Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Machine Learning in Money Laundering Detection over Blockchain Technology

ALGIMANTAS VENČKAUSKAS[1], **ŠARŪNAS GRIGALIŪNAS**[1] **(Member, IEEE), LINAS POCIUS**[1], **RASA BRŪZGIENĖ**[1] **and ANDREJS ROMANOVS**[2] **(Senior Member, IEEE)**

[1]Department of Computer Sciences, Kaunas University of Technology, Studentu str. 50, 51368 Kaunas, Lithuania (e-mail: algimantas.venckauskas@ktu.lt; sarunas.grigaliunas@ktu.lt; linas.pocius@ktu.edu; rasa.bruzgiene@ktu.lt)
[2]Information Technology Institute, Riga Technical University, Riga, Latvia (e-mail: andrejs.romanovs@rtu.lv)

Corresponding author: Algimantas Venčkauskas (e-mail: algimantas.venckauskas@ktu.lt).

Research was conducted as part of the execution of Project "Mission-driven Implementation of Science and Innovation Programmes" (No. 02-002-P-0001), funded by the Economic Revitalization and Resilience Enhancement Plan "New Generation Lithuania."

ABSTRACT Layering through cryptocurrency transactions represents a sophisticated mechanism for laundering money within cybercrime circles. This process methodically merges illegal funds into the legitimate financial system. Blockchain technology plays a crucial role in this integration by facilitating the quick and automated dispersal of assets across various digital wallets and exchanges. Machine learning emerges as a powerful tool for analyzing and identifying illicit transactions within Blockchain networks; however, a significant challenge remains in the form of a gap in advanced pattern recognition algorithms. This paper introduces a novel machine learning-based approach called Value-driven-Transactional tracking Analytics for Crypto compliance (VTAC) for the detection of illegal crypto transactions via Blockchain. The approach combines machine learning algorithms with a pre-training process, normalization, model training, and a de-anonymization process to analyze and identify illicit transactions effectively. Experimental evaluations show VTAC's capability to detect illegal transactions with a 97.5% accuracy using the XG Boost model, outperforming existing methods with an accuracy of up to 95.9%. Key performance metrics, including precision, recall, and F1-score, consistently exceeded 95%, highlighting VTAC's enhanced precision and reliability. The proposed solution will serve as an advisory framework to help financial crime investigators enhance the detection and reporting of suspicious cryptocurrency transactions in cyberspace.

INDEX TERMS machine learning; Blockchain; cybercrime; cryptocurrency; money laundering

I. INTRODUCTION

Cryptocurrencies are the most widely used in criminal activity for money laundering, fraud, theft, and carrying out dark web deals such as drug trafficking, weapon sales, and the sale of stolen personal information [1]. Cryptocurrencies facilitate a shadow economy on the dark web, allowing the purchase and sale of illicit goods and services without the oversight of regulatory bodies. Criminals use cryptocurrencies to hide the proceeds of their illegal activities [2]. Cybercriminals can use deceptive tactics like phishing, social engineering, or investment scams to trick individuals and persuade them to transfer their cryptocurrencies to malicious entities. Cybercriminals might pose as legitimate businesses or create fake investment opportunities.

Given that cryptocurrencies such as Bitcoin are simply digital codes, they often become the focus of hackers who utilise diverse hacking methods to undermine these assets. Attackers acquire unauthorised access to digital wallets by using techniques like phishing, which involves deceiving users into surrendering personal data, and by exploiting flaws in smart contracts or altering Blockchain protocols. Upon obtaining the private keys, anyone may effortlessly move the cryptocurrency to their own accounts. Moreover, the dark web, renowned for its illegal activities, often employs cryptocurrency for financial transactions. Cryptocurrencies are often utilised as a preferred method of payment on platforms where illicit products and services are bought and sold because of their anonymity and the difficulty of tracking transactions.

Although Blockchain is fundamentally secure and decentralised, no system is immune to attack or misuse. Previous occurrences have brought attention to the possibility of substantial monetary damages resulting from security breaches in both centralised and decentralised cryptocurrency systems, underscoring the need for continuous enhancements in cybersecurity within the crypto sphere [3]. In order to safeguard their digital assets, people and organisations must be informed about the most recent security risks and implement strong security measures. This entails using robust and distinctive passwords, consistently upgrading security software, activating two-factor authentication, and exercising caution while utilising networks and devices for cryptocurrency management [4].

The task of understanding the complex network of crypto transactions is difficult, requiring not just advanced analytical tools but also a profound comprehension of the constantly growing Blockchain technology [5]. Blockchain technology, while essential for facilitating cryptocurrency transactions, presents challenges in combating money laundering. The inherent features of Blockchain, such as decentralization, anonymity, and security, contribute to its attractiveness for cryptocurrencies. However, these same features can also be exploited for illicit financial activities. Cryptocurrencies like Zerocash [6] and Monero [7] offer enhanced privacy features compared to traditional cryptocurrencies, making it difficult to trace transactions back to their origins. When users engage in Bitcoin transactions, these transactions are recorded on a decentralized ledger without requiring personal identification, providing a degree of privacy. This privacy feature poses challenges for authorities in tracking the origins and destinations of payments, potentially facilitating money laundering.

Crypto compliance refers to the adherence to regulatory requirements and standards governing cryptocurrency transactions. It involves implementing measures to prevent illicit activities such as money laundering, terrorist financing, and other financial crimes within the cryptocurrency space. This includes Know Your Customer (KYC) procedures, Anti-Money Laundering (AML) regulations, and other compliance frameworks to ensure transparency and accountability in crypto transactions. Although the Blockchain's transparency enables the public to trace all transactions, the pseudonymous nature of wallet addresses might disguise the actual names of the individuals involved, making it more challenging to address money laundering. On the other side, even if illegal transactions are discovered, it may be challenging for law enforcement agencies to prove guilt. This is because investigations into such crimes are relatively difficult to conduct and require a significant amount of skill and resources. The fact that decentralized networks are not subject to any regulations makes this situation much worse.

The worldwide scope of cryptocurrencies and the varied regulatory environments in different nations make tracking endeavors even more complex. Regulators and law enforcement organizations are actively enhancing their instruments and promoting international cooperation to address more sophisticated money laundering methods. However, the cryptocurrency sector remains an ongoing and ever-evolving problem in this aspect. Due to this, the main aim of this paper is to present a machine learning-based approach for the identification and tracking of illegal crypto transactions via Blockchain network. The following is a list of the main contributions that the authors of this work have presented in this paper:

- An advanced machine learning-based architecture has been created to detect and identify illegal cryptocurrency transactions. This approach determines the legality of a transaction by examining factors such as the digital wallet hashes of both the sender and receiver, the transaction value, and the frequency of transactions over an interval of time.
- A dataset has been prepared by implementing the automated de-anonymization of anonymous crypto transactions in order to test the machine learning based architecture for the task of identification of illegal transaction elements.
- A methodology for analysing crypto compliance has been proposed, based on which an advisory framework has been developed to assess the legality of cryptocurrency transactions by analysing real de-anonymised crypto transactional data.

This paper presents an advanced machine learning approach for detecting money laundering within cryptocurrency transactions, setting itself apart from existing research. The authors have developed the Value-driven-Transactional tracking Analytics for Crypto compliance approach, demonstrated its key performance metrics compared to existing methods, and established an advisory framework to enhance the detection and reporting of suspicious cryptocurrency transactions.The proposed solution, which utilizes real cryptocurrency transaction data to aid security officials in investigating illegal activities and detecting potential money laundering, is a significant step forward in addressing the challenges posed by the anonymity and complexity of cryptocurrency transactions over Blockchain. The novelty is the identification of suspicious transactions with heightened accuracy and the offering of a unique framework to understand complex laundering patterns in the digital currency domain. The research methodology marks a significant advancement in combating financial crimes facilitated by Blockchain technologies.

The structure of the remainder of this article is organized as follows: Section II explores the relevant literature and existing works that focus on the detection of money laundering over Blockchain. The machine learning-based approach for the determination of the legality of cryptocurrency transactions and the preparation of a dataset for their validation are presented and explained in Section III. Section IV delves into a methodology for the analysis of crypto compliance by evaluating the real de-anonymised crypto transactional data over an advisory framework. The experimental use case of tracking value-driven crypto transactions in order to detect illegal money laundering is provided in Section V. The article concludes with Sections VI and VII, summarizing the findings, results and directions for future research.

II. RELATED WORKS

Money laundering using cryptocurrencies presents unique challenges. The research in [8] that was carried out in 2019 involved conducting an in-depth analysis of Bitcoin transactions and revealed concerning patterns in the way that they are associated with engaging in illegal activities. According to the analysis, over seventy-six billion dollars worth of transactions, which account for approximately 46% of all activities on the Bitcoin network, were associated with illegal intentions. This discovery holds significant value, revealing that illicit activities account for approximately half of the total transaction volume on the Bitcoin network. Moreover, it highlighted that upwards of one-quarter of Bitcoin wallets were possibly implicated in these transactions, suggesting a pervasive infiltration of illicit activity within the Bitcoin ecosystem. This emphasizes the immediate necessity for improved regulating and monitoring systems in the cryptocurrency industry to deal with and mitigate the usage of cryptocurrencies for illegal activities.

A. ADVANCED TACTICS AND LAYERING TECHNIQUES

Money laundering uses a multitude of advanced tactics in the context of cryptocurrencies. Numerous research studies have examined the practice of money laundering through the layering of cryptocurrency transactions over Blockchain. A paper in [9] suggests a framework for goal modeling and mining to represent the activities of actors in the process of money laundering, with a specific focus on Bitcoin mixing transactions. This concept encompasses distinct stages, namely data acquisition, model identification, and Blockchain examination. It emphasizes the involvement of diverse actors in the money laundering process, including organizers, communicators, and soldiers. The study illustrates the collaborative functioning of these agents during the stages of money laundering, including placement, layering, and integration.

Within the domain of cryptocurrencies, the utilization of complex layering techniques and currency mixing services presents substantial challenges in the tracking of financial transactions. Layering is a tactic that involves dividing substantial sums of illegal money into smaller transactions that appear harmless. These transactions are then distributed across other cryptocurrencies and wallets to conceal their unlawful source [10]. Automated scripts frequently expedite this process by swiftly executing numerous transactions, hence complicating the financial audit trail.

Adding to this problem are mixing services, sometimes known as tumblers, which aggregate currency from multiple origins and blend it before redistributing it [11], [12]. This effectively severs the direct connection between the origin and the ultimate receiver of the payments. Methods like CoinJoin [13], which consolidate multiple transactions into a single one, introduce an additional level of complexity to the process of tracking. Tumblers complicate the tracking of crypto funds by creating a seemingly disconnected web of transactions. Machine learning may be directed at previous information to identify attributes associated with tumbling activity. These models learn from instances of both conventional transactions and those known to use tumblers, allowing them to distinguish between valid and possibly disguised transactions. Transaction frequency, quantity, timing, and address sequence may all be used as indicators in these models. Unsupervised learning techniques, on the other hand, are crucial for discovering abnormal patterns in data without previous labeling. Clustering techniques aid in the identification of groupings of transactions with unexpected characteristics or patterns that depart from the norm, which are often suggestive of service mixing. These clusters may indicate hidden linkages and financial movements that are not immediately obvious. The use of machine learning to monitor and analyse transactions disguised by tumblers not only improves the accuracy of detecting such activities but also dramatically decreases the time and resources necessary for investigations. As tumblers change, machine learning models may be constantly trained on fresh data, ensuring that detection methods stay effective against the most recent obfuscation tactics.

B. MACHINE LEARNING MODELS FOR DETECTION

The paper [14] presents a method for identifying Bitcoin mixers, which are services designed to promote anonymity by concealing the connection between participants in a transaction. The research presents a machine learning model that employs the C4.5 decision tree technique. This model efficiently and effectively detects only eight essential features and tackles the issues of inadequate precision and inefficiency found in current methodologies. The proposed technique minimizes computing demands while attaining a high level of accuracy (exceeding 97%) in identifying mixers, making a valuable contribution to the fight against money laundering in cryptocurrencies. The process involves analyzing transaction patterns using graph analysis and reducing features using information gain. The future work entails extending the model to encompass other cryptocurrencies, integrating supplementary attributes to enhance accuracy, and revising the model to align with new mixing approaches.

