A Project report on

COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR FRAUD DETECTION IN BLOCKCHAIN

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

Bachelor of Technology

in

Computer Science and Engineering

Submitted by

Kumbala Pavan Reddy
(20H51A0515)

Mamidi Varun
(20H51A0516)

Thalari Nihith Novah
(20H51A0553)

Under the esteemed guidance of

Ms. T. Adarana
(Assistant Professor)

Department of Computer Science and Engineering

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

(UGC Autonomous)

*Approved by AICTE *Affiliated to JNTUH *NAAC Accredited with A^+ Grade

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

2020- 2024

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Major Project Phase I report entitled **"Comparitive Study of Machine Learning Algorithms for Fraud Detection in Blockchain"** being submitted by Kumbala Pavan Reddy (20H51A0515), Mamidi Varun (20H51A0516), Thalari Nihith Novah (20H51A0553) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

Ms. T. Adarana
Assistant Professor
Dept. of CSE

Dr. Siva Skandha Sanagala
Associate Professor and HOD
Dept. of CSE

ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Ms. T. Adarana, Assistant Professor**, Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

K. Pavan Reddy 20H51A0515
M. Varun 20H51A0516
T. Nihith Novah 20H51A0553

TABLE OF CONTENTS

List of Figures

List of Tables

ABSTRACT

The economy and trust in a blockchain network are significantly impacted by fraudulent transactions. Consensus methods like proof of work or proof of stake can confirm a transaction's validity, but they cannot confirm the identity of the persons that participated in the transaction or verified it. A blockchain network is still susceptible to fraud because of this. Use of machine learning algorithms is one method for eradicating fraud.

The existence of fraudulent exchanges in the economy discourages investors from investing in bitcoin and other blockchain-based businesses. False exchanges are regularly viewed with scepticism due to the gatherings in issue or the way they are put up. To prevent them from jeopardizing the trustworthiness of the neighborhood and the blockchain network, people endeavor to identify false exchanges wherever possible. Numerous other Machine Learning approaches have been suggested to address this issue, but none of them has clearly emerged as the best one, even though some of the results show promise. This study looks at how well a few controlled AI models and a few deep learning models do at spotting bogus transactions in a blockchain network. Our goal is to pinpoint the clients and transactions that will probably resort to extortion.

The machine learning techniques train the dataset based on the fraudulent and integrated transaction patterns and predict the new incoming transactions. The blockchain technology is integrated with machine learning algorithms to detect fraudulent transactions in the Bitcoin network.

CHAPTER 1
INTRODUCTION

CHAPTER 1
INTRODUCTION

1.1. Problem Statement

The issue of identifying fraudulent transactions has long been researched. The economy suffers from fraudulent transactions, which also make individuals less likely to buy bitcoins or have faith in other blockchain based products. Fraudulent transactions are frequently suspect, either because of the parties involved or because of the way they are structured. To keep fraudulent transactions from jeopardizing the community and integrity of the blockchain network, members of a blockchain network strive to identify them as quickly as feasible. Such comparison research will assist in selecting the optimum algorithm based on the trade-off between accuracy and computing speed. Blockchain transactions are constant because once they are recorded, they cannot be changed or reversed. Before a "block" of transactions is added to the blockchain, network clients must agree on the validity of each transaction.

The problem of detecting fraudulent transactions is being studied for a long time. Fraudulent transactions are harmful to the economy and discourage people from investing in bitcoins or even trusting other blockchain-based solutions. Fraudulent transactions are usually suspicious either in terms of participants involved in the transaction or the nature of the transaction. Members of a blockchain network want to detect Fraudulent transactions as soon as possible to prevent them from harming the blockchain network’s community and integrity. Many Machine Learning techniques have been proposed to deal with this problem, some results appear to be quite promising, but there is no obvious superior method.

1.2. Research Objective

The objective of the project is to create sophisticated machine learning models for analyzing and predicting the data set utilizing typical machine learning methods, statistics, and calculus to forecast the frequency and volume of fraudulent and legitimate transactions. We also provide an extensive comparative study of various supervised machine learning techniques like decision trees, Naive Bayes, logistic regression, multilayer perceptron, and so on for the above task.

To compare the performance of various supervised machine learning models like SVM, Decision Tree, Naive Bayes, Logistic Regression, and few deep learning models in detecting fraudulent transactions in a blockchain network. Such comparative study will help decide the best algorithm based on accuracy and computational speed trade-off. Our goal is to see which users and transactions have the highest probability of being involved in fraudulent transactions.

