A Machine Learning Framework for Detecting Illicit Cryptocurrency Transactions

E Pavithra[1], E Vasantha Kumar[2], K Dhanamjay[3]

[1]Assistant Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, E-mail: epavithra526@gmail.com, ORC-ID: https://orcid.org/0009-0006-8871-4551

[2]P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, E-mail: vasanthkumarv892@gmail.com, ORC-ID: https://orcid.org/0009-0004-8538-7108

[3]Assistant Professor, Department of CSE(AI & ML), Sri Venkatesa Perumal College of Engineering & Technology, Puttur, E-mail: kanipakkamdhanamjay@gmail.com

Abstract: Money laundering using cryptocurrency networks poses an escalating challenge owing to the intrinsic anonymity of blockchain transactions. Layering strategies employed on decentralized platforms conceal the identities of senders and receivers, hence complicating compliance and enforcement efforts. A dual-layered strategy integrating Machine Learning (ML) and Value-driven Transactional Tracking Analytics for Crypto-compliance (VTAC) is presented to resolve this issue. This system facilitates efficient de-anonymization and categorization of criminal transactions. The machine learning component encompasses data preprocessing, standard scaling normalization, model training with supervised classifiers, and hash address identification from transaction identifiers. VTAC improves detection through the analysis of transaction frequency and behavioral irregularities. Three foundational classifiers—Random Forest, XGBoost, and AdaBoost—are assessed. Additional enhancements are implemented by a hybrid ensemble of XGBoost and Random Forest, along with advanced algorithms such as LightGBM and CatBoost. Performance is evaluated by accuracy, precision, recall, F1-score, and confusion matrix on the Elliptic Bitcoin dataset. Results indicate enhanced detection rates using LightGBM. A web-based interface enables real-time prediction of unlawful transactions, hash traceability, and compliance reporting. This approach enhances blockchain forensics and regulatory monitoring by merging analytics with blockchain infrastructure.

"Index Terms: *Machine learning, blockchain, cybercrime, cryptocurrency, money laundering".*

1. INTRODUCTION

Cryptocurrencies have swiftly transformed into a prevalent financial instrument, extensively utilized for both legitimate and illicit activities. Their decentralized architecture, cryptographic security, and user anonymity render them appealing not only to investors and enterprises but also to nefarious individuals aiming to exploit financial systems devoid of regulatory supervision [1]. Among the myriad cybercrimes facilitated by digital currencies, money laundering has surfaced as one of the most significant and formidable threats. The lack of centralized oversight and Know Your Customer (KYC) rules at numerous cryptocurrency exchanges fosters an optimal setting for the laundering of criminal cash.

Criminals exploit the anonymity and pseudonymity provided by blockchain networks to obfuscate, conceal, and transfer funds acquired through illicit activities. These monies are frequently moved through several digital wallets, decentralized exchanges (DEXs), and privacy currencies to obscure transaction trails and evade discovery by financial regulators [3]. Cryptocurrencies such as Bitcoin, Monero, and Zerocash are commonly utilized in dark web transactions related to drug trafficking, identity theft, arms trade, and other illicit activities. Privacy-focused cryptocurrencies, specifically, use advanced obfuscation methods including ring signatures, stealth addresses, and zero-knowledge proofs, hence complicating forensic blockchain investigation significantly [5].

Although blockchain is fundamentally secure owing to its distributed ledger and cryptographic features, vulnerabilities persist inside the ecosystem. Malefactors frequently exploit vulnerabilities in smart contracts, execute phishing campaigns, and employ fraudulent malware or trojans to obtain private keys and digital wallets [6]. Upon the compromise of private keys, hackers can clandestinely transfer digital assets across various addresses, utilizing mixers, tumblers, and privacy coins to further conceal their activities.

Regulatory authorities and financial institutions encounter increasing challenges in tracking such activities owing to the pseudonymous characteristics of the majority of blockchain transactions. Despite the public recording of each transaction, the lack of association with real-world identities presents a considerable barrier. Conventional anti-money laundering (AML) instruments are inadequate for blockchain systems, particularly in monitoring intricate, multi-tiered cryptocurrency transactions [8].

