Healthcare Insurance Fraud Detection Using AI and Blockchain

[1]Dr M.A. Chaudhari, [2]Rushikesh Patil, [3]Monali Patkal, [4]Sanika Pagare, [5]Soniya Kolge

Information Technology
Amrutvahini College of Engineering
Sangamner - 422605, India.

Abstract-
In today's society, health insurance is indispensable for managing medical emergencies, yet the persistent threat of fraud within the industry necessitates robust detection mechanisms. This paper proposes a systematic approach to secure health insurance fraud detection, utilizing blockchain technology and artificial intelligence (AI). By categorizing security concerns such as privacy breaches and fraudulent claims, emphasizes the urgency for effective fraud detection methods. The suggested solution advocates for leveraging blockchain's data integrity and AI's analytical capabilities to identify fraudulent activities accurately. Through a practical case study, the efficacy of this approach in detecting health insurance fraud is demonstrated, underscoring its real-world applicability. Moreover, the paper discusses unresolved challenges and future research directions, including scalability and regulatory compliance, to further enhance fraud detection in health insurance systems. The ultimate goal of this all-encompassing framework is to improve the honesty and reliability of health insurance operations, which will benefit all parties involved—including corporations, governments, and individual policyholders.

Keywords: healthcare insurance, , SVM, J48, blockchain, proof of work, SHA-256.

Published in IJIRMPS (E-ISSN: 2349-7300), Volume 12, Issue 2, March- April 2024

I. INTRODUCTION

Healthcare insurance fraud has grown to be a serious problem in recent years, costing billions of dollars yearly and jeopardizing the integrity of healthcare systems throughout the globe. A contract for health insurance is made between the insurance provider and the policyholder. The insurance company pays for the person's medical costs. To tackle this widespread problem, innovative approaches leveraging Artificial Intelligence (AI) and blockchain technology have emerged. This research paper explores the fusion of AI algorithms and blockchain's secure ledger system, particularly utilizing SHA-256 encryption and Proof of Work consensus mechanism, to enhance fraud detection in healthcare insurance., providing patients with greater control and security over their personal and medical information.

Traditional fraud detection methods often struggle to keep pace with evolving fraudulent tactics and fail to provide timely intervention. AI algorithms, such as SVM and J48, offer a solution by analyzing large volumes of data to identify fraudulent patterns and anomalies efficiently. SVM excels in classifying complex data, while J48 decision trees provide interpretable insights into fraudulent behaviour, enhancing the transparency and effectiveness of fraud detection processes. Furthermore, integrating Proof of Work into Blockchain adds a layer of security by requiring network participants to demonstrate a computational effort to validate transactions. This consensus mechanism enhances the trustworthiness of the system and reduces the likelihood of fraudulent activities going undetected.

In summary, the combination of AI techniques like SVM and J48 with Blockchain technology, fortified by SHA-256 encryption and Proof of Work, offers a robust framework for healthcare insurance fraud detection.

This research seeks to explore the synergies between these technologies and their potential to enhance the security and reliability of insurance systems, ultimately benefiting both insurers and policyholders.

A. Motivation

Detecting healthcare insurance fraud is crucial for ensuring fair access to healthcare services and preventing financial losses. Utilising AI algorithms like can significantly enhance fraud detection accuracy. Integrating these algorithms with blockchain technology, ensures secure and transparent data storage. This research aims to explore the effectiveness in detecting healthcare insurance fraud, thereby fostering trust, reducing fraudulent activities, and ultimately benefiting both healthcare providers and patients. This will promote confidence, reduce deceptive practices, and ultimately benefit both healthcare providers and patients.

B. Objectives

- Identify fraudulent activities: Detect and flag instances of fraudulent behaviour within healthcare insurance claims and transactions.
- Minimize financial losses: Prevent financial losses resulting from fraudulent claims by promptly identifying and addressing suspicious activities.
- Protect patient interests: Ensure that healthcare resources are allocated fairly and efficiently, safeguarding patient access to quality care.
- Maintain trust and integrity: Uphold the trust and integrity of the healthcare system by deterring fraudulent behaviour and promoting transparency and accountability.
- Improve regulatory compliance: Comply with regulatory requirements and standards by implementing effective fraud detection mechanisms and reporting protocols.

