Combating money laundering with machine learning – applicability of supervised-learning algorithms at cryptocurrency exchanges

Eric Pettersson Ruiz
CGI, Stockholm, Sweden, and

Jannis Angelis
KTH Royal Institute of Technology, Stockholm, Sweden and
IFN Research Institute of Industrial Economics, Stockholm, Sweden

Abstract

Purpose – This study aims to explore how to deanonymize cryptocurrency money launderers with the help of machine learning (ML). Money is laundered through cryptocurrencies by distributing funds to multiple accounts and then reexchanging the crypto back. This process of exchanging currencies is done through cryptocurrency exchanges. Current preventive efforts are outdated, and ML may provide novel ways to identify illicit currency movements. Hence, this study investigates ML applicability for combatting money laundering activities using cryptocurrency.

Design/methodology/approach – Four supervised-learning algorithms were compared using the Bitcoin Elliptic Dataset. The method covered a quantitative analysis of the algorithmic performance, capturing differences in three key evaluation metrics of F1-scores, precision and recall. Two complementary qualitative interviews were performed at cryptocurrency exchanges to identify fit and applicability of the algorithms.

Findings – The study results show that the current implemented ML tools for preventing money laundering at cryptocurrency exchanges are all too slow and need to be optimized for the task. The results also show that while not one single algorithm is most suitable for detecting transactions related to money-laundering, the specific applicability of the decision tree algorithm is most suitable for adoption by cryptocurrency exchanges.

Originality/value – Given the growth of cryptocurrency use, this study explores the newly developed field of algorithmic tools to combat illicit currency movement, in particular in the growing arena of cryptocurrencies. The study results provide new insights into the applicability of ML as a tool to combat money laundering using cryptocurrency exchanges.

Keywords Anti-money laundering, Machine learning, Supervised learning, Algorithms, Cryptocurrency

Paper type Research paper

1. Introduction

Cryptocurrencies are a financial asset class increasing in use, with traded cryptocurrencies having a market capitalization of over $3 trilion in 2021 (Ossinger, 2021). They have an

Journal of Money Laundering Control
Emerald Publishing Limited
1368-5201
DOI 10.1108/JMLC-09-2021-0106

impact on financial transactions and payments (Hairudin *et al.*, 2020), investment strategies and portfolio diversification (Bouri *et al.*, 2016) and even on promoting entrepreneurship (Kshetri and Voas, 2018). However, as virtual and decentralized assets cryptocurrencies enable publicly validated transactions while its users remain anonymous and difficult to trace (Choo, 2015). This makes cryptocurrencies suitable for avoidance of anti-money laundering (AML) measures (Adam and Fazekas, 2018; World Bank, 2018; Mugarura and Ssali, 2021; Teichmann and Falker, 2021). For instance, Grauer and Updegrave (2021) estimate that in 2020 almost $10bn were laundered through cryptocurrencies. Given their nature, Campbell-Verduyn (2018) argues that the current AML efforts need to be modified as they do not address the potential wrong-doing of Bitcoin and similar alt-coins such as Ethereum, Ripple or Litecoin. Financial institutions, including cryptocurrency exchanges, rely on traditional rules-based systems to flag suspicious transactions. (Gonzálvez-Gallego and Pérez-Cárceles, 2021; Saiedi *et al.*, 2020). Cryptocurrency exchanges are entities that offer exchange services to cryptocurrency users, normally against payment of a certain fee. These exchanges allow the users to sell or buy cryptocurrencies with fiat currency (Houben and Snyers, 2018). This approach has showed high false positive rates and low detection rates, implying that the systems do not capture the entire complexity of the problem and have a tendency to be bias (Canhoto, 2021; Lorenz *et al.*, 2020; Savage *et al.*, 2016).

Several recent studies (Jullum *et al.*, 2020; Alarab *et al.*, 2020) exhibit promising results in the field of machine learning (ML) when detecting money laundering, which might be incentive enough to modernize current AML efforts (Chen *et al.*, 2018; Weber *et al.*, 2018). However, it remains unclear which ML algorithm is suitable to use for preventing money laundering. Therefore, this study investigated the research question of whether ML algorithms are suitable for combating money laundering activities using cryptocurrencies. This was explored by adapting different ML algorithms to the context and comparing their performance based on three identified key evaluation metrics: F1-score, recall and precision.

The study first investigated the performance differences between algorithms. It then explored how cryptocurrency exchanges work with ML and explored the fit of the investigated algorithms with the ongoing money laundering prevention.

2. Literature background

This section presents the relevant literature. It covers cryptocurrency characteristics and ML use to address money laundering and then presents supervised learning as a ML approach to combat money laundering. Finally, it discusses the literature around algorithm selection.