2. LITERATURE REVIEW

Scharfman [9] underscores the urgent necessity of incorporating anti-money laundering (AML) protocols into cryptocurrency platforms to prevent the misuse of digital assets for illicit activity. He examines the compliance obstacles encountered by virtual asset service providers (VASPs) and offers pragmatic strategies for maintaining anti-money laundering (AML) compliance within cryptocurrency exchanges. Scharfman delineates the structural disparities between conventional finance and blockchain-based assets, advocating for a compliance-oriented operational framework for cryptocurrency ecosystems. This framework encompasses customer due diligence (CDD), transaction monitoring, and suspicious activity reporting, with the objective of aligning cryptocurrency operations with international regulatory standards.

Chang et al. [10] investigated the influence of blockchain technology on financial services through an extensive study that included interviews with subject-matter experts. Their findings emphasize that although blockchain enhances transparency and security in financial systems, its decentralized and immutable characteristics provide significant obstacles for fraud detection and compliance enforcement. The report emphasizes the critical need for industry-wide standards and proposes a multi-stakeholder strategy that includes regulators, developers, and institutions to produce AML guidelines specifically designed for decentralized environments. The authors contend that the effective integration of blockchain in finance relies on harmonizing innovation with risk management.

Jullum et al. [11] suggested a machine learning system for detecting money laundering activities, emphasizing the identification of suspicious behavior within extensive financial transaction data. Their research utilized supervised learning algorithms trained on historical data categorized as either suspicious or non-suspicious. Experiments revealed that machine learning models substantially surpass traditional rule-based systems in accuracy and adaptability. The authors underscored the necessity for explainable models, particularly in regulatory contexts where auditability and transparency are critical. Their methodology provides a basis for implementing analogous techniques in cryptocurrency transaction surveillance.

Gerbrands et al. [12] examined the efficacy of AML policies by empirical network analysis. Their research sought to measure the effects of regulatory actions by analyzing transaction networks across different jurisdictions. The study revealed that whereas AML policies exhibit observable impacts, their efficacy is frequently constrained by inconsistent implementation and insufficient global coordination. Gerbrands and associates proposed that improved information-sharing mechanisms and international regulatory collaborations are crucial for effective AML enforcement. This research advocates that technology solutions require concurrent policy harmonization to achieve optimal efficacy.

Serena, Ferretti, and D'Angelo [13] proposed a graph-based methodology for examining cryptocurrency activity by representing transaction histories as intricate networks. Their research offered significant insights into the framework and development of digital financial conduct. Utilizing network science methods, the authors were able to discover anomalous transaction patterns, pinpoint central actors, and evaluate systemic risk. Their research substantiates the claim that blockchain's transparency can be utilized for forensic investigations when integrated with sophisticated graph analytics. The research serves as a fundamental foundation for developing automated systems that can do real-time risk assessments of wallets and transactions.

Pareja et al. [14] introduced EvolveGCN, an innovative approach for handling dynamic graph data with evolving graph convolutional networks. EvolveGCN, initially utilized in dynamic graph contexts like social networks, holds considerable relevance for transaction monitoring in blockchain, where the network architecture undergoes rapid evolution. Their research incorporated a temporal element into Graph Convolutional Networks (GCNs), allowing the model to assimilate information from both structural and temporal alterations in graphs. This dynamic adaptation is especially beneficial for detecting money laundering, as transactions frequently exhibit time-sensitive patterns intended to conceal tracking.

Lo et al. [15] presented Inspection-L, a self-supervised graph neural network (GNN) architecture that produces node embeddings for the identification of money laundering activities in Bitcoin. Their research employed blockchain data to represent financial activity as a graph and implemented unsupervised learning to detect unusual behavior without dependence on labeled datasets. This approach mitigates the prevalent challenge of label scarcity in financial fraud detection by enabling the system to discern significant patterns from unprocessed transaction data. The experimental findings indicated that Inspection-L surpassed other models for precision

and recall, highlighting the efficacy of GNN-based methodologies in bitcoin compliance.