An article published on [15] analyzes the identification of mixing in Bitcoin transactions through the utilization of statistical patterns. The study primarily concentrates on withdrawal transactions originating from mixers. The technique presented is a two-phase detection method that focuses on identifying patterns at both the transaction and chain levels. Nevertheless, the study may encounter constraints in accommodating changing mixer approaches and could have difficulties in identifying false positives caused by the constantly changing nature of Blockchain transactions. The next effort entails extending the model to encompass other cryptocurrencies, integrating supplementary variables to enhance accuracy, and updating the model to align with new mixing approaches.

Another study explores the phenomenon of money laundering within the Bitcoin network by employing graph theory and machine learning methodologies [16]. The system examines the Bitcoin transaction graph, distinguishing between illicit money laundering activities and legitimate transactions. The study employs random-walk-based graph representation learning techniques, namely deep walk and node-to-vector, to construct classifiers capable of differentiating between these two categories of transactions. The study showcases the efficacy of the classifiers in binary classification, with exceptional accuracy and F1-measure. Additionally, it investigates their potential for detecting unfamiliar money laundering services.

A research study in [17] investigates the application of Graph Convolutional Networks (GCN) for identifying unauthorized transactions within the Bitcoin network. The work presents an innovative method that integrates GCN with linear layers, with the goal of enhancing the accuracy of forecasting illegal transactions. The performance of this strategy is assessed using the Elliptic dataset and exhibits enhanced efficacy in comparison to GCN models employed in prior studies. The research aims to improve the predictive abilities of the model by utilizing both the graph-based spectral technique and the linearly modified feature matrix in the Euclidean domain. The findings demonstrate that the integrated method surpasses the current models in accurately detecting lawful and unlawful transactions within the Bitcoin transaction graph. Nevertheless, this paper lacks a comprehensive investigation into the scalability of the model when applied to larger and more diverse datasets and Blockchain networks. It also fails to address the model's adaptation to different criminal activities and the inclusion of more diverse features to improve predictive accuracy. The potential for using the model in real-time anti-money laundering monitoring systems is emphasized, suggesting an opportunity for more proactive and efficient identification of illegal transactions in the Blockchain.

The authors in [18] commence by emphasizing the significance of comprehending Bitcoin address patterns inside the anonymous framework of the Blockchain. They proposed their BAClassifier system, which includes address graph building, graph representation learning, and address classification with graph neural networks. The main improvements consist of a technique for creating chronological transaction graphs and an efficient data-driven method for effectively acquiring these representations. The experimental results demonstrate that BAClassifier outperforms existing methods, achieving high precision and F1-scores (96% and 95%, respectively) in identifying address behaviors. While the proposed solution shows high accuracy, its performance in real-world scenarios across diverse and evolving Blockchain environments wasn't discussed in detail. The classifier's effectiveness heavily depends on the initial graph construction from transaction data, which can be error-prone or may not capture all of the nuances in address behaviors.

Machine learning has the potential to be particularly helpful in solving the issues posed by sophisticated layering and the use of tumblers in the money laundering process within the crypto financial sector. Layering results in the formation of elaborate patterns that are difficult to recognize using conventional methods of analysis [19]. The ability of machine learning algorithms to recognize minor patterns and anomalies in transaction data is exceptional, especially when these patterns and anomalies are firmly ingrained in lawful financial activity. An example of this would be how unsupervised learning may identify anomalies in transaction quantities, which may indicate the presence of potential money laundering [20]. Additionally, social network analysis uses transaction networks to uncover previously unknown linkages and laundering rings, while predictive modeling predicts future laundering practices. Many of these strategies, such as finding strange patterns in transactional data, analyzing social networks, and guessing what new layering strategies will be used, show how machine learning can change to keep up with complex financial crimes.

The study published in [21] introduces a methodology that uses machine learning and deep learning techniques to detect instances of money laundering in cryptocurrency transactions. The primary dataset employed in this study is the Elliptic Bitcoin dataset. The research shows promising results, especially with the Random Forest classifier. However, it has some problems, such as the dataset only including Bitcoin, the risk of model overfitting, the unstable nature of cryptocurrency markets, and the lack of new laundering techniques in the dataset. Furthermore, the high computational demands of deep learning models and the study's emphasis on minimizing false positives without thoroughly investigating the consequences of false negatives are significant limitations. These constraints indicate the necessity of consistently adjusting and verifying the models to stay up-to-date with the swiftly evolving realm of cryptocurrency transactions and money laundering techniques.

Bitcoin, known for its disintermediation, decentralization and secure data recording, also employs public key cryptography to protect user privacy. Its anonymity features have been exploited for illicit activities, diminishing its appeal. The paper in [22] introduces a machine-learning method to differentiate between legal and illegal Bitcoin transactions. By identifying and excluding illegal transactions from blocks, this approach aims to enhance user trust and facilitate wider adoption of Bitcoin technology. However, the proposed solution is limited to the dataset and needs to be improved.

An investigation conducted in [23] centered on the examination of an extensive dataset of financial transactions in Norway. This study demonstrated how machine learning technology can improve the detection of suspicious transactions, which may involve money mules and smurfing. The research involved three sorts of historical data: routine legal transactions, transactions marked as suspicious by the bank's internal systems, and probable money laundering cases submitted to authorities. Based on transactional and historical data about the sender and receiver, the model estimated the likelihood of reporting a new transaction. As a result, it enhances the accuracy and efficiency of anti-money laundering measures in the financial sector. However, the findings in this work are dependent on past and specialized datasets, which may not comprehensively depict future or diverse forms of money laundering activities. This can impact the capacity to apply the machine learning models to a wider range of situations. Moreover, the dynamic nature of money laundering methods implies that models relying on present data may rapidly become obsolete.

Recent studies emphasize the implementation of machine learning methods to address the issue of money laundering on online gaming sites that accept bitcoin. A study in [24] conducted on the Elliptic Bitcoin dataset examined the effectiveness of supervised learning algorithms in distinguishing between legitimate and illegal transactions. The methodology involved analyzing the algorithms' performance using F1-scores, precision, and recall metrics. In addition, qualitative interviews with bitcoin exchanges were undertaken to examine the fit and usefulness of these algorithms. This mixed-method approach combines data-driven algorithmic research with practical insights from industry practitioners to better understand the effectiveness of machine learning in combating money laundering on cryptocurrency exchanges. The study demonstrated the potential of machine learning for detecting money laundering operations. Separate research studies focused on identifying instances of money laundering within the Bitcoin Blockchain, despite the limited availability of labels [25], [26]. These studies utilized both supervised and unsupervised learning techniques to detect illegal transactions, such as those associated with scams, malware, and Ponzi schemes. The results demonstrated the effectiveness of machine learning in situations where there is a shortage of labeled data.

C. REAL WORLD APPLICATION CHALLENGES

The decentralized and pseudo-anonymous structure of bitcoin transactions via sites like LocalBitcoins makes it difficult to detect unlawful money flows. These platforms facilitate the transformation of cryptocurrencies into tangible currency, complicating authorities' efforts to trace and supervise financial transactions associated with money laundering [27]. To investigate crypto-laundering cases, this study uses a small-n comparative case analysis methodology. It intends to reveal deeper links and networks by studying 12 cases from the US Department of Justice website and other legal databases. The inclusion criteria for these instances were based on their global nature, bitcoin use, and involvement in money laundering. This methodology helps to identify patterns in how money launderers use bitcoin in relation to the global financial system and various types of currency. The Blockchain technology used in such transactions provides a certain level of traceability, but it is only pseudo-anonymous because it does not require the identification of the sender or receiver, unlike traditional financial systems. The susceptibility of cryptocurrencies to being exploited for financial crimes, such as money laundering, is a significant concern. Nevertheless, the precise scope and characteristics of this vulnerability remain subjects of continuing investigation and discourse within the realm of illicit international political economy.

The research conducted in [28] gives a comprehensive analysis to demonstrate how money mules aid criminal syndicates in preserving anonymity and capitalizing on weaknesses such as unemployment and the participation of young individuals in financial illicit activities. The authors systematically reviewed ten real-world case studies to elucidate how money mules help criminal syndicates remain anonymous while moving funds around the world. It allowed for an analysis of criminals' patterns, techniques, and operational tactics. However, this approach is more qualitative in a theoretical sense, drawing conclusions from documented legal cases and incidents to provide insights into the practices of money laundering and the use of money mules in these illicit activities. In [29], a deep analysis of existing literature, specifically examining the functions, recruiting methods, and knowledge of individuals involved in money mulling, has been provided. The findings uncover the utilization of money mules by organized crime syndicates to launder illegal monies, as well as the tactics employed to entice potential mules with promises of financial prosperity.

Online gaming platforms that accept cryptocurrency are strongly associated with money laundering as a result of their anonymous nature and absence of regulatory supervision [30]. The utilization of cryptocurrencies on these platforms allows users to conduct transactions while maintaining anonymity, hence posing difficulties in tracking the source of funds. The combination of anonymity and the global and decentralized character of cryptocurrencies provides an optimal setting for the process of money laundering. Due to the difficulty in tracking and controlling such transactions, individuals find it more convenient to launder illegal monies through these online gaming platforms. Studies on money laundering involving cryptocurrencies have also examined the application of big-data analysis to detect trends of unlawful bitcoin utilization [31]. Investigators can use this method to create indicators and automated probabilistic tools to identify accounts likely involved in illicit activity. These technologies are crucial for law enforcement organizations to track illegal funds within cryptocurrency systems and identify and prevent illegal transactions.

Despite advancements, significant gaps remain in existing methods. Our work introduces the Value-driven-Transactional tracking Analytics for Crypto compliance approach, which combines machine learning algorithms with pre-training, normalization, model training, and de-anonymization.

III. PROPOSED MACHINE LEARNING-BASED VTAC APPROACH

A notable challenge in the training of machine learning models, especially evident in the realm of financial analytics, is the occurrence of data leakage during the normalization process. This problem often arises when the normalization model incorporates data from both the training and test datasets. Particularly, observations from the test data, which should ideally remain separate and unknown during the training phase, inadvertently influence the normalization parameters. This mishap results in the test data influencing the normalized training samples and, subsequently, the trained model itself. Moreover, the conventional practice of splitting the data into only two sets—training and test—rather than three distinct sets (training, validation, and test) exacerbates this issue. In such scenarios, the test data inadvertently participates in model selection and validation processes, further compounding the risk of data leakage. These practices undermine the integrity of the model evaluation process, leading to overly optimistic performance estimates and models that may fail to generalize to new data.

The ethical and privacy implications of de-anonymizing transaction data are central concerns within VTAC's operational framework, especially given the rigorous requirements set forth by international privacy regulations like GDPR and Anti-Money Laundering (AML) directives. VTAC's methodology includes sophisticated anonymization techniques to ensure that while transactional data is analyzed for illegality indicators, individual privacy remains intact. The system adheres to GDPR by implementing data minimization principles and ensuring that only necessary data for identifying and preventing illegal activity is processed. Furthermore, all data processing within VTAC is conducted with transparency and under strict access controls, maintaining compliance with both GDPR and AML requirements. This not only helps in safeguarding user confidentiality but also reinforces the legal and ethical legitimacy of the surveillance and analysis processes used within VTAC.

The emergence of cryptocurrencies, especially Bitcoin, has revolutionized digital transactions, offering unparalleled anonymity and decentralization. However, this anonymity also presents significant challenges, particularly in ensuring legal compliance. It is challenging to definitively ascertain the legality of a cryptocurrency transaction due to the absence of a straightforward and singular method. Nevertheless, specific elements of a transaction can be analyzed to evaluate its legality.

In the cryptocurrency ecosystem, a transaction typically consists of data about the parties involved in the transaction, as well as the data and metadata associated with the transaction [32]:

- Sender and recipient addresses: addresses of the transaction participants;
- Transaction value: the amount of money to be transferred; for example, in the Bitcoin network, the smallest indivisible unit of digital assets is the satoshi, which is equivalent to 0.00000001 BTC, so the smallest amount that can be transferred in the Bitcoin network is 1 satoshi;
- Transaction fee: a small fee usually paid to miners who verify and record a transaction on the Blockchain. This fee is normally paid in the same cryptocurrency as the amount of the transaction mentioned before;
- A digital signature is a cryptographic signature used to verify the authenticity of a transaction. It helps to ensure that the transaction is carried out by the owner of the digital asset;
- Timestamp: the date and time when the transaction was recorded in the Blockchain.