1.3. Project Scope and Limitations

The scope of the project is to develop advanced machine learning models that utilize typical machine learning methods, statistics, and calculus to analyze and predict the frequency and volume of both fraudulent and legitimate transactions within a blockchain network. The primary objective is to create predictive models capable of identifying and distinguishing between fraudulent and legitimate transactions. The project aims to achieve this through the application of various supervised machine learning techniques such as Support Vector Machines (SVM), Decision Trees, Naive Bayes, Logistic Regression, and deep learning models.

The comparative study of these models will provide valuable insights into their performance in detecting fraudulent transactions. By assessing accuracy and computational speed trade-offs, the project seeks to determine which algorithm is most suitable for this specific task.
Additionally, the project intends to identify users and transactions with the highest probability of involvement in fraudulent activities, which can be valuable for fraud detection and prevention in a blockchain network.

However, there are certain limitations to consider in this project. Firstly, the quality and quantity of the available data will significantly impact the model's performance, and obtaining a comprehensive and reliable dataset can be challenging. Moreover, the success of the project depends on the assumption that fraudulent activities leave distinct patterns in the data, which may not always be the case. Additionally, while the comparative study aims to find the most effective algorithm, it may not consider all possible factors relevant to practical implementation, such as the scalability and interpretability of the chosen models. Furthermore, the project may face ethical considerations regarding user privacy and data handling, which should be carefully addressed. Lastly, the project's results might not be directly applicable to all blockchain networks, as the characteristics of the data and the nature of fraud may vary between different systems.

CHAPTER 2
BACKGROUND WORK

CHAPTER 2
BACKGROUND WORK

2.1. A Machine Learning and Blockchain Based Efficient Fraud Detection Mechanism

2.1.1. Introduction

Every industry, including banking, education, health care, and others, has modernized as a result of technological growth. Moreover, with the advent of communication technology, online transactions and means of payment are also being modernized. Through this modernization, traditional currencies are being converted into digital currencies, and all financial transactions are being conducted digitally. However, these transactions are not fully secured and are vulnerable to various digital attacks, such as fraud issues, anomalies, and privacy breaches. Additionally, as the volume of transactions rises, there is an increase in fraud associated with financial transactions. As a result, billions of dollars are lost globally every year. Any suspicious activity on a network that behaves abnormally is called an anomaly. In cybersecurity and digital financial exchange, anomaly detection is used to detect fraud and network invasion. The goal of anomaly detection is to protect the network from illegal and fraudulent activities. In the financial sector, anomaly detection applications have investigated suspicious activity and identified hackers and fraudulent users. However, all anomaly detection methods in traditional financial systems are designed for centralized systems. Therefore, with the development of digital currencies, such as Bitcoin, anomaly detection methods using the blockchain are improving. Despite these advances, there are still many fraud occurrences.

Many artificial intelligences (AI) and machine learning techniques have been proposed to detect anomalies and fraud in digital transactions; however, there is no suitable solution for centralized systems. Blockchain is the most advanced and quickly evolving technology in many fields. It first became visible with the appearance of Bitcoin in 2008, which was introduced by Satoshi Nakamoto.

2.1.2. Merits, Demerits and Challenges

Blockchain is the latest and most secure technology that covers various research areas related to security. Blockchain development is based on digital currencies and is used to secure digital financial transactions. It protects financial systems from fraudulent attacks. Therefore, a blockchain-based machine learning algorithm is proposed to secure digital transactions. The proposed model predicts whether the incoming transaction in the blockchain is fraudulent or not. The proposed machine learning algorithms are trained and tested on a bitcoin-based dataset based on bitcoin transactions and predict the behavior of the incoming transactions. The given dataset is based on 30,047 entities, with smaller numbers of fraudulent entities. Due to the small amount of fraudulent data in the dataset, good results cannot be obtained because of the data imbalance problem. Therefore, we generate synthetic malicious data points through SMOTE to achieve better results. We use XGboost and random forest to classify the model and calculate the confusion matrix. This classification allows the model to distinguish between fraudulent and real data. The simulation results show that the proposed algorithm works adequately to find transaction fraud. Moreover, two attacker models are implemented to check the efficacy of the system against bugs and attacks. The proposed system is robust against double-spending and Sybil attacks

A major limitation of this proposal is that it can be affected by the adversarial attack it also address such threats.

2.1.3. Implementation

This section first presents the simulation results of our proposed model, then we present the results after inducing modern cyber attacks to the system, i.e., Sybil attack, and double-spending attack. The selected dataset is highly skewed, as shown in Figures 1.1 and 1.2. The classification models are biased toward the majority class due to the imbalance of the data.