Adewumi and Akinyelu [16] conducted an extensive survey on machine learning and nature-inspired methodologies for detecting credit card fraud. The approaches examined, including artificial neural networks, genetic algorithms, and swarm intelligence, while centered on conventional financial systems, provide applicable insights for the surveillance of digital assets. The authors emphasized the significance of integrating different models (ensemble learning) and modifying algorithms to address advancing fraud methodologies. Their analysis underscores the necessity for hybrid models and continuous learning frameworks, which are becoming increasingly pertinent in the turbulent and swiftly changing realm of bitcoin transactions.

3. MATERIALS AND METHODS

The suggested method seeks to identify money laundering activities on blockchain by incorporating advanced Machine Learning techniques with Value-driven Transactional Tracking Analytics for Crypto-compliance (VTAC). The system initiates by loading and prepping the Elliptic Bitcoin dataset, which involves addressing missing values and converting non-numeric properties. Data normalization is executed by Standard Scaling to ready it for model training. Various machine learning techniques are employed to categorize transactions as legitimate or illicit depending on their behavior. A de-anonymization module correlates transaction IDs with hash addresses, facilitating the identification of prospective participants in unlawful activity. The VTAC component enhances detection by identifying transactions characterized by atypically high frequency or dubious value patterns. Moreover, the system integrates improvements via hybrid machine learning models and cutting-edge algorithms, like LightGBM and CatBoost. An intuitive interface is available for uploading and processing fresh transaction data, assessing its legality, and obtaining associated hash addresses for additional blockchain-level analysis.

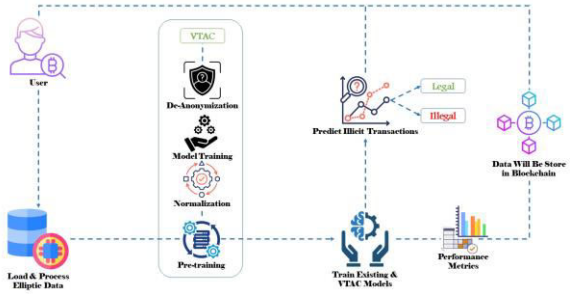


Fig.1 Proposed Architecture

The diagram depicts a decision-tree approach for identifying illicit cryptocurrency transactions. It analyzes cryptocurrency transactional data via various tree models to discern patterns. Attributes such as wallet hashes, value, and transaction frequency inform predictions. The technology identifies possible illegal activities and generates findings for recognition, improving precision in fraud detection and prevention.

a) Modules:

Signup: This module enables new users to register by providing essential information, like username, email, and password. Following successful registration, user information is securely saved in the database, facilitating access to system features and assuring individualized engagement for monitoring blockchain transactions and model predictions.

User Login: This module allows registered users to authenticate using their credentials. It authenticates users by cross-referencing information with the

database. Upon successful authentication, users obtain access to diverse features such as data loading, model training, and transaction prediction, thereby assuring secure and regulated access to the system.

Load & Process Elliptic Dataset: This module enables the uploading and preprocessing of the Elliptic Bitcoin Dataset. It manages data cleansing, imputation of missing values, and normalization. Subsequent to processing, the data is partitioned into training and testing sets, facilitating machine learning activities and enabling effective data management for precise analysis.

Train Existing & VTAC Models: This module instructs multiple machine learning models utilizing the processed dataset. It encompasses classic, proposed, and expanded models, with and without VTAC integration. The module assesses the performance of each model through criteria like as accuracy, precision, recall, and F-score to determine the most effective strategy.

Predict Illicit Transactions: This module enables users to upload new transaction data for analysis. The trained model forecasts the legality of each transaction, recognizes hash addresses, and presents transaction data. It aids in de-anonymizing dubious activity and facilitates proactive identification of money laundering.

Logout: This module securely terminates the user session, erasing access tokens and session data. It guarantees the protection of user credentials and activity data, thwarting unauthorized access and preserving system integrity following a user's contact with the system.

b) Methods/ Algorithms:

Existing Random Forest: Random Forest classifies transactions as legitimate or illegal by generating several decision trees and aggregating their outputs. It effectively manages unbalanced data and mitigates overfitting. The ensemble approach enhances accuracy by averaging predictions from many decision pathways, rendering it effective for identifying trends in blockchain transaction behavior.

