Detecting Illicit Entities in Bitcoin using Supervised Learning of Ensemble Decision Trees
Pranav Nerurkar, Yann Busnel, Romaric Ludinard, Kunjal Shah, Sunil Bhirud, Dhiren Patel

To cite this version:

HAL Id: hal-02952081
https://hal-imt-atlantique.archives-ouvertes.fr/hal-02952081
Submitted on 27 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Detecting Illicit Entities in Bitcoin using Supervised Learning of Ensemble Decision Trees

Pranav Nerurkar  
panerurkar_p16@ce.vjti.ac.in  
pranav.n@nmims.edu  
Dept. of CE&IT, VJTI  
MPSTME, NMIMS University  
Mumbai, India

Yann Busnel  
yann.busnel@imt-atlantique.fr  
IMT Atlantique, IRISA  
Cesson-Sévigné, France

Romaric Ludinard  
romaric.ludinard@imt-atlantique.fr  
IMT Atlantique, IRISA  
Cesson-Sévigné, France

Kunjal Shah  
kunjal1999@gmail.com  
VJTI Mumbai  
Mumbai, India

Sunil Bhirud  
sgbhirud@ce.vjti.ac.in  
VJTI Mumbai  
Mumbai, India

Dhiren Patel  
dhiren29p@gmail.com  
VJTI Mumbai  
Mumbai, India

Abstract

Since its inception in 2009, Bitcoin has been mired in controversies for providing a haven for illegal activities. Several types of illicit users hide behind the blanket of anonymity. Uncovering these entities is key for forensic investigations. Current methods utilize machine learning for identifying these illicit entities. However, the existing approaches only focus on a limited category of illicit users. The current paper proposes to address the issue by implementing an ensemble of decision trees for supervised learning. More parameters allow the ensemble model to learn discriminating features that can categorize multiple groups of illicit users from licit users. To evaluate the model, a dataset of 2059 real-life entities on Bitcoin was extracted from the Blockchain. Nine features were engineered to train the model for segregating 28 different licit-illicit categories of users. The proposed model provided a reliable tool for forensic study. Empirical evaluation of the proposed model vis-a-vis three existing benchmark models was performed to highlight its efficacy. Experiments showed that the specificity and sensitivity of the proposed model were comparable to other models.

CCS Concepts:
• Computing methodologies → Artificial intelligence; Boosting.

Keywords: Bitcoin, Fraud detection, Supervised Learning, Data Mining

1 Introduction

Bitcoin 1 platform has attracted anti-social elements [1] as it creates hurdles for law enforcement to trace suspicious transactions due to the anonymity and privacy [2]. As Bitcoin became financially significant, the emergence of Ponzi schemes, money laundering, frauds, embezzlements, extortion, and tax evasion [3] practices were seen. These businesses used the blanket of secrecy afforded by Bitcoin to mislead the audit trail. It was speculated that in 2017, BTCs worth $770 million were traded for illicit activities [4], a quarter of bitcoin users were malicious and 46% of all bitcoin activity was illegal [5].

Due to voluminous data generated about bitcoin transactions on the Blockchain, machine learning became a popular technique for tracking and scrutinizing illicit users or transactions. Existing literature surveyed on detecting illegal activities using Machine Learning (ML) had focused on illegal transactions, identifying suspicious bitcoin users (extortionists, ponzi scams, darknet markets, ransomware, human traffickers, frauds), detecting money laundering, identifying mixing services, identifying bitcoin exchanges, identifying illegal transactions, identifying bitcoin wallets and bitcoin miners.

The issues faced in the application of machine learning in identifying illegal activities are lack of benchmark, public datasets (see Table 2), full information of Blockchain, and lack of ground truth information on the identities of bitcoin users. Apart from these issues, cryptocurrencies offer their users pseudo-anonymity by allowing users to transact with each other through hash address. These addresses can be created and discarded countless times, complicating the task of linking a transaction to a user. The target of interest was restricted to limited categories of illicit users to reduce the computational complexity of machine learning models. Additionally, the time interval for which data was collected from the Blockchain for feature engineering was restricted to shorter spans. Due to these, the models obtained after training were not generalized.

1.1 Motivation

The current paper aimed to build upon and extend work in detecting illegal entities in Bitcoin. Features of bitcoin users were derived by scrutinizing the Blockchain from 03
Jan 2009 12:45:05 GMT to 08 May 2020 at 13:21:33 GMT. This was to provide the model with features suitable for
generalized learning of entity behavior. Additionally, this
avoided a model trained for recognizing a limited category
of entities.