2.1 Machine learning and anti-money laundering

As defined by several authors (Choo, 2015; Lansky, 2018), cryptocurrencies must meet several criteria cantered around a lack of central authority, established system of creating and owning the currency based on cryptography and transactions verified through a third party. Linking transactions to specific individuals or entities is difficult, which means that cryptocurrencies can facilitate the move, launder and hiding of funds from illegal activities (Adam and Fazekas, 2018). Cryptocurrencies pose a serious threat to AML efforts which has made supranational organizations such as the Financial Action Task Force (FATF) and the EU join forces to develop new techniques to combat money laundering (Canhoto, 2021; Nanyun and Nasiri, 2021). The European Banking Authority (EBA) have noted that cryptocurrencies can be misused for tax evasion, money laundering, regulation avoidance and illegal trade of weapons or drugs, among other illicit practices (EBA, 2014; Maupin, 2017). Even when criminal activities are detected, enforcement against cross-border

jurisdictions is difficult, making cryptocurrency use for illegal purposes a global challenge in terms of monitoring and law enforcement (Buchanan, 2004; Nanyun and Nasiri, 2021).

ML is a prominent technique as it can analyze large amounts of data to find undiscovered patterns of illicit financial behavior. ML is defined as a branch of computer science and artificial intelligence that makes use of data and algorithms to imitate the way humans learn, progressively improving its accuracy and acting without being explicitly programmed. Canhoto (2021) explains that ML outperforms the traditional ways of addressing money laundering of predefined rules, as it relies on deductive reasoning. Predefined rules, on the other hand, are pre-established by an analyst that reasons how suspicious activities of money laundering look like. This implies that the human cantered approach cannot solely be used to uncover new topologies (Savage *et al.*, 2016). Lorenz *et al.* (2020) further explain that the rules-based approach leads to high false positive rates and low detection rates, which in other words, means it is prone to be bias. Despite the known ML advantages, many financial institutions still rely on a rules-based approach. Jullum *et al.* (2020) explain that this is because ML algorithms mostly work as a black box, not revealing why a specific transaction has been flagged as suspicious. Due to compliance regulations, institutions are obligated to explain the reason behind a flagging and hence prefer to not use ML. This has three negative consequences. First, it requires firms to recurrently update and adapt the rules. As time passes, the difficulty of evaluating the performance of the rules and identifying every single exception increases. Second, the rules-based approach is typically too simplistic and does not capture the complexity of money laundering. Finally, the number of transactions increases exponentially over time, so it is imperative to reduce the number of false positives to ensure a reliable system (Chen *et al.*, 2018). Given these limitations, Weber *et al.* (2018) and Alarab *et al.* (2020) encourage the development of new methods for addressing money laundering and Savage *et al.* (2016) suggest supervised learning classifications based on identified ability.

2.2 Supervised learning

There are several ML approaches used for predicting licit from illicit transactions. One is supervised learning, which is further divided in two branches, one of them being classification. Classification is a commonly used technique for identifying the belonging of a data-point to a particular class. It identifies group memberships of different inputs. Classification models have performed well in the past when detecting illicit transactions related to money laundering as showed by Lorenz *et al.* (2020), Savage *et al.* (2016) and Weber *et al.* (2018). Furthermore, Canhoto (2021) explains that classification models are suitable when input and output are known, e.g. when licit and illicit transactions have been preidentified. But financial institutions do not typically reveal such information, which makes it difficult to perform classifications on real data. This follows the study by Zhang and Trubey (2018), where results on the advantages of supervised learning were inconclusive. In contrast, Feng *et al.* (2019) classification model, trained on real data showed good results accuracy. To identify the best performing classification model, the correct algorithm must first be known. Alarab *et al.* (2020), Chen *et al.* (2018), Lorenz *et al.* (2020), Rivera *et al.* (2015), Savage *et al.* (2016), Weber *et al.* (2018) and Zhang and Trubey (2018) have all shown that for problems related to classifying licit and illicit transactions, four algorithms have achieved high performance: logistic regression, random forest, support vector machines and decision tree. Logistic regression is commonly used for problems where the input data is comprised of multiple variables and when there are many outliers in the data set (Sperandei, 2014). However, the algorithm tends to underperform when the data set is heavily imbalanced as the boundary decision gets skewed toward the majority class

(Alarab *et al.*, 2020). Decision tree is simple to understand and to interpret and is readily visualized (Weber *et al.*, 2018). Outputs can be easily motivated with simple Boolean logic. A key reason for why financial institutions do not use ML is because of algorithm's black box nature. However, decision tree can lead to overfitting, since the models can be overly complex and not generalize well (Alarab *et al.*, 2020). Furthermore, if the classes are imbalanced the models might be biased. In contrast, random forest algorithms are based on decision tree, but are composed of several trees and is less interpretative and reduces the bias and over-fitting of the decision tree (Breiman, 2001). Finally, support vector machines perform well in high dimension spaces where there are many parameters within the data-points (Zhang and Trubey, 2018). This flexibility means that they perform well when the data set is unknown (Chen *et al.*, 2018), but that they tend to underperform when the data set is large and there are more outliers than in small data sets.

