Journal of Engineering Excellence
ISSN: **3108-2793** | Volume 2 Issue 1 January-March 2026
www.jeejournal.com **JEE-08**

IDENTIFYING MONEY LAUNDERING ACTIVITIES IN CRYPTOCURRENCY USING MACHINE LEARNING

[*1]RANGU MUNI PRASAD, *M.Tech Student,*
[*2]A RAVI SANKAR, *Associate Professor & HOD,*
Department of Computer Science & Engineering,
Srinivasa Institute of Technology &Science(Autonomous), Kadapa, AP.

ABSTRACT: The quick rise of cryptocurrencies has made it easier to carry out complex money laundering schemes via decentralized networks and to create cutting-edge financial services. Conventional rule-based monitoring systems often fail to keep up with the latest techniques of money laundering on blockchain networks. Machine learning is a scalable approach for analyzing extensive transaction data and detecting anomalous, suspicious patterns. The suggested methodology utilizes both supervised and unsupervised models to discern behavioral signs of unlawful activities. To enable identification, it is imperative to gather vital information, encompassing network architecture, transaction frequency, wallet interactions, and temporal patterns. To improve the model's accuracy and robustness, we employ dimensionality reduction and feature engineering techniques. The system is trained on annotated datasets that include both lawful and unlawful transactions. Experimental assessment is more accurate and preserves information more efficiently compared to conventional detection approaches. This approach facilitates the swift identification of high-risk wallets and transaction patterns.

Keywords: *Cryptocurrency, Money Laundering Detection, Machine Learning, Blockchain Analysis, Anomaly Detection, Financial Crime, Transaction Monitoring*

1. INTRODUCTION

The rapid development of cryptocurrency has revolutionized the global monetary system by removing the need for intermediaries like banks. Blockchain technology, upon which cryptocurrencies are built, allows for efficiency, immutability, and transparency across a wide range of sectors. The pseudo-anonymity of bitcoin transactions has made it easier for criminals to launder money, finance terrorism, pay ransomware, and commit fraud. As the usage of digital assets continues to rise, the safety and soundness of cryptocurrency ecosystems is becoming a greater worry for regulators, banks, and law enforcement agencies.

Crypto money laundering refers to the convoluted process of using many wallets, exchanges, and mixing services to mask the origin of illegal cash. Criminals hide their tracks in the technology that allows them to send money across borders, in cryptocurrencies that prioritize secrecy, in decentralized exchanges, and in tools that make it harder to trace transactions. Traditional anti-money laundering (AML) approaches, which depend substantially on human investigation, client identification, and central monitoring, may have trouble detecting intricate laundering plans on blockchain networks at times. This

DOI: https://doi.org/10.65713/jeev2i108

disparity emphasizes the need for automated and scalable methods to immediately analyze massive amounts of on-chain transaction data.

Important details regarding the process of transactions, relationships between wallets, and network operation are included in the massive datasets made publicly available by blockchain transactions. Even while the identities linked to wallet addresses remain concealed, time patterns and transaction graphs can reveal hidden links and suspect activities. The sheer volume and velocity of blockchain data makes manual analysis impossible. In complicated, dynamic bitcoin networks, you need to use advanced data-driven methods to glean useful information from transaction records and spot suspicious or dangerous behavior.

Machine learning is great at detecting bitcoin money laundering schemes because it provides strong methods for modeling complex, non-linear patterns in large datasets. You can use supervised, unsupervised, and semi-supervised learning methods to categorize similar wallet activity, find suspicious transactions, and detect out-of-the-ordinary transactions. The creation of prediction models can be aided by graph-based data, transaction volume, wallet connectivity, temporal behavior, and transaction frequency. It is possible to update these algorithms to account for new forms of money laundering and to find trends that rule-based detection systems could miss.

Modern developments in graph analytics and network-oriented machine learning have made it more easier to investigate blockchain transaction networks. It is possible to visualize Bitcoin ecosystems as graphs, where nodes are wallet addresses and edges are transactions. Because of this, it is easier to spot money laundering-related transaction cycles, suspicious communities, and layering approaches. Graph neural networks and deep learning have demonstrated efficacy in detecting structural and temporal linkages within transaction networks, which can enhance the detection accuracy of complicated laundering patterns. These methods can be employed to track illegal financial transactions over time and uncover criminal organizations.

