Detection of Credit Card Fraud Utilizing Blockchain Technology and Machine Learning

Atef Zaki Ghalwash [1], Sameh Gamal Khalil Sayed Ahmed [2], Amr Galal Mohamed[3]

[1] *Professor, Faculty of Computers & Artificial Intelligence, Helwan University, Cairo, Egypt.*

[2] *[1]Department of Business Information System, Faculty of Commerce and Business Administration, Helwan University ,Cairo, Egypt*

[3] *Faculty of Computers & Artificial Intelligence, Cairo University, Cairo, Egypt*

Abstract:

Despite the advances made in technology, the rate of cybercrime is rising. It is difficult to get reliable data on bank transactions since there is a lack of synergy across banking and there are privacy issues. However, data-driven technologies such as machine learning cannot function well in operational settings unless they are provided with accurate data. The purpose of this work is to propose a blockchain-based method for constructing a strong machine learning system that can detect fraudulent online transactions. The use of blockchain technology within the framework that has been presented guarantees the privacy of the information. A criterion has been devised for mining the block in the most effective manner. In the last step of the evaluation process, the performance of the blockchain network is evaluated by subjecting it to a range of data loads and escalating levels of difficulty.

Introduction:

The Internet facilitates everything we do now, from purchasing to investing to banking. To commit credit card fraud, one must use another person's card fraudulently. Credit card fraud hurts financial companies' bottom lines. Thanks to software and technological advancements, individuals may remain anonymous while conducting unlawful financial transactions online.

Blockchain reduces fraud and improves security [1,2]. This technology detects and stops fraudulent financial transactions. A secure, shared ledger tracks transaction blocks in a secure, shared ledger. Decentralized networks are safe because they cannot be attacked. Only confirmed contributors have access, increasing network security. Permissioned blockchain networks are secure. Allowing select individuals to undertake jobs adds security.

Predicting financial crises [3] and detecting credit card fraud [4] are two areas where machine learning algorithms have shown their effectiveness.

Justin D. Harris and Bo Waggoner [5] created a framework for collaborative machine learning that uses ML and blockchain to train a machine learning algorithm in real time using a distributed consensus process. Corporate privacy, adaptive incentives, and focused mining criteria [5] were not considered in their construction. Customer data and company insights are in the same e-commerce dataset [6]. This study proposes a privacy-preserving e-commerce solution based on blockchain technology and smart contracts. A fraud detection model is trained, maintained, and updated in micro increments.

Our strategy involves evaluating many machine learning algorithms on a seed dataset. Once the machine learning algorithm is chosen, it's preserved on the blockchain so any bank may use it and enhance it over time. Smart contracts are immutable once established using blockchain technology. These agreements maintain the incentive mechanism and automatically update the ML model.

This paper valuable contributions. To facilitate the development of reliable ML models within the banking sector, a proposed model a method that protects clients' privacy. The proposed model has been tried out a variety of mining times to evaluate the effectiveness of the system with varying amounts of data for model training and blockchain mining difficulties.

Background:

An efficient system to identify fraudulent transactions must be implemented by every financial institution that offers credit or debit cards. RF, ANN, SVM, k-nearest neighbors, and other hybrid and privacy-preserving algorithms have been recognized as useful for credit card fraud detection. Ostapowicz and Zbikowski [7] studied how blockchain's security and reliability might detect fake bank accounts. They examined the effectiveness of several supervised learning algorithms on the Ethereum blockchain and exhibited their potential. Fashoto et al. [8] used k-means clustering to detect credit card fraud. The k-means clustering method has been evaluated using MLPs and HMMs. Thang et al. [9] presented a technique to detect small-business tax fraud. Fuzzy inferences in this model identify the company's business type; neural networks determine whether it's fake. Financial information, industry and market data, and the company's fraud record are fed into a neural network. To automatically sign blockchain transactions and detect fraud, Podgorelec et al. [10] used machine learning to sign blockchain transactions and identify fraud.