3. Method

The study was operationalized by quantitatively exploring the performance of supervised learning algorithms when classifying licit from illicit transactions that are stored in the Elliptic Bitcoin data set. The explored algorithms were run on a specific data set and their respective F1, recall and precision scores compared. To ensure applicability of the results, the study also explored how cryptocurrency exchanges work with ML and the fit of the investigated algorithms with the ongoing money laundering prevention. This was captured through semi-structured interviews at two Swedish cryptocurrency exchanges and used as complementary to the main quantitative data-based analysis. The research process had two parts: a quantitative part based on an algorithmic experiment and a second complementary and qualitative part based on interviews. The algorithmic experimental approach was adopted to explore how analysis of the investigated supervised learning algorithms are conducted when detecting transactions related to money laundering. To analyze the performance of the investigated algorithms when detecting money laundering related transactions, the *Elliptic Bitcoin Dataset* (kaggle.com/ellipticco/ellipticdataset) was used. It is the largest sampled publicly available data set containing information of transactions being licit or illicit and used by law enforcement and financial institutions (Elliptic, 2021). The data set used is composed of 49 graphs obtained directly from the Bitcoin blockchain at different moments in time. Each graph represents a directed acyclic graph containing two weeks of data from the transactions made in that time period. Furthermore, each graph starts from one transaction and includes subsequent related transactions of the blockchain. Each transaction represents a real transaction of the Bitcoin blockchain and has a unique ID that is determined by its predecessor. All transaction IDs correspond to the transactions in the Bitcoin blockchain given that the data set is extracted from the publicly available Bitcoin blockchain (blockchain.com/explorer). This makes it possible to ensure the validity of data set transactions. In turn, it shows the consistency of the data set as it represents the same data that can be obtained directly from the Bitcoin blockchain. Focusing on transactions, licit transactions are those belonging to exchanges, wallet providers, miners and other licit services. Illicit transactions are those related to scams, malware, terrorist organizations, ransomware, Ponzi schemes and other fraudulent activities. Each category has 166 associated features that determine whether they are licit or illicit. Note that licit and illicit transactions are already flagged by the data set. Although the data set is divided to three categories of licit, illicit and unknown, the latter category was not included in the creation or testing of the ML model. The investigated algorithms are based on supervised learning, so there is a requirement of ground truth to each datapoint used. Since there are transactions with an unknown label, supervised learning cannot be used. The distribution of the adapted

Elliptic Bitcoin Dataset had 46,564 transactions – 4,545 illicit and 42,019 licit. Once the distribution was identified, cross-validation methods were selected for training and testing the supervised learning model.

3.1 Preprocessing of data

With the objective of classifying transactions as licit or illicit and given the hardware resource limitations, the classification task was only performed on a subset of the entire data set. The unknown transactions were disregarded when preparing, validating and testing the ML learning model. To further obtain reliable results, the bias variance trade-off was addressed by implementing a Stratified k-fold Cross Validation technique. A stratified version was used because of the data set skewness. When using a stratified version every fold is guaranteed to have a proportional split of the two categories. By not having a proportional split may lead to folds including transaction of a single category. That would negatively impact the ML-model performance as it does not identify how a transaction from the missing category looks like. Furthermore, the k value was set to 10, based on other studies using similar sized data sets (Brownlee, 2018; Hastie *et al.*, 2008). The data set was composed of 46 564 nodes, including licit and illicit transactions. 90% of those nodes were used for training the ML model and 10% used for testing. The training set was then split into ten folds. Each fold was composed of a proportional amount of licit and illicit transactions. Every fold held in total 4,191 nodes, where 9.8% were illicit and 90.2% licit. The model was then trained in ten iterations, each iteration using a different fold as validation set to identify model performance. Once all iterations were finished, the model used an average score of every iteration to avoid unwillingly choosing a split that by chance is beneficial for certain algorithms. The last step was to test the model by classifying unseen data from the Test set.

3.2 Choice of algorithms and hyperparameters

Several studies have shown that there is no ideal algorithm for classification purposes. Chen *et al.* (2018), Rivera *et al.* (2015), Savage *et al.* (2016) and Zhang and Trubey (2018) all encourage to try several algorithms on the same data set and then decide which one to choose. As there is no conclusive research done on the Elliptic Bitcoin dataset, four different algorithms, that have performed well in other similar data sets are used in this study (Alarab *et al.*, 2020; Canhoto, 2021; Weber *et al.*, 2018). These are logistic regression, random forest, support vector machines and decision tree. The hyper-parameters used for each algorithm were as follows: logistic regression – default; random forest - number of trees in the forest = 100, default; support vector machines – linear kernel, c = 1, default; decision tree – default. To identify the highest performing hyper-parameters, they were tuned several times. However, for the majority of the algorithms, the default values given by sci-kit returned the best scores. For random forest and support vector machines, a couple of hyper-parameters had to be slightly modified. The evaluation metrics used for analyzing algorithm performance are: Precision, Recall and F1-Score. These metrics are commonly used when the data is highly distributed (Brownlee, 2021), as is the Elliptic Bitcoin dataset. The metrics are based on the confusion matrix shown in Figure 1.