Existing XGBoost: XGBoost is utilized for high-performance classification owing to its capacity to efficiently manage extensive datasets with precision. It employs gradient boosting with regularization, thereby diminishing both bias and variation. It effectively detects anomalous transaction patterns through recurrent data analysis and performance optimization, resulting in reduced mistakes.

Existing AdaBoost: AdaBoost functions by integrating several weak learners to create a robust classifier. It modifies weights following each iteration, emphasizing more challenging transactions for classification. This renders it efficient for discerning nuanced distinctions between lawful and unlawful blockchain activities, however it may be susceptible to data noise.

Propose VTAC Random Forest: This approach augments Random Forest with VTAC, which identifies recurrent high-value transfers. VTAC flags transactions according to volume and frequency, implementing a rule-based filter prior to classification. It enhances the model's precision in detecting recurring laundering patterns and augments its dependability in identifying high-risk transactions.

Propose VTAC XGBoost: The integration of XGBoost with VTAC facilitates the preliminary filtration of potentially illicit transfers prior to

classification. VTAC detects anomalous frequency and value patterns, whereas XGBoost classifies them with precision. This integration greatly enhances detection capability by correlating transaction attributes with sophisticated model learning.

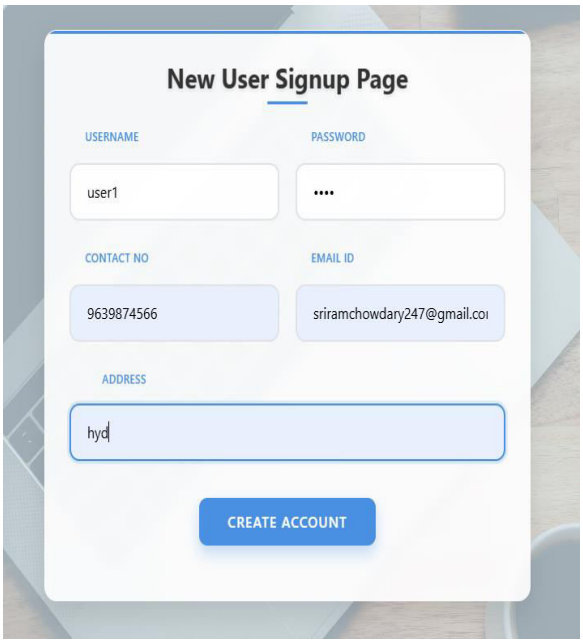
Propose VTAC AdaBoost: AdaBoost is augmented with VTAC to favor dubious transactions according to volume. VTAC preprocesses data, while AdaBoost emphasizes learning from high-risk instances. This enhances the model's sensitivity to laundering activity and elevates overall classification quality, notwithstanding AdaBoost's usual susceptibility to noisy data.

Extension1 Hybrid Model: The hybrid model integrates XGBoost and Random Forest to leverage the advantages of both, enhancing accuracy and diminishing variance. The optimization of XGBoost combined with the ensemble stability of Random Forest produces a robust classifier that surpasses individual models, particularly in the context of intricate transaction patterns associated with concealed laundering activities.

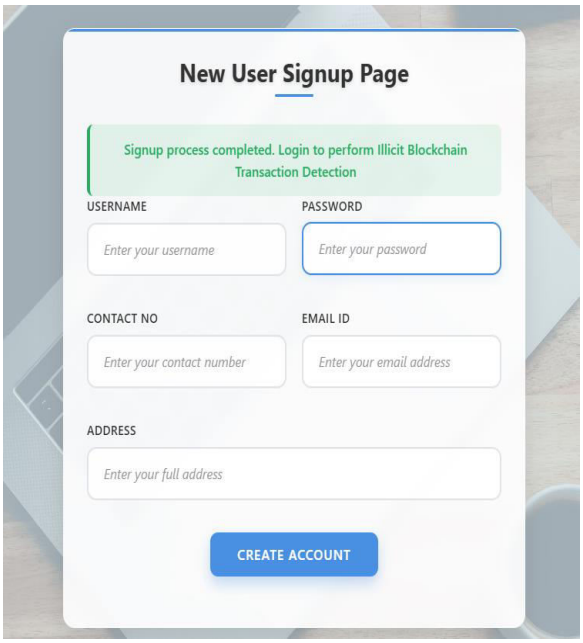
Extension2 LightGBM: LightGBM is a rapid, high-performance gradient boosting technique adept at efficiently managing large-scale datasets. It employs leaf-wise tree development, enhancing accuracy and decreasing training duration. It excelled at identifying illicit transfers owing to its capacity to discern intricate patterns and scale efficiently.