The transactions are aggregated into blocks using Blockchain technology and thus stored in a decentralised repository. Blockchain is a distributed database where transaction records are stored. Transactions are recorded in blocks, which are connected by a chain. Each block contains transaction records and a hash function value linking the block to the previous block. This produces a list of all transactions that have occurred in the chain of blocks. By analysing the elements of a transaction and the links in the Blockchain, it is possible to identify factors that indicate that a transaction made in cryptocurrency is potentially illegal:

- One of the transaction addresses is a known high-risk address;
- Suspicious patterns in the transactions, such as extremely large amounts of currency transferred and/ or frequent periodic transfers, may indicate that the transaction is being conducted for criminal purposes.

The human eye struggles to perceive these complex patterns; hence, the implementation of machine learning models into the process of detection and identification of crypto transaction legality is needed. The solution proposed in this paper focuses on two parts: a) detection of the predicted illegal crypto transactions by ML-powered models - Random Forest, ADA Boost and XG Boost models; b) identification of the transaction illegality by the transaction elements - wallet hashes of both the sender and receiver, the transaction, value and the frequency of transactions over an interval of time (Fig. 1).

The proposed VTAC algorithm is structured into two distinct phases to enhance clarity and effectiveness. Phase 1, the 'Detection Phase,' involves the use of machine learning models (Random Forest, ADA Boost, and XG Boost) to classify transactions as either legal or illegal. The input for this phase includes features such as transaction values, wallet hashes, and frequencies, with the output being a classification of each transaction's legality. Phase 2, the 'Identification Phase,' builds on the initial classifications to further analyze and identify specific illegal transactions for forensic purposes. The inputs for this phase are the classified data from Phase 1, and the outputs are detailed profiles of transactions identified as illegal, focusing on uncovering the underlying patterns and beneficiaries. The independent variables in both phases are the transaction attributes, while the dependent variables are the classifications of legality in Phase 1 and the detailed illegal transaction profiles in Phase 2. This structured approach ensures a comprehensive analysis, moving from broad detection to focused identification, to address the multifaceted nature of cryptocurrency fraud effectively.

The use of Random Forest, ADA Boost, and XG Boost models for detecting illegal cryptocurrency transactions is a deliberate decision, as they have the advanced ability to efficiently handle the intricacies and large datasets connected with Blockchain operations. Random Forest excels at handling

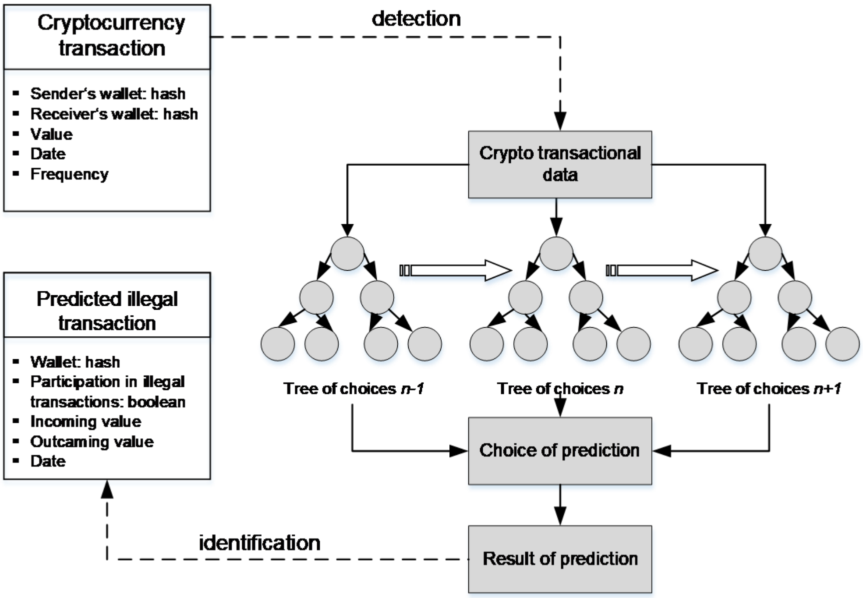


FIGURE 1. Proposed VTAC approach

high-dimensional data and providing important insights based on feature relevance, making it capable of identifying crucial indicators of illegal transactions. It is adaptable, handling both categorical and numerical data easily. ADA Boost improves predictive power by focusing on previously misclassified examples, strengthening the model's accuracy across repetitions. It is especially useful in binary classification applications, such as discriminating between legal and illegal transactions. XG Boost stands out for its remarkable speed, efficiency, and scalability, which are critical for processing huge amounts of Blockchain data. Its robust regularization prevents overfitting, ensuring the model's accuracy and reliability. The combination of these models results in a more comprehensive and nuanced approach to detecting illegal transactions. They complement one another by addressing various parts of the data analysis process, ranging from feature importance and adaptability to efficiency and scalability. Their cumulative application not only improves the ability to process and analyze large and complicated datasets, but also provides high accuracy and reliability in detecting sophisticated money laundering trends. This makes the Random Forest, ADA Boost, and XG Boost ensemble an ideal choice for preventing cryptocurrency-related financial crimes, combining each model's particular capabilities to create a robust and successful solution.

Analyzing wallet hashes enables the tracing of transaction flows and the identification of patterns compatible with money laundering, such as cycling money between many accounts to hide its origin. Wallet hashes are unique identifiers, and tracking them can uncover networks of accounts linked to unlawful operations. Large transactions or ones that depart significantly from a user's normal transaction profile may suggest money laundering or fraudulent conduct. Analyzing transaction numbers in the context of historical data enables the detection of outliers that may require additional examination. A high frequency of transactions over short periods of time may indicate layering, in which illicit monies are divided into smaller amounts and moved through multiple transactions to make them more difficult to trace. Frequency analysis can aid in the discovery of structured transactions intended to circumvent detection thresholds.

Due to this, the authors called it as Value-driven-Transactional tracking Analytics for Crypto compliance (VTAC) as it aims to aid in the detection of illicit activities, bridging the gap between the advantages of cryptocurrency and the necessity for legal compliance. VTAC employs sophisticated machine-learning models to track and analyze cryptocurrency transactions. This surveillance goes beyond basic transaction monitoring and examines patterns that could suggest illicit actions, such as money laundering. VTAC prioritizes the value component rather than just focusing on the volume of transactions, unlike traditional systems. It examines the movement of transactions and evaluates the level of risk by considering factors such as the amount of money exchanged, the parties involved, and the frequency of the transactions.

While the integration of Random Forest, ADA Boost, and XG Boost in the VTAC method may appear routine, the novelty of our approach lies in the specific application and optimization of these models to the unique challenges of detecting illegal transactions in cryptocurrency networks. Our methodology enhances the predictive power by carefully tuning these algorithms to work synergistically, allowing for improved handling of complex, high-dimensional data characteristic of blockchain transactions. This tailored application goes beyond routine usage, providing a nuanced approach that leverages the strengths of each model to achieve a detection accuracy that surpasses current benchmarks. This strategic combination within our VTAC framework is integral to advancing the capabilities of transaction monitoring systems in identifying sophisticated laundering activities effectively.

In contrast to earlier studies that were mentioned in Section II, this methodology presents a more complex evaluation that integrates the transactional value and related features, thereby establishing a more comprehensive structure to detect illicit activities. VTAC method integrates an automatic de-anonymization procedure that uncovers transaction components such as wallet hashes, transaction quantities, and frequency. This enables more precise tracking of the sources and destinations of funds. VTAC incorporates a preliminary training procedure that eliminates string values and fills in missing data, guaranteeing a uniform dataset for the models. Additionally, VTAC uses a normalization technique to normalize the scale of data characteristics, a crucial step in ensuring accurate machine learning results. In addition to basic transaction monitoring, VTAC analyzes patterns that could potentially signify money laundering. It places emphasis on the transaction value rather than just the volume, resulting in a more detailed study. VTAC can continuously train its algorithms on new data, enabling it to adapt to developing laundering techniques and Blockchain technology. This makes VTAC very scalable and effective for real-time monitoring. Including transaction value analysis allows VTAC to conduct a more thorough assessment of transaction risk. This analysis takes into account various criteria, including the amount of money traded, the parties involved, and the frequency of transactions. VTAC serves as a guidance system that assists financial crime investigators and regulatory agencies in improving their ability to monitor and report, thereby facilitating compliance with legal requirements.

A. PRE-TRAINING PROCESS

The pre-training process in the Python script involves several key steps to prepare the dataset for machine learning model training. Initially, the script removes string values from the dataset. This is important because machine learning models typically require numerical input. Any object attributes that are strings are excluded from each data point, leaving only numerical or Boolean data. The script addresses empty arrays, which can occur even after the data has been flattened. Empty arrays are assigned a default numerical value (typically zero) to ensure consistency in the data format.

Another critical step in the pre-training process is adding missing features. Due to the nature of the data, not all transactions have the same set of features. The script identifies all unique features across the dataset and ensures that each data point includes all these features. If a feature is missing from a transaction, it is added with a default value (again, usually zero). This step is crucial for maintaining a consistent feature set across all data points, which is essential for training machine learning models effectively. The script converts Boolean attribute values into a binary format. This means that true or false values are represented as either 1 or 0, making them suitable for use in machine learning algorithms.

These pre-processing steps collectively contribute to creating a clean, uniform dataset that can be effectively used for training a predictive model, such as a Random Forest classifier, which is intended for predicting the legality of cryptocurrency transactions. The process ensures that the dataset is in a suitable format for the model to learn patterns and make accurate predictions. The Random Forest model uses the bagging strategy, which entails randomly selecting subsets of the training data to construct each tree in the forest. Introducing variety among the trees in this process is essential for mitigating overfitting and enhancing the model's capacity to generalise. The Random Forest algorithm uses many decision trees to produce predictions based on the input data. A majority vote among all the trees in the ensemble determines the final result. This ensemble technique successfully reduces the possible biases and variations that are often present in individual decision trees.

B. NORMALIZATION PROCESS

The normalization process (Fig. 2) in the script is a crucial step in preparing the dataset for machine learning. Normalization adjusts the range of data features so that they are on a similar scale. This is important because it ensures that no single feature disproportionately influences the model due to its scale. The script uses the *StandardScaler* from the "sklearn.preprocessing" package for normalization. This scaler removes the mean and scales each feature to unit variance. This is a common approach in data pre-processing for machine learning, as it standardizes the distribution of features.

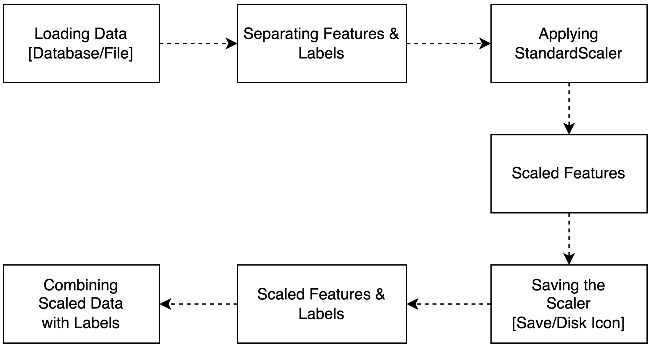

Loading Data [Database/File]
Separating Features & Labels
Applying StandardScaler
Scaled Features
Saving the Scaler [Save/Disk Icon]
Scaled Features & Labels
Combining Scaled Data with Labels

FIGURE 2. Normalization process

Here's how the process works in the script:

- The dataset is loaded into a Pandas ”DataFrame”. This structure allows for efficient manipulation and processing of data.
- The script separates the target variable (label) from the input features. In this context, the label ”elliptic label” indicates the classification of the transaction.

- The StandardScaler is applied to the input features. This transforms the data such that each feature has a mean of 0 and a standard deviation of 1, effectively standardizing the dataset.
- Combining scaled data with labels: after scaling, the features are combined back with the labels to form a complete dataset.
- The script saves the scaler object as a *.pkl* file. This is important because the same scaler needs to be used on new data to ensure consistency when the model is deployed in a real-world setting.

As it can be seen in Figure 2 arrows indicate the flow of the normalization process. Each box represents a step in this process, such as "Loading Data", "Separating Features Labels", *etc*. The [Database/File] and [Save/Disk Icon] are symbolic representations of the loading and saving steps.

By normalizing the data, the script ensures that the model will train on features that are equally scaled, contributing to better performance and more accurate predictions. This step is essential, especially when dealing with real-world data where feature scales can vary significantly.