II. LITERATURE REVIEW

EMAN NABRAWI AND ABDULLAH ALANAZI (2023) .“Fraud Detection in Healthcare Insurance Claims Using Machine Learning”. In this research, supervised machine learning is used to construct a model for detecting healthcare insurance fraud in Saudi Arabia. With the use of logistic regression, randomized forests, and artificial neural networks, the model shows excellent accuracy.. [1]

FATIMA , MAMOONA HUMAYUN ,SIDRA (2023). “ Towards a Secure Technology-Driven Architecture for Smart Health Insurance Systems: An Empirical Study”. By facilitating early fraud detection and demonstrating its value through case studies and industry focus groups, the SHINFDP framework leverages cutting-edge technology to enhance health insurance security.. [2]

KHYATI KAPADIYA, USHA PATEL, RAJESH GUPTA, MOHAMMAD DAHMAN ALSHEHRI, SUDEEP TANWAR, GULSHAN SHARMA, AND PITSHOU N. BOKORO (2022). “Blockchain and AI-Empowered Healthcare Insurance Fraud Detection: An Analysis, Architecture, and Future Prospects”. Through a survey, system design, case study, and discussion of research problems, this paper analyzes security concerns in the domain. It provides an AI and blockchain system for determining health insurance fraud. [3]

LEILA ISMAIL, SHERALI ZEADALLY (2021). “Healthcare Insurance Frauds: Taxonomy and Blockchain-Based Detection Framework (Block-HI)”. This paper tackles healthcare insurance fraud in the United States and suggests a blockchain-based system for safe claim validation to minimize losses and guarantee process transparency. [4]

SEYEDNIMA KHEZR, MD MONIRUZZAMAN, ABDULSALAM YASSINE AND RACHID BENLAMRI (2019), “ Blockchain Technology in Healthcare: A Comprehensive Review and Directions for Future Research”. This survey investigates how blockchain technology is increasingly being used for healthcare programs, addressing unresolved research concerns and presenting novel developments that might have revolutionary implications.. [5]

III. DESIGN SYSTEM

The proposed approach is based on blockchain technology and predicts insurance fraud using predictive machine learning algorithms. The system works by collecting data from insurance claims, which is then preprocessed and used to train the machine learning algorithms. The trained algorithms are then used to classify new claims as either fraudulent or legitimate. The system also includes a blockchain layer, which is used to store the insurance claims data and the predictions made by the machine learning algorithms. The blockchain technology layer makes sure the data is safe and irreversible while it is being stored.. Additionally, it provides a transparent and traceable record of all transactions, making it auditable.

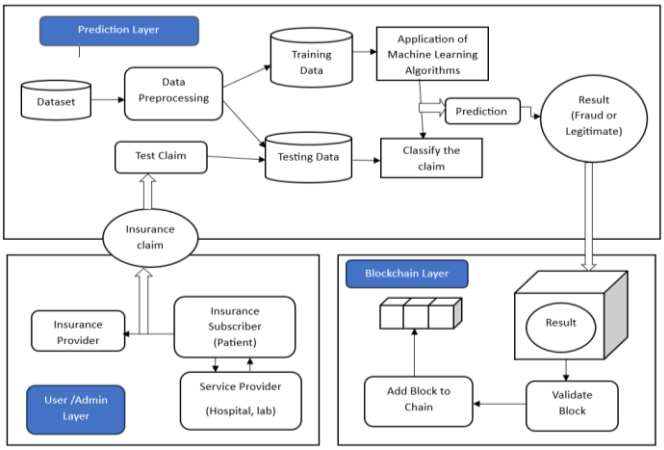

Figure 3.1. System Architecture

The system works by first collecting data from users in the user layer. This information may contain specifics on the insurance claim, such as the kind of procedure, its price, and the patient's medical background.. The data is then prepared for analysis in the prediction layer. Here, the data is formatted and cleaned to ensure consistency and accuracy. After the data is preprocessed, it is fed into a machine-learning model. This model has been previously trained on a large dataset of historical insurance claims, which are labelled as either fraudulent or legitimate. Through the examination of this past data, the machine learning model can spot trends that set false claims apart from real ones.. These patterns can include things like inconsistencies in the claim data, a history of fraudulent claims from the same policyholder, or procedures that are statistically more likely to be fraudulent.

