**Insurance Fraud Detection using Blockchain and ML**

Arul Arasu N (Roll No: 20Z306)

Dhanaseelan V (Roll No: 20Z313)

Dinesh Baabu R (Roll No: 20Z315)

Lokajit G (Roll No: 20Z328)

Sudarshan S (Roll No: 20Z350)

Dissertation submitted in partial fulfillment of the requirements for the degree of

**BACHELOR OF ENGINEERING**

**Branch: COMPUTER SCIENCE AND ENGINEERING**



OCTOBER 2023

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PSG COLLEGE OF TECHNOLOGY

(Autonomous Institution)

COIMBATORE – 641 004

# 

**CHAPTER Page No.**

1. [**ABSTRACT 2**](#_u1eso6f97g2b)
2. [**INTRODUCTION 3**](#_xqi2v4yxtle9)
3. **PROBLEM STATEMENT** [**4**](#_mfr3d8uh4m8x)
4. **LITERATURE REVIEW………………………………………………………………...9**
5. [**SYSTEM REQUIREMENTS**](#_do2khhlbry2f) **10**

5[.1 Software Requirements](#_xzcxmu2vh8uz) 10

5[.2 Hardware Requirements](#_62effme1l5tr) **10**

1. **SYSTEM ARCHITECTURE****11**
2. [**TIMELINE**](#_e4895270d9tf) **12**
3. [**BIBLIOGRAPHY**](#_p34d93ai7wmo) **13**

# 

# 

# CHAPTER 1

# ABSTRACT

The widespread nature of fraudulent activities in today’s insurance landscape has led to the search for alternative solutions. Our proposal involves blending state-of-the-art blockchain technology with sophisticated machine learning techniques to effectively address this issue. The 2015 comprehensive report highlighted the enormous loss from healthcare fraud of £303.8 million, stretching across the alarming spectrum of £3.73 to £5.74 billion when considering potentially hidden losses. Our strategic approach component of complex historical feedback data analysis, using advanced machine learning algorithms . These algorithms exhibit adaptive capabilities, gradually improving their ability to detect incorrect patterns from past cases of fraud and consequently a powerful tool for early fraud detection, and retention an incredible investment in time and resources. The integrated blockchain further supports our system by establishing a secure repository of virtual insurance transactions. This builds trust by establishing clear actions. Specifically, our ambition is to change the landscape to combat insurance fraud, influence operational efficiency, inspire trust and achieve security. These efforts accelerate claims, reduce financial bleeding, foster a culture of integrity, together serve the interests of insurers and policyholders, and absorb focus is more on the unmanageable future

# 

# CHAPTER 2

# INTRODUCTION

# Blockchain is a groundbreaking distributed ledger technology that facilitates the secure, transparent, and tamper-resistant recording of transactions and information across a network of computers. Operating in a decentralized manner, it eliminates the need for intermediaries and central authorities, fostering trust among participants. Originally conceptualized for cryptocurrencies like Bitcoin, blockchain's applications have expanded across diverse industries. Its potential benefits include enhanced efficiency, improved supply chain management, increased data security, and streamlined processes in finance, healthcare, logistics, and more. By offering a novel approach to data management, blockchain promises to revolutionize how we exchange value and conduct business in the digital age. Healthcare plays a vital role in society, offering essential medical services and treatments to improve people's well-being. However, the healthcare industry is not immune to the pervasive issue of healthcare fraud. This type of fraud involves dishonest practices aimed at obtaining unauthorized benefits or financial gains at the expense of patients, insurers, or government healthcare programs. Examples include false billing, upcoding, kickbacks, and identity theft.

Healthcare fraud poses significant challenges and consequences for the healthcare system and its stakeholders. It leads to inflated costs, reduced access to care, compromised patient safety, and diminished trust in healthcare providers. To combat healthcare fraud, various measures are employed, such as stringent regulations, advanced technology for fraud detection, and whistleblower programs that encourage individuals to report fraudulent activities confidentially.

