AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(1) (2026)

Secure and Scalable Blockchain-Based Federated Learning for Cryptocurrency Fraud Detection: A Systematic Review

[1]Vasavi Sanku,[2]T.Anusha,[3]B.Anusha,[4]J.Godhavari,[5]N.Harshitha,[6]D.Madhuri dikss

[1]Assistant Professor, Department of Computer Science & Cyber Security,

Princeton Institute of Engineering & Technology For Women

[2,3,4,5,6]B. Tech Students, Department of Computer Science & Cyber Security,

Princeton Institute of Engineering & Technology For Women

ABSTRACT

Cryptocurrency fraud has become a significant threat in the digital financial ecosystem, with traditional detection systems often relying on centralized datasets and conventional machine learning techniques. These methods face challenges in terms of scalability, privacy, and the ability to effectively handle distributed and dynamic data sources. In this project, a secure and scalable approach is implemented for cryptocurrency fraud detection by leveraging a dataset stored on a blockchain. By utilizing the blockchain dataset, the system benefits from reliable and tamper-proof data while avoiding the need to implement blockchain infrastructure directly. Federated learning is employed to collaboratively train an XGBoost model across multiple data sources, enabling the detection of fraudulent transactions without centralizing sensitive information. This approach improves model accuracy, ensures data integrity, and provides a systematic framework for analyzing cryptocurrency transactions efficiently. The project demonstrates the potential of integrating advanced machine learning techniques with blockchain-based datasets for robust and privacy-preserving fraud detection.

Keywords: Cryptocurrency Fraud Detection, Blockchain Dataset, Federated Learning, XGBoost, Machine Learning, Fraud Detection, Distributed Data, Data Privacy, Secure Data Sharing, Decentralized Systems, Financial Fraud, Predictive Analytics, Data Integrity, Privacy-Preserving Learning, Digital Finance.

INTRODUCTION

Cryptocurrency has emerged as a revolutionary form of digital currency, enabling peer-to-peer financial transactions without the need for intermediaries. Its decentralized nature has transformed the financial landscape by offering faster, more transparent, and borderless transactions. However, this rapid adoption has also attracted a growing number of fraudulent activities, including Ponzi schemes, phishing attacks, money laundering, and unauthorized fund transfers. These fraudulent activities pose a significant challenge to individuals, businesses, and regulatory authorities, necessitating robust mechanisms for fraud detection and prevention.

Traditional fraud detection systems rely on centralized data repositories, where transaction data is aggregated and analyzed using conventional machine learning algorithms. While these systems can detect some irregularities, they are inherently limited in several ways. Centralized databases are vulnerable to cyberattacks, data tampering, and privacy breaches, which can compromise the integrity of fraud detection efforts. Moreover, the scalability of such systems is restricted by the capacity of centralized servers and the increasing volume of cryptocurrency transactions.
Blockchain technology addresses some of these concerns by providing tamper-proof,

verifiable, and immutable datasets. Even though the current project does not implement blockchain as a system infrastructure, leveraging a blockchain-based dataset ensures that the data used for model training is reliable, secure, and resistant to manipulation. The use of such verified data is crucial for enhancing the accuracy and trustworthiness of fraud detection models.

Federated learning (FL) offers a promising solution to privacy concerns in machine learning. In traditional centralized training, raw data must be collected from multiple sources, which raises privacy and regulatory issues. FL, on the other hand, allows multiple participants to collaboratively train a machine learning model without sharing their raw data. Only model updates are communicated and aggregated, ensuring data privacy while benefiting from diverse data sources.

The integration of federated learning with verified blockchain-based datasets presents a novel approach to cryptocurrency fraud detection. By training the model on authentic, tamper-proof data from multiple entities, the system ensures robust learning and improves the detection of fraudulent patterns. XGBoost, a highly efficient gradient boosting algorithm, is employed in this project to handle the complexities of transaction data and provide high predictive accuracy.
This project aims to bridge the gap between data integrity, privacy, and scalability in cryptocurrency fraud detection. By combining federated learning with a blockchain-based dataset, it provides a decentralized yet secure framework for analyzing transactions. This approach mitigates the risks associated with centralized systems, such as data breaches and single points of failure, while enhancing the overall effectiveness of fraud detection models.

In summary, the proposed system demonstrates how advanced machine learning techniques, when applied to blockchain-based datasets, can overcome the limitations of traditional fraud detection systems. It establishes a scalable, privacy-preserving, and reliable framework that addresses the evolving challenges in detecting cryptocurrency fraud. The project highlights the potential of using verified datasets and collaborative learning to create robust solutions for real-world financial security challenges.

