for creating a robust illicit detector.

Analysis of F1 metrics across time periods revealed that early periods with few illicit transactions presented greater classification challenges. We observed notable performance drifts in periods 18 and 43, potentially indicating significant events that altered transaction patterns. This temporal variation highlights Bitcoin's evolving nature with new participants and use cases, emphasizing the importance of near-continuous model retraining.

The model outputs can be integrated into compliance workflows in several ways. Transactions can be ranked by illicit probability for focused investigations, while SHAP explanations provide transparency for regulatory reviews. Compliance teams can create alerts based on the most important features, and threshold values and SHAP analysis can help authorities identify emerging illicit patterns. This system enables compliance teams to detect complex relationships in transaction data while optimizing investigation resources.

To use this algorithm in a production setting it would depend if its use is intended for online monitoring or offline as its design depends on the transaction having inflows and outflows of money. So for that reason, a neighborhood of transactions would need to be available for the feature generation and model inference to be possible. In terms of speed we believe the pipeline is swift enough to scale.