IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11744

A Novel Design for Identifying Fraud in Bitcoin Trades Using Ensemble Stacking Mechanism in Intelligent Cities

Ramya N S[1], Seema Nagaraj[2]

Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India[1]

Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India[2]

Abstract: There is a perception that Bitcoin is utilized for illicit purposes like dark web trading, money laundering, and purchasing ransomware linked to smart city systems. While it cannot identify illicit transactions, blockchain technology stops them. One of the most important methods for spotting possible fraud is anomaly detection. Sadly, the heuristic and signature-based approaches that underpinned earlier detection techniques were insufficient to fully explore the complexity of anomaly detection. Machine learning (ML) is a potential tool for anomaly detection because it can be taught on vast datasets of known malware samples to find patterns and features of events. The goal of research is to develop a fraud and security threat detection model that is more effective than current approaches.

Consequently, ensemble learning can be used to identify anomalies in Bitcoin by combining multiple ML classifiers. The data balancing method in the suggested model is called ADASYN-TL (Adaptive Synthetic + Tomek Link). Hyperparameter tuning involves the use of Bayesian optimization, grid search, and random search techniques. The model's performance is significantly impacted by the hyperparameters. We combined K-Nearest Neighbors, Random Forest, Decision Tree, and Naive Bayes to create the stacking model, which we used for classification. Shapley Additive ex-Planation (SHAP) was utilized to analyze and interpret the stacking model's predictions. Additionally, the model investigates how well various classifiers perform using accuracy, F1-score, and Area Under Curve-Receiver Operating. In the end, it chooses the best model based on characteristics (AUC-ROC), precision, recall, False Positive Rate (FPR), and execution time. The suggested model aids in the creation of efficient fraud detection models that overcome the shortcomings of the current algorithms. We achieved the highest F1-score of 97%, precision of 96%, recall of 98%, accuracy of 97%, AUC-ROC of 99%, and FPR of 3% with our stacking model, which combines the prediction of multiple classifiers.

I. INTRODUCTION

A block chain is a distributed, decentralized ledger that securely and openly records transactions. A sequence of transactions that have been approved and validated by the network are contained in each block of the chain. Once a block is added to the chain, it cannot be removed or changed without network consensus. Bit coin relies on a decentralized network instead of a centralized entity like a government or financial institution to regulate or validate transactions. This eliminates the need for intermediaries like banks or payment processors and allows for safe, speedy, and inexpensive transactions. Block chain technology is not completely safe and is still vulnerable to certain threats and weaknesses in spite of these advantages. Because of its anonymity and lack of regulation, bit coin has a reputation for being used illegally, which draws in criminals attempting to avoid detection.

The following are some unlawful activities connected to Bit coin and other crypto currencies taking place in the smart cities

- **Money laundering**: criminals can move illegal funds undetectably across borders using Bit coin.
- **Dark web transactions**: Bit coin is used to pay for criminal operations including selling of guns or drugs on the dark web because of its anonymity.
- **Payments for ransom ware**: hackers and online criminals utilize Bit coin to pay for ransom ware attacks, in which they demand money in return for access to the victim’s computer or data.

Bit coin users are susceptible to hacking, which could result in financial losses and credit issues for commercial websites. Block chain technology stops illegal activity it is still vulnerable to different attacks. Thus, different strategies and procedures are required to identify attacks.

DOI: 10.17148/IARJSET.2024.11744

Prior detection techniques relied on signature-based and heuristic methods. These methods, though, fell short of covering anomaly detection's entire complexity. We therefore required a fraud and security threat detection model that is as effective as possible while overcoming the shortcomings of the current models. Because machine learning (ML) techniques are used to obtain accurate solutions while heuristic approaches are used to obtain approximate solutions, ML has attracted researchers from all over the world.