Once the machine learning model has analyzed the claim data, it generates a prediction for the new claim. The prediction is either ‘fraudulent’ or ‘legitimate’. This prediction is then stored on a blockchain, along with the claim data, in the blockchain layer. A blockchain is a distributed, trustworthy ledger that many parties may access. Each entry on the blockchain is verified and cryptographically linked to the entries before and after it, making it tamper-proof. In this system, authorized users, such as insurance providers, hospitals and labs (service providers) can all access and validate the claim data and the predicted outcome stored on the blockchain.

A. Software/Hardware Specification Requirements

Software Requirement

- Operating System: Microsoft Windows 7 and Above
- Programming Language: Java
- IDE: Netbeans

Hardware Requirement

- RAM - 8 GB or Higher
- HDD - 100 GB (min3. Processor – Intel Core I3 or Higher)

B. Working

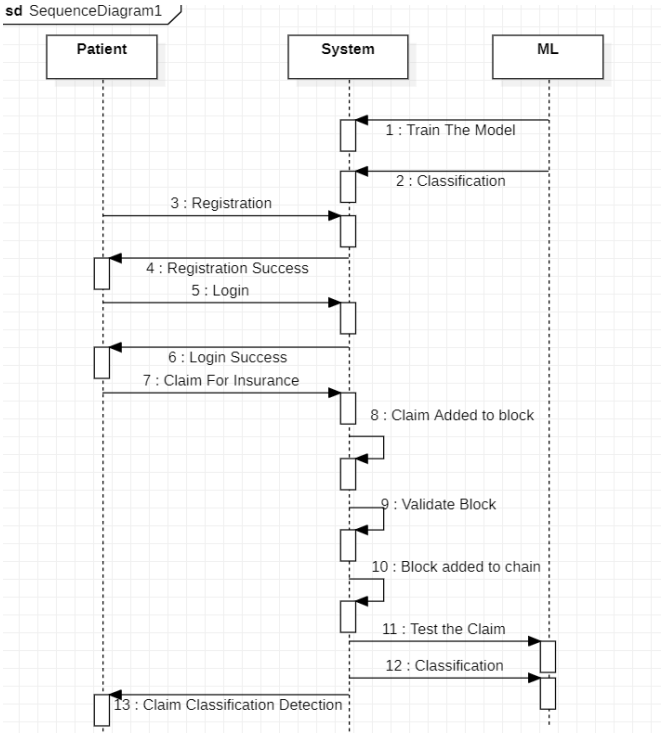

Figure 3.2. Sequence Diagram

1. User Layer

The user layer is where a user, which can be an insurance provider, administrator, or healthcare provider enters the claim data. This data could include information about the patient, the service provided, and the cost of the service. Once the data is entered, it is then sent to the prediction layer for analysis.

2. Prediction Layer

The prediction layer is where the system uses machine learning algorithms to determine whether a claim is fraudulent or legitimate. two algorithms, SVM and J48, are being used. These algorithms are trained on a large dataset of historical insurance claims, which have been labelled as fraudulent or legitimate. When a new claim is submitted, the algorithms compare the new claim data to the data in the training set and generate a prediction about whether the new claim is fraudulent.

The data is clustered by disease before being fed into the machine-learning models. This means that the system may be able to identify patterns of fraud that are specific to certain diseases.

3. Blockchain Layer

The prediction layer's results are kept in the blockchain layer. The claim data, the machine learning model's forecast, and the ultimate determination of whether the claim is fraudulent or valid are all kept on the blockchain in the system depicted in the figure. The proof of work consensus method and the SHA-256 hashing algorithm are used by the blockchain. A cryptographic hash function called SHA-256 is used to generate a block of data's unique identification. Verification of transactions on a blockchain is accomplished by a consensus technique called proof of work.

C. Algorithms

Support Vector Machine(SVM)

The Support Vector Machine(SVM) aims to classify medical insurance claims as either legitimate or not, utilizing machine learning techniques. Initially, it parses input data such as patient age, gender, medical bill, and claim amount. It imports a dataset from a CSV file and separates it into sets for testing and training

