Table of Contents

[Executive Summary 2](#_Toc155176981)

[Introduction 2](#_Toc155176982)

[Real Life Cryptocurrency Money Laundering Case 3](#_Toc155176983)

[Solution 3](#_Toc155176984)

[Real-Time Data 4](#_Toc155176985)

[Legacy Integration 4](#_Toc155176986)

[Trade-Off Contemplation 5](#_Toc155176987)

[Data 5](#_Toc155176988)

[Adaptable for Future Changes 5](#_Toc155176989)

[Challenges 5](#_Toc155176990)

[Ethical Considerations 6](#_Toc155176991)

[Conclusion 6](#_Toc155176992)

[References 6](#_Toc155176993)

[Appendix 7](#_Toc155176994)

Tackling Cryptocurrency Money Laundering with Machine Learning Algorithm

# Executive Summary

The issue of money laundering has become extremely important in the digital age, especially since it includes cryptocurrency. Criminals use digital currencies and internet banking to conceal illicit funds and make them seem as if they originated from reliable sources. The high level of anonymity and confidentiality provided through cryptocurrencies facilitates an effective concealment of assets and identities. Despite the fact that crypto transactions are recorded on the blockchain, the identities of the parties involved remain concealed, making it more challenging for law enforcement to track down and identify those who are engaged in illegal conduct. This project suggests a dual approach combining machine learning predictive models and clustering, also real-time monitoring to effectively prevent bitcoin money laundering. By classifying transactions into threat or non-threat categories based on factors which includes the quantity and frequency of transfers in a day, the machine learning model aims to detect and identify potential money laundering activities. Clustering is used to categorise similar data points based on similar characteristics. Clustering and machine learning classification combination enables a deeper examination of cryptocurrency transactions, improving the detection of suspicious clusters. The CoinGecko API will be used for this test-run because it provides real-time cryptocurrency market data, such as trade volumes, price changes, and market values, each of them is crucial for the machine learning algorithm's efficiency. Legacy integration is ensuring seamless compatibility with current systems, procedures, and infrastructure. With this strategy, businesses can deploy the solution without needing to make major infrastructure changes. The proposed solution also addresses ethical considerations by minimising bias in machine learning through good quality of data and human supervision. Involving people in the process ensures that crucial choices, for instance, contacting law enforcement, must be thoroughly examined, and validated to prevent misconceptions.

# Introduction

Money laundering is a method of concealing illegal money from criminal activities such as drug trafficking, corruption, bribery, fraud, human trafficking, etc, so it will look like it appeared to come from a legitimate source. In this digital era, criminals can send and withdraw money without being detected by using online banking and cryptocurrency. Because cryptocurrencies provide a high level of confidentiality, it is possible to disguise identities and assets for the purpose of money laundering (Institute of International Finance, 2017). The identities of the persons involved are kept private in a digital currency’s transaction, and the transaction is recorded on a blockchain. Every transaction has unique address, and it is generated in order to track down every transaction happening inside the blockchain. Furthermore, untraceable cross-border money transfers become possible by cryptocurrencies without banks and government agencies being aware. Eliminating intermediaries or third parties is one of the reasons why it can happen, and the regulations that regulate cryptocurrency is not the same as financial institutions. In addition, criminals can utilise cryptocurrency exchanges to turn launder their money into other cryptocurrencies or fiat, therefore, following the movement of the illegal money can be difficult. Despite the fact that cryptocurrencies offer a high level of anonymity, it is important to understand that they cannot guarantee total confidentiality.

## Real Life Cryptocurrency Money Laundering Case

The BBC reports that through 2021, criminals have laundered $8.6 billion (£6.4 billion) in cryptocurrencies, an increase of 30% from the previous year. The elimination of specific companies, including specific exchanges that criminals recall as being commonplace, would have a significant impact because most digital currencies are laundered through these. Certain aspects of the digital currency structure are being used to launder money from illegal activity, especially drug sales. In addition, with the increase of ransomware threat, cryptocurrencies are used by the perpetrator as its payment method, with them receiving payment in cryptocurrencies. The usage of cryptocurrencies in money laundering operations has been growing, and many criminal organisations depended on the digital currency as a payment method during the Covid-19 outbreak, according to a recent trend report (BBC, 2022).

The US Department of Justice (DOJ) arrested two Chinese nationals in the year 2020 for laundering the cryptocurrency worth more than $100 million that North Korean hackers had stolen. According to the DOJ, in order to wash the money, they used eCurrency transactions. A certain amount of the money stolen was utilised to pay for the infrastructure needed for North Korean campaign (Leuprecht et al., 2022).

# Solution

In order to prevent future problem regarding the use of cryptocurrency for money laundering, 3 solutions will be implemented in this project, first is to do a classification, second is clustering, and third is real-time monitoring.