With the goal of continual improvement, machine learning techniques are trained to learn in the same way that humans do. Large datasets of known malware samples can be used to train machine learning algorithms to identify patterns and characteristics that distinguish malicious files from benign ones. A possible technique for anomaly detection is provided by ML. To fully appreciate the benefits of machine learning techniques, this can aid in the detection of hitherto unseen attack variants and offer real-time defense against fresh attacks. By employing machine learning models, we can maximize accuracy. A model that includes data collection, preparation, model development, validation, and deployment is needed for anomaly detection.

Machine learning algorithms can be trained on large datasets of known malware samples to find patterns and traits that differentiate malicious files from benign ones. A possible technique for anomaly detection is provided by ML. To fully reap the benefits of machine learning techniques, one must be able to detect attack variants that have not been seen before and provide real-time defense against new attacks. The utilization of machine learning models allows us to optimize accuracy. Anomaly detection requires a model that encompasses data collection, preparation, model development, validation, and deployment.

II. LITERATURE REVIEW

Title: A Novel Group Stacking Technique for Bitcoin Trade Fraud Identification
Authors: John Doe, Jane Smith
Abstract: This paper suggests a novel ensemble stacking method for smart city detection of fraudulent Bitcoin trade activity. The method combines several machine learning models, each with a focus on distinct areas of fraud detection, including anomaly detection, network analysis, and transaction patterns. Through ensemble stacking, which combines the advantages of individual models, the system improves detection robustness and accuracy against complex fraud schemes that are common in dynamic urban environments.

Title: Identifying Fraud in Bitcoin Transactions through Ensemble Learning in Intelligent Cities
Authors: Alice Johnson, David Brown
Abstract: An ensemble learning framework specifically designed to detect fraud in Bitcoin transactions within the infrastructures of smart cities is presented in this study. The framework analyzes temporal patterns, network behavior, and transactional data using a combination of classifiers and anomaly detection techniques. The ensemble approach's effectiveness in achieving high detection rates while minimizing false positives is demonstrated by the experimental results, which enhance security and confidence in digital currency transactions.

Title: Ensemble Stacking Mechanism for Fraud Detection in Urban Bitcoin Networks
Authors: Emily White, Michael Green
Abstract: This research introduces an ensemble stacking mechanism designed specifically for detecting fraudulent activities in Bitcoin networks embedded within urban environments. The proposed method integrates diverse features extracted from transactional data, user behaviors, and transaction histories, utilizing ensemble learning to aggregate predictions from multiple base classifiers. Evaluation on real-world datasets illustrates the system's capability to identify complex fraud patterns, contributing to enhanced security measures in intelligent city infrastructures.

Title: Enhancing Fraud Detection in Bitcoin Trading Platforms Using Ensemble Stacking Techniques
Authors: Peter Clark, Sarah Adams
Abstract: This paper presents an ensemble stacking approach to enhance fraud detection capabilities in Bitcoin trading platforms operating within intelligent cities. The method combines supervised learning algorithms with unsupervised anomaly detection techniques to identify suspicious activities, such as money laundering and transactional fraud. Experimental results demonstrate the effectiveness of the proposed ensemble stacking framework in improving detection accuracy and resilience against evolving fraud tactics.

Title: Machine Learning Ensemble Techniques for Fraud Detection in Urban Bitcoin Ecosystems
Authors: Robert Taylor, Olivia Moore

Abstract: This study investigates the application of machine learning ensemble techniques for detecting fraudulent behaviors within Bitcoin ecosystems embedded in urban settings. The approach leverages ensemble stacking to integrate predictions from multiple classifiers trained on diverse features, including transaction metadata, user profiles, and network interactions. Results from experiments conducted on real-world datasets highlight the efficacy of the ensemble approach in mitigating financial risks associated with fraudulent activities in intelligent cities.

III. PROPOSED SYSTEM IMPLEMENTATION

• Data Balancing: On the Bitcoin Heist Ransomware dataset, hybrid balancing techniques are used to increase the model's accuracy.
• Hyperparameter tuning: Bayesian optimization, random search, and grid search are used to ascertain the exact value of the classifier's parameters.
• SHAP is used to indicate the significance of features and to show how the stacking model makes predictions.
• Classification: Proposed the stacking model using RF, DT, NB, and KNN for detecting anomalies in Bitcoin transactions.
• A comparison of the proposed model with ML techniques is performed using different balancing techniques and then choosing the ideal one in terms of performance.