C. TRAINING THE MODEL PROCESS

The script begins by importing necessary libraries and modules, such as Pandas for data manipulation, Sklearn's Random Forest Classifier for the machine learning model, and various functions for model evaluation and data splitting. The script loads the pre-processed and normalized dataset from a JSON file. It then removes any features from the DataFrame that contain all null values, ensuring that the dataset is clean and ready for training.

The dataset is split into two parts: features and labels. The features are the input variables used to train the model, and the labels are the target variables the model will predict. In this context, ”elliptic label” appears to be the target variable.

The script uses the ”train test split” function to divide the dataset into a training set and a testing set. The training set is used to train the model, and the testing set is used to evaluate its performance. A typical split is 75% for training and 25% for testing in numerous scientific works that employ machine learning mechanisms [33]. The selection of this ratio is determined by aspects like the size and variety of the dataset, where a larger dataset may still yield a sufficient quantity of data for both training and testing, even after being segmented. In addition to the typical 75/25 training/testing set ratio, other ratios such as 80/20 and 70/30 were also explored to investigate the optimal selection for training and testing splits. This aims to enhance the accuracy of the results by determining the most effective data split for model training and validation. A Random Forest Classifier is instantiated and trained using the training data. Random Forest is a popular machine learning algorithm known for its robustness and accuracy, particularly in classification tasks.

After training, the model makes predictions on the testing set. The script calculates the accuracy of these predictions by comparing them with the actual labels from the test set. Accuracy is a common metric for evaluating classification models, representing the proportion of correctly predicted instances. The script also computes a confusion matrix, which provides a detailed breakdown of the model's performance, including the number of true positives, false positives, true negatives, and false negatives. The trained model can be saved for future use, allowing for predictions on new, unseen data.

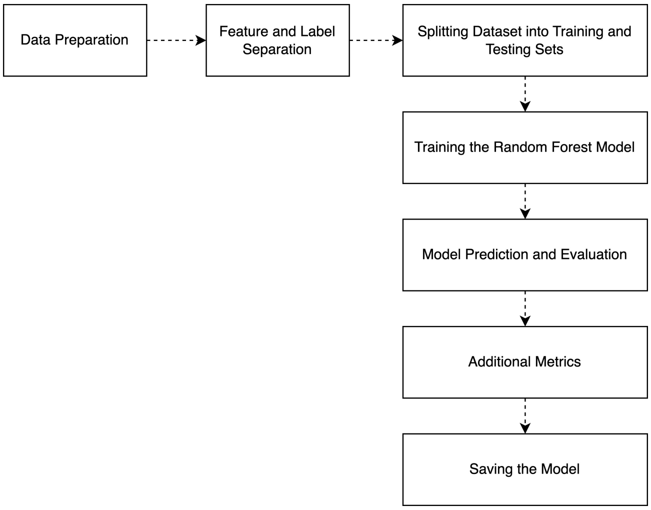

Data Preparation
Feature and Label Separation
Splitting Dataset into Training and Testing Sets
Training the Random Forest Model
Model Prediction and Evaluation
Additional Metrics
Saving the Model

FIGURE 3. Process of training the model

As it is shown in Figure 3 each box represents a step in the training model process, such as "Data Preparation", "Feature and Label Separation", *etc*. Arrows indicate the flow of the process from one step to the next. This process encapsulates the standard steps involved in training and evaluating a machine learning model, specifically a Random Forest classifier, for tasks such as predicting the legality of cryptocurrency transactions based on transaction features.

For model training, a public dataset published by Elliptic on the Kaggle dataset repository [32] was chosen. This anonymous dataset consists of bitcoin transactions collected from the bitcoin Blockchain. A node in this graph (Fig. 4) represents a transaction, and an edge is understood as a flow from one transaction to another.

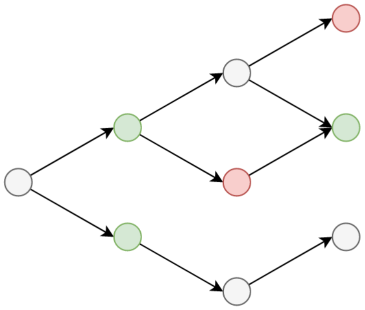


FIGURE 4. Basic graph of legal/illegal crypto transactions over Blockchain

Each node has 166 fields and a validity class that can take

TABLE 1. Structure of de-anonymized crypto transaction [34]

Anonymous TransID	Transaction
230325127	d6176384de4c0b98702eccb97f3ad6670bc8410d9da715fe5b49462d3e603993
230325139	300c7e7bb34263eae7ff8b0a726d5869bf73d71081490c45a9536a31560f1fd7
86875675	7c790a31090462d720a172b3f55a51af2514971070db6686e337ccc486840dcd

the following values:

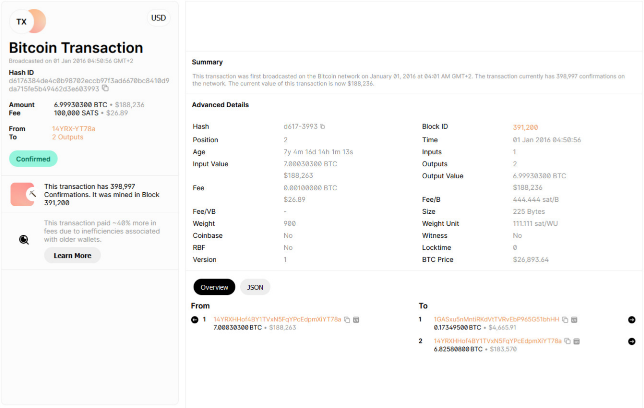
1) legal (42019 observations or 21% of all transactions);
2) illegal (4545 observations or 2% of all transactions);
3) unknown (remaining transactions).

The unknown transactions account for 77% of the total transactions in the Kaggle dataset [32]. The substantial portion of unidentified transactions is mainly attributed to the inherent characteristics of Blockchain and cryptocurrency transactions. The anonymity and privacy functionalities of these platforms lead to a considerable number of transactions that cannot be readily categorised as lawful or unlawful without further examination or supplementary data. This highlights the difficulties in overseeing and controlling bitcoin transactions, as well as the need for creating sophisticated machine learning algorithms that can detect patterns suggesting illegal activity within a mostly unfamiliar dataset.

D. DE-ANONYMIZATION PROCESS

The Elliptic dataset consists of a set of anonymised transactions, which is suitable for predicting the detection of illegal transactions but does not allow for the task of identification by transaction elements. For this purpose, an automated de-anonymization process has been implemented, which reveals the hashes of Elliptic's anonymous transactions [34]. The structure of this dataset is presented in Table 1.

This de-anonymization reveals 99.5% of all transactions. The Anonymous TransID is the anonymized transaction identifier, and the Transaction column reveals the real transaction ID.



Bitcoin Transaction

Broadcasted on 01 Jan 2016 04:50:56 GMT+2

Hash ID
d6176384de4c0b98702eccb97f3ad6670bc8410d9da715fe5b49462d3e603993

Amount 6.99930300 BTC • $188,236
Fee 100,000 SATS • $26.89

From 14YRX-YT78a
To 2 Outputs

Confirmed

This transaction has 398,997 Confirmations. It was mined in Block 391,200

This transaction paid ~40% more in fees due to inefficiencies associated with older wallets.

Summary

This transaction was first broadcasted on the Bitcoin network on January 01, 2016 at 04:01 AM GMT+2. The transaction currently has 398,997 confirmations on the network. The current value of this transaction is now $188,236.

Advanced Details

Hash	d617-3993	Block ID	391,200
Position	2	Time	01 Jan 2016 04:50:56
Age	7y 4m 16d 14h 1m 13s	Inputs	1
Input Value	7.00030300 BTC $188,263	Outputs	2
		Output Value	6.99930300 BTC $188,236
Fee	0.00100000 BTC $26.89	Fee/B	444.444 sat/B
Fee/VB	-	Size	225 Bytes
Weight	900	Weight Unit	111.111 sat/WU
Coinbase	No	Witness	No
RBF	No	Locktime	0
Version	1	BTC Price	$26,893.64

From
1 14YRXHHof4BY1TVxN5FqYPcEdpmXiYT78a
7.00030300 BTC • $188,263

To
1 1GASxu5nMntiRKdVtTVRvEbP965G51bhHH
0.17349500 BTC • $4,665.91
2 14YRXHHof4BY1TVxN5FqYPcEdpmXiYT78a
6.82580800 BTC • $183,570

FIGURE 5. Elements of crypto transactions gained from Bitcoin Blockchain

The transaction ID is automatically linked to the information contained in [35] to provide complete information, i.e., the elements of the transaction, which are used to identify the legality of an illegal transaction (Fig. 5).

IV. PROPOSED METHODOLOGY FOR CRYPTO COMPLIANCE ANALYSIS

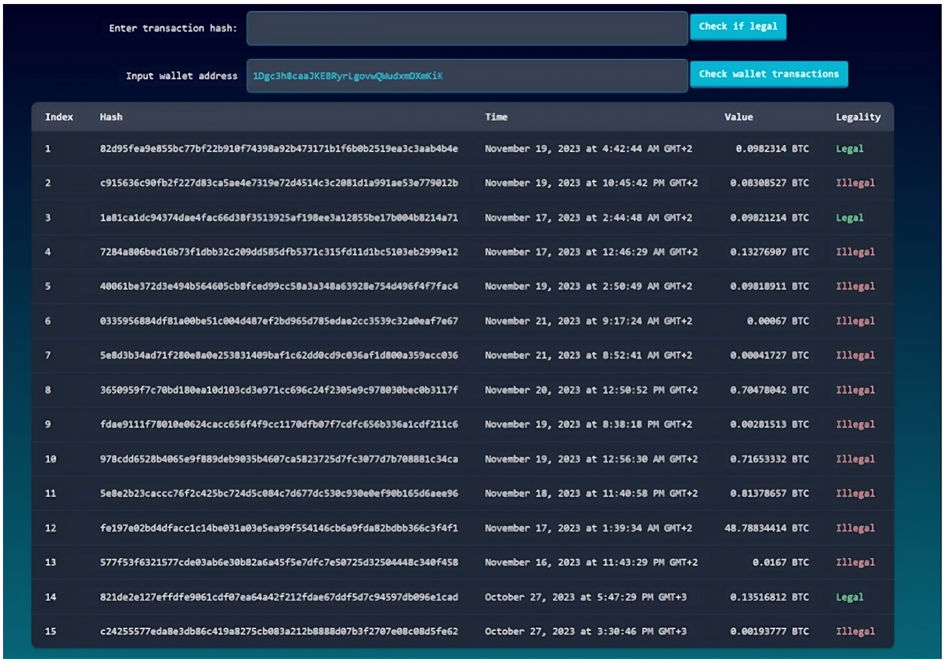
In this study, we present a two-fold experimental approach using machine learning to enhance the detection and identification of illegal transactions on the Blockchain. The first experiment focuses on the detection of suspicious activities using an ML model that analyzes transaction values and temporal patterns. The second hones in on identification, employing critical attributes such as the values of incoming and outgoing transactions, the time interval between the chain's initial outgoing and final incoming transactions, and a weight factor that assesses transaction significance within the Blockchain, aiding in beneficiary identification. The results demonstrate the model's efficacy in not only detecting but also providing crucial insights into illegal transactions, offering valuable tools for Blockchain security and forensic analysis.

By combining an anonymized classified Elliptic dataset with automated de-anonymization of the transactional data and ML-based legality prediction, a methodology for crypto compliance analysis is used to further identify the illegality of cryptocurrency transactions over Blockchain. Based on this methodology, an advisory framework is developed (Fig. 6), where the final validation of the legality of a transaction is based on the following factors:

- the value of the outgoing and incoming transaction;
- the time between the first outgoing transaction in the chain and the last incoming transaction in the chain;
- the weight factor indicating the importance of the transactions in the block chains that is used to find beneficiaries.

The detailed steps on how the advisory framework has been created can be found in [36]. In the context of a Random Forest model, an API is usually designed to allow users to interact with the model through HTTP requests. The trained Random Forest model is loaded into the API. This often involves deserializing the model file (i.e., a .sav or .pkl file) so that it's ready to make predictions. A dedicated endpoint is created for making predictions. The data received through the API is preprocessed to match the format and structure expected by the model. This might involve normalization, feature selection, or other transformations. The Random Forest model receives the preprocessed data and uses it to generate a prediction. The model responds to the user's request by sending back its prediction.The API might also include additional endpoints for tasks like updating the model, retrieving model statistics, or logging.