4. EXPERIMENTAL RESULTS

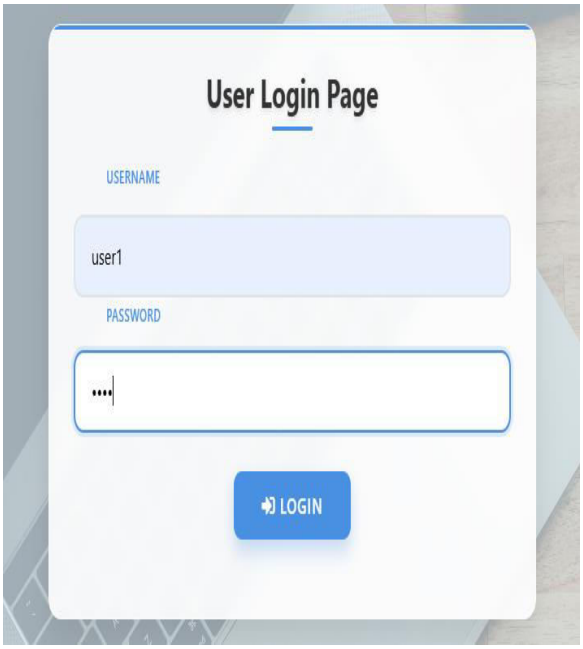
Consequently, the project interface is displayed as depicted in the preceding screen.



Input the necessary new user information for the signup procedure presented on the screen below.

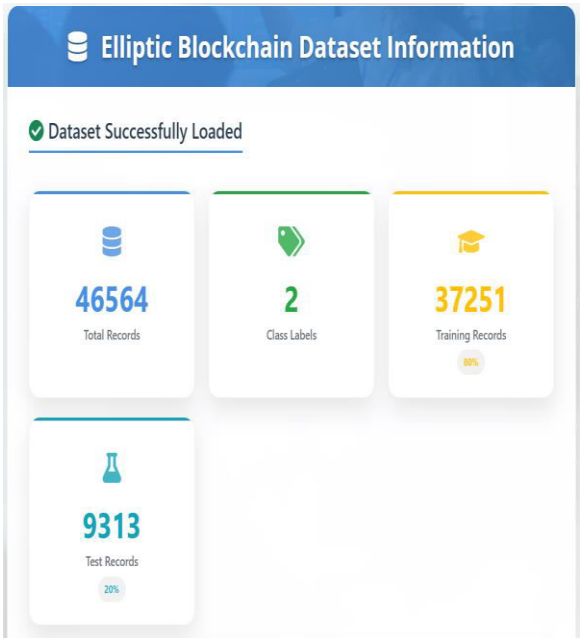


The user registration has been successfully finalized, and the information is securely documented on the blockchain.

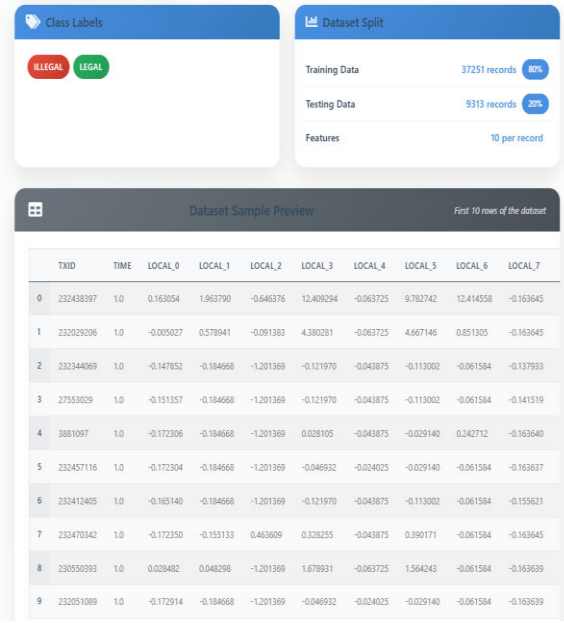


Users verify their identity by inputting their login and password into the specified sections.