Blockchain transactions are simplified without a user signature. Using Ethereum network data, they tested the model. In order to enhance the effectiveness and performance of the k-means clustering method, Brown and Huntley [11] tried to formulate an unsupervised clustering technique as an optimization issue. He clustered information using a simulated annealing process.

Methodology:

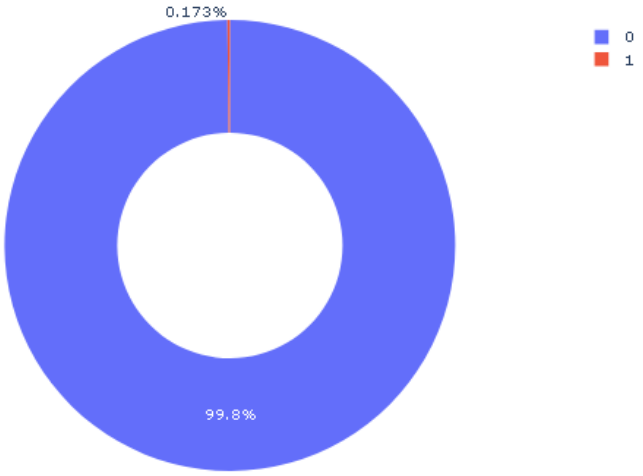
There are two stages to this paper. In the first stage, thorough experimentation is conducted with five machine learning algorithms to select the best model. In the second stage, the necessary implementation of blockchain and smart contract is done to accommodate the ML algorithm.

3.1. Machine Learning Selection:

In this section, the experiments with five machine learning algorithms and their results have been shown to describe how the final model was selected.

Banking systems handle a massive number of payments. In this paper used the data included all transactions made by European cardholders. There were 492 instances of fraud out of a total of 284,807 transactions. In comparison to the huge amounts of transaction data, there are fewer occurrences of fraud because that data is not balanced.

The initial phase in the machine learning experiments is to find ways to deal with the huge inconsistency of the data. This dataset is first normalized, then standardized, and then oversampled with imbalanced data in order to achieve balance before being put through its tests with several machine learning algorithms.



Check Balance and analysis the target

Within the dataset, the training and testing datasets are separated. 70 % the data is being used for training, while the remaining 30% is being used for test performance. The algorithms include Support Vector Machine, Logistic Regression, Decision Tree, and Random Forest with a boosting approach.

In terms of evaluation, each algorithm has its own set of performance metrics, which have been built to evaluate a broad variety of items. Therefore, several methods are being evaluated, and this should be one of the characteristics considered FP and FP, as well as their connections, are often used in credit card fraud detection to assess the accuracy of various methodologies.

True Positive: A "true positive" occurs when a transaction is properly detected as fraudulent (TP).

True positive = TP/TP+FN

True Negative: The true negative rate is determined as the percentage of normal transactions that are properly classified as normal transactions.

True negative=TN/TN+FP

False Positive (FP): The percentage of non-fraudulent transactions that are wrongly classified as fraudulent.

False positive=FP/FP+TN

False Negative (FN): It indicates that a percentage of non-fraudulent transactions are incorrectly categorized as regular transactions.

False negative = FN/FN+TP

The following results indicate that models such as logistic regression, SVM, decision trees, and random forest with boosting approach performed well against the data.

This dataset is difficult to group since fraudulent and legal transactions are so similar. Therefore, it is very difficult to distinguish between fraud instances and legitimate organizations.

The comparison table was prepared using the findings of the computer simulation. Comparisons are made between accuracy, precision, and recall. The table demonstrates that the Logistic Regression model has the greatest precision, recall, and accuracy.

A table comparing the precision, recall, and accuracy of several machine learning algorithms.