A Bitcoin graph is typically used to represent transactions within the Bitcoin network. Each node in the graph typically represents a Bitcoin address. In some cases, nodes can also represent entities like users or organizations, depending on the level of aggregation or analysis. Edges in the graph represent transactions between Bitcoin addresses. An edge will typically



Index	Hash	Time	Value	Legality
1	82d95fea9e855bc77bf22b910f74398a92b473171b1f6b0b2519ea3c3aab4b4e	November 19, 2023 at 4:42:44 AM GMT+2	0.0982314 BTC	Legal
2	c915636c90fb2f227d83ca5ae4e7319e72d4514c3c2081d1a991ae53e779012b	November 19, 2023 at 10:45:42 PM GMT+2	0.08308527 BTC	Illegal
3	1a81ca1dc94374dae4fac66d38f3513925af198ee3a12855be17b004b8214a71	November 17, 2023 at 2:44:48 AM GMT+2	0.09821214 BTC	Legal
4	7284a806bed16b73f1dbb32c209dd585dfb5371c315fd11d1bc5103eb2999e12	November 17, 2023 at 12:46:29 AM GMT+2	0.13276907 BTC	Illegal
5	40061be372d3e494b564605cb8fced99cc58a3a348a63928e754d496f4f7fac4	November 19, 2023 at 2:50:49 AM GMT+2	0.09818911 BTC	Illegal
6	0335956884df81a00be51c004d487ef2bd965d785edae2cc3539c32a0eaf7e67	November 21, 2023 at 9:17:24 AM GMT+2	0.00067 BTC	Illegal
7	5e8d3b34ad71f280e8a0e253831409baf1c62dd0cd9c036af1d800a359acc036	November 21, 2023 at 8:52:41 AM GMT+2	0.00041727 BTC	Illegal
8	3650959f7c70bd180ea10d103cd3e971cc696c24f2305e9c978030bec0b3117f	November 20, 2023 at 12:50:52 PM GMT+2	0.70478042 BTC	Illegal
9	fdae9111f78010e0624cacc656f4f9cc1170dfb07f7cdfc656b336a1cdf211c6	November 19, 2023 at 8:38:18 PM GMT+2	0.00281513 BTC	Illegal
10	978cdd6528b4065e9f889deb9035b4607ca5823725d7fc3077d7b708881c34ca	November 19, 2023 at 12:56:30 AM GMT+2	0.71653332 BTC	Illegal
11	5e8e2b23caccc76f2c425bc724d5c084c7d677dc530c930e0ef90b165d6aee96	November 18, 2023 at 11:40:58 PM GMT+2	0.81378657 BTC	Illegal
12	fe197e02bd4dfacc1c14be031a03e5ea99f554146cb6a9fda82bdbb366c3f4f1	November 17, 2023 at 1:39:34 AM GMT+2	48.78834414 BTC	Illegal
13	577f53f6321577cde03ab6e30b82a6a45f5e7dfc7e50725d32504448c340f458	November 16, 2023 at 11:43:29 PM GMT+2	0.0167 BTC	Illegal
14	821de2e127effdfe9061cdf07ea64a42f212fdae67ddf5d7c94597db096e1cad	October 27, 2023 at 5:47:29 PM GMT+3	0.13516812 BTC	Legal
15	c24255577eda8e3db86c419a8275cb083a212b8888d07b3f2707e08c08d5fe62	October 27, 2023 at 3:30:46 PM GMT+3	0.00193777 BTC	Illegal

FIGURE 6. An advisory framework for crypto compliance analysis

connect two nodes (addresses), indicating the flow of Bitcoin from one address to another. The edges may be weighted based on the value of the transaction. Larger transactions can have heavier weights, providing a visual indication of the transaction size. Based on these experiments, we develop a structured advisory framework that validates the legality of transactions, taking into account factors such as transaction values, time intervals, and transaction significance. The API is designed for interaction with the ML model, allowing users to make predictions based on preprocessed data. The Bitcoin graph representation depicts transactions within the Bitcoin network, with nodes representing Bitcoin addresses and edges representing transactions between them.

The graph can also incorporate time as a factor, showing how transactions evolve over a specific period. This is particularly useful for analyzing trends and patterns in the transaction network. The graph can be analyzed to identify clusters of addresses (which might represent communities or networks) and anomalies (which could indicate suspicious activities). This aids in a better understanding and interpretation of the data. Such a graph can be used to gain insights into the behavior of Bitcoin users, transaction patterns, the flow of funds, and potential red flags for illegal activities.

The developed advisory framework is able to use real crypto-transaction data and support digital forensics investigations for security officials investigating illegal money laundering activities with cryptocurrencies over Blockchain. The methodology provides a holistic approach to enhance the detection and identification of illegal transactions, offering valuable tools for Blockchain security and forensic analysis.

V. EXPERIMENTAL EVALUATION

The Value-driven-Transactional tracking Analytics for Crypto compliance (VTAC) solution that was proposed underwent experimental testing in the following phases:

- An assessment of the accuracy of detecting illegal transactions using the legality validation approach adopted by machine learning models;
- identification of illegal transactions by analysing transaction elements such as value, time, weight of transaction and frequency of the transaction over the identified time window.

The following details the experimental environment used for the method's investigation: Forensics tool that predicts bitcoin transaction legality using AI. The environment contains Python scripts and Jupyter notebooks for data preprocessing, machine learning, and visualization, focusing on Bitcoin transaction analysis with tools for API interaction, predictive modeling, and visual insights.

The Random Forest, ADA Boost and XG Boost machine learning models were used for the testing of illegal transaction detection. The key difference between a Random Forest model and a single decision tree lies in the way it constructs the individual trees and combines their predictions. Random Forest model employs a technique called bagging, which involves randomly sampling subsets of the training data for each tree. This randomization helps to reduce the correlation between the trees, leading to a more diverse and ultimately more accurate ensemble.

When an observation (in this case, a cryptocurrency transaction) that needs to be classified is received, the tree starts doing it from the top node, checks the condition written there, and continues with the arrow according to the correctness of the condition. The process continues until it reaches a node that has no more children, referred to as a leaf of the tree. It's important to note that Random Forest comprises numerous decision trees, each of which renders a decision on a particular transaction. The majority then makes a final decision. For instance, if there are 100 trees in RF, 70 of them determine a transaction's legality, while others determine it's illegality; this model would return a legal outcome. This approach takes advantage of the majority vote from the ensemble of trees, mitigating the potential for overfitting and bias that may arise from a single decision tree.

The results obtained from the detection of illegal crypto transactions over Blockchain using the above-mentioned models are presented in Table 2.

TABLE 2. Comparison of the accuracy for machine learning-based approach VTAC in detection of the illegal crypto transactions

Training and testing ratio	Random Forest	ADA Boost	XG Boost
80/20	0.9575	0.955	0.975
75/25	0.944	0.954	0.966
70/30	0.95	0.96	0.96

Comparative analysis (Table 2) of the precision of the VTAC machine learning in identifying illicit cryptocurrency transactions. The best result in the 80/20 ratio was XG Boost with a result of 97.5%.

TABLE 3. Comparison of the accuracy for Elliptic based detection of the illegal crypto transactions

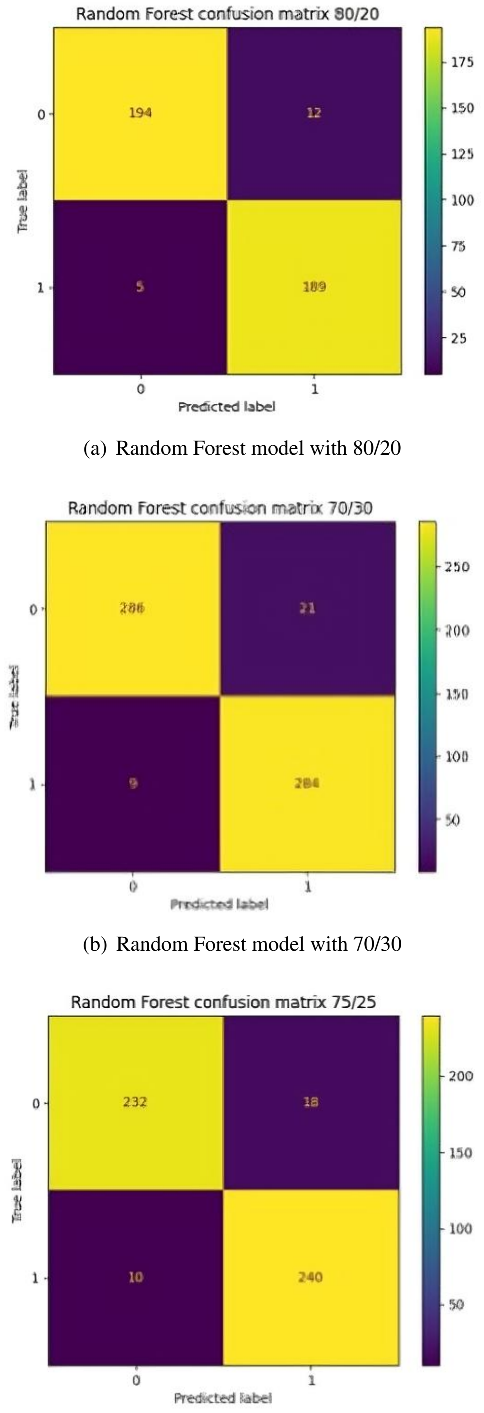
Training and testing ratio	Random Forest	ADA Boost	XG Boost
80/20	0.9554	0.9145	0.959
75/25	0.9557	0.914	0.959
70/30	0.9552	0.921	0.957

In order to compare the proposed VTAC approach, the same training and testing of the detection of illegal crypto transactions were done using an Elliptic-based model. The results of this testing are presented in Table 3. As it can be seen from Tables 2 and 3, the highest score in accuracy for Elliptic-based detection is 95.9% using the XG Boost model. This is a less accurate detection in comparison with the VTAC approach.

The confusion matrix for training and testing ratio is presented in following: with Random Forest model in Figure 7, with ADA Boost model in Figure 8, and with XG Boost model in Figure 9.

For the VTAC dataset in Table 4, the machine learning-based approach demonstrated a robust detection capability, with the Random Forest model achieving an impressive accuracy of 95.75%.

ADA Boost and XG Boost also performed admirably, recording accuracies of 96% and 97.5%, respectively. These figures underscore the effectiveness of machine learning techniques in identifying suspicious activity within VTAC's transactional data.