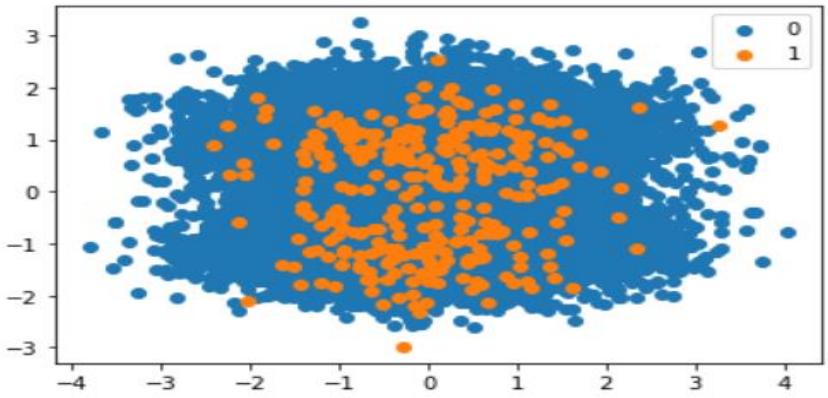

Figure.1.1: Imbalanced data

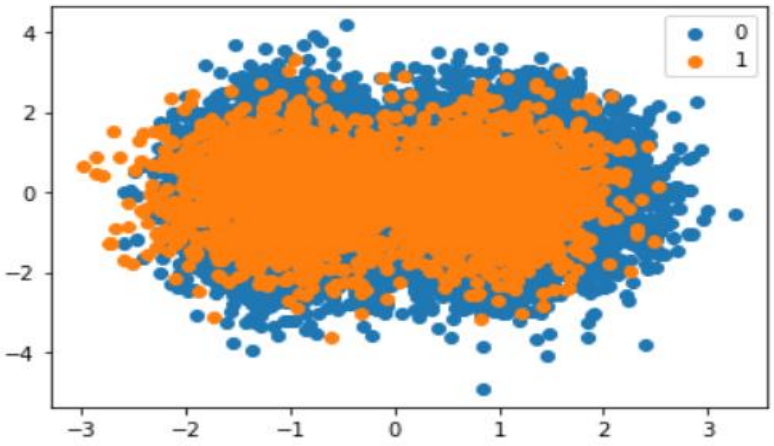

Figure.1.2: Balanced data

Figure 1.1 shows the presence of malicious and honest transactions in the dataset. It can be seen from the figure that the number of honest transactions is higher than the number of malicious transactions. This imbalanced nature of the data leads to a bias in the classification. Synthetic data are used to solve this problem. The malicious entities are oversampled using SMOTE. The synthesized transactions are added to the dataset to limit the bias of the model during classification. The results obtained after using SMOTE are shown in Figure 1.2. The observed log loss of XGBoost during training is shown in Figure 1.3. The log loss is observed for both the training data and the test data. From the figure, it can be seen that at a count of 10 iterations, a drastic drop is observed for both the training and test data. Moreover, the smoothness of the curves indicates that the model efficiently captures the nonlinear patterns of the data. For the test data, the log loss is higher than for the training data. However, the difference is not too large. The smaller difference between the training and test curves indicates that the model is well trained on unseen data. The trained

model can be applied to real-world scenarios for anomaly detection in blockchain networks.

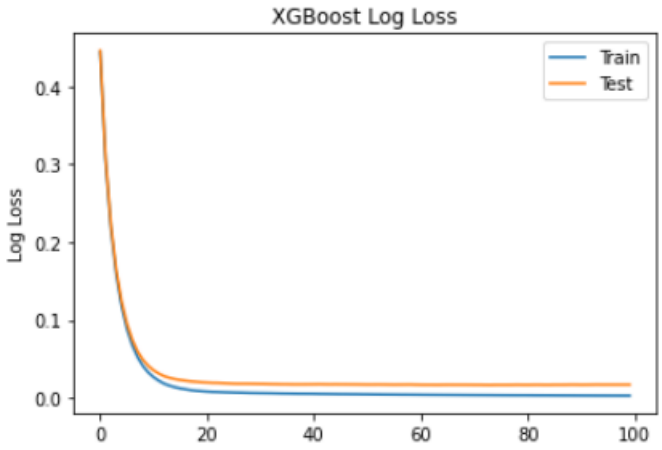

Figure.1.3:Logloss of XGboost.

Figure 1.4 shows the correlation between the fraudulent and non-fraudulent class. Meanwhile, the value almost equal to 0, in the case of mean ni nb tc, shows the minimum correlation between fraudulent and non-fraudulent.