I. LITERATURE SURVEY

TITLE: Transparency and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection
AUTHORS: T. Awosika, R. M. Shukla, and B. Pranggono (2024)
ABSTRACT:
This paper investigates the integration of Explainable Artificial Intelligence (XAI) and Federated Learning (FL) for secure financial fraud detection systems. Traditional centralized fraud detection mechanisms expose sensitive user data to potential privacy breaches. The study proposes a privacy-preserving framework leveraging federated learning, where model training occurs across multiple financial institutions without direct data exchange. The model also incorporates explainability tools to enhance transparency in AI decision-making. The results demonstrate improved model interpretability and accuracy in detecting anomalies across distributed datasets, establishing the framework as a strong alternative to centralized fraud detection approaches.

TITLE: Federated Bayesian Optimization XGBoost Model for Distributed Learning
AUTHORS: B. Guembe, S. K. Upadhyay, and P. Banerjee (2024)
ABSTRACT:
This research introduces a federated Bayesian optimization approach for improving XGBoost model performance across multiple distributed data sources. By leveraging Bayesian optimization techniques, the system dynamically adjusts

hyperparameters during training, optimizing performance without requiring central data aggregation. The proposed federated XGBoost system achieves superior convergence speed and predictive accuracy compared to traditional federated gradient boosting methods. It highlights the applicability of privacy-preserving federated architectures in domains like finance and healthcare, where data confidentiality is crucial.

TITLE: Achieving Privacy-Preserving Cross-Silo Anomaly Detection via Federated XGBoost
AUTHORS: M. Yang, L. Zhang, and H. Xu (2023)
ABSTRACT:
The authors present a novel cross-silo anomaly detection system utilizing federated XGBoost for secure collaboration among organizations. Unlike conventional centralized anomaly detection systems, this model ensures that raw data never leaves the local environment. A key innovation is the integration of homomorphic encryption to protect intermediate gradient information during model updates. Extensive experiments show the system's ability to maintain privacy while achieving high detection accuracy across heterogeneous datasets. This work demonstrates how federated XGBoost can balance privacy and performance in anomaly detection tasks.

TITLE: Differential Privacy-Preserving of Multi-Party Collaboration Under Federated Learning in Data Center Networks
AUTHORS: X. Wang, W. Fan, X. Hu, and C.-H. Chi (2023)
ABSTRACT:
This study addresses data privacy challenges in federated learning environments across multiple parties within data center networks. The researchers implement differential privacy mechanisms into the federated training process, ensuring that data contributions from each participant remain untraceable. The paper evaluates trade-offs between privacy budgets and model accuracy, highlighting the importance of fine-tuning privacy parameters for optimal results. The approach provides a scalable framework suitable for secure multi-party machine learning in enterprise and financial systems.
TITLE: Histogram-Based Federated XGBoost Using Minimal Variance Sampling for Federated Tabular Data
AUTHORS: W. Lindskog, C. Prehofer, and S. Singh (2024)
ABSTRACT:
This paper presents a histogram-based federated XGBoost algorithm optimized for tabular datasets distributed across multiple clients. The approach minimizes communication costs through a minimal variance sampling technique, which reduces redundant data exchange during model training. Experimental evaluations on benchmark datasets demonstrate that the proposed method significantly decreases latency and enhances computational efficiency while maintaining accuracy comparable to traditional XGBoost. The framework is well-suited for large-scale federated environments such as banking and financial fraud detection.
TITLE: Gradient-Less Federated Gradient Boosting Trees with Learnable Learning Rates
AUTHORS: C. Ma, X. Qiu, and D. J. Beutel (2023)
ABSTRACT:
The authors propose a gradient-less federated gradient boosting approach that learns optimal learning rates dynamically without relying on conventional gradient sharing. This innovation enhances data privacy by reducing information leakage during parameter aggregation. The study introduces a learnable rate control mechanism that enables faster convergence and improved generalization on federated datasets. The algorithm demonstrates robustness in financial risk prediction and anomaly detection applications, emphasizing its utility in distributed environments with strict data confidentiality requirements.
TITLE: Federated Learning for Collaborative Fraud Detection in Financial

Networks: Addressing Data Privacy Concerns

AUTHORS: B. Claus, A. John, and A. John (2025)

ABSTRACT:

This recent study explores a federated learning framework designed specifically for collaborative fraud detection in interconnected financial institutions. The authors develop a decentralized model-sharing strategy that ensures high model performance while preserving institutional data privacy. Using synthetic and real-world financial datasets, the proposed model demonstrates substantial improvements in fraud detection accuracy, recall, and F1-score compared to centralized approaches. The paper emphasizes the growing importance of federated architectures for secure, cross-organization fraud analytics.