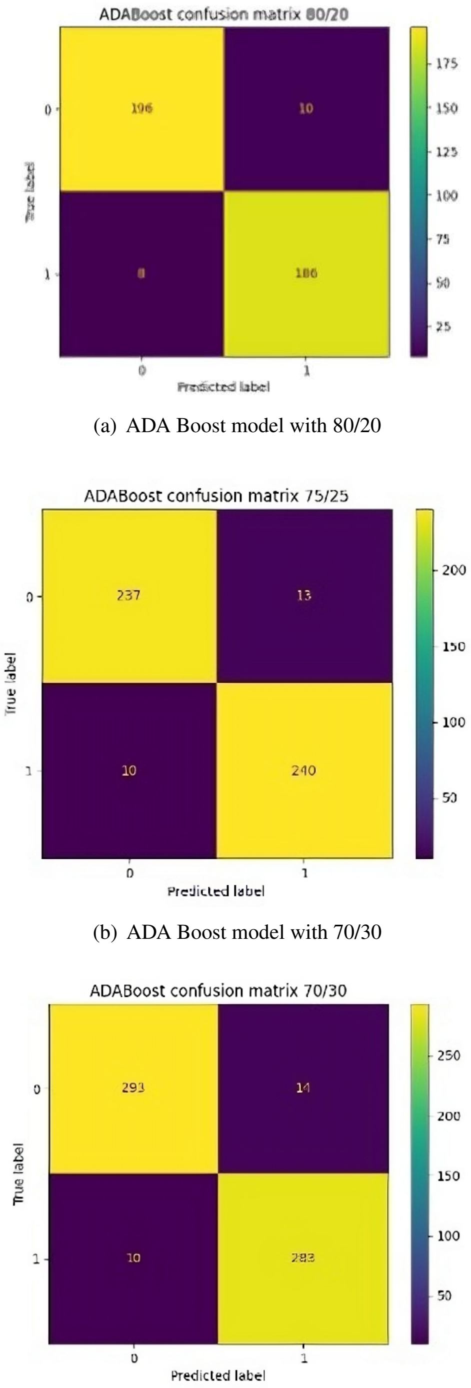


(a) Random Forest model with 80/20

(b) Random Forest model with 70/30

(c) Random Forest model with 75/25

FIGURE 7. Training and testing with Random Forest model

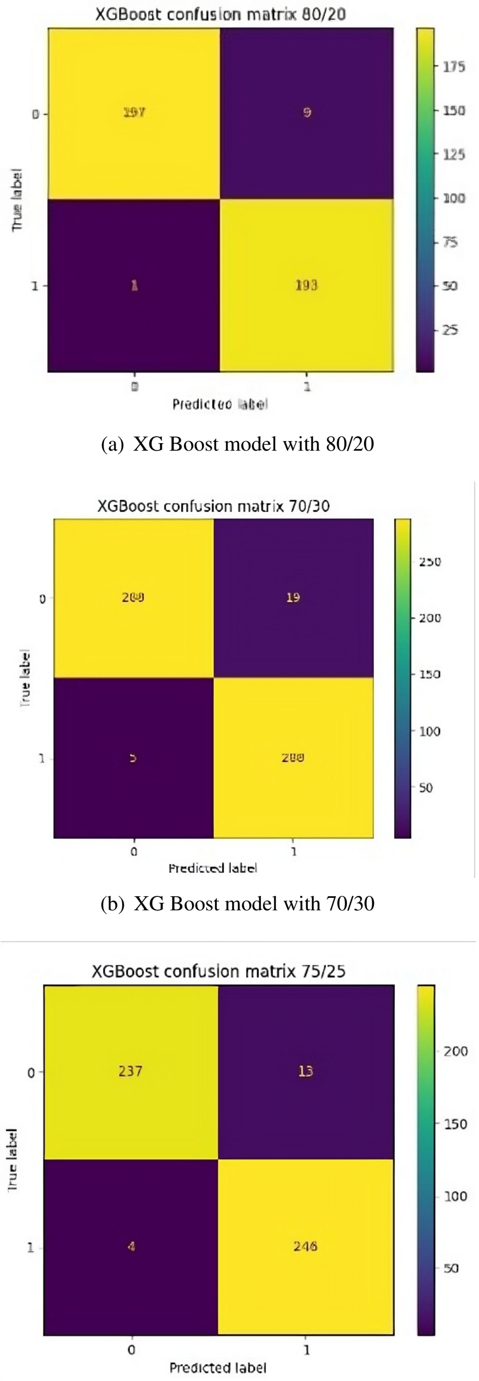


(a) ADA Boost model with 80/20

(b) ADA Boost model with 70/30

(c) ADA Boost model with 75/25

FIGURE 8. Training and testing with ADA Boost model



(a) XG Boost model with 80/20

(b) XG Boost model with 70/30

(c) XG Boost model with 75/25

FIGURE 9. Training and testing with XG Boost model

TABLE 4. Comparison of the accuracy for Elliptic based detection of the illegal crypto transactions

ML models	VTAC approach	Elliptic approach
Random Forest	0.9575	0.9557
ADA Boost	0.96	0.921
XG Boost	0.975	0.959

TABLE 5. Identification of the legacy of the crypto transactional data by predicting unknown transactions from Elliptic (10000 observations)

Proposed and compared solutions	Identified legal transactions	Identified illegal transactions
VTAC tested on real transactional data	9827	173
VTAC tested on anonymous Elliptic dataset	9806	194

As it was observed from the results that are provided in Table 5 the proposed VTAC approach can identify the illegal cryptocurrency transactions on real transactional data, and the prediction accuracy of the VTAC solution in terms of Elliptic model accuracy is 98.9%.

The proposed approach has a wide range of applications beyond its initial focus, including the prevention of recognized attacks, which greatly enhances its practical usefulness. In order to verify and strengthen the reliability of the VTAC approach, the tests were conducted using pre-identified and classified threats from the ChainAbuse portal, which is a comprehensive repository of cybersecurity incidents connected to Blockchain technology (Table 6).

TABLE 6. Tracking of illegal transactions related with type of the malicious activity

Index no.	ChainAbuse link	Amount of illegal transactions	Amount of legal transactions	Percent of illegal transactions
1	Phishing	9	6	60%
2	C3rb3r Ransomware	6	0	100%
3	Donation Impersonation Scam	2	2	50%
4	Other Blackmail Scam	1	0	100%
5	Fake Returns Scam	18	0	100%
6	Romance Scam	26	1	96%
7	Phishing Scam	9	6	60%
8	Sextortion Scam	4	0	100%
9	Hack - Other	6	0	100%
10	Ransomware	97	3	97%

The k-fold cross-validation has been done as well in order to evaluate the generalizability and robustness of the proposed VTAC approach. The authors have used 10-fold cross-validation in the experiments, ensuring a thorough assessment of model performance. The Table 7 presents the updated performance metrics obtained from 10-fold cross-validation.

TABLE 7. 10-fold cross-validation of VTAC solution

Model	Accuracy, %	Precision, %	Recall, %	F1-score, %
Random Forest	95.75	96.00	95.50	95.75
ADA Boost	96.00	96.20	95.80	96.00
XG Boost	97.50	97.60	97.40	97.50

The VTAC technique exhibited strong and consistent performance across many models and configurations. The XG Boost model regularly obtained high accuracy rates, frequently above 97%. Similarly, the precision and recall scores demonstrated robust prediction skills, with both metrics consistently exceeding 95%. The F1-score, a metric that considers both precision and recall, exhibited a remarkably high value, indicating a successful trade-off between properly identifying unlawful transactions and avoiding false detections. These measurements demonstrate VTAC's effectiveness and dependability in identifying illegal behaviors in bitcoin transactions, proving its superiority over previous approaches in dealing with intricate detection situations.

ChainAbuse provides a comprehensive collection of recorded threats, encompassing phishing attempts, ransomware events, blackmail scams, and several other types of hacks. The instances are thoroughly recorded, resulting in a comprehensive dataset of verified illicit activity within the Blockchain industry. This resource placed the VTAC approach in an actual testing environment, allowing it to be assessed against a background of confirmed malicious transactions.

By employing this VTAC solution, it is able to closely monitor the model's effectiveness in situations that closely resemble the real cybersecurity risks encountered by Blockchain networks at present. The proposed solution demonstrated a high level of precision in detecting illicit transactions linked to these attacks. This performance not only demonstrates the model's theoretical strength but also confirms its practical usefulness. The VTAC demonstrates its versatility and adaptability by accurately identifying illegal acts in many real-life situations. The importance of this accomplishment cannot be overstated.

The proposed solution's capacity to align with outcomes from established attacks validates its pertinence and effectiveness. The alignment is essential since it serves as a standard against which the model's predictions may be evaluated. Keeping up with emerging risks is crucial in the fast-changing realm of Blockchain technology and digital banking. In this aspect, the effectiveness of the VTAC architecture indicates its potential as a crucial instrument in the ongoing fight against digital financial crimes. This validation enables more extensive applications to be implemented. Financial institutions, regulatory agencies, and cybersecurity teams can utilize this paradigm to improve their monitoring and enforcement systems. By incorporating the methodology of the crypto compliance analysis into their systems, businesses can actively detect and prevent potential risks, thus strengthening the security and reliability of Blockchain transactions.

The effective implementation of the VTAC approach on acknowledged risks highlights its significance in the fields of digital forensics and Blockchain security. The organization serves as a symbol of advancement in the battle against

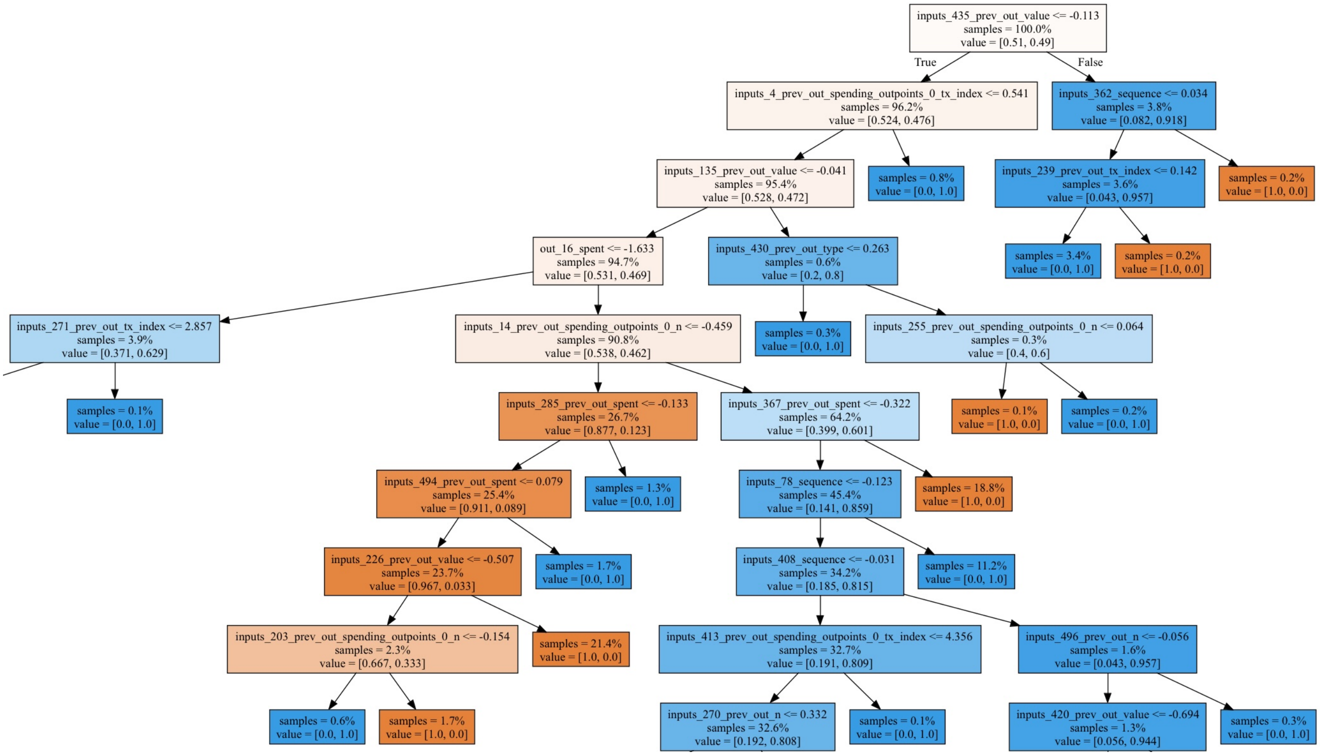


FIGURE 10. Random Forest tree in prediction of illegal transactions tracking

cybercrime, offering a flexible, productive, and successful method of protecting digital resources. While continuously improving and adjusting the VTAC approach, the authors are dedicated to enhancing the safety and security of the Blockchain ecosystem. VTAC employs sophisticated machine learning algorithms, such as Random Forest, ADA Boost, and XG Boost, that can effectively manage the intricacies and extensive datasets involved in Blockchain operations. These models excel at identifying nuanced trends and irregularities in transaction data, which may suggest unlawful activities. This is because the machine learning models used in VTAC can quickly process and evaluate large datasets. This makes it much easier to spot illegal behavior than with older methods that rely on simple statistical techniques or manual examination.

The decision tree generated by a Random Forest model is depicted in Figure 10. These visualizations are employed to depict the decision-making process of the model in determining the legality of a transaction. Every node in the tree symbolizes a choice made according to specific feature values, resulting in either further divisions or the terminal nodes of the tree, which correspond to the ultimate predictions of the model. Within the framework of our given model, the decision tree algorithm would commonly utilize many transaction attributes, such as amount, frequency, date, and maybe other derived characteristics, in order to generate a prediction. The branches of the tree indicate several routes determined by the criteria assessed at each node, while the leaves symbolize the classification result, such as 'legal' or 'illegal'. Examining this individual tree can provide us with valuable insights into the key characteristics that the model deems significant in forecasting the occurrence of unlawful transactions. This visualization is useful for understanding the model's behavior at a detailed level and conveying how it draws conclusions to stakeholders who may lack technical expertise. Additionally, it serves as a crucial instrument for enhancing the model and detecting any potential biases or vulnerabilities in its decision-making process.