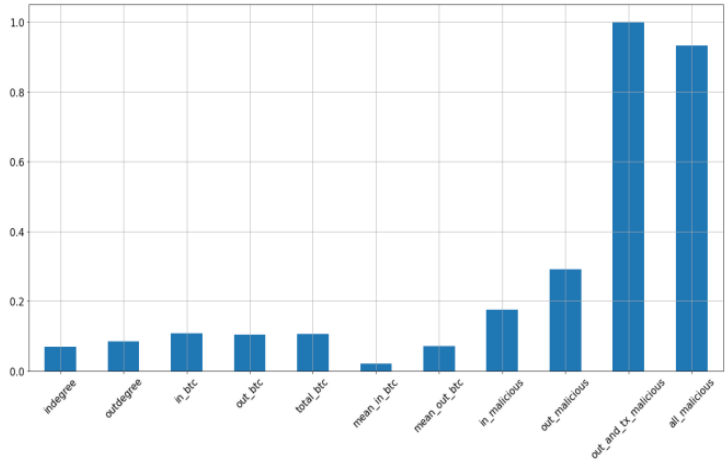

Figure.1.4: Correlation with class fraudulent or not.

Figure 1.4 shows the error that occurs when classifying with XGBoost. It shows the error for both training and test data. It can be observed that the classification error decreases as the number of iterations increases. The error is high for training data, and the figure shows a gradual decrease, while it is lower for test data and decreases rapidly. The precision–recall curve of the XGboost model is visualized in Figure 1.5. This curve predicts the harmonic mean of both precision and recall. It is seen that a very slight decrease is observed, starting from 1. As soon as the recall value reaches more than 0.9, there is a sudden drop in the precision value. Figure 1.6 shows the accuracy when XGBoost is used. It shows that the highest peak of 0 to 1 indicates that the model achieves optimal accuracy in classifying blockchain transactions as legitimate or malicious. After reaching the maximum value of 0.9, the accuracy remains constant throughout the training.

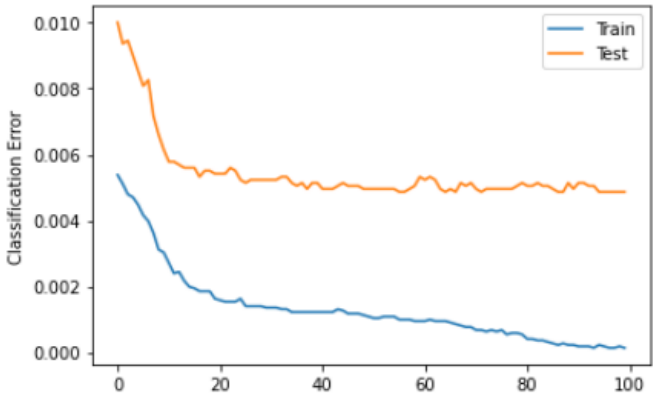

Figure.1.5: Classification error of XGboost.

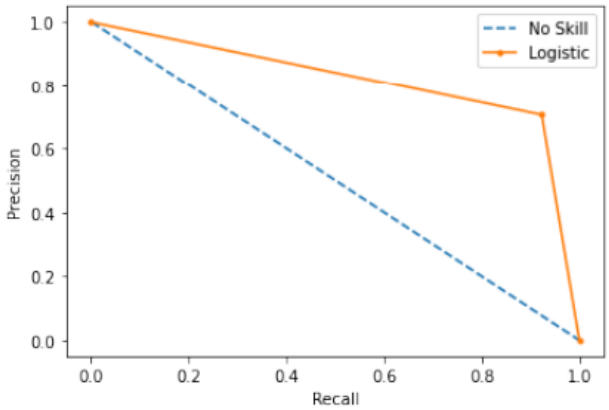

Figure.1.6: Precision of RF.

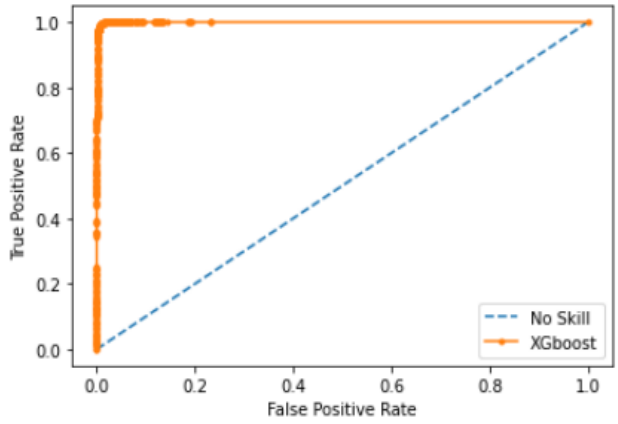

Figure.1.7: Accuracy of XGboost.