After a successful login, the user is brought to the aforementioned home page to proceed with the process.



The Elliptic dataset has been loaded, consisting of two classes and a total of 46,564 records.

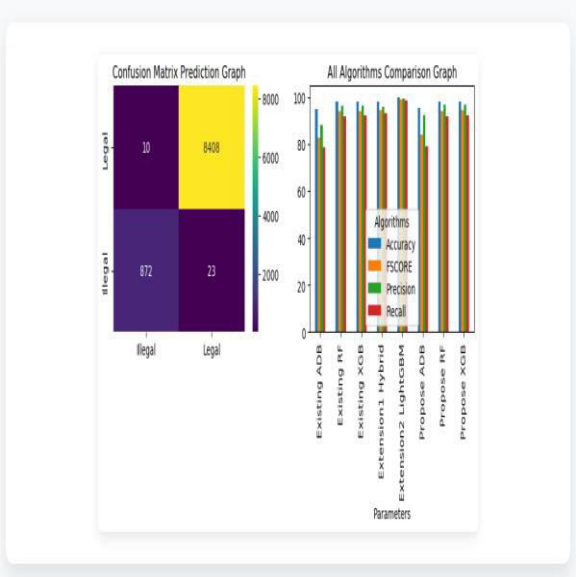


	TXID	TIME	LOCAL_0	LOCAL_1	LOCAL_2	LOCAL_3	LOCAL_4	LOCAL_5	LOCAL_6	LOCAL_7
0	232438397	1.0	0.163054	1.963790	-0.646376	12.409294	-0.063725	9.782742	12.414558	-0.163645
1	232029206	1.0	-0.005027	0.578941	-0.091383	4.380281	-0.063725	4.667146	0.851305	-0.163645
2	232344069	1.0	-0.147852	-0.184668	-1.201369	-0.121970	-0.043875	-0.113002	-0.061584	-0.137933
3	27553029	1.0	-0.151357	-0.184668	-1.201369	-0.121970	-0.043875	-0.113002	-0.061584	-0.141519
4	3881097	1.0	-0.172306	-0.184668	-1.201369	0.028105	-0.043875	-0.029140	0.242712	-0.163640
5	232457116	1.0	-0.172304	-0.184668	-1.201369	-0.046932	-0.024025	-0.029140	-0.061584	-0.163637
6	232412405	1.0	-0.165140	-0.184668	-1.201369	-0.121970	-0.043875	-0.113002	-0.061584	-0.155621
7	232470342	1.0	-0.172350	-0.155133	0.463609	0.328255	-0.043875	0.390171	-0.061584	-0.163645
8	230550393	1.0	0.028482	0.048298	-1.201369	1.678931	-0.063725	1.564243	-0.061584	-0.163639
9	232051089	1.0	-0.172914	-0.184668	-1.201369	-0.046932	-0.024025	-0.029140	-0.061584	-0.163639

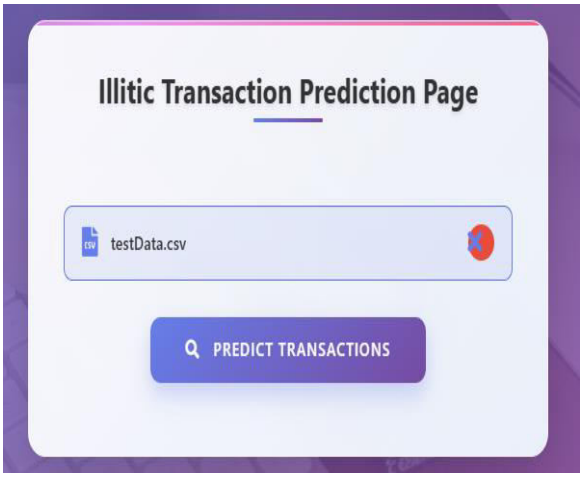
The dataset has class labels, train-test data splits, and associated feature values for each column.