	Accuracy	Precision	Recall
SVM	93%	88%	98%
Random Forest	95%	93%	1
Logistic Regression	96%	95%	1
Decision Tree	92%	92%	91%

3.2. Blockchain Architecture

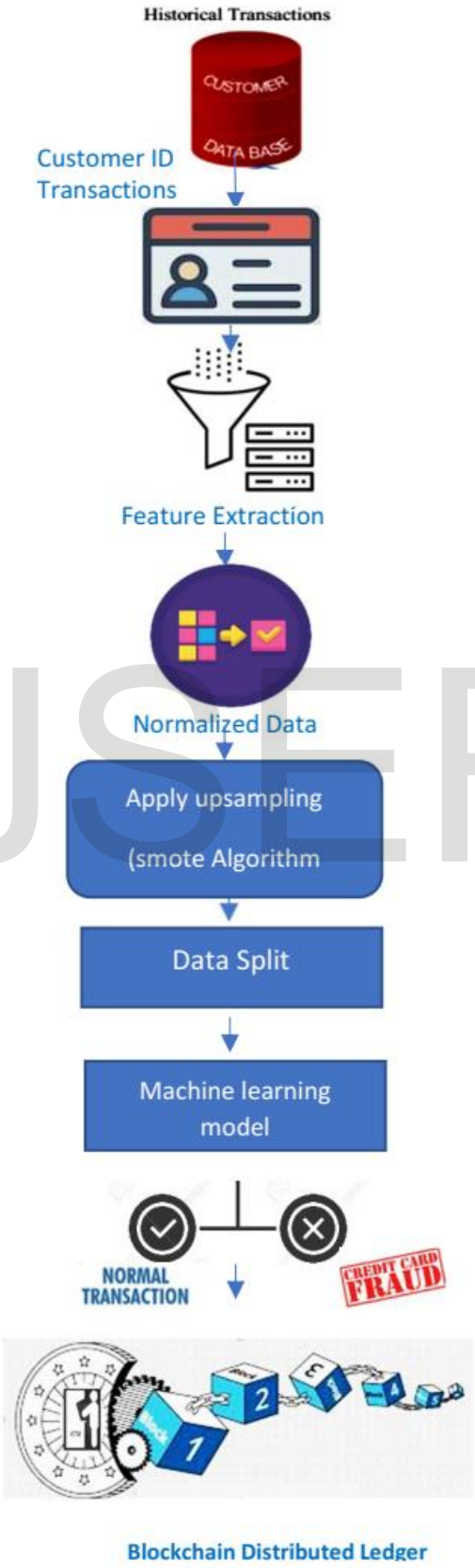
Unfortunately, the complexity of financial transaction data makes open and collaborative research difficult [12]. For reasons of privacy, market competitiveness, and confidentiality, transaction information is kept private. Financial fraud, on the other hand, causes massive losses for the industry. Proposed solution, which makes use of smart contracts and the blockchain, creates a framework in which businesses can collaborate to gradually train a machine-learning algorithm while protecting their own data.

Three layers form the system architecture: the application layer, the off-chain machine learning layer, and the blockchain layer. The application layer is the user interface where contributors and users may register, use the machine learning model, and contribute their organization's transaction data to enhance the ML model.

In the application layer, a decentralized application links the actor to the blockchain network. The chosen ML model is implemented on the blockchain at the network's start. When a client node enters the network, they will apply the ML model to execute requests regardless of the system role.

To contribute data to the system for substantially training the model, the participant node must still be a contributor. After a contributor uploads a new instance of the dataset, the request is delivered to the interface through the distributed application. The interface then transmits the data to our application's ML layer.

As seen in Figure below, the ML layer consists of all the required components for machine learning, starting with data preparation and ending with partial model training. Before processing data, the application layer puts it via a data preparation filter. This filter evaluates the dataset for typical issues, including inappropriate data form, the lack of needed characteristics, and null values.