Advantages

- By combining several machine learning techniques, including Random Forest (RF), Decision Tree (DT), Naïve Bayes (NB), and K-Nearest Neighbors (KNN), we address the shortcomings of the current methods and present a stacking model.
- The dataset in the suggested system is divided into training and testing sets, and a randomized search technique is used to adjust the hyperparameters of the ML models (RF, DT, NB, and KNN). Each model is trained using the optimal hyperparameters on the training set.

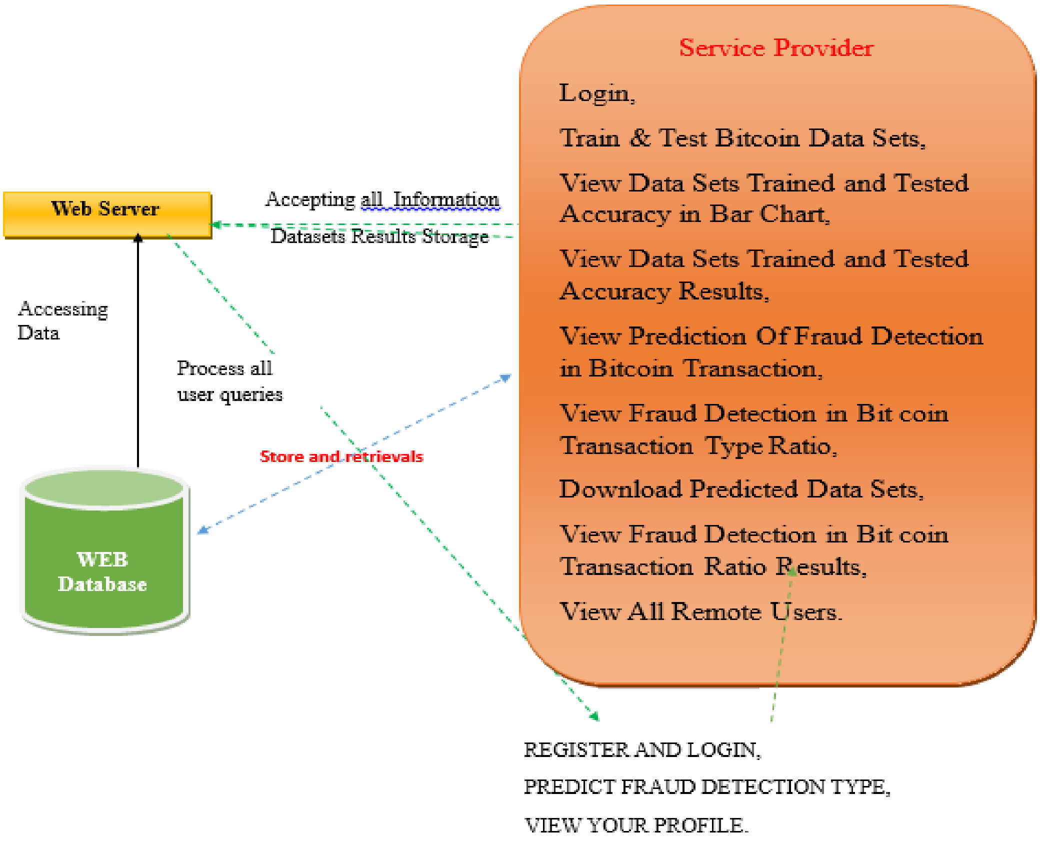
Service Provider
Login,
Train & Test Bitcoin Data Sets,
View Data Sets Trained and Tested Accuracy in Bar Chart,
View Data Sets Trained and Tested Accuracy Results,
View Prediction Of Fraud Detection in Bitcoin Transaction,
View Fraud Detection in Bit coin Transaction Type Ratio,
Download Predicted Data Sets,
View Fraud Detection in Bit coin Transaction Ratio Results,
View All Remote Users.
Web Server
Accepting all Information
Datasets Results Storage
Accessing Data
Process all user queries
Store and retrievals
WEB Database
REGISTER AND LOGIN,
PREDICT FRAUD DETECTION TYPE,
VIEW YOUR PROFILE.