ALGORITHM NAME	ACCURACY (%)	PRECISION (%)	RECALL (%)	F-SCORE (%)
Existing Random Forest	97.992	96.397	91.75	93.919
Existing XGBoost	98.046	96.148	92.329	94.134
Existing AdaBoost	94.717	88.31	78.755	82.668
Propose VTAC Random Forest	97.992	96.6	91.55	93.893
Propose VTAC XGBoost	98.099	96.538	92.259	94.268
Propose VTAC AdaBoost	95.329	92.31	78.993	84.104
Extension1 Hybrid Model	98.142	95.999	93.081	94.479
Extension2 LightGBM	99.646	99.297	98.656	98.974

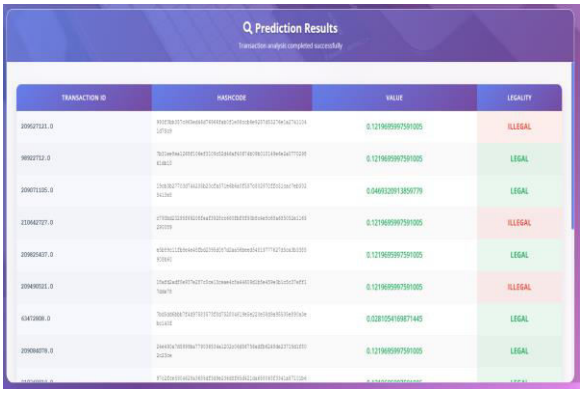
The performance metrics of all models are assessed, and the optimal model is identified below.



The confusion matrix of the extension model and the performance metrics comparison graph for all implemented models are presented below.



Select the test dataset from the test cases directory to evaluate the model's performance.



The model's predictions for the test dataset, along with the associated output classes, are displayed below.

5. CONCLUSION

This study presents an efficient and scalable method for identifying money laundering in blockchain environments by integrating sophisticated Machine Learning (ML) techniques with the Value-driven Transactional Tracking Analytics for Crypto-compliance (VTAC) framework. The method use the Elliptic Bitcoin dataset, a detailed, annotated graph dataset depicting actual cryptocurrency transactions, to conduct data normalization and preprocessing aimed at improving model performance. It executes and evaluates many classification algorithms, including independent models such as Random Forest, XGBoost, AdaBoost, LightGBM, and CatBoost, in addition to a proposed hybrid ensemble model that integrates XGBoost and Random Forest. The use of VTAC facilitates a more profound study of transactional behavior by integrating frequency, directionality, and flow patterns of transactions, hence enhancing the differentiation between legitimate and questionable activities. The de-anonymization of transaction identifiers exposes high-risk hash addresses, facilitating traceability and compliance enforcement. Experimental findings demonstrate that the hybrid model markedly enhances detection accuracy and robustness, surpassing conventional classifiers in the identification of unlawful transactions. A web-based user interface enhances real-time prediction and performance visualization. The technology offers a technically robust and practically feasible alternative to improve blockchain transparency and bolster anti-money laundering (AML) initiatives.

Future enhancements may improve this system by the integration of real-time blockchain monitoring, the expansion of support to many cryptocurrencies beyond Bitcoin, and the use of sophisticated deep learning methodologies such as Graph Neural Networks (GNNs) for superior pattern identification. Augmenting the VTAC framework with dynamic behavioral profiling and anomaly detection might enhance detection accuracy. Furthermore, cooperation with regulatory agencies and financial institutions may facilitate the creation of standardized compliance tools, while enhanced de-anonymization methods could assist law enforcement in more efficiently tracking criminal financial networks.

REFERENCES

[1] Ferretti, S., D'Angelo, G., & Ghini, V. (2025). Enhancing anti-money laundering frameworks: An application of graph neural networks in cryptocurrency transaction classification. IEEE Access.

[2] Dr, K, Pushpa Latha., Mr, M, N, Mallikarjuna Reddy., Dr, B, Rajalingam., Malleswari Akurati., Dr, G, Swapna., Bakkala Santha Kumar., (2026). Blockchain-Enabled Trade Finance Framework for Secure Drug Supply Chain Transactions., International Journal of Drug Delivery Technology, 16(3s), 884-889.