1.2 Contributions
Contribution of the current paper are as follows:

- A public dataset of addresses and features of illicit
  Bitcoin entities
- Empirical analysis of different learning strategies for
classifying illicit Bitcoin entities;
- Implementing a supervised learning approach that esti-
  mated the most discriminating features for detecting
categories of illicit Bitcoin users.

1.3 Novelty
An extensive literature survey could find studies focusing
on only a subset of illicit activities viz. botnets, extortionists,
ponzi scams, darknet markets, ransomware, human traffick-
ers, frauds, money laundering, and mixing services. At the
time of writing (May 2020), there has not been any research
focusing on a broad spectrum of illegal activities.

1.4 Outline
The rest of the paper is organized into four sections. Section
2 provides the preliminaries needed for the paper, along with
a critique of the current literature. Materials and methods
detail the data collection and preparation strategy in Section
3. The proposed work is described in Section 4 followed by
Experimental study in Section 5 and Conclusion and future
works in Section 6.

2 Related work
Fundamental concepts of Bitcoin are described in Sections
2.1 and 2.2. Followed by critical analysis of published studies
on detecting illegal users (see Section 2.3), issues in avail-
able datasets (see Table 2) and popular ML models used in
published studies (see Table 3).

2.1 Description of Bitcoin system
Bitcoin transactions are added to “Blocks” and recorded into
a distributed public ledger “Blockchain.” Each transaction
has several inputs (senders) and outputs (receivers). The
metadata ¹ associated with blocks, transactions, inputs, and
outputs provides scope for analysis.

2.2 Common types of services on Bitcoin
- Gambling: Allow placing of bets using BTCs
- Darknet markets: Selling and buying goods using BTCs
- Mixers: Remove traceability of BTCs from source
- Cyber-criminals: Blacklisted by governments
- Ponzi: High yield investment scams

Other illicit users are ransom calls for extortions, bombs
threats, sextortionists, scams, and blackmail.

2.3 Studies on detecting illegal activities in Bitcoin

<table>
<thead>
<tr>
<th>Authors</th>
<th>Description</th>
<th>Features extracted</th>
</tr>
</thead>
<tbody>
<tr>
<td>C Lee et al. [4]</td>
<td>Detecting Illegal Transactions</td>
<td>Transaction features</td>
</tr>
<tr>
<td>Y Wu et al. [7]</td>
<td>Tracing suspicious bitcoin entities</td>
<td>Transaction features</td>
</tr>
<tr>
<td>M Weber et al. [8]</td>
<td>Identifying illicit bitcoin users</td>
<td>Transaction features</td>
</tr>
<tr>
<td>Y Hu et al. [9]</td>
<td>Detecting Money Laundering</td>
<td>Graph embeddings</td>
</tr>
<tr>
<td>H Yin et al. [10]</td>
<td>Identifying illicit bitcoin users</td>
<td>Transaction features</td>
</tr>
<tr>
<td>L Nan et al. [11]</td>
<td>Mixing service detection</td>
<td>Graph embeddings</td>
</tr>
</tbody>
</table>

Majority researchers have focused on limited categories of
illicit users and shorter periods.
Table 2. Datasets used in published bitcoin studies

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Accessibility</th>
<th>Features</th>
<th>Categories</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chainalysis [12]</td>
<td>Private</td>
<td>9</td>
<td>exchange, gambling, hosted wallet, merchant services, mining, mixing,</td>
<td>198,097,356</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ransomware, scam, tor market or other</td>
<td></td>
</tr>
<tr>
<td>UIUC [13]</td>
<td>Public</td>
<td>0</td>
<td></td>
<td>37,450,461</td>
</tr>
<tr>
<td>BitcoinPonzi [14]</td>
<td>Public</td>
<td>11</td>
<td>Ponzi, Non-ponzi</td>
<td>6432</td>
</tr>
<tr>
<td>R Portnoff et al. [15]</td>
<td>Private</td>
<td>2</td>
<td>Sex offender, Ordinary</td>
<td>753,929</td>
</tr>
<tr>
<td>D Ermilov et al. [16]</td>
<td>Private</td>
<td>238</td>
<td>Service, gambling, mixer, exchange, pool, darknet</td>
<td>244,030,115</td>
</tr>
</tbody>
</table>

Table 3 gives the popular ML models for bitcoin studies.