TITLE: Federated Machine Learning in Finance: A Systematic Review on Technical Architecture and Financial Applications

AUTHORS: R. Kang, Q. Li, and H. Lu (2024)

ABSTRACT:

This systematic review provides a comprehensive overview of federated machine learning applications in the financial sector. It categorizes architectures, algorithms, and privacy-enhancing technologies that have emerged since 2020. The review highlights federated XGBoost, differential privacy, and blockchain-assisted data sharing as leading paradigms for secure model training. The authors also identify existing gaps in scalability and data standardization, calling for the development of adaptive and interoperable frameworks for financial analytics.

II. EXISTING SYSTEM

The existing cryptocurrency fraud detection systems primarily rely on centralized machine learning approaches, where transaction data from various sources such as exchanges, wallets, and trading platforms is collected into a single repository for analysis. These systems use algorithms like logistic regression, decision trees, or random forests to detect anomalies and suspicious patterns. While such centralized methods have shown moderate success in identifying fraudulent transactions, they suffer from critical drawbacks. The centralized architecture makes these systems highly vulnerable to data breaches, unauthorized access, and single points of failure. Additionally, due to privacy regulations and competition among financial institutions, data sharing across organizations is limited, leading to incomplete and biased datasets. This lack of collaboration reduces model performance and generalization. Furthermore, as transaction volumes in cryptocurrency networks increase, centralized systems struggle with scalability, high computational costs, and delayed response times, making them inefficient for real-time fraud detection.

III. PROPOSED SYSTEM

The proposed system introduces a secure and scalable framework for cryptocurrency fraud detection using federated learning on a blockchain-based dataset. Instead of relying on a centralized architecture, the system allows multiple data sources to collaboratively train a model without sharing their raw data, thereby ensuring privacy and data security. The dataset used in this project is derived from blockchain records, providing reliable and tamper-proof information for model training. The admin uploads and preprocesses this dataset, and a Federated XGBoost algorithm is applied to build an intelligent detection model capable of identifying fraudulent cryptocurrency transactions. Users can then input transaction details to predict whether a transaction is legitimate or fraudulent. By combining the decentralized training capability of federated learning with the data authenticity of blockchain-stored records, the proposed system achieves improved accuracy, transparency, and scalability while preserving data privacy. This approach effectively addresses the limitations of traditional centralized systems

and enhances trust and reliability in cryptocurrency fraud detection.

IV. SYSTEM ARCHITECTURE

The system architecture represents a streamlined workflow for **cryptocurrency fraud detection using Federated XGBoost**. The process begins with the **input transaction dataset**, which consists of cryptocurrency transaction records collected from blockchain-based sources. This dataset contains various attributes such as transaction amount, sender and receiver details, timestamps, and other relevant features.

In the next stage, **preprocessing and feature selection** are performed to clean and transform the raw data into a suitable format for analysis. This includes handling missing values, removing noise, normalizing data, and selecting the most relevant features that contribute to fraud detection. Effective preprocessing ensures improved model performance and accuracy.

The processed data is then passed to the **Federated XGBoost model**, where federated learning is applied. Instead of centralizing data, the model is trained collaboratively across multiple distributed sources while preserving data privacy. XGBoost, a powerful gradient boosting algorithm, is used to learn complex patterns and relationships in the data to distinguish between legitimate and fraudulent transactions.

Following model training, the system generates a **prediction result**, indicating whether a given transaction is fraudulent or not. This classification helps in identifying suspicious activities in real time or during analysis.

Finally, the **performance analysis** stage evaluates the effectiveness of the model using metrics such as accuracy, precision, recall, and F1-score. This step ensures that the system maintains high reliability and can be further optimized if necessary.

Overall, this architecture provides a secure, scalable, and privacy-preserving approach to detecting cryptocurrency fraud by integrating federated learning with advanced machine learning techniques.