Figure 1.7 shows the confusion matrix obtained using RF. In this matrix, random forest selects 9014 random samples, correctly identifying 9009 predictions. This means that the proposed model efficiently discriminates between malicious and legitimate transactions. The matrix shows that the highest values are obtained in the case of true negatives, namely 99%. In the other three cases, the number of values is lower. This shows that the proposed model is efficient in detecting true negative transactions. Moreover, the phenomenon of majority voting in the random forest increases the performance of the model during classification. Figure 1.8 shows the AUC of a random forest. The AUC describes how well the model distinguishes between the positive and negative classes. It can be seen that the value of the AUC increases dramatically at the beginning to almost 0.85. Thereafter, a gradual increase is observed until the maximum value of 0.92 AUC is reached. The random forest model achieves an AUC of 0.92, which means that it performs well in capturing legitimate and malicious transactions.

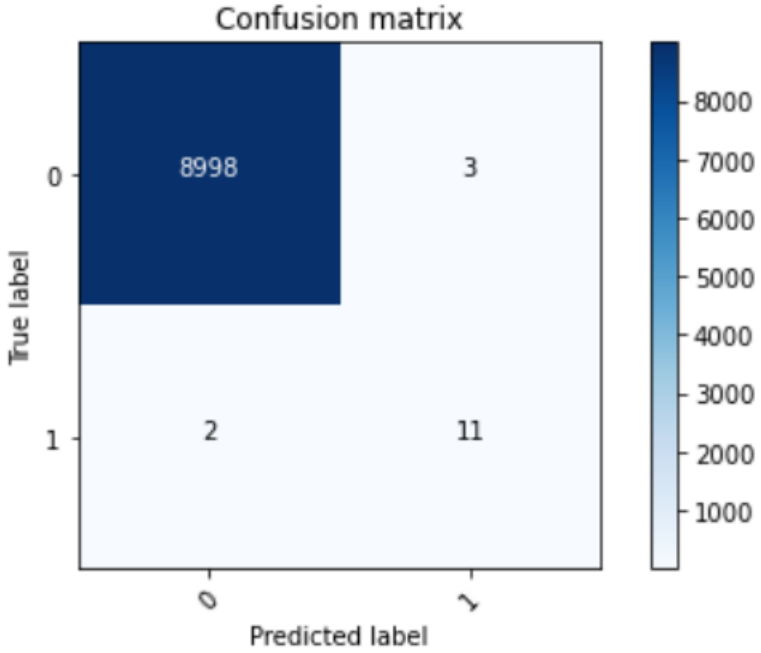

Figure.1.8: Confusion matrix with random forest.

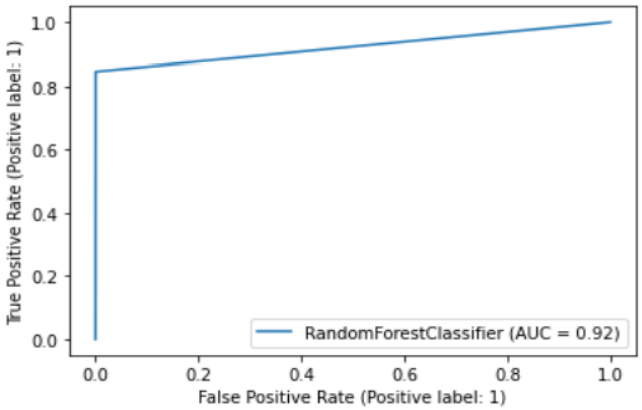

Figure.1.9: Accuracy of Random Forest.

Figure 1.10 shows the transaction and execution costs incurred in executing the functions involved in the blockchain smart contract. The costs are expressed in terms of gas, a basic unit of gas consumption in the blockchain network. From the figure, it can be seen that the transaction costs of all functions remain the same, while the execution costs of the publish transaction function are the highest, as mining costs are also included. Overall, the transaction costs are higher than the execution costs for all functions. The reason for this is that the former includes the processing costs of entire transactions, while the latter includes only the execution costs of some operations in a given function.