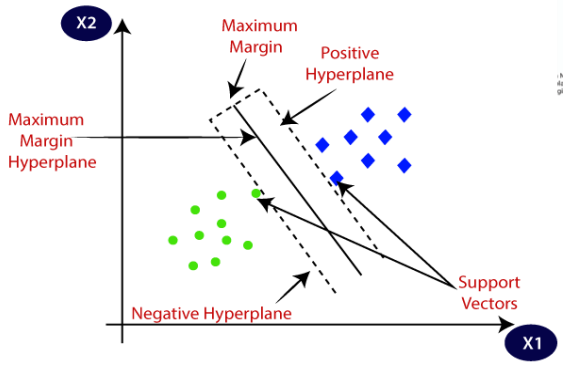

Figure 3.3. SVM

.. The Support Vector Machine (SVM) algorithm is then applied to build a classifier. An evaluation is conducted to assess the classifier's performance. Subsequently, a new instance is created based on user input and classified using the trained model. Depending on the classification result, a status variable is set to indicate the legitimacy of the claim. Finally, the details of the transaction, including the classification result, are stored in a database. If a claim is deemed illegitimate, a message is displayed to notify the user. This system helps in automating the process of identifying potentially fraudulent insurance claims.

SHA-256

The SHA256 algorithm is a condensed version of the blockchain consensus and structure. A blockchain stores data in blocks, each of which is comprised of a cryptographic hash of the one before it, forming a chain of blocks. A single blockchain block is represented by the Block class. It includes parameters like data, timestamp, nonce, hash of the current block, and hash of the prior block. The SHA-256 hash of an input is determined using the applySha256 method. By combining the data from the current block with the hash, timestamp, and nonce of the previous block, the calculateHash method calculates the hash of the current block.

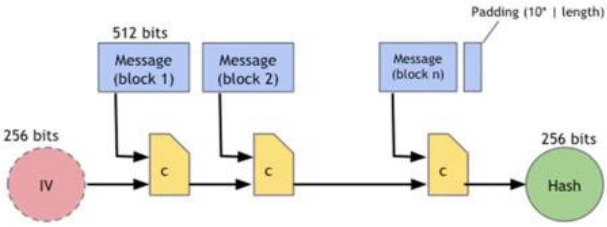

Figure 3.4. SHA-256

The SHA256 algorithm is a condensed version of the blockchain consensus and structure. A blockchain stores data in blocks, each of which is comprised of a cryptographic hash of the one before it, forming a chain of blocks. A single blockchain block is represented by the Block class. It includes parameters like data, timestamp, nonce, hash of the current block, and hash of the prior block. The SHA-256 hash of an input is determined using the applySha256 method. By combining the data from the current block with the hash, timestamp, and nonce of the previous block, the calculateHash method calculates the hash of the current block. In order to add a layer of security to the blockchain, the mineBlock1 method uses a proof-of-work algorithm to discover a nonce value that produces a hash with a particular amount of leading zeros. Reaching consensus amongst several blockchain systems is the main goal of the ChainConsensus class. It gathers the corresponding block validations for each blockchain system as it iterates over them. Subsequently, it ascertains the majority validation to guarantee consensus amongst the systems.

In the ProofOfWork method, the code iterates through four blockchain systems, mining blocks and collecting their validations. Afterwards, it determines the majority validation and adjusts the blockchain accordingly, ensuring consistency among the systems. Additionally, it handles data recovery in case of discrepancies among the blockchains. Finally, it inserts new transactions into each blockchain system.

IV. RESULT AND SUMMARY

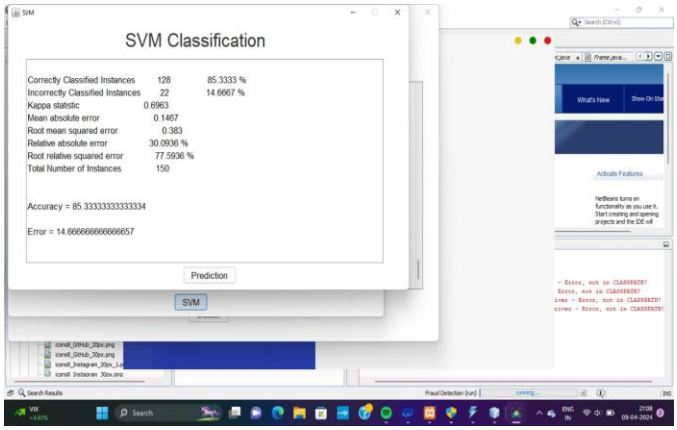

Fig 4.1. SVM Classification Result

The SVM algorithm used here seems to be quite effective at distinguishing between legitimate and fraudulent insurance claims. It correctly classified 85.33% of instances, meaning it accurately identified whether a transaction was genuine or suspicious most of the time. Although it made mistakes in about 14.67% of cases, its overall performance is solid. The Kappa Statistic, which measures agreement with the actual labels, indicates a good level of consistency beyond what would be expected by chance alone. Additionally, the model's predictions were generally near to the actual labels, as indicated by the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics.. With a dataset of 150 instances, the SVM demonstrated its capability in analyzing insurance claims, providing valuable insights for fraud detection efforts.