The matrix summarizes to what extent a specific ML algorithm is able to predict the outcome of a specific data point. The top left cell represents the number of data points that were correctly classified as positive. The top right cell represents those inputs that were classified as negative but were in reality positive. The false positives are those data points that are negative but were incorrectly classified as positive. Finally, the true

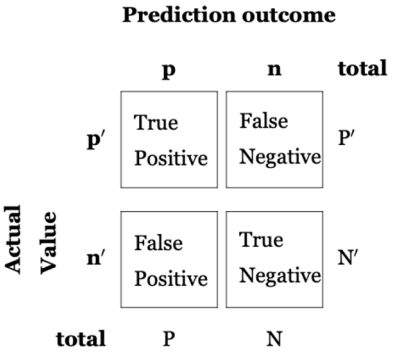

Figure 1.
Confusion matrix

negatives represent the number of negative data points correctly classified as negative. Precision, Recall and F1-score are metrics that make use of this information in the following way:

Precision is the quantification of the number of positive class predictions that belong to the positive class (Flach and Kull, 2015). This is expressed as:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Recall is the quantifiable number of positive class predictions made out of all positive examples in the data set (ibid.). This is expressed as:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

F1-score is the harmonic mean of Precision and Recall (ibid.). This is expressed as:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

3.3 Complementary qualitative method element

Exploring the applicability of algorithms at cryptocurrency exchanges, complementary qualitative interviews were made. These were semi-structured, performed on two managers (interviewee A1, B1) from two different exchanges (Company A, B). The interviews were used to capture views from senior managers (Goffin *et al.*, 2019) on how money laundering preventions was being conducted to identify applicability of the supervised learning algorithms. The two respondents were selected based on their professional roles at the currency exchanges. They were contacted through email and interviewed in person. The interviews lasted about 1 h each and were recorded as well as notes taken. Questions covered views and efforts of money laundering prevention and the use of ML tools to do so.

4. Results and analysis

The study explored the use of ML to combat money laundering activities using cryptocurrencies. This was done by adapting different supervised learning algorithms and comparing their performance based on three established key evaluation metrics: F1-score, recall and precision. The performance of the algorithms is presented in Figure 2. There are three plots on the graph, where each line represents one of the three metrics Precision, Recall and F1-Score. Precision measures the percentage of transactions flagged as illicit that were correctly classified. Recall measures the percentage of actual illicit transactions that were correctly classified. F1-score is the harmonic mean of Precision and Recall. The *y*-axis of the graph indicates the score, ranging from 0 to 1, where 0 indicates 0% and 1, 100%. Finally, the *x*-axis shows the algorithms used to train the ML model: logistic regression, random forest, support vector machines and decision tree.

On precision, the random forest classifier had the highest score of 0.998. This indicates ability to correctly classify the flagged illicit transactions. support vector machines and decision tree algorithms both obtained a similar score of around 0.9. In comparison, the logistic regression classifier compared underperforms slightly with a score difference of 0.1–0.2, which in practice is considered non-significant. It is reasonable to assume that the four algorithms perform equally well for this particular data set. For recall scores, the algorithms showed different results than for the precision score. decision tree achieved the highest score, followed closely by the other tree-based algorithms, random forest. support vector machines obtained a lower score of 0.81 and logistic regression an even lower score of 0.767. Although the ranking of the algorithms differs slightly from the precision scores, it has little impact on choosing the algorithm suitability. However, it is noteworthy that for all algorithms the precision score is higher than the corresponding recall score for the same algorithm, except for the decision tree classifier. For the decision tree, the recall score is slightly higher than the precision score, indicating that this algorithm is better at avoiding false-negative results. Moreover, F1-scores from the algorithms were similar to the results obtained from the recall score. The difference in performance is not greater than 0.13, which indicates that the choice of algorithm has little impact when classifying data transactions. The two tree-based algorithms random forest and decision tree performed better than the other algorithms, but it is a marginal difference. In practice the choice of algorithm has little impact, possibly because of the chosen split used during the cross-validation step where the two algorithms perform better. With a different split another algorithm may have fared better.