Machine learning predictive model will be used to detect and identify potential money laundering activity. In order to solve that, a classification will be performed by defining class feature with threat or non-threat, and some of features that need to be in the dataset will be amounts of money transferred and number of times transferred (Jullum et al,. 2020). After collecting the dataset, the machine learning model will be prepared with several steps starting from data preprocessing, splitting the dataset for training and testing the model, training the model, validating the model, evaluate and fine-tuning to improve the accuracy of the model. After it finished, the model will be saved and used for classification.

The next step is clustering where similar data points can be grouped together based on similar characteristics. Clustering will be used to cluster or combine addresses in the same cluster that are most likely controlled by the same entity. This can be a powerful tool for identifying money laundering activity. Clustering will be used on unlabelled data, so it can detect hidden characteristics from the data, and in this project, it will choose numbers of transfer and amount of transferred in a day as the feature to be studied (Gao, 2009).

In order to do a real-time monitoring, the machine learning that has been saved will be connected to an API, then run a data pre-processing, so the data used for predicting is in the same characteristics as in training. After that, an alert will be placed, and the criteria are unusual amount of transaction or high activity on transfer. CoinGecko, a website that provides real-time cryptocurrency data and information, is the API used for the test run. It provides data for machine learning algorithm, such as trading volume, quantities, market value, and the changes of price over time. This project will utilise the API documentation provided by CoinGecko (Appendix 1) in order to access the CoinGecko API. In the documentation, it will show how to use the API, how to extract the data from the API, and how to request for authentication.

The steps are first, machine learning model will be trained using historical data regarding cryptocurrency exchange and transaction to learn about the patterns of money laundering, the ML model must have the best ability to distinguish threat or non-threat activity. Next, the CoinGecko API will be used for the machine learning model and clustering to do a real-time monitoring, and to clarify, after being trained, ML does not need label for prediction, so unlabelled data can be used. After the ML model detect suspicious activity, in order to track the transaction, clustering result will be analysed to find the address of the transaction based on the suspicious cluster. For instance, in 1 case the ML model detect a suspicious activity, the address of that transaction will be collected and used to find the other address in the cluster that might be connected to pinpoint the perpetrator. By combining clustering and machine learning it can give more complete picture of the data and identify suspicious crypto transactions.

By implementing the solution above, there are several parties that can be benefited (Comply Advantage, 2021):

* Financial Institutions: It supports them in identifying and mitigating the risks linked to cryptocurrency-related money laundering operations, ensuring compliance with anti-money laundering (AML) regulations.
* Regulatory Authorities: It provides more efficient control of the cryptocurrency market by helping in the detection of suspicious activities and tracing a possible illegal transaction.

# Real-Time Data

The solution will utilise a real-time data in order to do a real-time monitoring. The reason for gathering real-time data is because it will allow detecting any suspicious activity immediately and take action to prevent next potential crime activity. The data will be collected using the API of the client, however, if the client does not have an API, web scrapping method will be used to get the real-time data for the solution to work. In the test run, CoinGecko API is used, the step of collecting the data is already provided in solution section.

# Legacy Integration

The solution is considering the legacy integration. The compatibility and integration with existing systems, processes, and infrastructure are taken into consideration throughout implementation. The objective is to provide a solution that supports the organization's processes in the present time and enhances their anti-money laundering initiatives in the cryptocurrency area by taking legacy integration into account. One thing that might be incompatible might be the data (Ashraf, 2020). Data in current financial systems may come in a variety of structures and formats, however, in machine learning, the data can be transformed by using data preprocessing method, so the data can be used in this solution. In addition, machine learning models need training data to identify patterns and generate accurate predictions. The model's performance can be improved, and its ability to recognise changing money laundering strategies can be made better by utilising historical data which are transaction data from legacy systems into the training process.

# Trade-Off Contemplation

The proposed solution will consider the trade-off regarding the implementation and the frictionless onboarding. The entity that will potentially use this solution are financial institutions, cryptocurrency exchange, fintech startups, regulatory or law authorities, so most of the client must have a modern system that can adapt to this solution and compatible with their technology. If they have legacy system, it will be quite challenging. For example, the API, if they did not have an API, it will be quite time-consuming to do a real-time monitoring, the solution for that is to collect the data manually or use a method called web scrapping, to extract the data from the website. Data also will be pre-processed after scrapping. Therefore, adopting this solution for entity that have modern technology will not be a problem, but those who have old legacy system, could be challenging, but it is possible even though it might take a while.

# Data

The data from the CoinGecko API used for the solution is bulletproof according to Andrey Balyasnikov, Head of Product at Zerion because the data is unbiased, reliable, and accurate. While CoinGecko is offering real-time the cryptocurrency data, the CoinGecko API does not utilise blockchain technology. The information in their website or application is subjected to be change when new transactions are added to the database. The API gives users access to the most recent market data, but it cannot deliver the same level of immutability and trust as blockchain technology do. The data in CoinGecko API will keep changing because it gives a real-time data, so it will be updated every time, for instance, when the price going up or down, market changes, etc. However, the trustworthiness of the data that will be used for clients for their monitoring is based on their system.