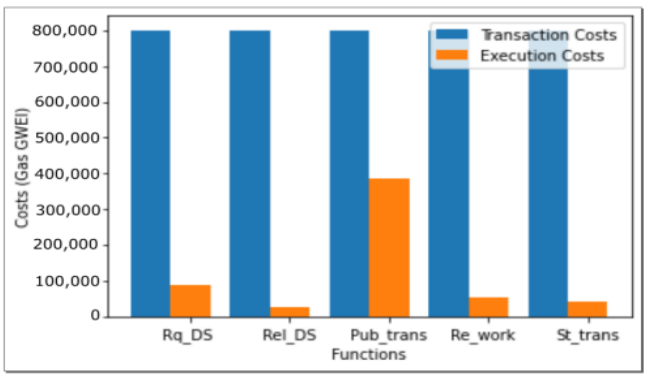

Figure.1.10: Transactions published and stored on Blockchain.

Validation of Proposed Model Based on Modern Cyber Attacks:

Nowadays, blockchain technology is considered the most secure technology for financial transactions due to its advances; however, it is still vulnerable to current cyber-attacks. Despite all the advances and security measures, some advanced cyber criminals find strong attacks against the blockchain. The security features of blockchain cannot maintain its security measures against modern cyber-attacks, such as selfish mining attacks, Sybil attacks, double-spending attacks, and replay attacks [38]. Therefore, this section explicitly presents results of our proposed model when modern cyber-attacks are induced in the system.

Double-Spending Attack:

In the blockchain, a transaction is only confirmed after the agreement/verification of all nodes. This verification takes a specific period, which creates a chance for cyber attacks. Double spending is one of these attacks that exploit the transaction verification time. Every transaction on the blockchain takes time for verification, and attackers use this time to their advantage. During the transaction verification delay, the attacker uses the same coin at two places as the verification of both transactions takes place simultaneously. In this way, digital currency is duplicated and falsified easily. In Ref. [33], the authors worked on the two double-spending attacker models. They enhance the two existing attacker models of Satoshi Nakamoto and Rosenfield for double spending. The first proposed model is called the “generalized model”, in which authors added a time parameter. This parameter is used to calculate the time advantage of an attacker. The second proposed model is known as the time-based model. This model counts the time when an attacker and honest node mined their last blocks.

2.2 Fraud Detection: A Review on Blockchain

2.2.1. Introduction

In recent years, the world of blockchain technology has captured widespread attention. This dynamic field has witnessed the introduction of multiple measures aimed at detecting and mitigating fraudulent transactions and unusual activities that deviate from established behavioral patterns. To bolster the accuracy of fraud detection, a range of outlier analysis techniques have been employed, including decision trees, support vector machines, evolutionary algorithms, Bayesian belief networks, and the formidable neural networks. Despite considerable progress, only a limited number of models have proven capable of reliably identifying all forms of fraudulent behavior.

Even after unmasking deception and accurately tracing the origins of deceitful inputs, the quest for truth and well-informed decision-making remains an elusive endeavor. As our interactions increasingly transition to the digital realm, upholding fundamental principles of security, precision, reliability, and transparency among market participants becomes paramount. This research scrutinizes the global landscape of operational fraud detection systems, spotlighting their demonstrated effectiveness in combating deceptive practices. We delve into ten diverse domains where blockchain technology has wielded transformative influence and propose practical solutions to the everyday challenges confronting both individuals and organizations.

2.2.2. Merits, Demerits and Challenges

Blockchain technology is described in this paper as a way to decrease the layers of corruption in government procedures. This study focuses on employing a novel form of encryption method to overcome security and privacy concerns in blockchain. The proposed model ensures that everyone linked to a soon-to-be implemented blockchain network may see all government procedures. This enables ordinary citizens to investigate the operation of any government scheme, track its progress, and track financial transfers.
The following are some of the drawbacks of this model:

- Stagnation of funds caused by middle-level authorities.
- Money is misappropriated in the middle levels, with everyone blaming each other.
- Schemes are executed slowly.
- Identifying the true needy/beneficiary is a challenge.
- Inappropriate financial allocation.
- The beneficiary's illiteracy and stupidity.

They also review state-of-the-art technologies for detecting online fraud and intrusions, identify certain fraud and malicious activities that blockchain technology can effectively prevent, and make recommendations for strategically fighting various attacks to which blockchain technology may be vulnerable. Existing machine learning and data-mining algorithms could find new uses in identifying fraud and intrusions in blockchain-based transactions. Guided machine learning methods like deep-learning neural networks, support vector machines, and Bayesian belief networks may help detect outlier behaviors by profiling, monitoring, and detecting behavioral trends based on people's transaction histories. Despite the advancement in technology, still, the problems regarding Video Fraudulence are faced and there is no concrete solution for this problem.