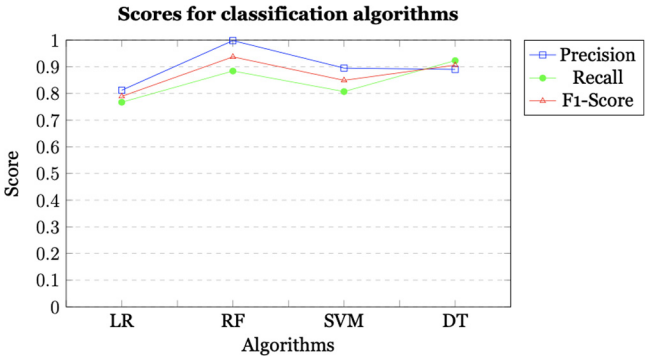

Figure 2.
Performance of four different algorithms

4.1 Exploring cryptocurrency exchanges

Illicit transactions are regarded by the interviewees as transactions that are originated or received from/by an address with a connection to criminal activity, made by an individual or organization. This includes terrorist organizations, cyber criminals, human trafficking, etc. Several authorities including, FATF and the US Treasury department have their own lists of illegal cryptocurrency addresses. As noted by the interviewed managers, these lists are in turn used by their companies to stop illegal activity on their platforms. In addition, they also make use of tools such as rules-based systems, Chainalysis and Valega to analyze the blockchain. The two respondents explained that with these tools they were able to flag, identify flagged addresses and freeze the coins that are sent or received from the same. Respondent B1 explained that, apart from the flagged addresses by the US treasury department, Chainalysis uses ML to help them flag many more addresses. They do not use ML for any other purpose partly because they do not have that need and partly due to compliance regulations. However, they expect to use it more. As for Company A, who is a bigger, they also make use of supervised learning to further analyze and discover patterns of criminal behavior. A1 noted that they have been helping the national authorities by sharing information about their activities and through coaching. They have been doing this for the past few years through their information sharing events and workshops. Both respondents claimed there was room for improvement for ML. A1 explained that investigations performed by compliance teams take too long as they are most likely performed by individuals who are working for different exchange platforms or geographies, so sometimes the funds are able to be swiftly transferred to an account before they are being flagged as fraudulent. The exchange cannot react fast enough.

5. Discussion

This study has showed that the tools used for addressing transactions with illicit nature vary. For company B, the rules-based system was prioritized because they had issues with national compliance regulations as they could not motivate to the Financial Authority (2021) why they should use ML. This supports the study by Jullum *et al.* (2020), where because the majority of algorithms work as a black box, it is not always possible to explain why a certain transaction is flagged and frozen. Due to the legal requirements financial institutions abide by, they are required to be able to explain the reasons behind any frozen transactions. The typical preference is to adopt a simpler approach that performs relatively worse but meets the compliance regulations. The reasoning of using rules-based systems because of compliance fulfilment is paradoxical, as shown by Weber *et al.* (2018). They illustrate the money laundering scandals occurred at 1MDB's and Danske Bank, who used this approach and were still sanctioned with millions of euros. Therefore, while rules-based systems allow exchanges to justify why transaction flagging occurs and hence meet any compliance regulations, it does not address the fact that they are not doing everything they can to prevent money laundering from happening. Importantly, they still get sanctioned. Furthermore, as highlighted by Chen *et al.* (2018), the rules-based system is prone to return high rates of false positive transactions as the number of transactions increase exponentially over time. These issues are properly addressed by ML as showed by Alarab *et al.* (2020).

As for the performance of the classification model, the scores for all the algorithms were almost the same for the evaluation metrics used. As previously mentioned, when the algorithmic performance is as narrow as in this case, it is practically impossible to draw a definitive conclusion on which algorithm is best for the classification model to predict licit and illicit transactions. Similarly, Alarab *et al.* (2020), Chen *et al.* (2018), Lorenz *et al.* (2020), Rivera *et al.* (2015), Savage *et al.* (2016), Zhang and Trubey (2018) and Weber *et al.* (2018)

have all found good performance from the used algorithms, which reinforces the point that all algorithms are usable. In any case, given that one reason an exchanges had chosen to use a rules-based system because of the algorithms' black box nature, decision tree is a prime candidate as a solution and to reduce any money laundering. It is simple to understand and visualize as outputs easily are motivated with Boolean logic. This classifies decision tree as a white box in practice, making exchanges able to explain why a certain transaction has been flagged.

Despite the implemented decision tree having a lower F1-score than the Random Forest algorithm and its tendency to overfit data compared (Breiman, 2001), the trade-off may be worthwhile between a smaller score and being able to use ML and overall improve the detection of money laundering related transactions. As for the other two algorithms, support vector machines performs well in high dimensional spaces where the input data has multiple features. In the data set, the inputs have 166 associated features which motivates the high performance. The logistic regression algorithm is also known to perform well with multiple variables and when there are many outliers (Sperandei, 2014). The high performance is thus motivated for similar reasons as for support vector machines. However, because of the black box nature of these algorithms, it might not be suitable for exchanges. However, if the regulations change in the future and the requirements for motivating why a transaction is frozen are abolished, then logistic regression could be an attractive algorithm to use since it is good at eliminating outliers. This means it is good at reducing the false positives, which was one of the big pitfalls of rules-based system as noted by Chen *et al.* (2018).