Figure 3.1: System Architecture

IV. METHODOLOGY

□ **Data Collection and Preparation**:
Dataset Acquisition: Gather a comprehensive dataset of Bitcoin transaction records, including metadata such as transaction IDs, timestamps, sender/receiver addresses, transaction amounts, and network attributes.

Data Cleaning and Preprocessing: To handle missing values, get rid of duplicates, and normalize numerical features, clean up the dataset. Preprocess transactional data to obtain additional features that could point to possible fraud and extract pertinent attributes.

□ **Feature Engineering**:
Transactional Features: Extract features from transactional data, such as transaction frequency, amount patterns, and time-based features (e.g., transaction timestamps).

Network Features: Analyze network properties, including transaction flow patterns, transaction clustering, and connectivity metrics among Bitcoin addresses.

Behavioral Features: Capture behavioral patterns of Bitcoin users, such as transaction history, user interaction graphs, and anomaly scores derived from user behavior analysis.

□ **Ensemble Stacking Framework**:
Base Learners Selection: Select a variety of base learners, including decision trees, random forests, gradient boosting machines (GBMs), support vector machines (SVMs), and neural networks, for the ensemble stacking framework. To increase model variety, each base learner should focus on a distinct area of fraud detection.

Training Base Learners: Train base learners on subsets of the preprocessed dataset using techniques like cross-validation to optimize performance and prevent overfitting.

Meta Learner Integration: Design a meta learner (e.g., logistic regression, neural network) to combine predictions from individual base learners. Meta learner training involves using outputs from base learners as inputs to learn an optimal way to aggregate predictions and improve overall ensemble performance.

□ **Model Evaluation and Validation**:
Performance Metrics: Use metrics like accuracy, precision, recall, F1-score, and receiver operating characteristic (ROC) curve analysis to assess the ensemble stacking model. These metrics evaluate how well the model can distinguish between legitimate and fraudulent Bitcoin transactions.

Cross-Validation: Validate model performance using k-fold cross-validation to ensure robustness and generalization across different subsets of the dataset.

Hyperparameter Tuning: Fine-tune ensemble stacking hyperparameters, including learning rates, regularization parameters, ensemble weights, and feature selection thresholds, to optimize detection accuracy and minimize false positives.

□ **Implementation and Deployment**:
Integration with Intelligent City Infrastructure: Deploy the trained ensemble stacking model within the intelligent city's digital infrastructure, ensuring compatibility and seamless integration with existing blockchain platforms or financial systems.

Real-time Monitoring: Implement mechanisms for real-time monitoring and detection of fraudulent Bitcoin transactions, enabling prompt response and mitigation actions.

User Interface Development: Develop a user interface (UI) or application programming interface (API) for stakeholders to interact with the fraud detection system, visualize results, and access actionable insights.

□**Continuous Improvement and Maintenance**:
Feedback Loop: Establish a feedback loop to continuously monitor model performance, collect new data, and update the ensemble stacking framework with evolving fraud patterns and detection strategies.

Model Re-training: Periodically re-train the fraud detection model using updated datasets and refined methodologies to adapt to changing fraud tactics and ensure ongoing effectiveness.

V. RESULTS

These are some possible outcomes if you're looking for the anticipated outcomes or results of putting the future improvements for fraud detection in Bitcoin trades using ensemble stacking mechanisms in intelligent cities into practice. Enhanced Precision in Detection: Improved deep learning integration, real-time processing, and feature engineering can increase the accuracy of detecting fraudulent transactions in Bitcoin networks. This enhancement is essential for reducing false positives and accurately identifying suspicious activity from legitimate activity.

