**De-anonymisation of illegal cryptocurrency transactions**

**Background for the Methodology**

Bitcoin allows end-users to create (pseudo-)anonymous financial transactions without the need for disclosing their personal information. This is done by generating a pseudonym for the user, also called “address.” The apparent anonymity and ease to create pseudo-anonymous financial transactions attracted users who value their privacy on one hand and, on the other hand, it has also attracted cybercriminals who want to use it for ransom-ware and other illegal activities.

Therefore, analyzing the pseudo-anonymity and understanding the traceability of Bitcoin flows to investigate the use of it for criminal or fraudulent purposes is of high academic importance as well as practical relevance.

**Methodology**

1. **Dataset Preprocessing**

We used an effective 70/30 train-test split with respect to time steps for both the transactions and actors datasets, with time steps 1 to 34 for training and time steps 35 to 49 for testing. we genearate graph which graphically shows the distribution of data points (top for transactions, bottom for actors) of all three classes by time step for both the training and testing sets. Due to the underlying class imbalance across illicit and licit classes, normalization and standardization transformations are applied.

The augmented features in the transactions dataset and all features in the actors dataset are transformed by scaling each feature using the MinMaxScaler to the range (0, 1), reducing imbalance and assisting with model convergence.

1. **Machine Learning Models**

The ML models used for evaluation included Random Forest (RF) , Multilayer Perceptrons (MLP) , Long Short-Term Memory (LSTM) , and Extreme Gradient Boosting (XGB) . We also include Logistic Regression (LR) as the baseline. LR is a single layer neural network which estimates the probability of an event, such as licit or illicit, based on independent variables. RF is an ensemble of classification trees trained on samples with random subsets of features for each decision, evaluating a final decision from averaging all decision trees.

1. **Fraud Detection Evaluation Metrics**

The metrics used to verify the models were Precision (ratio of correct classifications), Recall (proportion of actual positive labels correctly classified), F1 Score (harmonic mean of precision and recall), and Micro-Avg F1 (Micro-F1) Score (ratio of correct classifications to total classifications)