# Adaptable for Future Changes

The machine learning classification algorithm and clustering method is adaptable because it can be improved with new dataset and new technique in the future. Retraining the model with a whole new data regularly can increase the capability of the model and adapt to a new pattern in money laundering activities because every improvement not only introduce more efficiency and benefit, but also newer of way of doing crimes. This model is also scalable, so handling cryptocurrency transaction that growing in volumes will not become a problem. The development team or the company that manages the solution is mainly responsible for integrating the future modifications. The development team is in charge of maintaining the latest technology and making adjustments to solve new issues. In addition, feedbacks from customers also helps for the integration for addressing specific needs.

# Challenges

Some challenges that may be faced in the future include regulatory compliance and false positives and negatives. First, the solution must comply with a lot of financial regulations and in the future, more regulations will be implemented, so maintaining this solution in order to follow the regulations will be an ongoing challenge. Second is the false positives and negatives in machine learning outcome (McKinsey & Company, 2022). False positives are where the legal transaction reported as illegal and false negatives are where illegal transaction is undetected. Minimizing these 2 will be crucial because too many false positives could lead to false alarm, while false negative can allow money laundering to go unnoticed.

# Ethical Considerations

The ethical considerations from this project (UK Statistics Authority, 2021):

* Bias in Machine Learning: The machine learning will be trained with data in good quality to avoid bias, so machine learning can judge more fairly. In this case, the machine learning will be used to classify transaction that might be connected to money laundering association. Avoiding the machine learning to be bias is very crucial because it might trigger a wrong alert and wrongfully judging the transaction that does not associated with illegal activities as a cluster that poses a threat.
* Human Supervision: Although machine learning algorithms are capable of generating useful outcomes, human involvement is still required to double check and confirm the results from the system. The decisions that have big impact or consequences, for instance, alerting law authorities for possible money laundering situations, must be reviewed and include human investigation and their judgement, therefore, there will be no misunderstanding and capturing or accusing the wrong person for something that they did not do.
* Web Scrapping: As for clients who do not have an API, permission will be requested and ensured to web scrap their website to capture real-time data.

# Conclusion

Cryptocurrencies money laundering poses significant difficulties because it provides a high degree of anonymity. The proposed solution combines machine learning classification and clustering to identify possible money laundering activity by leveraging real-time analysis. In order to bring the model to its fullest capability, it must be run regularly and train under different dataset, therefore, the model will recognise suspicious activity easily in the future. Ethical considerations to minimise biases and guarantee human supervision must be applied for more fair decisions. By increasing security and compliance in the cryptocurrency market, it will benefits businesses, investors, regulatory agencies, financial institutions, and society as a whole.

# References

Ashraf, S. (2020, June 23). *Data Integration with Legacy Systems – White Elephants of Business*. Retrieved from DataIntegrationInfo: https://dataintegrationinfo.com/data-integration-legacy-systems/

BBC. (2022, January 26). *Crypto money laundering rises 30%, report finds*. Retrieved from BBC News: https://www.bbc.co.uk/news/technology-60072195

Comply Advantage. (2021, April 22). *Crypto Laundering - Money Laundering Through Cryptocurrency Exchanges*. Retrieved from Comply Advantage: https://complyadvantage.com/insights/money-laundering-crypto-exchanges/

Gao, Z. (2009, September). Application of cluster-based local outlier factor algorithm in anti- money laundering. In 2009 International Conference on Management and Service Science (pp. 1-4). IEEE.

Institute of International Finance. (2017). *DEPLOYING REGTECH AGAINST FINANCIAL CRIME.* *https://www.iif.com/portals/0/Files/private/32370132\_aml\_final\_id.pdf*

Jullum, M., Løland, A., Huseby, R. B., Ånonsen, G., & Lorentzen, J. (2020). Detecting money laundering transactions with machine learning. Journal of Money Laundering Control, 23(1), 173-186.

Leuprecht, C., Jenkins, C., & Hamilton, R. (2023). Virtual money laundering: policy implications of the proliferation in the illicit use of cryptocurrency. Journal of Financial Crime, 30(4), 1036-1054.

McKinsey & Company. (2022, October 7). *The fight against money laundering: Machine learning is a game changer*. Retrieved from McKinsey & Company: https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/the-fight-against-money-laundering-machine-learning-is-a-game-changer

UK Statistics Authority. (2021, October 26). *Ethical considerations in the use of Machine Learning for research and statistics*. Retrieved from UK Statistics Authority: https://uksa.statisticsauthority.gov.uk/publication/ethical-considerations-in-the-use-of-machine-learning-for-research-and-statistics/pages/7/#pid-general-ethical-principles

# Appendix

Appendix 1

CoinGecko API Documentation Website: <https://www.coingecko.com/en/api/documentation>