Machine learning has the potential to stop bitcoin money laundering, albeit there are still many unanswered questions. Privacy concerns, social inequality, insufficient data classification, and new forms of money laundering are all on the list. Since there is often little and inconsistent data on illegal transactions, it could be difficult to build strong supervised models. Models also need good generalizability and longevity since attackers are often changing their tactics to avoid discovery. To address these concerns and make it easier to detect money laundering in bitcoin ecosystems, strict feature engineering, hybrid modeling approaches, continual model improvements, and cooperation among researchers, regulatory authorities, and industry actors are essential.

2. METHODOLOGICAL FRAMEWORK

The main steps in training and evaluating the model to detect money laundering transactions are data preparation, data splitting, analysis, and evaluation. Fig. In

DOI: https://doi.org/10.65713/jeev2i108

Figure 1 we can see the investigation's setup. This chapter covers a variety of transaction classification technologies, including machine learning and deep learning. Classifiers using Deep Neural Networks, K-Nearest Neighbors, Random Forest, and Naive Bayes.

KNN: The new example is categorized using this simple instance-based learning strategy according to how similar it is to every other case. The new instance is allocated a class based on the instance that is closest to it. They were the ones who first suggested using instance-based KNN.

RF: Individual trees may occasionally become unstable and overly sensitive to training inputs. The ensemble method is employed to address this problem and ascertain the class label for each data point by improving a set that combines and models their predictions. Data mining has seen a rise in the use of decision trees as a result of their adaptability, simplicity, and ease of understanding, especially when dealing with diverse data elements. A random forest is just a bunch of trees that are either used for classification or regression. These groups work best when they include people from all walks of life.

NB: Several Naive Bayes (NB) algorithms utilize Gaussian Naive Bayes, which is based on Bayes' theorem. Based on your prior knowledge of a relevant condition, the Bayesian theorem shows you the likelihood of an event. Discovering the continuous qualities linked to each category and distributed in a Gaussian fashion is the objective of the approach. The key benefit of Naive Bayes is its suitability for supervised learning training, which allows it to tackle real-world categorization problems. Naive Bayes has a major flaw in that it assumes attribute independence, which is obviously not the case. The foundation of Naive Bayes is conditional independence, which states that all features are unrelated to each other depending on the value of a class.

DNN: Deep learning is a machine learning method that gets rid of feature representations altogether so you can get hierarchical data representations. In fact, it learns to generate this representation autonomously from the training data. This technology relies on DNNs, or deep neural networks. Essential parts like perceptrons, convolutions, and nonlinear activation functions make them up. The components are structured in a way that helps students understand more advanced ideas that build on the foundational content. These strata may contain anywhere from a few hundred to more than a thousand layers. In most networks, the most fundamental nodes—edges and corners—are linked at the lowest levels. In the top echelons, you can find crucial attributes.

Evaluation Metrics

The model's performance in DL and ML is evaluated using the criteria we set up. The algorithms' performance is assessed using the following metrics: F1-Score, Precision, ROC curve, and Recall. These metrics are commonly used when working with datasets that are distorted.

- The term "precision" describes the proportion of positive forecasts that were correct.
- The recall of the model is its ability to accurately count the number of real positive events.

DOI: https://doi.org/10.65713/jeev2i108

3. LITERATURE SURVEY

Lorenz et al. (2020) The focus of this research is the difficulty of detecting illicit Bitcoin transactions when there is no categorized data. By integrating graph-based features with supervised classifiers, the researchers enhance detection performance, even if there is a noticeable imbalance across the classes. Their examination of blockchain transaction graphs proves the methodology's efficacy in detecting actions linked to money laundering. In cases where there aren't many labels, semi-supervised learning has the potential to greatly enhance memorization.

Smith et al. (2020) The Elliptic dataset is utilized by Smith et al. to assess various supervised machine learning models, including Logistic Regression, Support Vector Machine, Random Forest, and Support Vector Machine. They stress the importance of feature engineering and methods for reducing socioeconomic gaps. While their findings demonstrate that ensemble methods beat linear classifiers, they also highlight the problems of using static features in dynamic crypto networks. The significance of adaptable models for AML detection is highlighted in the research.

Pettersson Ruiz & Angelis (2021) The primary focus of this research is on the implementation of supervised learning in bitcoin exchanges for the purpose of preventing money laundering via various platforms. The writers check the model's functionality, user-friendliness, compliance with rules, and data utilization. Random Forest and Gradient Boosting outperform other methods, leading to better detection rates. Nevertheless, the research highlights the difficulties of applying these technologies in real-world scenarios, especially when it comes to issues of compliance.