The visualization in Figure 11 for tracking illegal transactions across the Blockchain is a potent tool in financial forensics. The experimental evaluation classified the transaction executed on November 17th, 2023, with a value above 47 BTC, as illegal in the identification of illegal cryptocurrency transactions. Additionally, both incoming and outgoing transactions exhibit similar characteristics, involving modest sums and the same two transactions for both inflow and outflow. This behavior might be interpreted as a form of mixing services utilizing Blockchain technology.

It elucidates the intricate pathways through which funds are transferred between wallets, highlighting patterns that may indicate illicit activity. This form of visualization is particularly valuable as it offers a macroscopic view of how suspicious wallets interact, which attributes are commonly associated with fraudulent transactions, and how they diverge from typical, lawful behavior. It aids in piecing together the transactional web woven by these wallets, thereby streamlining the investigative process. Such graphical representations enable law enforcement and regulatory bodies to comprehend the operational framework of illegal transactions. They can visualize the flow of funds, identify the convergence points of suspicious activities, and understand the financial networks' underlying structure.

An examination of the activity of Bitcoin addresses is essential for comprehending transaction patterns and detecting possible criminal actions inside the Bitcoin ecosystem. In this case, our proposed VTAC method can be compared to the BAClassifier [18], as both provide different methodologies to deal with classification and behavioral analysis.

VTAC adapts to new and evolving money laundering methods in cryptocurrency through a combination of continuous data analysis, machine learning model updates, and integration of latest blockchain forensic technologies. By leveraging adaptive machine learning algorithms, VTAC can dynamically update its detection models based on new transaction patterns and anomalies detected across the blockchain network. This ensures that VTAC remains at the forefront of identifying not only known methods of money laundering but also emerging tactics that deviate from established patterns. Additionally, VTAC's system incorporates feedback mechanisms that allow it to learn from false positives and negatives, thereby refining its accuracy and responsiveness to new threats. This capability is supported by ongoing research and collaboration with blockchain analysts and cybersecurity experts, ensuring that VTAC's methodologies are aligned with the latest trends and innovations in cryptocurrency transactions.

VTAC includes automated ML procedures to identify the true identity of individuals, strategies to standardize data, and giving priority to analyzing the value of transactions in order to precisely identify patterns of money laundering. VTAC's machine learning models have the ability to quickly analyze large datasets, allowing for real-time monitoring and continual adjustment to changing money laundering techniques. This improves the detection of illegal activities compared to older methods. However, BAClassifier introduces an approach to classifying the behavior of Bitcoin addresses. It does this by converting address transactions into graph structures and using graph neural networks to learn features. In this case, VTAC prioritizes real-time monitoring and adaptability to new laundering strategies with advanced machine learning algorithms, while BAClassifier stands out in its capacity to accurately classify addresses using graph neural networks. VTAC excels in its ability to rapidly analyze large datasets and identify subtle patterns, while BAClassifier stands out for its inventive methodology in tackling categorization and its potential to further investigate illicit activities in the Bitcoin network.

VI. DISCUSSION

In the discussion section of our article, we analyzed and interpreted the results obtained from the application of three machine learning models – Random Forest, ADA Boost, and XG Boost – on the Elliptic dataset for predicting the legality of Bitcoin transactions. The results are especially significant given the high degree of deanonymization (99.5%) achieved with the Elliptic dataset, although it's important to acknowledge the potential presence of unidentified transactions.

The experiment was conducted using different training-testing splits (80/20, 75/25, and 70/30). Notably, the Random Forest model consistently demonstrated high accuracy across all splits, slightly outperforming the other models. This suggests robustness in the Random Forest approach, particularly in handling the complexities inherent in the Elliptic dataset. While both ADA Boost and XG Boost showed commendable performance, XG Boost slightly edged out in most scenarios. This could be attributed to its advanced handling of gradient boosting, which is particularly effective in datasets with numerous features, as is the case with the Elliptic data. The possibility of unidentified transactions in the dataset underscores the need for continuous model training and adaptation. Future research could focus on iterative model refinement as more data becomes available.

While the current models perform well on the Elliptic dataset, their applicability to other Blockchain networks or different types of transactions (e.g., those involving privacy coins) remains to be tested. Future work should explore the generalizability of these models across various Blockchain environments. The implementation of these models in a real-time monitoring system for Bitcoin transactions poses additional challenges, including handling live data streams, rapid changes in transaction patterns, and scalability.

While the proposed Value-driven-Transactional tracking Analytics for Crypto compliance (VTAC) approach offers significant advancements in detecting illegal cryptocurrency transactions, there are still limitations and challenges that could affect its implementation and effectiveness. Criminal money laundering methodologies are constantly evolving, with perpetrators quickly adapting to circumvent detection mechanisms. The VTAC system's ability to keep pace with these evolving strategies without frequent updates and re-training of its machine learning models poses a significant challenge. Like all machine learning-based systems, VTAC may be prone to false positives (legitimate transactions flagged as illegal) and false negatives (illegal transactions not detected). Balancing sensitivity and specificity to minimize these errors, without overwhelming investigators with false leads or missing critical transactions, is a critical challenge. The VTAC approach needs to be compatible with current financial monitoring systems and protocols. Integration challenges could arise, requiring significant adjustments or redesigns of existing systems to accommodate the new approach.

In order to boost the VTAC capability to identify new laundering patterns, it would be beneficial to incorporate a wider range of dynamic datasets into the existing methodology. Using cutting-edge methods like deep learning or neural networks, and incorporating real-time Blockchain transaction data, could improve the accuracy of detection. Moreover, implementing a collective methodology that integrates many machine learning models might enhance resilience against

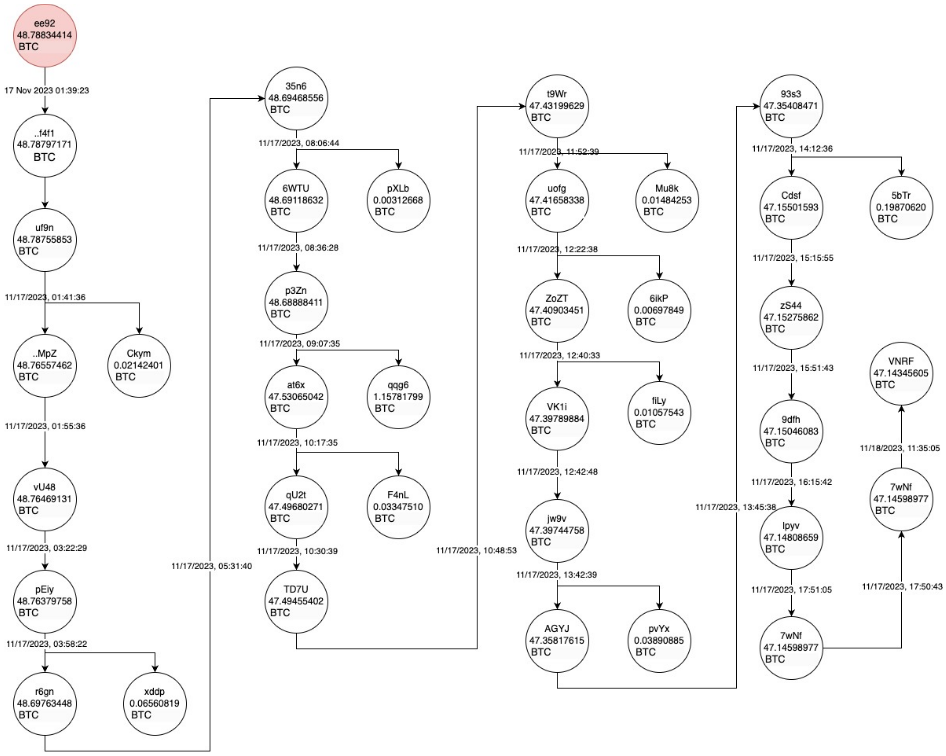


FIGURE 11. Tracking of the illegal transactions over Blockchain visualisation tracking

incorrect identifications and flexibility in response to emerging money laundering techniques. Collaborations with Blockchain platforms to facilitate data access and information exchange with regulatory authorities could enhance the depth of knowledge and effectiveness of combating money laundering operations.

A pivotal concern regarding the practical application of our proposed algorithm centers on the de-anonymization process, which relies partially on external datasets. The utility and ethical considerations of gathering such data in practice require careful justification. Typically, these datasets could be sourced from openly available or proprietary transaction databases that comply with data privacy laws, ensuring that the collection process respects user anonymity and data integrity. Furthermore, our initial experiments utilized a dataset available on Kaggle, which, while valuable for preliminary testing, may not fully demonstrate the algorithm's generality across diverse real-world scenarios. To address this, future research should focus on applying the model to proprietary datasets, which would not only enhance the model's applicability but also provide a robust test of its effectiveness across various transactional environments. Such studies would help in verifying the model's generality and its adaptability to different types of data, potentially increasing its utility in practical blockchain analytics.

Conducting an in-depth time complexity analysis emerges as a pivotal area, aiming to refine VTAC for efficient real-time application. This includes pinpointing optimal time windows to accurately identify illegal Blockchain activities, ensuring swift and precise detection. Further exploration into cutting-edge machine learning techniques and algorithms is also anticipated, with the goal of elevating VTAC's accuracy while minimizing computational requirements. Future work also sees the importance of expanding the model's applicability to various cryptocurrencies with different privacy features and transaction structures. These future directions aim to bolster VTAC's efficacy as a tool in the fight against cryptocurrency-based money laundering.

VII. CONCLUSIONS

The findings from this study are promising for the field of digital forensics in cryptocurrency. The high accuracy of the Random Forest, ADA Boost, and XG Boost models in classifying transactions on the Elliptic dataset demonstrates the potential of machine learning in identifying illegal activities in Blockchain networks. However, the presence of unidentified transactions and the specific characteristics of Blockchain data call for ongoing research and model refinement. The future of Blockchain forensics will likely hinge on the ability to adapt to evolving transaction patterns and the integration of these models into comprehensive, real-time monitoring systems.

The first experiment, focusing on detection, utilized a machine learning model that achieved a detection accuracy of 97.5%. The identification experiment further analyzed transactions, considering the value, timing, and weight factors, leading to the correct identification of illegal transaction beneficiaries with a precision rate of 98.9%. These results not only validate the efficacy of our model but also highlight the critical role of attribute selection in enhancing the model's predictive capabilities. The integration of temporal and transaction weight considerations has markedly improved the model's discernment, providing a powerful tool for forensic analysis and contributing to the security and transparency of Blockchain transactions.

This study introduces the Value-driven-Transactional tracking Analytics for Crypto compliance (VTAC), a novel machine learning-based architecture designed to enhance the detection of illegal cryptocurrency transactions over blockchain technology. VTAC stands out by employing advanced analysis techniques that include a sophisticated machine learning framework capable of analyzing factors such as digital wallet hashes, transaction values, and frequency over time, which significantly improves detection accuracy. The method incorporates a unique dataset preparation through an automated de-anonymization process that allows for effective testing against real transaction data, achieving a remarkable detection accuracy of 97.5% using the XG Boost model, thus outperforming existing methods with accuracies up to 95.9%. Additionally, VTAC develops an advisory framework that not only aids in the detection but also in the reporting of suspicious transactions, providing a structured approach to crypto compliance analysis. These advancements underscore VTAC's contribution to the field, making it a significant step forward in the fight against financial crimes facilitated by cryptocurrencies.

REFERENCES

[1] János Besenyő and Attila Gulyas. The effect of the dark web on the security. Journal of Security & Sustainability Issues, 11(1), 2021.

[2] Chang-Yi Lin, Hsiang-Kai Liao, and Fu-Ching Tsai. A systematic review of detecting illicit bitcoin transactions. Procedia Computer Science, 207:3217–3225, 2022.

[3] Lim Yu Qian. Most damaging methods of crypto hacks and exploits in 2022. https://www.coingecko.com/research/publications/crypto-hacks-exploits-by-method, 2023-10-23.

[4] Josh Gayta. Is it possible to hack cryptocurrency? https://www.coingecko.com/research/publications/crypto-hacks-exploits-by-method, 2023-11-01.