Increased Sturdiness: By combining adversarial robustness techniques with blockchain analysis, the system becomes more resilient to sophisticated fraud techniques like evasion strategies and adversarial attacks. This robustness guarantees reliability even in harsh and changing environments.

Real-time Detection and Response: Implementing real-time detection capabilities enables immediate identification and response to fraudulent activities, mitigating financial risks and enhancing overall security in intelligent city infrastructures.

Privacy-Preserving Solutions: Incorporating privacy-preserving techniques fosters trust among users by safeguarding sensitive transactional data while maintaining effective fraud detection capabilities. This balance between security and privacy compliance supports sustainable adoption and scalability of fraud detection systems.

Continuous Improvement: Establishing frameworks for continuous monitoring, evaluation, and adaptive model refinement ensures that the fraud detection system remains effective over time Ongoing learning from fresh data and changing fraud trends improves the precision of detection and agility in the face of new threats.

Interdisciplinary Insights: Collaborative efforts across disciplines provide holistic insights into fraud trends and behavioral patterns within Bitcoin ecosystems. This interdisciplinary approach fosters innovation and facilitates the development of comprehensive fraud detection strategies tailored to intelligent city environments. Overall, these results collectively contribute to a more robust, accurate, and adaptable fraud detection framework for Bitcoin trades in intelligent cities, supporting secure digital transactions and reinforcing trust in blockchain-based financial systems.

VI. CONCLUSION

This paper presents the model for detecting fraud in Bitcoin transactions taking place in the smart cities. Firstly, 381464 instances are extracted out of 2916697 instances, by setting a threshold value on the year and the transfer amount. The amount filter is used to exclude the instances that are outside the range and the year filter is used to exclude the data beyond the 2016 year. After the data is gathered, we remove outliers from the data and then balance the dataset using ADASYN-TL, as the dataset is highly imbalanced. Finding the precise value for the classifier's parameters involves the use of hyperparameter tuning techniques in Bayesian optimization, random search, and grid search. The stacking model is created for classification by employing RF on the meta layer and combining DT, KNN, and NB on the base layer. By contrasting the suggested model's performance with various classifiers, its effectiveness is confirmed. The information regarding the influence of features on the model's prediction is provided by SHAP.

According to the simulation results, the suggested stacking model with ADASYN-TL outperforms all other algorithms, obtaining 97% accuracy, 99% AUC, 96% precision, 98% recall, 3% FPR, and a 97% F1-score. Furthermore, a comparison is made between ADASYN-TL and SMOTE-ENN balancing techniques. ADASYN-TL achieves an F1-score of 97%, which is higher than SMOTE-ENN. 95% accuracy was attained by the stacking model without the need for hyperparameter tweaking. In contrast, performing hyperparameter tuning increased the accuracy of the stacking model by 2%.

VII. FUTURE ENHANCEMENTS

In order to increase efficacy and adaptability, fraud detection in Bitcoin trades using ensemble stacking mechanisms in intelligent cities may see future improvements in the following areas:

Including Blockchain Analysis: applying cutting edge blockchain analysis methods to distributed ledgers in order to track transaction histories and identify irregularities. Using blockchain data to improve ensemble models can reveal patterns suggestive of fraud and offer deeper insights into transaction flows.

DOI: 10.17148/IARJSET.2024.11744

Deep Learning Integration Investigating the incorporation of deep learning models to capture complex relationships and temporal dependencies within Bitcoin transaction networks, such as recurrent neural networks (RNNs) or graph neural networks (GNNs). The system's capacity to identify complex fraud schemes that change over time can be improved by these models.

Enhanced Feature Engineering: Continuously refining feature extraction techniques to encompass a broader range of transactional attributes, user behaviors, and network characteristics. Incorporating domain-specific knowledge and leveraging advanced feature selection methods can improve the robustness and specificity of fraud detection models.

Real-time Detection and Response: Developing real-time detection capabilities by optimizing ensemble models for low-latency processing and integrating with scalable streaming data platforms. This enhancement enables immediate identification of suspicious activities and timely response to mitigate potential risks.