2.2.3. Implementation

blockchain applications have attracted a lot of attention. They are more valuable than money and can be used to replace fiat money and traditional banking. The ability to trade wealth on a blockchain, on the other hand, is at the heart of the system and must be reliable. Blockchains have built-in features that assure the system's stability and durability. Malicious actors can still use well-known tactics to steal money, such as virus software or falsified emails. We also undertake a sensitivity analysis to show how the models offered rely on specific attributes and how the lack of some of them impacts overall system performance. Blockchain can be used to fight and prevent fraud in a business network. One of the fundamental characteristics that determines blockchain's worth is its ability to share data rapidly and securely without relying on a single institution to assume responsibility for data security. One of the most significant advantages of blockchain technology is increased security. The increased security provided by blockchain is due to the way the technology works: With end-to-end encryption, blockchain generates an unalterable record of transactions, preventing fraud and unauthorized activity. Furthermore, blockchain data is kept across a network of computers, making it nearly impossible to attack (unlike conventional computer systems that store data together in servers). Furthermore, by anonymizing data and requiring permissions to limit access, blockchain can solve privacy concerns better than traditional computer systems. Because blockchain transactions cannot be removed or modified, they are immutable. Before a "block" of transactions can be added to the blockchain, network participants must agree that the transaction is valid via a consensus process.

2.3. Analysis of Fraud Detection in Blockchain system using Machine Learning Algorithms

2.3.1 Introduction

Blockchain uses end-to-end encryption to produce a changeless record of exchanges, eliminating fraud and other illegal activity. Information is stored on the blockchain using a network of PCs, making it virtually impossible to hack (in contrast to ordinary PC frameworks that stores information together in servers). Additionally, blockchain can more easily address security concerns than conventional PC frameworks by encrypting information and requiring consents to limit access. Blockchain transactions are constant because once they are recorded, they cannot be changed or reversed. Before a "block" of transactions is added to the blockchain, network clients must agree on the validity of each transaction. Fraudsters employ a variety of techniques to hide their illegal activities, including the production of fictitious records, the alteration of physical or electronic records, and the manipulation of data in an association's bookkeeping frameworks. Using a shared electronic record can help reduce extortion because it increases the openness and clarity of communications between members of a company organization and within a production network. False trades are easier to spot since groups can track the evolution of resources and experiences.

2.3.2. Merits, Demerits and Challenges

Proactive Approach: The research proactively addresses the challenge of identifying fraudulent transactions in blockchain networks, acknowledging the potential harm they can inflict on the economy and user confidence. This forward-looking approach is essential for maintaining trust in the blockchain ecosystem.

Exploring Machine Learning and AI: The study explores both traditional Machine Learning models and deep learning models, indicating a comprehensive examination of AI methods to detect fraudulent transactions. This breadth of analysis can help identify the most effective techniques, balancing precision and processing efficiency.

Potential for Fraud Prevention: By investigating AI algorithms, the research aims to identify clients and transactions with a higher likelihood of engaging in fraudulent activities. This has the potential to significantly reduce the impact of fraudulent exchanges and bolster the security of blockchain networks.

Demerits:

Lack of Concrete Findings: The abstract does not provide specific findings or outcomes of the research, making it challenging to assess the effectiveness of the AI models discussed. Readers are left wondering about the practical impact of the study.

Complexity and Technical Language: The abstract uses technical language and concepts that may be challenging for a non-technical audience to understand. This could limit the accessibility of the research to a broader readership.

Challenges:

Evolving Nature of Fraud: The blockchain landscape is continually evolving, and fraudsters adapt to new methods and technologies. Detecting fraudulent transactions requires staying ahead of these evolving tactics, which poses a substantial challenge.

Data Quality and Availability: Effective AI models for fraud detection require high-quality, labeled data. Gathering such data in the blockchain context can be a challenge due to the decentralized and often anonymous nature of transactions.

Privacy Concerns: Balancing fraud detection with user privacy is a significant challenge. The use of AI to identify fraudulent behavior may inadvertently infringe on user privacy rights, raising ethical and legal concerns.

In summary, while the research outlined in the abstract takes a proactive approach to address the issue of fraudulent exchanges in blockchain networks, it faces the challenge of delivering concrete findings, navigating the complex technical landscape, and addressing the ever-evolving nature of fraud in the blockchain ecosystem. It also needs to balance the effectiveness of fraud detection with the privacy and ethical considerations inherent in AI-based solutions.

2.2.3. Implementation

1. Pre-handling stage We preprocess using network node embedding and the node2vec method. The combined ratings dataset is then read to produce a data frame. The node2vec method's outputs are then normalized, and the normalized values are then saved in a file. When a transaction is discovered to be fraudulent, we assign it a score of 1, and when it is not, we assign it a score of 0. Then the mean and SD of the node features are calculated, and the outcomes are saved to a CSV file. Next, train and test sets of the gathered data were created.