Historical Transactions
CUSTOMER
DATA BASE
Customer ID
Transactions
Feature Extraction
Normalized Data
Apply upsampling
(smote Algorithm
Data Split
Machine learning
model
NORMAL
TRANSACTION
CREDIT CARD
FRAUD
Blockchain Distributed Ledger

The primary contract compares the current measurements to the previous ones and determines whether they represent an improvement. The decision is then sent to the model contract. If the decision shows an improvement, a new block containing the model's measurements and outcomes is generated and broadcast around the network to be added to the main chain. The model contract additionally revises the existing ML model in the event of enhancements.

The blockchain network further stores the learning history of the ML model. The cryptographic hash of the modified model is saved on the blockchain, and the ML layer and application layer will utilize this updated model in subsequent phases. The application layer displays real-time updates of the blockchain, machine learning model, transactions, etc.

Discussion and Result:

As a result of using blockchain, the system can preserve users' privacy as they collaborate to train an ML model without requiring them to divulge any sensitive information. Access to pertinent data is crucial in the modern banking sector. Clearly, this is a highly competitive industry where cooperation is difficult [13]. Over the last decade, machine learning strategies have been used in a vast array of real-world settings where they have continuously shown their precision and reliability. For machine learning models to train and develop successfully, they need an accurate representation of real data [14]. In contrast, financial information is secret and cannot be divulged. To solve this problem, our platform uses blockchain technology and smart contracts to cooperatively and progressively train a machine learning algorithm for fraud detection without needing enterprises to compromise data confidentiality or privacy. No data is transferred to a centralized server. Instead, the data is processed locally at each node (off-chain), and only the metrics and results are stored on a central server.

To broadcast a new node, a series of limitations known as "mining criteria" must be met. As such, it is a crucial decision for the whole network [15]. The new data is used to partially retrain the model, and the system then decides whether to use the new knowledge (creating a new block). The mining criteria prevent this by comparing the other three indicators to the historical statistical norm of the best outputs. Developing a revolutionary incentive system that addresses the obstacles of machine learning Our system's computational demands are not incurred all at once, but rather through a series of phases. Based on our time research, we decided that the mechanism is sufficiently quick.

The proposed solution employs blockchain technology to guarantee the transaction identity's secrecy, immutability, and recoverability [16]. As the training data for the model is stored on the blockchain, the whole process may be automated by a smart contract while retaining the network's inherent security. As a result, transaction may preserve their mutual trust, and the model's capacity to learn from shared data can be made far more precise and rigorous. Using our proposed strategy, organizations can take the necessary steps toward cooperation. The real-world implications are summarized as follows: banking is anticipated to grow as a result of the advancement of current technology via cooperation. It is a generic strategy that can be used anytime commercial concerns limit essential collaboration.

Conclusion:

This study proposed an approach for banking to combine their resources in order to create high-quality machine learning algorithms while maintaining vital business strategies and minimizing privacy issues. This study employs state-of-the-art blockchain technology to create a platform for training machine learning (ML) algorithms for fraud detection in a collaborative context while protecting the anonymity of contributing financial transactions. In this study, an intelligent contract was used to automate the whole system in a reliable and consistent manner. There is no way for anybody to circumvent the system's security, allowing companies to feel secure exchanging information. To encourage banking, a system of incentives that is both flexible and egalitarian has been devised to reward excellent performance. If performance increases are difficult to achieve, the incentive mechanism will reward you more. In proportion to the complexity of updating a model, our strategy will deliver more benefits. Due to the increased potential impact of legitimate data on the performance of the model, the provider of authentic data gets considerable incentives. Credit card fraud was fought using a variety of machine learning techniques, including logistic regression, random forests, decision trees, and support vector machines. Several metrics, including error rate, sensitivity, and specificity, are utilized to assess the effectiveness of the proposed system. The overall accuracy of the four models was 96 %, 92 %, 95 %, and 93 %, respectively. The outcomes of logistic regression are superior to those of the other three models and the support vector machine. Our methodology is sufficiently flexible to be utilized in banking sector where data privacy and security are crucial, yet cooperation across businesses may greatly improve performance and precision.