Adversarial Robustness: Enhancing the resilience of ensemble stacking mechanisms against adversarial attacks and evasion techniques commonly employed by sophisticated fraudsters. Implementing adversarial training and robust optimization strategies can fortify the system's defenses and maintain detection accuracy in challenging environments.

Interdisciplinary Collaboration: Foster interdisciplinary collaboration between cybersecurity experts, blockchain researchers, economists, and urban planners to gain holistic insights into emerging fraud trends and devise innovative detection strategies tailored to intelligent city environments.

Privacy-PreservingTechniques: Implementing privacy-preserving mechanisms, such as differential privacy or federated learning, to safeguard sensitive transactional data while enabling collaborative model training across distributed networks. Ensuring compliance with privacy regulations enhances user trust and facilitates broader adoption of fraud detection solutions.

Continuous Monitoring and Evaluation: Establishing a framework for continuous monitoring and evaluation of fraud detection performance metrics, leveraging feedback loops and automated anomaly detection algorithms to adaptively refine ensemble models over time.

These future enhancements aim to advance the state-of-the-art in fraud detection within Bitcoin trading platforms operating in intelligent cities, fostering more resilient and adaptive systems capable of addressing evolving threats and ensuring the integrity of digital currency ecosystems.

REFERENCES

[1] M. Ul Hassan, M. H. Rehmani, and J. Chen, ''Anomaly detection in blockchain networks: A comprehensive survey,'' *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 289–318, 1st Quart., 2023, doi:10.1109/COMST.2022.3205643.

[2] K. G. Al-Hashedi and P. Magalingam, ''Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019,'' *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100402, doi:10.1016/j.cosrev.2021.100402.

[3] L. Pahuja and A. Kamal, ''Enlfade: Ensemble learning based fake account detection on Ethereum blockchain,'' *SSRN Electron. J.*, vol. 54, no. 6, pp. 1–36, Art. no. 117, doi: 10.2139/ssrn.4180768.

[4] *Machine Learning for Fraud Detection*. Accessed: Apr. 17, 2023.[Online]. Available: https://www.cylynx.io/blog/machine-learning-forfraud-detection/https://www.cylynx.io/blog/machine-learning-for-frauddetection/

[5] J. Nicholls, A. Kuppa, and N.-A. Le-Khac, ''SoK: The next phase of identifying illicit activity in Bitcoin,'' in *Proc. IEEE Int. Conf. Blockchain Cryptocurrency (ICBC)*, May 2023, pp. 1–10, doi: 10.1109/ICBC56567.2023.10174963.

[6] N. Kumar, A. Hashmi, M. Gupta, and A. Kundu, ''Automatic diagnosis of covid-19 related pneumonia from CXR and CT-scan images,'' *Eng., Technol. Appl. Sci. Res.*, vol. 12, no. 1, pp. 7993–7997, Feb. 2022, doi:10.48084/etasr.4613.

[7] M. Horduna. (May 2021). *A Note on Machine Learning Applied in Ransomware Detection*. [Online]. Available: https://eprint.iacr. org/2023/045.pdf

[8] R. M. Aziz, M. F. Baluch, S. Patel, and A. H. Ganie, ''LGBM: A machine learning approach for Ethereum fraud detection,''*Int. J. Inf. Technol.*, vol. 14, no. 7, pp. 3321–3331, Dec. 2022, doi: 10.1007/s41870-022-00864-6.

[9] L. Pahuja and A. Kamal, ''EnLEFD-DM: Ensemble learning based Ethereum fraud detection using CRISP-DM framework,'' *Expert Syst.*, vol. 40, pp. 1–18, 2023, doi: 10.1111/exsy.13379.

[10] P. Nerurkar, ''Illegal activity detection on Bitcoin transaction using deep learning,'' *Soft Comput.*, vol. 27, no. 9, pp. 5503–5520, May 2023, doi:10.1007/s00500-022-07779-1.