2. Building and preparing different models Test (0.2) and train (0.8) data were included in our analysis. Then, in our train and test sets, we evaluate the ratio of honest to dishonest transactions. We use machine learning and deep learning techniques to predict the likelihood that a transaction will be successful.

3. Performance assessment of the relative multitude of models We evaluate each of our classification models. In order to estimate a parameter, sampling in machine learning involves selecting a sample of data from the dataset with replacement. So, we start by choosing the bootstrap sample size. The model's efficacy is next evaluated using the mean of all accuracy values obtained in this way, after which the sample size is determined

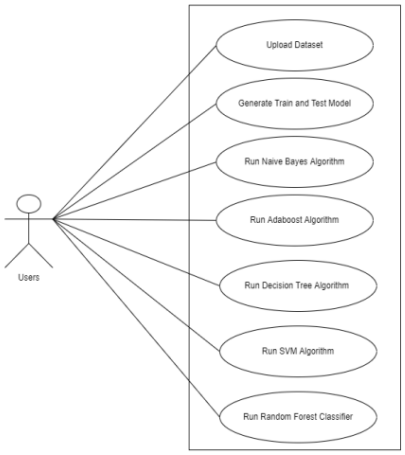

Figure.3.1: processing steps.

CHAPTER 3
RESULTS AND DISCUSSION

CHAPTER 3

RESULTS AND DISCUSSION

3.1. Performance Metrics

Dataset Overview: The dataset used for training and testing comprises 30,047 entities, with a relatively small number of fraudulent transactions. This data distribution introduces a data imbalance problem.

Data Augmentation: Synthetic Minority Over-sampling Technique (SMOTE) is employed to address the data imbalance issue. This technique generates synthetic malicious data points to rebalance the dataset.

Machine Learning Algorithms: Two classification algorithms, XGBoost and Random Forest, are utilized to train the model. These algorithms are designed to distinguish between fraudulent and legitimate transactions.

Confusion Matrix: To evaluate the performance of the model, a confusion matrix is calculated. This matrix provides insights into the model's ability to correctly classify transactions into true positives (correctly identified fraudulent transactions), true negatives (correctly identified legitimate transactions), false positives (legitimate transactions misclassified as fraudulent), and false negatives (fraudulent transactions misclassified as legitimate).

Simulation Results: The study reports that the proposed algorithm demonstrates adequacy in detecting transaction fraud. However, specific performance metrics such as accuracy, precision, recall, and F1 score, which provide a more comprehensive assessment of the model's effectiveness, are not mentioned in the provided content.

Security Testing: Two attacker models are implemented to assess the system's resilience against potential attacks, including double-spending and Sybil attacks. While it is stated that the system is robust, specific metrics related to its ability to withstand these attacks are not provided.

CHAPTER 4
CONCLUSION

CHAPTER 4

CONCLUSION

3.1. Conclusion

Dataset Overview: The dataset used for training and testing comprises 30,047 entities, with a relatively small number of fraudulent transactions. This data distribution introduces a data imbalance problem.

Data Augmentation: Synthetic Minority Over-sampling Technique (SMOTE) is employed to address the data imbalance issue. This technique generates synthetic malicious data points to rebalance the dataset.

Machine Learning Algorithms: Two classification algorithms, XGBoost and Random Forest, are utilized to train the model. These algorithms are designed to distinguish between fraudulent and legitimate transactions.

Confusion Matrix: To evaluate the performance of the model, a confusion matrix is calculated. This matrix provides insights into the model's ability to correctly classify transactions into true positives (correctly identified fraudulent transactions), true negatives (correctly identified legitimate transactions), false positives (legitimate transactions misclassified as fraudulent), and false negatives (fraudulent transactions misclassified as legitimate).

Simulation Results: The study reports that the proposed algorithm demonstrates adequacy in detecting transaction fraud. However, specific performance metrics such as accuracy, precision, recall, and F1 score, which provide a more comprehensive assessment of the model's effectiveness, are not mentioned in the provided content.

Security Testing: Two attacker models are implemented to assess the system's resilience against potential attacks, including double-spending and Sybil attacks. While it is stated that the system is robust, specific metrics related to its ability to withstand these attacks are not provided.

REFERENCES

REFERENCES

[1].https://www.mdpi.com/1424-8220/22/19/7162

[2].https://ijarcce.com/papers/comparative-study-of-machine-learning-algorithms-for-fraud-detection-in-blockchain

[3].https://www.irjet.net/archives/V9/i1/IRJET-V9I1185.pdf

[4].https://chat.openai.com/

GitHub Link

1.