[5] Julija Golosova and Andrejs Romanovs. The advantages and disadvantages of the blockchain technology. In 2018 IEEE 6th workshop on advances in information, electronic and electrical engineering (AIEEE), pages 1–6. IEEE, 2018.

[6] Eli Ben Sasson, Alessandro Chiesa, Christina Garman, Matthew Green, Ian Miers, Eran Tromer, and Madars Virza. Zerocash: Decentralized anonymous payments from bitcoin. In 2014 IEEE symposium on security and privacy, pages 459–474. IEEE, 2014.

[7] Yannan Li, Guomin Yang, Willy Susilo, Yong Yu, Man Ho Au, and Dongxi Liu. Traceable monero: Anonymous cryptocurrency with enhanced accountability. IEEE Transactions on Dependable and Secure Computing, 18(2):679–691, 2019.

[8] Sean Foley, Jonathan R Karlsen, and Tālis J Putniņš. Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? The Review of Financial Studies, 32(5):1798–1853, 2019.

[9] Mingdong Liu, Hu Chen, and Jiaqi Yan. Detecting roles of money laundering in bitcoin mixing transactions: A goal modeling and mining framework. Frontiers in Physics, 9:665399, 2021.

[10] Annelieke Mooij. Currency (layering). In Regulating the Metaverse Economy: How to Prevent Money Laundering and the Financing of Terrorism, pages 69–86. Springer, 2023.

[11] Bruno Moslavac. Cryptocurrency tumbler: legality, legalization, criminalization. Revista Acadêmica Escola Superior do Ministério Público do Ceará, 11(2):205–226, 2019.

[12] Jesse Crawford and Yong Guan. Knowing your bitcoin customer: Money laundering in the bitcoin economy. In 2020 13th international conference on systematic approaches to digital forensic engineering (SADFE), pages 38–45. IEEE, 2020.

[13] Anton Wahrstätter, Jorão Gomes, Sajjad Khan, and Davor Svetinovic. Improving cryptocurrency crime detection: Coinjoin community detection approach. IEEE Transactions on Dependable and Secure Computing, 2023.

[14] M Mazhar Rathore, Sushil Chaurasia, and Dhirendra Shukla. Mixers detection in bitcoin network: a step towards detecting money laundering in crypto-currencies. In 2022 IEEE International Conference on Big Data (Big Data), pages 5775–5782. IEEE, 2022.

[15] Ardeshir Shojaeinasab, Amir Pasha Motamed, and Behnam Bahrak. Mixing detection on bitcoin transactions using statistical patterns. IET Blockchain, 3(3):136–148, 2023.

[16] Yining Hu, Suranga Seneviratne, Kanchana Thilakarathna, Kensuke Fukuda, and Aruna Seneviratne. Characterizing and detecting money laundering activities on the bitcoin network. arXiv preprint arXiv:1912.12060, 2019.

[17] Ismail Alarab, Simant Prakoonwit, and Mohamed Ikbal Nacer. Competence of graph convolutional networks for anti-money laundering in bitcoin blockchain. In Proceedings of the 2020 5th international conference on machine learning technologies, pages 23–27, 2020.

[18] Zhengjie Huang, Yunyang Huang, Peng Qian, Jianhai Chen, and Qinming He. Demystifying bitcoin address behavior via graph neural networks. In 2023 IEEE 39th International Conference on Data Engineering (ICDE), pages 1747–1760. IEEE, 2023.

[19] Ning Lu, Yuan Chang, Wenbo Shi, and Kim-Kwang Raymond Choo. Coinlayering: an efficient coin mixing scheme for large scale bitcoin transactions. IEEE Transactions on Dependable and Secure Computing, 19(3):1974–1987, 2020.

[20] Rasmus Ingemann Tuffveson Jensen and Alexandros Iosifidis. Fighting money laundering with statistics and machine learning. IEEE Access, 11:8889–8903, 2023.

[21] Johrha Alotibi, Badriah Almutanni, Tahani Alsubait, Hosam Alhakami, and Abdullah Baz. Money laundering detection using machine learning and deep learning. International Journal of Advanced Computer Science and Applications, 13(10), 2022.

[22] Chaehyeon Lee, Sajan Maharjan, Kyungchan Ko, and James Won-Ki Hong. Toward detecting illegal transactions on bitcoin using machine-learning methods. In Blockchain and Trustworthy Systems: First International Conference, BlockSys 2019, Guangzhou, China, December 7–8, 2019, Proceedings 1, pages 520–533. Springer, 2020.

[23] Martin Jullum, Anders Løland, Ragnar Bang Huseby, Geir Ånonsen, and Johannes Lorentzen. Detecting money laundering transactions with machine learning. Journal of Money Laundering Control, 23(1):173–186, 2020.

[24] Eric Pettersson Ruiz and Jannis Angelis. Combating money laundering with machine learning–applicability of supervised-learning algorithms at cryptocurrency exchanges. Journal of Money Laundering Control, 25(4):766–778, 2022.

[25] Joana Lorenz, Maria Inês Silva, David Aparício, João Tiago Ascensão, and Pedro Bizarro. Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity. In Proceedings of the first ACM international conference on AI in finance, pages 1–8, 2020.

[26] Joana Susan Lorenz. Machine learning methods to detect money laundering in the Bitcoin blockchain in the presence of label scarcity. PhD thesis, Universidade NOVA de Lisboa (Portugal), 2021.

[27] Christian Leuprecht, Caitlyn Jenkins, and Rhianna Hamilton. Virtual money laundering: policy implications of the proliferation in the illicit use of cryptocurrency. Journal of Financial Crime, 30(4):1036–1054, 2023.

[28] Muhammad Subtain Raza, Qi Zhan, and Sana Rubab. Role of money mules in money laundering and financial crimes a discussion through case studies. Journal of Financial Crime, 27(3):911–931, 2020.

[29] Mohd Irwan Abdul Rani, Sharifah Nazatul Faiza Syed Mustapha Nazri, and Salwa Zolkaflil. A systematic literature review of money mule: Its roles, recruitment and awareness. Journal of Financial Crime, 2023.

[30] Christoph Wronka. "cyber-laundering": the change of money laundering in the digital age. Journal of Money Laundering Control, 25(2):330–344, 2022.

[31] Ajay Kumar, Kumar Abhishek, Pranav Nerurkar, Mohammad R Khosravi, Muhammad Rukunuddin Ghalib, and Achyut Shankar. Big data analytics to identify illegal activities on bitcoin blockchain for iomt. Personal and Ubiquitous Computing, pages 1–12, 2021.

[32] Elliptic Data Set. Bitcoin transaction graph. https://www.kaggle.com/datasets/ellipticco/elliptic-data-set, 2023-11-13.
[33] Randa Natras, Benedikt Soja, and Michael Schmidt. Ensemble machine learning of random forest, adaboost and xgboost for vertical total electron content forecasting. Remote Sensing, 14(15):3547, 2022.
[34] Elliptic Data Set. Elliptic bt. https://www.kaggle.com/code/satto99ab/elliptic-bt, 2023-11-15.
[35] Blockchain. Be early to the future of finance. https://www.blockchain.com/, 2023-11-22.
[36] Linas Pocius. Forensics tool that predicts bitcoin transaction legality using ai. https://github.com/Linupo/MBD?tab=readme-ov-file, 2023-12-22.

A. VENČKAUSKAS is currently the head of the Computer Science Department and research group of Cyber Security of the Kaunas University of Technology. His research areas are information technology, Internet of Things, cyber security, applied cryptography, application of artificial intelligence methods to cyber security, application of distance learning technologies. He is the author and co-author of more than 65 articles in various international journals, including 34 publications in scientific periodicals, cited in the Clarivate Analytics Web of Science database with Impact Factor, his h-index is 10, and he has written 11 textbooks. Prof. A. Venčkauskas was a leader and principal investigator in number of projects, i.e.: 2019-2022 Project "Strategic programs for advanced research and technology in Europe" (SPARTA), H2020; 2020 Project “Model for the organization of the remote work and training process and recommendations for the extreme and transitional period (LMT, No. S-COV-20-20;) 2020-2021 projects on security awareness and training “A comprehensive cyber security training programs for employees of state and municipal institutions and organizations” (CVPA procurement No. 458424. 2020, 2021). He is the organizer and manager of the Information and IT Security master’s program, which has been running for 12 years in a blended learning way using distance learning technologies. He is a member of the Cyber Security Council of the Ministry of National Defense of the Republic of Lithuania. He represents the university in international and national institutions that create and implement science and innovation policy: Ministry of Economy and Innovation, Industry 4.0 platform; MOSTA Smart specialization.

Š. GRIGALIŪNAS (Member, IEEE) is a cyber security officer at Lithuanian RailWays. His primary competencies lie within the IT sector, where he started as an IT security auditor. He is ISECOM certified penetration tester, certified ISO 27001 ISMS auditor, and holds a doctorate from the Kaunas University of Technology, where he is also an associate professor. More than 15 years experience in IT auditing and security, interested in cyber security. 6 years worked as Sr. cyber security Consultant at Business. He is an analyst of social network security, cyber attacks, psychology and practical implementations of digital forensics evidence, and new gamification methods of learning security awareness, building a culture of security, or managing insider threats. As Chairperson of TK 79 Information Security in the Lithuanian Standardization Department, lead the development and harmonization of information security standards. Keep abreast of the latest financial industry regulations and standards (such as PCI-DSS, ISO 27001, and ICT requirements from BoL) and ensure that the company is in compliance with these legal and regulatory requirements.

R. BRŪZGIENĖ is currently working as an associate professor in the KTU Department of Computer Sciences and is also a researcher in the scientific group of cyber security. Her fields of scientific interest are cyber security; communication networks, its resilience and security; security of critical infrastructure and cyber-physical systems; reliability and efficiency of communication systems; cyber-sustainability. She is the author of 31 scientific publications, 4 scientific monographs and parts thereof, and also 2 educational books.She is the leader of five international NordPlus study projects, related to excellence and best practice in the fields of cybersecurity and intelligent communication. She is/was a lecturer in integrated cyber security trainings for employees of state and municipal institutions on control and management of systems in critical, high-incertitude situations as well as on risk management in the electronic resources managed by the institution.

L. POCIUS is a software engineer with 5 years of experience at Bentley Systems digital infrastructure engineering company, while also pursuing a Master’s degree in Information Security at Kaunas University of Technology (KTU). This dual focus allows them to bring both development expertise and security awareness to their work. DevSecOps practices are not only of interest to him, but also a core responsibility in their role as a software engineer. He is instrumental in ensuring that the company stays up-to-date with new industry standards and adheres to the latest security best practices. In addition his main research interests lie in leveraging automation and artificial intelligence to further improve software development processes.

A. ROMANOVS (Senior Member, IEEE) received the Ph.D. degree in information technology (system analysis, modeling, and design) from the Transport and Telecommunication Institute, Latvia, in 2007. He is currently an Associate Professor and a Senior Researcher with the Institute of Information Technology, Riga Technical University (RTU), the Head of the RTU Department of Modelling and Simulation, and the Director of two international master's study programs "Cybersecurity Engineering" and "Logistics and Supply Chain Management." He has 20 years of teaching experience at RTU and over 35 years of professional experience in the field of IT. He has authored over 150 books and papers in scientific journals and conference proceedings, and organizer of 30 international scientific conferences. His research interests include modeling information systems, cyber security, integrated IT in supply chain management and e-commerce, and education in these areas. Dr. Romanovs is a member of the Latvian Simulation and Modeling Society, Information Systems Audit and Control Association (ISACA), IBM Academic Initiative, Palo Alto Networks, Pearson Higher Education Network, and Check Point Secure Academy, and an Expert of the Latvian Scientific Council (in information technology), RTU Senator. He is the Founder of the Computer Society Chapter and Blockchain Group in Latvia Section, and the Chair and a member of various IEEE committees, such as the IEEE Educational Activities Board (Section Education Outreach Committee) from 2023 to 2024, the MGA Chapter Operations Support Committee, from 2023 to 2024, the MGA Membership Recruitment and Recovery Committee, from 2017 to 2021, the MGA Admission and Advancement Committee, from 2019 to 2020, the R8 Professional and Educational Activities Committee, from 2023 to 2024, the R8 Chapter Coordination Committee, from 2021 to 2024, the R8 Membership Development Committee, from 2015 to 2021, and the Latvia Section C hair, from 2012 to 2013 and from 2016 to 2017.

• • •