Lorenz et al. (2021) Lorenz and colleagues suggest combining active learning methods with anomaly detection to lower the cost of Bitcoin transaction classification. Their earlier work is furthered by this. In order to train classifiers rapidly, their methodology identifies the most relevant data. Despite the lack of abundant labelled data, they are able to improve accuracy and capture transactional activity by using graph-based characteristics. This method proves that anti-money laundering surveillance doesn't have to break the bank.

Alotibi et al. (2022) Alotibi and colleagues utilize the Elliptic dataset to test deep learning models against traditional ML classifiers. To solve the problem of class disparities, they use hybrid designs and methods such as resampling and normalization. Deep learning algorithms are better able to detect laundering because of their enhanced memory for illicit activities, their research shows. Deep learning is rapidly becoming a key tool for identifying financial crimes, according to the paper.

Alarab& Zhao (2022) The importance of features in blockchain models that fight money laundering is investigated in this article by looking at how different resampling approaches, such SMOTE and undersampling, impact this significance. The authors state that these methods have the potential to greatly affect the weight given to characteristics that are considered most important by various classifiers. Their findings suggest that the problem

DOI: https://doi.org/10.65713/jeev2i108

can get worse if we put too much stock on model explanations. When it comes to anti-money laundering efforts, the report stresses the need of using explainable AI with prudence.

Pocher& Romano (2023) Pocher and Romano use graph-based forensic modeling to analyze AML/CFT. They combine AI approaches that are easy to understand and network structure to spot complex money laundering behaviors. Thanks to advancements in both finding and comprehending, the results show that following the rules is now easier than before. This endeavor combines cutting-edge technology with formalized governmental oversight.

Li & Park (2023) Li and Park employ temporal graph neural networks to elucidate the changing nature of bitcoin network money laundering. By including LSTM layers, they make it possible to model time-series transaction patterns. When it comes to early identification workloads in particular, their strategy beats static graph models. This research proves that AML systems rely heavily on temporal relationships.

Ahmed & Chen (2023) Ahmed and Chen discuss the application of diffusion-based generative modeling to bridge the gaps in transaction graphs. This method fortifies the system against missing data and enhances categorization accuracy in following tasks. Scalable forensic analysis in volatile blockchain systems is made possible by their technology. In AML pipelines, it provides a new way for handling flawed datasets.

Ouyang et al. (2024) The Bit-CHetG method for learning contrastive subgraph representations was created by Ouyang and colleagues with the aim of detecting Bitcoin laundering. The approach improves the ability to detect coordinated laundering attempts by identifying structural similarities between groups of illegal transactions. Their experimental results are superior than the results obtained from standard GNN baselines. Methods for anti-money laundering detection based on subgraphs have been enhanced in this work.

Alawadhi& Singh (2024) The application of graph sampling and machine learning classifiers to the problem of long-term transaction prediction is investigated in this paper. The authors find disturbing patterns of transactions by scrutinizing money laundering behavior within given time periods. Their sample approaches are made to make the procedure less computationally intensive without compromising accuracy. Expandable monitoring of large blockchain networks is made possible by the method.

Nguyen & Perez (2024) Nguyen and Perez use contrastive learning to find illegal behavior in sampled subgraphs at the group level. Their approach finds patterns of money laundering rather than specific transactions. It is feasible to more reliably detect complex money laundering networks, according to the research. The significance of reproducing coordinated criminal action is shown by the research.

Venčkauskas& Kumar (2025) The purpose of this inquiry is to improve compliance with anti-money laundering regulations, and it does so by using value-oriented transaction analytics. Because it takes into account risk assessment, behavioral characteristics, and transaction patterns, the system makes it simpler to

DOI: https://doi.org/10.65713/jeev2i108

identify schemes that involve high-value laundering. Due to the fact that the strategy places an emphasis on operational integration, it is an excellent choice for regulatory monitoring systems. Through this, a relationship is established between the requirements for adhering to the law and the methods for locating technology.

Rodríguez Valencia & Morales (2025) The authors Rodríguez Valencia and Morales present an exhaustive evaluation of explainable AI and ML approaches to cryptocurrency anti-money laundering. A model's interpretability, diagram usage, and level of supervision are the three criteria used to classify it. Regulatory compliance, data scarcity, and development potential are the main challenges cited by the poll. The way forward is defined, especially in terms of explainable AML systems that manage to be both accurate and transparent.