Table 3. ML classifier used in published bitcoin studies

<table>
<thead>
<tr>
<th>ML models</th>
<th>Research Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Trees (DT)</td>
<td>[17]</td>
</tr>
<tr>
<td>Bagging Classifier</td>
<td>[18]</td>
</tr>
<tr>
<td>Gradient Boosting</td>
<td>[12]</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>[19]</td>
</tr>
<tr>
<td>Logistic regression (LogReg)</td>
<td>[17]</td>
</tr>
</tbody>
</table>

From Table 1 in literature, it is evident that the implementation of a reliable and secure illegal user detection system is a major concern for privacy and security in Bitcoin. Existing works have not focused on a broad spectrum of illegal activities that are conducted on Bitcoin. Additionally, existing datasets to are unsuitable for machine learning as their focus is narrow or are proprietary. In this respect, in Section 3, describes the data collection methodology to overcome the issue of data availability in public datasets. Section 4 discusses the proposed methodology for a classifier that could identify a broad spectrum of illicit users on Bitcoin.

3 Materials and Methods

As available datasets in literature (see Table 2) have shortcomings, procedure described in Section 3.1 was used to extract features mentioned in Section 3.2 of Bitcoin entities. Hardware and software configuration used for data collection is given in Section 3.3.

3.1 Data collection and preprocessing

Bitcoin blockchain dataset in raw form was obtained from VJTI Blockchain lab. This raw data was then converted to CSV files using the blockchain parser. Table 4 shows the three "csv" files of the processed dataset.

Table 5. List of Features

<table>
<thead>
<tr>
<th>Feature symbol</th>
<th>Feature description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>Total active life of the wallet</td>
</tr>
<tr>
<td>$A_w$</td>
<td>Total addresses of the wallet</td>
</tr>
<tr>
<td>$A_v$</td>
<td>Average number of incoming transactions received by an address of a wallet</td>
</tr>
<tr>
<td>$R$</td>
<td>Ratio of Transaction count and address count: gives the average number of times an address of the wallet was reused for a transaction</td>
</tr>
</tbody>
</table>

Table 5. List of Features

Features extracted for each entity would allow a classifier to learn the categorization of each wallet into one of the 28 categories. Bitcoin entities were identified using an API [20]. The flowchart for the proposed work is illustrated in Figure 2.

Table 6. Types of Bitcoin entities in dataset

<table>
<thead>
<tr>
<th>affiliatemarketing</th>
<th>blackmail</th>
<th>bomb</th>
<th>bond</th>
<th>criminals</th>
<th>cybersec</th>
<th>darkmarket</th>
<th>donations</th>
<th>exchange</th>
<th>explorer</th>
<th>faucet</th>
<th>gambling</th>
<th>laundering</th>
<th>microworker</th>
<th>miner</th>
<th>mixer</th>
<th>p2plender</th>
<th>p2pmarket</th>
<th>paymentgateway</th>
<th>ponzi</th>
<th>pools</th>
<th>ransomwares</th>
<th>scams</th>
<th>sextortionist</th>
<th>trading</th>
<th>Unclassified</th>
<th>videosharing</th>
<th>wallets</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>53</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>16</td>
<td>48</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>53</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>28</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>9</td>
<td>1592</td>
<td>1</td>
<td>8</td>
</tr>
</tbody>
</table>

3.2 Feature extraction

Based on the structure of Bitcoin (see Figure 1), features to train the classifier were extracted (see Table 5).

3.3 Experimental setup

Experiments were performed on a single core 1 TB Intel(R) Xeon(R) Silver 4114 CPU@2.20GHz.

---

4https://www.vjti-bct.in/
5https://github.com/pranavn91/blockchain
6https://github.com/pranavn91/blockchain/blob/master/walletexplorer-api
4 Classification of Illicit entities in Bitcoin

The mathematical model of the proposed classifier is given in Section 4.1 with steps used for training it listed in Section 5.3.

4.1 Mathematical model

Each training sample \((x_i, y_i)\) of dataset \(D\) is with \(m\) features and total \(n\) samples are present in the dataset. Hence, \(D = \{(x_i, y_i)\} (|D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R})\). The proposed tree ensemble model uses \(K\) additive functions to predict the output.

\[
\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F},
\]

where

- \(\mathcal{F} = \{f(x) = wq(x) \mid q : \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T\}\): set of regression trees
- \(q\): tree structure mapping an \(x^{(i)}\) to its leaf index
- \(T\): leaves count in the tree
- \(f_k\): \(q\) having leaf weights \(w\)
- \(w_i\): score on \(i^{th}\) leaf

For each \(x^{(i)}\) the decision rules of \(q\) classify it into the leaf nodes and calculate the final prediction by \(\sum w\) i.e. summing up the score in the corresponding leaves. The obtain the optimal model \(\phi\), the loss \(\mathcal{L}(\phi)\) is minimized by following regularized objective.

\[
\mathcal{L}(\phi) = \sum_{i} l(\hat{y}_i, y_i) + \sum_{k} \Omega(f_k)
\]

where \(\Omega(f) = yT + \frac{1}{2} \|w\|^2\)

4.1.1 Optimization. As the loss function given in Eq. 2 cannot be optimized using standard optimization techniques, the model is trained in an additive manner specified in [21]. Eq. 2 is modified as Eq 3.

\[
\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)
\]

where,

- \(\hat{y}_i^{(t)}\): prediction of the \(i\)-th instance at the \(t\)-th iteration
- \(f_t\): \(q\) having leaf weights \(w\)

5 Experimental study

The proposed model in Section 4 was evaluated on dataset described in Table 6 for the experimental study with metrics (Section 5.2).