For the applicability of the investigated algorithms, Company A, which is a mature company in terms of life-span and resources, is currently employing ML for several reasons. In addition to using ML tools such as Chainalysis and Valega for analyzing the blockchain, they perform supervised learning techniques to find deeper patterns in the data to spot more fraudulent activities performed at the platforms. Such use of supervised learning is encouraged by Canhoto (2021), Feng *et al.* (2019), Lorenz *et al.* (2020), Savage *et al.* (2016) and Weber *et al.* (2018) and shows an awareness at the exchange of the advantages of ML. Despite Company A's extensive use of ML for preventing money laundering, the respondent's view was that there is significant room of improvement for ML in the area of AML. There had been issues with flagging transactions in real time, which needs urgent improvement. As explained by Interviewee A1, criminals are able to quickly clean money by sending them to multiple addresses and then cashing out in an exchange.

5.1 Conclusion

To summarize, the study explored which ML algorithms are suitable for combating money laundering activities using cryptocurrency. In answering the research question, the study's findings showed that the Random Forrest algorithm outperformed the others and is the most effective identifier of illicit transactions. It had a 3% better F1-score and 11% better Precision score compared to the second best performing algorithm Decision Tree. However, given that the performance differences between the four algorithms are very low, it would be an overstatement to conclude that one algorithm is more suitable in all circumstances. Despite using a stratified ten-cross validation method to minimize potential overfitting of the classification model, it would be overstating the fact to assume that the model is completely free from overfitting. No current method guarantees the absolute absence of overfitting. With respect to exchange applicability of the algorithms, the decision tree algorithm is advantageous since it is not a black box algorithm, the exchange is able to motivate to the authorities why a certain transaction has been flagged as illegal. Furthermore, if such requirement to motivate is removed, then the random forest algorithm

would be the most suitable for exchanges. Regardless, the results show that all investigated algorithms outperform the traditional rules-based systems. Given that the original data set contained unlabeled data, a suitable approach would be to perform a semi-supervised learning model, where new algorithms have been developed such as Label Propagation (Kontonatsios *et al.*, 2017). Also, external metrics such as NMI and AMI would not be possible to use as the data set lacks ground truth. This can turn out to be a problem when evaluating how well the clustering algorithms perform as there is no previous knowledge on which class the unknown datapoints belong to. Nevertheless, internal metrics are usable for assessment of cohesion and separation.

5.2 Contribution

On the theoretical contribution, the study results show that the used algorithms for training the supervised learning model have all performed well on an unexplored data set- the Elliptic Bitcoin dataset. The obtained results show similar results to the ones obtained on other data sets, investigated by Canhoto (2021) and Savage *et al.* (2016). This indicates that the use of supervised learning techniques is encouraged to be used in AML efforts, as also specified by Feng *et al.* (2019). Furthermore, the aim of this study was to identify the best performing algorithm when detecting illegal bitcoin transactions. Given that this study's implemented algorithms, as well as Canhoto (2021) and Savage *et al.* (2016), have similar Precision, Recall and F1-score, it might be more convenient to look at different evaluation metrics to define which is the most suitable algorithm for the study's purpose, as the current metrics are not sufficient. Perhaps looking at evaluation metrics based on computational cost and time complexity would give a more accurate result. Given that the algorithms are equally good at categorizing licit from illicit transactions it might be more interesting to see which algorithm performs the job with fewer resources in terms of computational cost and time complexity.

On practical implications of using ML to prevent money laundering at cryptocurrency exchanges, despite the usage of rules-based systems at one of the exchanges, the study results indicate that ML is a better tool for spotting transaction irregularities. In particular, the Decision tree algorithm seems to be the most attractive approach as it is a white box algorithm and follows the compliance regulations and achieves high classification scores. Nevertheless, the current ML techniques used to flag illicit transactions at the exchanges are slow and do not react to made transactions in real time. Practically, it is advisable to focus and resources on improving the reaction time of the algorithms, rather than finding the most suitable algorithm, as they all perform more or less the same.

5.3 Limitations and further research

The Classification model produced in this research was trained on about 50,000 transactions over the Bitcoin network. Although the model found successful performance for all algorithms, it would be prudent to rerun the experiment on a larger data set over a longer time. This could incorporate a semi-supervised learning model including further unknown labelled datapoints using cloud computing for greater computational power.