Addressing healthcare fraud is a multifaceted endeavor that requires collaboration and vigilance from all stakeholders. By implementing robust fraud detection systems, fostering a culture of ethical conduct within the healthcare industry, and encouraging individuals to report suspicious activities, we can strive towards a more secure and trustworthy healthcare system. Moreover, investing in cutting-edge technologies and data analytics will continue to be instrumental in identifying patterns indicative of fraud, enabling swift action against wrongdoers. By effectively combating healthcare fraud, we not only protect patients, insurers, and government programs from financial losses but also uphold the integrity of the entire healthcare system.

# CHAPTER 3

# PROBLEM STATEMENT

In the insurance industry, the rise of fraudulent activities has become a significant concern, leading to financial losses for both insurers and legitimate policyholders. The main objective of this project is to develop an advanced system that utilizes a Machine Learning algorithm to effectively detect fake insurance claims while ensuring the secure storage of genuine transactions in a blockchain.

# CHAPTER 4

# LITERATURE REVIEW

[1] **“** A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology Anokye Acheampong Amponsah ∗ , Adebayo Felix Adekoya, Benjamin Asubam Weyori “ Link - [https://pdf.sciencedirectassets.com/780449](https://pdf.sciencedirectassets.com/780449/1-s2.0-S2772662222X00039/1-s2.0-S2772662222000534/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEO3%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJHMEUCIQD2t9hMp4GuLHzXhzwvRHBPdbF97%2Fbalhq4bDpRQzZ%2BhgIgHe0so4z7Py%2FvnDAxx3R7vpCpqwoufNNcRvCf%2FeObE7UqsgUIdhAFGgwwNTkwMDM1NDY4NjUiDC7loWIiGXHPJNNAaiqPBbh9Zd4PwezdbIKJe%2Fdt8EVwehjrw6OgWs9QspQ1kuvOFcRajdGTJxqrDm%2FoLmnltcLHWWzBKv2ssp%2FcOKeA2p%2FTmRcAkZmNXLi0gg3YS4DOViK1bpKT8Ge7ly96rbap0bCP5amXo6l3wVcZdZRUodGnopr07%2F016kQQCQCEPg9iLS6xoBueiQA9W64JQbY4J0gWSA9RmIHoyMtFrVLIf0hnTBHu41nHiJK5Hrptoeg54Gahmhh%2FUd%2FbmJZZkbMQ4XjQRVIFn3%2B3IFLpgyaTHZ9UlLtFWtQjSwqnzJW95RMonsjB8ymwI8zPDgs5i6vLhoinxAgKObRy%2B2h%2FI9tcG4MBtgsstjocb9JiBZI5T7b5LxdDNzIT7lwgYXVa9t4zqWRtG6tTOr5v0%2FMkFe5jtt%2BpnUfjlSptz7d3hK7Wn4rGeG%2FcOws4PCqVAxCMxKKSW%2F%2FNHauUemon1MaEA2WtH6uslZJOrHoBhR6UZHApNr40xIhHp0eockBMxup30noksfPuFClhNYdcHN8DUIZj6K9PFYyXFCZUv1jMF7OWeF8BXvEGYHc0nYSOqr8KK79WlrL2STaq%2FB553oL1fV8tuZOnbT5xQ5CQK%2F4QHnD%2BIx2fEoYq5wD9rO1TCTdu3l5kPjMGAMLzJBxSGUvaZNq1H0OFSYLcseUjFV3abpNpdh%2Fjl6boUcTfI23WGiccL2DgzFyUMdfIk4PxEwxYp5Yq5xrNH%2BgY9IQEIKt3vHLBHW4%2Fn2toJyprXiBo46G04YRq5Uc96qcWYtCMmbfbNg%2FfzwXfRGNGXEP2GiUTTUVVJ71L1IUtUP%2F4hFW%2FqgXwE51GrMTJ7QOzJqCDkkkU8jLSS2j%2Fwz1mfp5biAvJ59PYkdIwmprapQY6sQFKlEUg%2Bfgqtqxrkf0TC%2BslEN%2F0Mr0ByBQhxtjdLMWyaLjhxMDUFC9pePtz%2BuKBqh0uiWbzU43w5OcPHiAzPAK84IfnnW7lLEgDof4j7CPp1t6TvNNsT7ry9rHivElKswDg6RiSXvb8IdruUd37KHMdpzdNMiQM9ylnudTELDQxdwduWh1AUANrESUkDbByRdsZbsOKs0As5zAnq8%2BeJo9UKjD1wyOX9%2BPr6znulpEHWCY%3D&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20230718T135702Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTYS5VTQZNH%2F20230718%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=5091478f0af8ddbd4817716d6d504b3b0e157028f761221d8e0704579e4e8c85&hash=afdd8dc43a1b80e659bf822f4c7d9cf65552626218e81c321ca0b9cd25354881&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S2772662222000534&tid=spdf-1e172e8f-ff61-41b3-9599-1cc42f6e972f&sid=661779612d3cd3445b3b72276e7d88d6ca35gxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=0c08520a5c0052075552&rr=7e8b3b40fb9206a8&cc=in)