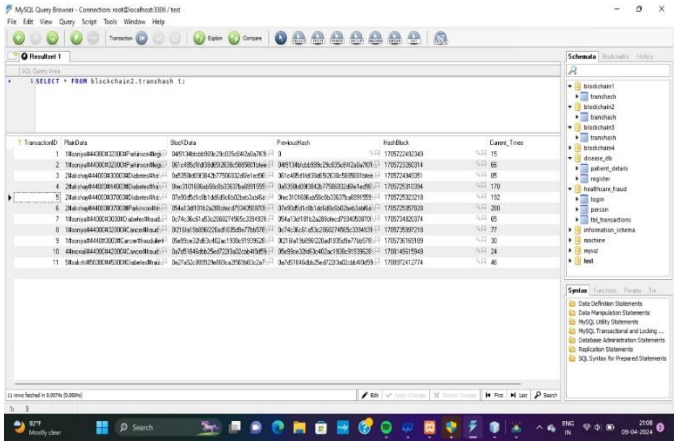

Fig 4.2. Transaction Frame

The screenshot depicts a MySQL Query Browser window showing transactions, likely related to insurance claims, with details such as TransactionID, Sender Name, Claim Amount, Disease, and Additional Data. These entries suggest health insurance claims, but some values like "fraud" raise questions. The inclusion of BlockData, PreviousHash, and HashBlock indicates blockchain integration for data integrity and security. While the sender names appear to be individual claimants, the exact context and system conventions remain unclear. This glimpse into the database showcases a structured approach to managing insurance transactions, leveraging blockchain for added trustworthiness. However, without a comprehensive understanding of the entire schema and system-specific conventions, it's challenging to draw definitive conclusions about the data's real-world significance or accuracy.

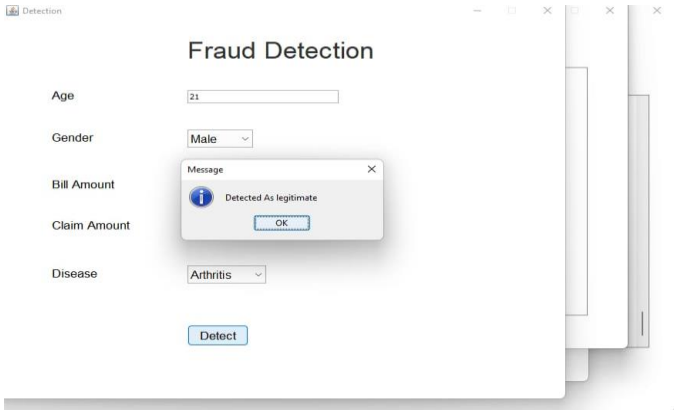

Fig. 4.3. Prediction frame

The image likely shows a fraud detection system's interface for assessing an insurance claim. It displays key details such as the bill amount, claim amount, disease (Arthritis), and the policyholder's age. The system's SVM model predicts the claim as legitimate despite the claim amount being lower than the bill amount. This indicates that the system considers various factors beyond just the amounts involved. It might use other data like medical conditions and policyholder demographics to make its decision. The interface probably offers options for users to review or override the model's prediction. Overall, it showcases how technology, like machine learning algorithms, is used to analyze insurance claims and detect potential fraud, helping insurance companies make informed decisions.

V. CONCLUSION

Ultimately, it should be mentioned that protecting the rights of insurers, customers, and providers as well as preserving the integrity of the healthcare system depends on the discovery of healthcare fraud insurance. Insurance companies may effectively identify and minimize fraudulent conduct in transactions and claims by utilizing advanced techniques such as data analysis tools, cryptography methods, and machine learning algorithms. This helps to minimize financial losses and ensure fair resource allocation while also upholding ethical norms, regulatory compliance, and the level of patient care. In the future, collaboration among industry participants and investments in fraud detection technology will be critical to the healthcare insurance sector's survival and resilience in the face of evolving fraudulent schemes.