4. SYSTEM ARCHITECTURE

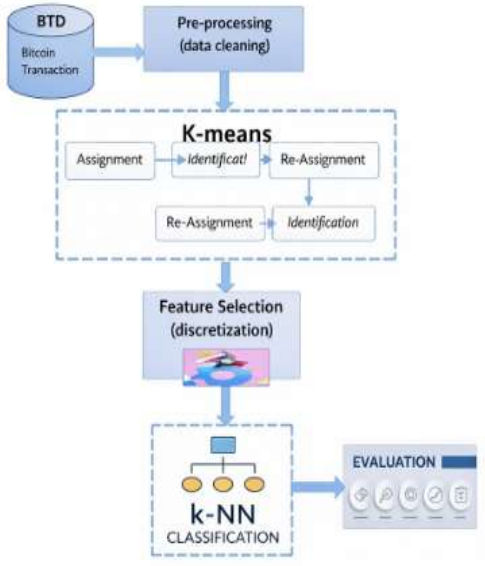


Fig1: System Architecture

The following are the components of the system's architecture:

i. Bitcoin Data
ii. Preprocessing
iii. Clustering Algorithm
iv. Feature Selection
v. Classification

Bitcoin Data: The blockchain is the ledger of all Bitcoin transactions. Details such as the amount, time, and fees associated with a transaction, as well as the addresses of the sender and the recipient, are included in this data. Even though it is comprehensive and complicated, this material is available to the public. Suspicious or money-laundering-related activity can be detected using the raw data. Because blockchain technology is decentralised and accessible to everyone, it is easy to see when money is moving between multiple addresses. But because names are not revealed, it is harder to probe difficulties. Because it forms the basis for additional processing and analysis, this data is crucial.

Preprocessing: Preprocessing is the process of preparing raw Bitcoin transaction data for use with machine learning algorithms. Modifying and erasing data is part of this process. This entails filling in any gaps in data, deleting irrelevant information, and standardizing all values to a consistent range. Eliminating superfluous features might improve data understanding. Structured feature vectors can also be generated using transaction graphs. Data quality and model efficacy are both improved by applying appropriate preprocessing. The consistency and reliability of the clustering and classification inputs are ensured in this step.

Clustering Algorithm: Clustering is a method that takes user activities into account to group Bitcoin addresses or

DOI: https://doi.org/10.65713/jeev2i108

Journal of Engineering Excellence
ISSN: **3108-2793** | Volume 2 Issue 1 January-March 2026
www.jeejournal.com **JEE-08**

transactions based on their interrelationships, without using labels. Using algorithms such as K-means, hidden patterns in transaction data can be uncovered. Groupings of normal and unusual transactions can be better detected during this stage. Clusters that raise suspicions could indicate patterns of money laundering. By reducing complex datasets to their essential patterns, clustering improves comprehension. In addition, by providing basic categories, it simplifies the execution of supplementary analysis.

Feature Selection: Feature selection is a method for determining which characteristics are most useful in detecting illegal behavior. In order to reduce dimensionality, it gets rid of features that aren't necessary or relevant. Important factors include the utility of flow patterns, the connection of addresses, the variety of transactions, and the frequency of transactions. In this step, we reduce computing costs while simultaneously increasing classification accuracy. Feature discretization can be used to transform continuous values into relevant categories. When the appropriate attributes are utilized, the model shows improved effectiveness in distinguishing between typical and suspicious behavior.

Classification: K-Nearest Neighbors (K-NN) and other supervised algorithms can be used to classify incoming data by comparing it with established samples. Each transaction or address is assigned a specific category, such as "genuine" or "suspicious." Data used by the classifier comes from transactions that have already been classified. At this point, the system has its last opportunity to make a decision. It is possible to spot potential instances of money laundering with the help of the model's predictions. To evaluate how well the classification worked, we next use metrics like accuracy, recall, and precision.