References

Adam, I. and Fazekas, M. (2018), "Are emerging technologies helping win the fight against corruption in developing countries?", *Pathways for Prosperity Commission Background Paper Series*, Vol. 21, pp. 2-28.

Alarab, I., Prakoonwit, S. and Nacer, M. (2020), "Comparative analysis using supervised learning methods for anti-money laundering in bitcoin", *Proceedings of the 5th International*.

Bouri, E., Molnár, P., Azzi, G., Roubaud, D. and Hagfors, L. (2017), "On the hedge and safe haven properties of Bitcoin: is it really more than a diversifier?", *Finance Research Letters*, Vol. 20, pp. 192-198.

Breiman, L. (2001), "Random forests", *Machine Learning*, Vol. 45 No. 1, pp. 5-32.

Brownlee, J. (2018), "A gentle introduction to k-fold cross-validation", available at: https://machinelearningmastery.com/k-fold-cross-validation/ (accessed 11 June 2021).

Brownlee, J. (2021), "Tour of evaluation metrics for imbalanced classification", available at: https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification (accessed 11 June 2021).

Buchanan, B. (2004), "Money laundering – a global obstacle", *Research in International Business and Finance*, Vol. 18 No. 1, pp. 115-127.

Campbell-Verduyn, M. (2018), "Bitcoin, crypto-coins and global anti-money laundering governance", *Crime, Law and Social Change*, Vol. 69 No. 1, pp. 283-305.

Canhoto, A. (2021), "Leveraging machine learning in the global fight against money laundering and terrorism financing: an affordances perspective", *Journal of Business Research*, Vol. 131, pp. 441-452.

Chen, Z., Van Khoa, L., Na Teoh, E., Nazir, A., Kandasamy Karuppiah, E. and Sim Lam, K. (2018), "Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review", *Knowledge and Information Systems*, Vol. 57 No. 2, pp. 245-285.

Choo, K. (2015), "Cryptocurrency and virtual currency", in Lee Kuo Chuen, D. (Ed.), *Handbook of Digital Currency*, Elsevier, New York, NY, pp. 283-307.

EBA (2014), "EBA opinion on 'virtual currencies'", available at: www.eba.europa.eu/sites/default/documents/files/documents/10180/657547/81409b94-4222-45d7-ba3b-7deb5863ab57/EBA-Op-2014-08%20Opinion%20on%20Virtual%20Currencies.pdf?retry=1 (accessed 9 November 2021).

Elliptic (2021), "Elliptic data set- bitcoin transaction graph", available at: www.kaggle.com/ellipticco/elliptic-data-set (accessed 11 June 2021).

Feng, Y., Li, C., Wang, Y., Wang, J., Zhang, G., Xing, C., Li, Z. and Lian, Z. (2019), "Anti-money laundering (AML) research: a system for identification and multi-classification", in Ni, W., Wang, X., Song, W. and Li, Y. (Eds), *International Conference on Web Information Systems and Applications*, Qingdao, pp. 169-175.

Financial Authority (2021), "Periodic financial information", available at: www.fi.se/en/markets/issuers/periodic-financial-information/ (accessed 9 November 2021).

Flach, P. and Kull, M. (2015), "Precision-recall-gain curves: pr analysis done right", in Cortes, C., Lee, D., Sugiyama, M. and Garnett, R. (Eds), *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems*, *Montreal*, pp. 838-846.

Goffin, K., Åhlström, P., Bianchi, M. and Richtner, A. (2019), "The quality of case study research in innovation management", *Journal of Product Innovation Management*, Vol. 36 No. 5, pp. 586-615.

Gonzálvez-Gallego, N. and Pérez-Cárceles, M. (2021), "Does goodness of governance dissuade citizens from using cryptocurrencies?", *Economics and Sociology*, Vol. 14 No. 1, pp. 11-27.

Grauer, K. and Updegrave, H. (2021), "The 2021 crypto crime report", available at: https://go.chainalysis.com/2021-Crypto-Crime-Report.html (accessed 21 September 2021).

Hairudin, A., Sifat, I., Mohamad, A. and Yusof, Y. (2020), "Cryptocurrencies: a survey on acceptance, governance and market dynamics", *International Journal of Financial Econnomics*, pp. 1-27.

Hastie, T., Tibshirani, R. and Friedman, J. (2008), *The Elements of Statistical Learning: data Mining, Inference and Prediction*, Springer, New York, NY.

Houben, R. and Snyers, A. (2018), "Cryptocurrencies and blockchain", available at: www.europarl.europa.eu/cmsdata/150761/TAX3%20Study%20on%20cryptocurrencies%20and%20blockchain.pdf (accessed 21 September 2021).

Jullum, M., Løland, A., Huseby, R., Ånonsen, G. and Lorentzen, J. (2020), "Detecting money laundering transactions with machine learning", *Journal of Money Laundering Control*, Vol. 23 No. 1, pp. 173-186.