VI. FUTURE SCOPE

One appealing approach to tackling the escalating issues facing the healthcare sector is research on blockchain and artificial intelligence in the identification of healthcare insurance fraud. By integrating AI algorithms and blockchain technology, researchers can develop advanced systems capable of identifying fraudulent activities efficiently and accurately. Additionally, incorporating IoT devices allows for the collection of real-time data, enhancing the system's ability to detect anomalies promptly. Big data analytics further strengthens the fraud detection process by enabling the processing and analysis of large volumes of healthcare data to uncover suspicious patterns and trends. Additionally, real-time analysis guarantees prompt fraud identification and action, reducing possible financial losses for insurance companies and enhancing the provision of healthcare services overall. Future research could explore novel techniques for enhancing the synergy between AI, blockchain, IoT, and big data analytics to develop even more robust and adaptive fraud detection systems, ultimately contributing to the sustainability and integrity of the healthcare insurance ecosystem.

ACKNOWLEDGEMENT

We would like to extend our heartfelt thanks to Dr M.A. Chaudhari from the Department of Information Technology at Amrutvahini College of Engineering in Sangamner, Ahmednagar, Maharashtra, for his crucial advice and assistance over the course of our project's growth on Health Insurance Fraud Detection using AI and Blockchain. We also want to acknowledge the collaborative efforts of our team members, whose dedication and contributions played a significant role in the success of this endeavour. We also acknowledge and express our sincere thanks to our institution for providing the required resources, in addition to our friends,

family, and colleagues for their unwavering support and encouragement along this journey. We couldn't have completed our job without their assistance.

REFERENCES:

1.Nabrawi, Eman & Alanazi, Abdullah. (2023). Fraud Detection in Healthcare Insurance Claims Using Machine Learning. Risks. 11. 160. 10.3390/risks11090160.
2. Al-Quayed, Fatima & Humayun, Mamoona & Tahir, Sidra. (2023). Towards a Secure Technology-Driven Architecture for Smart Health Insurance Systems: An Empirical Study. Healthcare. 11. 27. 10.3390/healthcare11162257.
3. Kapadiya, Khyati & Patel, Usha & Gupta, Rajesh & Alshehri, Mohammad & Tanwar, Sudeep & Sharma, Gulshan & Bokoro, Pitshou. (2022). Blockchain and AI-empowered Healthcare Insurance Fraud Detection: An Analysis, Architecture, and Future Prospects. IEEE Access. 10. 10.1109/ACCESS.2022.3194569.
4. Haque, A.K.M. & Bhushan, Bharat. (2023). Blockchain for medical insurance: Synthesizing current knowledge and problematizing it for future research avenues. 10.1016/B978-0-323-99199-5.00002-1.
5. Khezr, Seyednima & Moniruzzaman, Md & Yassine, Abdulsalam & Benlamri, Rachid. (2019). Blockchain Technology in Healthcare: A Comprehensive Review and Directions for Future Research. Applied Sciences. 9. 1736. 10.3390/app9091736.
6. Liu, Wei & Yu, Qinyong & Li, Zesong & Li, Zeyuan & Su, Yu & Zhou, Jian. (2019). A Blockchain-Based System for Anti-Fraud of Healthcare Insurance. 1264-1268. 10.1109/ICCC47050.2019.9064274.
7. Saeed, Huma & Malik, Hassaan & Bashir, Umair & Ahmad, Aiesha & Riaz, Shafia & Ilyas, Maheen & Bukhari, Wajahat & Khan, Muhammad. (2022). Blockchain technology in healthcare: A systematic review. PLOS ONE. 17. e0266462. 10.1371/journal.pone.0266462.
8. Dutt, Rajeev. (2020). The impact of artificial intelligence on healthcare insurances. 10.1016/B978-0-12-818438-7.00011-3.
9. Mary, A. & Claret, Angelin. (2021). Imbalanced Classification Problems: Systematic Study and Challenges in Healthcare Insurance Fraud Detection. 1049-1055. 10.1109/ICOEI51242.2021.9452828.
10. Yang, Wenyi & Hu, Wenhui & Liu, Yingjie & Huang, Yu & Liu, Xueyang & Zhang, Shikun. (2021). Research on Bootstrapping Algorithm for Health Insurance Data Fraud Detection Based on Decision Tree. 57-62. 10.1109/BigDataSecurityHPSCIDS52275.2021.00021.