5.1 Description of experiment

Comparative study was performed of popular non-parametric ML models in literature - Decision trees (DT) and Random Forest (RF) (see Section 3) with proposed model for evaluating classification accuracy on dataset (Table 6).

5.2 Metrics

Given the true positives \(t_p\), true negatives \(t_n\), type I error \(f_p\) and type II error \(f_n\) obtained from observing \((\hat{y}_i, y_i)\), following metrics were used.

\[
\text{Sensitivity}(S) = \frac{t_p}{t_p + f_n}
\]

\[
\text{Specificity}(S_p) = \frac{t_n}{t_n + f_p}
\]

\[
\text{Accuracy}(A) = \frac{t_p + t_n}{t_p + t_n + f_p + f_n}
\]

\[
\text{Prevalence}(P) = \frac{t_p + f_n}{t_p + t_n + f_p + f_n}
\]

5.3 Experimental results and discussion

Dataset was split in ratio 4:1 for training and evaluation. Optimal classifier parameters were identified using random grid search (caret package in R) [k-fold cross-validation, up sampling]. Table 7, and 8 give performance of the classifiers on the train set and Table 9 gives performance of the classifier on the Test set.

<table>
<thead>
<tr>
<th>Model</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision Trees</td>
<td>0.94</td>
<td>0.06</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.62</td>
<td>0.07</td>
</tr>
<tr>
<td>Proposed Model</td>
<td>0.79</td>
<td>0.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision Trees</td>
<td>0.02</td>
<td>0.0004</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Proposed Model</td>
<td>0.02</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision Trees</td>
<td>0.73</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.77</td>
</tr>
<tr>
<td>Proposed Model</td>
<td>0.66</td>
</tr>
</tbody>
</table>

\[
\text{Detection rate} = \frac{t_p}{t_p + f_n}
\]

\[
\text{Mean Specificity} = \frac{t_n}{t_n + f_p}
\]

\[
\text{No Information rate} = \frac{f_p + f_n}{t_p + t_n + f_p + f_n}
\]

\[
P\text{-value} = 0.05 - 0.95
\]

\[
\text{Kappa} = \frac{t_p + t_n - f_p - f_n}{t_p + t_n + f_p + f_n}
\]
Table 10. Time taken by classifiers for training

<table>
<thead>
<tr>
<th>Model</th>
<th>Type of Timing</th>
<th>User</th>
<th>System</th>
<th>Elapsed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree</td>
<td>Everything</td>
<td>357.2</td>
<td>0.29</td>
<td>357.5</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td>3.66</td>
<td>0</td>
<td>3.66</td>
</tr>
<tr>
<td>RF</td>
<td>Everything</td>
<td>2011.6</td>
<td>65.4</td>
<td>20382.2</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td>17.14</td>
<td>0.38</td>
<td>17.5</td>
</tr>
<tr>
<td>Proposed model</td>
<td>Everything</td>
<td>130928.1</td>
<td>50.6</td>
<td>16359.4</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td>821.3</td>
<td>0.3</td>
<td>102.8</td>
</tr>
</tbody>
</table>

The proposed model outperforms other popular tree-based classifiers on 13 out of 14 metrics with the exception of metric “accuracy”. The performance of the test set is slightly below par, albeit with a lower standard deviation than other models (see Table 9). As the proposed model has a higher number of parameters, it needs additional training time (see Table 10).

6 Conclusion and Future works

Various services and entities have become active in the Bitcoin space. Although Bitcoin has created new avenues for business models, the pseudo-anonymity has attracted even illegal operations to operate using Bitcoins. Individual differences between licit-illicit are observed in their transactions viz. amount transferred/received, and so on. Using Blockchain features were engineered and extracted to train an ensemble tree-based model to classify licit from illicit. Additionally, a dataset was created of 2059 real-life entities using Bitcoin. These included 28 categories of users who were involved in illegal and legal businesses on Bitcoin. The proposed tree-based classifier could identify 66% of users in the correct category. Additional features need to be engineered to improve accuracy and training time, which are tasks earmarked for future work.

References