Kontonatsios, G., Brockmeier, A., Przybyła, P., McNaught, J., Mu, T., Goulermas, J. and Ananiadou, S. (2017), "A semi-supervised approach using label propagation to support citation screening", *Journal of Biomedical Informatics*, Vol. 72, pp. 67-76.

Kshetri, N. and Voas, J. (2018), "Blockchain in developing countries", *IT Professional*, Vol. 20 No. 2, pp. 11-14.

Lansky, J. (2018), "Possible state approaches to cryptocurrencies", *Journal of Systems Integration*, Vol. 9 No. 1, pp. 19-31.

Lorenz, J., Silva, M., Aparicio, D., Ascensao, J. and Bizarro, P. (2020), "Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity", *2020 ACM International Conference on AI in Finance*, *New York, NY*, 15–16 October 2020, pp. 1-8.

Maupin, J. (2017), "Mapping the global legal landscape of blockchain and other distributed ledger technologies", *Planck Institute for Comparative Public Law & International Law*.

Mugarura, N. and Ssali, E. (2021), "Intricacies of anti-money laundering and cyber-crimes regulation in a fluid global system", *Journal of Money Laundering Control*, Vol. 24 No. 1, pp. 10-28.

Nanyun, N. and Nasiri, A. (2021), " "Role of FATF on financial systems of countries: successes and challenges", *Journal of Money Laundering Control*, Vol. 24 No. 2, pp. 234-245.

Ossinger, J. (2021), "Crypto world hits $3 trillion market cap as Ether, Bitcoin gain", available at: www.bloomberg.com/news/articles/2021-11-08/crypto-world-hits-3-trillion-market-cap-as-ether-bitcoin-gain (accessed 9 November 2021).

Rivera, E. West, J. and Suplee, C. (2015), " Addressing AML regulatory pressures by creating customer risk rating models with ordinal logistic regression", available at: https://support.sas.com/resources/papers/proceedings15/SAS2603-2015.pdf (accessed 22 September 2021).

Saiedi, E., Broström, A. and Ruiz, F. (2020), "Global drivers of cryptocurrency infrastructure adoption", *Small Business Economics*, Vol. 57No No. 1, pp. 353-406.

Savage, D. Wang, Q. Chou, P. Zhang, X. and Yu, X. (2016), " Detection of money laundering groups using supervised learning in networks", available at: www.researchgate.net/publication/305779558_Detection_of_money_laundering_groups_using_supervised_learning_in_networks (accessed 22 September 2021).

Sperandei, S. (2014), "Understanding logistic regression analysis", *Biochemia Medica*, Vol. 24 No. 1, pp. 12-18.

Teichmann, F. and Falker, M. (2021), "Money laundering via cryptocurrencies", *Journal of Money Laundering Control*, Vol. 24 No. 1, pp. 91-101.

Weber, M., Giacomo Domeniconi, G., Chen, J., Weidele, D., Bellei, C., Robinson, T., Zhang, Y. and Trubey, P. (2018), "Machine learning and sampling scheme", *Computational Economics*, Vol. 54, pp. 1043-1063.

World Bank (2018), "Blockchain governance and regulation as an enabler for market creation inemerging markets", available at: https://documents1.worldbank.org/curated/en/636421540530725523/pdf/131343-BRI-EMCompass-Note-57-Blockchain-Governance-v1-PUBLIC.pdf (accessed 9 November 2021).

Zhang, Y. and Trubey, P. (2018), "Machine learning and sampling scheme: an empirical study of money laundering detection", *Computational Economics*, Vol. 54, pp. 1043-1063.

Further reading

Bellei, C. (2019), "The elliptic data set", available at: https://medium.com/elliptic/the-elliptic-data-set-opening-up-machine-learning-on-the-blockchain-e0a343d99a14 (accessed 11 June 2021).

Dyntu, V. and Dykyi, O. (2018), "Cryptocurrency in the system of money laundering", *Baltic Journal of Economic Studies*, Vol. 4 No. 5, pp. 75-81.

European Commission (2018), "Anti-money laundering and counter terrorist financing", available at: https://ec.europa.eu/info/business-economy-euro/banking-and-finance (accessed 11 June 2021).

Evans, R. (2021), "Bitcoin: UK banks are getting tough on crypto, but money-laundering rules are the real problem", available at: https://theconversation.com/bitcoin-uk-banks-are-getting-tough-on-crypto-but-money-laundering-rules-are-the-real-problem-159651 (accessed 11 June 2021).

Murrar, F. and Barakat, K. (2021), "Role of FATF in spearheading AML and CFT", *Journal of Money Laundering Control*, Vol. 24 No. 1, pp. 77-90.

Corresponding author
Jannis Angelis can be contacted at: jannis.angelis@indek.kth.se