Reference:

[1] P. Rani, A. Balyan, V. Jain, D. Sangwan, P.P. Singh, J. Shokeen, A probabilistic routing-based secure approach for opportunistic IoT network using blockchain, in 17th India Council International Conference (INDICON 2020) (IEEE, 2020)

[2] P. Rani, P.P. Singh, A. Balyan, J. Shokeen, V. Jain, D. Sangwan, A secure epidemic routing using blockchain in opportunistic internet of things, in International Conference on Data Analytics and Management (Springer, Berlin, 2020)

[3] S. Sankhwar, D. Gupta, K. Ramya, S.S. Rani, K. Shankar, S. Lakshmanaprabu, Improved grey wolf optimization-based feature subset selection with fuzzy neural classifier for financial crisis prediction. Soft Comput. 24(1), 101–110 (2020)

[4] S.G. Fashoto, O. Owolabi, O. Adeleye, J. Wandera, Hybrid methods for credit card fraud detection using k-means clustering with hidden Markov model and multilayer perceptron algorithm. Curr. J. Appl. Sci. Technol. 1–11 (2016)

[5] J. D. Harris and B. Waggoner, "Decentralized and collaborative ai on blockchain," in 2019 IEEE international conference on blockchain (Blockchain), pp. 368–375, IEEE,2019.

[6] C. Ma, X. Kong, Q. Lan, and Z. Zhou, "The privacy protection mechanism of Hyperledger fabric and its application in supply chain finance," Cybersecurity, vol. 2, no. 1, pp. 1–9, 2019.

[7] S. A. Assefa, D. Dervovic, M. Mahfouz, R. E. Tillman, P.Reddy, and M. Veloso, "Generating synthetic data in finance: opportunities, challenges and pitfalls," in Proceedings of the First ACM International Conference on AI in Finance, pp. 1–8, 2020.

[8] "ONLINE PAYMENT FRAUD LOSSES TO EXCEED $206 BILLION OVER THE NEXT FIVE YEARS; DRIVEN BY IDENTITY FRAUD," Juniper Research, 2021. [Online]. Available: https://www.juniperresearch.com/press/online-payment-fraud-lossesexceed-206-bn [Accessed: 18-Apr-2022].

[9] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," Neural Networks, vol. 113, pp. 54–71, 2019

[10] T. Amarasinghe, A. Aponso, and N. Krishnarajah, "Critical analysis of machine learning based approaches for fraud detection in financial transactions," in Proceedings of the 2018 International Conference on Machine Learning Technologies, pp. 12–17, 2018.

[11] F. Beena, I. Mearaj, V. K. Shukla, and S. Anwar, "Mitigating financial fraud using data science - 'A case study on credit card frauds,'" 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM), 2021

[12] A. Assefa, D. Dervovic, M. Mahfouz, R. E. Tillman, P.Reddy, and M. Veloso, "Generating synthetic data in finance: opportunities, challenges and pitfalls," in Proceedings of the First ACM International Conference on AI in Finance, pp. 1–8, 2020

[13] K. D. Martin, J. J. Kim, R. W. Palmatier, L. Steinhoff, D. W. Stewart, B. A. Walker, Y. Wang, and S. K. Weaven, “Data Privacy in retail,” Journal of Retailing, vol. 96, no. 4, pp. 474–489, 2020.

[14] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," Journal of Big Data, vol. 7, no. 1, 2020.

[15] Y. Wang, C.-R. Chen, P.-Q. Huang, and K. Wang, "A new differential evolution algorithm for joint mining decision and Resource Allocation in a MEC-enabled wireless blockchain network," Computers & Industrial Engineering, vol. 155, p. 107186, 2021

[16] A. B. Haque, A. K. Islam, S. Hyrynsalmi, B. Naqvi, and K. Smolander, "GDPR compliant blockchains–A systematic literature review," IEEE Access, vol. 9, pp. 50593–50606, 2021