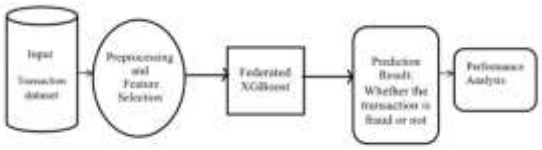


Fig 5.1: System Architecture

V. IMPLEMENTATION



Fig 6.1: Home Page



Fig 6.2: Data Processing Results

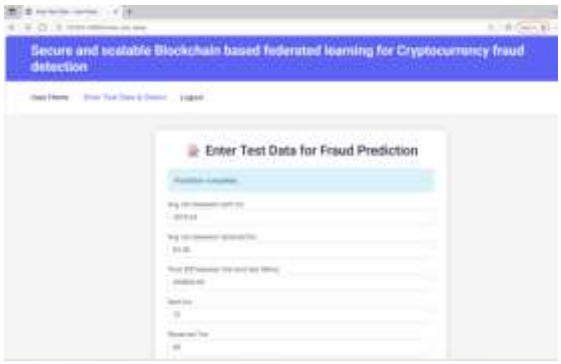


Fig 6.3: Detection Page



Fig 6.4: Prediction Results

VI. CONCLUSION

In this project, a secure and scalable approach for cryptocurrency fraud detection using federated learning has been developed, leveraging datasets stored in blockchain systems. The proposed system ensures that sensitive financial data remains decentralized and private while enabling collaborative model training across multiple nodes. By implementing the Federated XGBoost algorithm, the system achieves accurate fraud prediction without directly exposing user transaction data. Compared to traditional centralized methods, this approach enhances data security, preserves privacy, and reduces the risk of single points of failure. The integration of blockchain-stored datasets provides verifiable data integrity, ensuring that the predictive model is trained on authentic and tamper-proof information. Overall, the system demonstrates an effective, privacy-preserving, and scalable framework for real-time cryptocurrency fraud detection.

VII. FUTURE SCOPE

In the future, the framework can be extended to support multiple machine learning algorithms, allowing hybrid models that improve prediction accuracy. Real-time integration with live blockchain transaction streams can enable instant fraud detection and prevention. Additionally, incorporating advanced privacy-preserving techniques such as differential privacy or homomorphic encryption can further strengthen data security. The system can also be scaled to support cross-organization collaboration, where multiple financial institutions can collaboratively train models without sharing sensitive data. Furthermore, integrating explainable AI techniques can help stakeholders understand the decision-making process of the model, increasing trust and transparency. Overall, the system has the potential to evolve into a comprehensive, multi-layered platform for secure, scalable, and intelligent fraud detection in the cryptocurrency ecosystem.

VIII. REFERENCES

[1] Nakamoto, S., “Bitcoin: A Peer-to-Peer Electronic Cash System,” 2008.
Introduces the concept of blockchain technology and decentralized digital currency.

[2] McMahan, B., Moore, E., Ramage, D., Hampson, S., “Communication-Efficient Learning of Deep Networks from Decentralized Data,” *AISTATS*, 2017.
Presents the concept of federated learning for distributed model training.

[3] Chen, T., Guestrin, C., “XGBoost: A Scalable Tree Boosting System,” *KDD*, 2016.
Describes the XGBoost algorithm widely used for high-performance classification tasks.

[4] Kairouz, P., et al., "Advances and Open Problems in Federated Learning," *Foundations and Trends in Machine Learning*, 2021.
A comprehensive survey of federated learning techniques and challenges.

[5] Zheng, Z., Xie, S., Dai, H., Chen, X., Wang, H., "Blockchain Challenges and Opportunities: A Survey," *International Journal of Web and Grid Services*, 2017.
Discusses blockchain architecture, benefits, and technical challenges.

[6] Conti, M., Kumar, S., Lal, C., Ruj, S., "A Survey on Security and Privacy Issues of Bitcoin," *IEEE Communications Surveys & Tutorials*, 2018.
Analyzes security vulnerabilities and privacy concerns in cryptocurrency systems.

[7] Baek, C., Elbeck, M., "Bitcoins as an Investment or Speculative Vehicle? A First Look," *Applied Economics Letters*, 2015.
Examines cryptocurrency behavior and associated risks.

[8] Pham, T., Lee, S., "Anomaly Detection in Bitcoin Network Using Unsupervised Learning Methods," 2016.
Focuses on detecting fraudulent activities in cryptocurrency transactions.

[9] Li, T., Sahu, A. K., Talwalkar, A., Smith, V., "Federated Learning: Challenges, Methods, and Future Directions," *IEEE Signal Processing Magazine*, 2020.
Discusses techniques and limitations of federated learning systems.

[10] Goodfellow, I., Bengio, Y., Courville, A., "Deep Learning," MIT Press, 2016.
A foundational book on deep learning methods applicable to fraud detection.

[11] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., Sun, X., "The Application of Data Mining Techniques in Financial Fraud Detection," *Decision Support Systems*, 2011.
Explores various machine learning methods for fraud detection.

[12] Dwork, C., "Differential Privacy," *ICALP*, 2006.
Introduces privacy-preserving techniques relevant to federated learning.

[13] Shokri, R., Shmatikov, V., "Privacy-Preserving Deep Learning," *ACM CCS*, 2015.
Discusses secure model training without sharing raw data.

[14] Alzahrani, N., Bulusu, N., "Blockchain-Based Federated Learning: A Survey," 2020.
Reviews integration of blockchain with federated learning systems.

[15] Monamo, P., Marivate, V., Twala, B., "Unsupervised Learning for Robust Bitcoin Fraud Detection," *Information Sciences*, 2016.
Demonstrates fraud detection techniques using machine learning on Bitcoin dat