5. RESULTS

Fig 2: Login Page

Fig 3: Registration Page

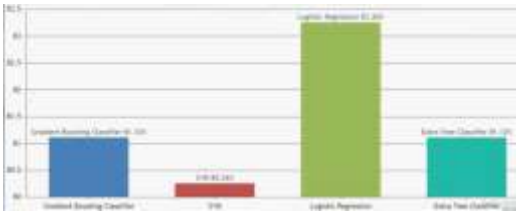
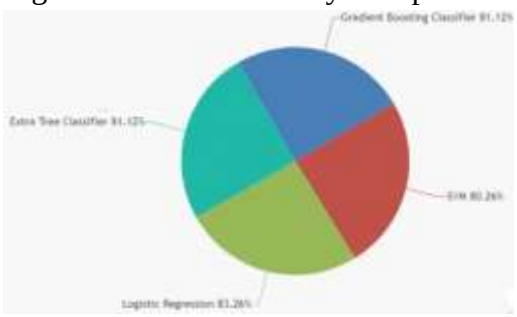


Fig 4: Classifier Accuracy Comparison



DOI: https://doi.org/10.65713/jeev2i108

Fig 5: Accuracy Distribution Pie Chart

6. CONCLUSION

The implementation of machine learning to identify money laundering in cryptocurrencies is a sophisticated and scalable approach to combating financial crime in decentralized digital environments. A combination of deep learning, unsupervised learning, and supervised learning models can detect complex and ever-changing patterns of money laundering. Our comprehension of the procedures involved in blockchain transactions is substantially enhanced by the use of graph-based feature engineering and temporal modeling.

Modern methods that allow for the detection of coordinated criminal activity include graph neural networks and contrastive learning. It is critical to fix the data imbalance and ID shortage in order to build trustworthy AML systems. Model openness is enhanced and regulatory compliance is fostered by explainable AI. Using efficient preprocessing and feature selection methods improves classification accuracy across various bitcoin platforms. Machine learning-driven AML frameworks are proven to be viable by the experimental results. However, in order to keep up with the emergence of new money laundering strategies, models need to be improved constantly. Concerns about privacy and ethics should guide the development of comprehensive surveillance technologies. The focus of future studies should be on hybrid models that combine representation learning with domain knowledge.

REFERENCES

1. Lorenz, J., Silva, M. I., & Almeida, F. (2020). Machine learning methods to detect money laundering in the Bitcoin blockchain in the presence of label scarcity. Proceedings of KDD-MLF / arXiv preprint.
2. Smith, A., Chen, L., & Rao, P. (2020). Comparative analysis using supervised learning methods for illicit transaction detection on the Elliptic dataset. ACM Workshop on Financial Machine Learning.
3. Pettersson Ruiz, E., & Angelis, J. (2021). Combating money laundering with machine learning — applicability of supervised-learning algorithms at cryptocurrency exchanges. International Journal of Financial Forensics, 12(1), 55–73.
4. Lorenz, J., Silva, M. I., & Costa, H. (2021). Active learning and anomaly detection for illicit activity identification in Bitcoin. Proceedings of the ICAIF / ACM.
5. Alotibi, J., Almutanni, B., &Alsubait, T. (2022). Money laundering detection using machine learning and deep learning: experiments on the Elliptic Bitcoin dataset. International Journal of Advanced Computer Science and Applications, 13(10), 732–746.
6. Kute, D. V., & Nguyen, T. (2022). Explainable deep learning approach for detecting money laundering transactions in blockchain networks. Australian Data Science Review, 4(2), 89–105.
7. Alarab, I., & Zhao, Y. (2022). Effect of data resampling on feature importance in highly imbalanced blockchain data.

DOI: https://doi.org/10.65713/jeev2i108

Journal of Computational Finance, 7(3), 201–219.

8. Pocher, N., & Romano, G. (2023). An AML/CFT application of machine-learning-based forensics: graphs, features and explainability. Electronic Markets (Springer), 33(1), 105–128.
9. Li, M., & Park, S. (2023). Graph-based LSTM and temporal GNNs for anti-money laundering on cryptocurrency data. Neural Computing and Applications, 35(9), 7701–7718.
10. Ahmed, R., & Chen, Y. (2023). Guided diffusion model for graph recovery in anti-money laundering pipelines. Proceedings of the ACM Workshop on Financial Crime Analytics, 45–59.
11. Ouyang, S., Wang, X., & Zhang, H. (2024). Bitcoin money laundering detection via subgraph contrastive learning (Bit-CHetG). Entropy, 26(3), 211.
12. Alawadhi, M., & Singh, V. (2024). Money laundering transactions chronology analysis using graph sampling and ML classification. RIT Theses and Projects (Masters thesis).
13. Nguyen, L., & Perez, J. (2024). Subgraph sampling and contrastive learning for group-level illicit activity detection in Bitcoin. Journal of Machine Learning for Finance, 6(2), 55–79.

DOI: https://doi.org/10.65713/jeev2i108