[3] Li, G., Mi, Y., Zhou, J., Zheng, X., & Wu, W. (2025). Group Based Detection of Cryptocurrency Laundering Using Multi-Persona Analysis. IEEE Transactions on Information Forensics and Security.

[4] Yu, Q., Xu, Z., & Ke, Z. (2024, November). Deep learning for cross-border transaction anomaly detection in anti-money laundering systems. In 2024 6th International Conference on Machine Learning,

Big Data and Business Intelligence (MLBDBI) (pp. 244-248). IEEE.

[5] Gudditti, V., & Krishna, P. V. (2021). Adaptive Light Weight Encryption Algorithm for Securing Multi-Cloud Storage. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(9), 545–554.

[6] Wu, J., Liu, J., Zhao, Y., & Zheng, Z. (2021). Analysis of cryptocurrency transactions from a network perspective: An overview. Journal of Network and Computer Applications, 190, Article 103139. https://www.sciencedirect.com/science/article/pii/S1084804521001557

[7] Marasi, S., & Ferretti, S. (2024, January). Anti-money laundering in cryptocurrencies through graph neural networks: A comparative study. In Proceedings of the IEEE 21st Consumer Communications & Networking Conference (CCNC) (pp. 272–277).

[8] Rathore, M. M., Chaurasia, S., & Shukla, D. (2022, December). Mixers detection in Bitcoin network: A step towards detecting money laundering in crypto-currencies. In Proceedings of the IEEE International Conference on Big Data (Big Data) (pp. 5775–5782).

[9] Scharfman, J. (2022). Anti-money laundering compliance for cryptocurrencies. In Cryptocurrency Compliance and Operations (pp. 91–114). Springer, Berlin, Germany.

[10] Chang, V., Baudier, P., Zhang, H., Xu, Q., Zhang, J., & Arami, M. (2020). How blockchain can impact financial services—The overview, challenges and recommendations from expert interviewees. Technological Forecasting and Social Change, 158, Article 120166.

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

[12] Singh, M., Tiwari, S. K., Swapna, G., Verma, K., Prasad, V., Patidar, V., Sharma, D. & Mewada, H. (2023). A Drug-Target Interaction Prediction Based on Supervised Probabilistic Classification. Journal of Computer Science, 19(10), 1203-1211. https://doi.org/10.3844/jcssp.2023.1203.1211

[13] Serena, L., Ferretti, S., & D'Angelo, G. (2022). Cryptocurrencies activity as a complex network: Analysis of transactions graphs. Peer-to-Peer Networking and Applications, 15(2), 839–853.

[14] Pareja, A., Domeniconi, G., Chen, J., Ma, T., Suzumura, T., Kanezashi, H., Kaler, T., Schardl, T. B., & Leiserson, C. E. (2020, April). EvolveGCN: Evolving graph convolutional networks for dynamic graphs. In Proceedings of the AAAI Conference on Artificial Intelligence, 34(4), 5363–5370.

[15] Lo, W. W., Kulatilleke, G. K., Sarhan, M., Layeghy, S., & Portmann, M. (2023). Inspection-L: Self-supervised GNN node embeddings for money laundering detection in Bitcoin. Applied Intelligence, 53(16), 19406–19417.

[16] Adewumi, A. O., & Akinyelu, A. A. (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. International Journal of System Assurance Engineering and Management, 8(2), 937–953.

[17] Popat, R. R., & Chaudhary, J. (2018, May). A survey on credit card fraud detection using machine learning. In Proceedings of the 2nd International

Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1120–1125).

[18] Sinayobye, J. O., Kiwanuka, F., & Kyanda, S. K. (2018, May). A state-of-the-art review of machine learning techniques for fraud detection research. In Proceedings of the IEEE/ACM Symposium on Software Engineering in Africa (SEiA) (pp. 11–19).

[19] G, Viswanath., N, Madhvik., K, Bhaskar., K, Supriya. (2024). Machine-Learning-Based Cloud Intrusion Detection. International Journal of Mechanical Engineering Research and Technology, 16(9), 38-52. [20] Sadgali, I., Sael, N., & Benabbou, F. (2018). Detection of credit card fraud: State of art. International Journal of Computer Science and Network Security, 18(11), 76–83.