Fraudulent Transaction Prediction in Bitcoin Network

Bhavay Aggarwal
2018384
bhavay18384@iiid.ac.in

Prasham Narayan
2018359
prasham18359@iiid.ac.in

Saad Ahmad
2018409
saad18409@iiid.ac.in

1. Motivation

Bitcoin is a cryptocurrency which is a popular method for transactions. Among all the transactions that take place some of them are maybe used as malware attacks or as ransoms. This project interests us because it gives us the possibility to work on graph based models alongside conventional machine learning models. In this project, we would work on Machine Learning models to distinguish between licit and illicit transactions.

2. Related Works

- Pham, T. and Lee, S., 2016. Anomaly detection in bitcoin network using unsupervised learning methods. arXiv preprint arXiv:1611.03941.
  They used three unsupervised learning methods, namely k-means clustering, Mahalanobis distance and Unsupervised support vector machine.

- M. Bartoletti, B. Pes and S. Serusi, "Data Mining for Detecting Bitcoin Ponzi Schemes," 2018 Crypto Valley Conference on Blockchain Technology (CVCBT), Zug, 2018, pp. 75-84, doi: 10.1109/CVCBT.2018.00014.
  Trained various Machine Learning (ML) algorithms and identified the algorithm that performs the best in detecting which accounts are malicious.

- S. SAYADI, S. BEN REJEB and Z. CHOUKAIR, "Anomaly Detection Model Over Blockchain Electronic Transactions," 2019 15th International Wireless Communications and Mobile Computing Conference (IWCMC), Tangier, Morocco, 2019, pp. 895-900, doi: 10.1109/IWCMC.2019.8766765.
  Used two machine learning algorithms, one-class support vector machines algorithm to detect the outliers and the k-means algorithm to group similar outliers with same anomalies.

3. Timeline

- September - October- Preprocessing and of extracting features from the graph and using the CSV file from Kaggle. During this period we will try to embed and vectorize nodes and will be performing initial statistical analysis.

- October - November - During this period we will be trying out different Machine learning techniques based on the data and will try to reason out our methodologies.

- November - December - This period will be mostly used for improving on the techniques that performed the best in our initial findings. We would be reasoning out the performances of models and will try hyperparameter optimization.

4. Individual Tasks

- Bhavay Aggarwal
  Performing statistical analysis on the data and researching about machine learning techniques which can work best with the data.

- Prasham Narayan
  Implementing the Machine learning algorithms and coming up with reasoning about the methodologies adopted in the project.

- Saad Ahmad
  Researching about preprocessing methods, vectorizing the graph and performing feature extraction

5. Final Outcome

With this project, we intend to come up with reasonable algorithms and pipelines which can be used with graphical data and will try to compare different machine learning techniques. Our work would include trying to come up with reasonable questions and answering them based upon statistical tests. We will also include unsupervised learning on non-labelled data. We shall come up with reasonable and interpretable feature extraction pipelines which help ML models to accurately distinguish between licit and illicit transactions.