This research paper provides a detailed explanation of a new approach to healthcare fraud detection and prevention using machine learning and blockchain technology. The authors explain that healthcare fraud is a serious problem in many countries, causing significant economic losses to insurance companies and government agencies The proposed method uses decision tree classification algorithms to analyze healthcare information and they are used to identify potentially misleading information and to include fraudulent information.

The authors also discuss the potential benefits of using blockchain technology in healthcare claims. They explain that blockchain provides a secure and transparent way to store and share health information, which can help prevent fraud and ensure that the rightful beneficiaries receive fair compensation. The use of blockchain technology can help reduce administrative costs and improve health care coverage.

However, the authors also acknowledge that there are some challenges in applying this approach to real-world health care settings. For example, there may be concerns about data privacy and security, as well as resistance from healthcare providers and insurance companies accustomed to traditional ways of handling data The authors suggest that these challenges can be addressed through policy with care and collaboration between stakeholders.

[2] Sathya, M. ., & Balakumar , B. . (2022). Insurance Fraud Detection Using Novel Machine Learning Technique. International Journal of Intelligent Systems and Applications in Engineering, 10(3), 374–381. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/2178>

This paper discusses the problem of fraudulent claims in the insurance industry and the limitations of traditional fraud detection techniques. The paper proposes a novel approach for detecting insurance fraud using a hybrid machine learning classifier called eRFSVM, which combines Random Forest and Support Vector Machine algorithms. The proposed approach also utilizes block chain technology for secure information sharing among insurance agencies.

The paper evaluates the proposed approach using a confusion matrix and various classification metrics. The results show that the eRFSVM classifier outperforms traditional fraud detection techniques in terms of accuracy and efficiency. The proposed approach shows remarkable performance in evaluating the authenticity of customer claims, with an exceptional accuracy of 97.176%. Additionally, the value of specificity and sensitivity is 96.158% and 96.634% respectively.

The paper highlights the potential of machine learning techniques for fraud detection in the insurance industry and the importance of secure information sharing among insurance agencies. The proposed approach can help insurance companies to detect fraudulent activities effectively and reduce unwarranted expenses. The paper also discusses of the proposed approach for detecting insurance fraud include the requirement for a large amount of data to train the machine learning model effectively, reliance on the accuracy of the data provided by insurance agencies, potential difficulty in detecting new types of fraud, and the need for significant computational resources to process large amounts of data.

[3]N. R. Bhamidipati et al., "ClaimChain: Secure Blockchain Platform for Handling Insurance Claims Processing," 2021 IEEE International Conference on Blockchain (Blockchain), Melbourne, Australia, 2021, pp. 55-64, doi: 10.1109/Blockchain53845.2021.00019. <https://par.nsf.gov/servlets/purl/10315023>

This paper discusses the benefits of using blockchain technology for insurance claims processing and introduces ClaimChain, a secure blockchain platform designed to replace traditional NICB/ISO database architecture used in the auto-insurance industry. The paper is organized into several sections, including a discussion of related works, an overview of the ClaimChain approach, and a description of the platform's security features.

The authors explain how blockchain technology offers transparency and auditability, enabling distributed trust among participating peers, and how smart contracts can reduce operation and maintenance costs while improving processing time. They also describe how ClaimChain uses machine learning and NICB-identified red flags to detect and prevent fraud, and how the platform's threat modeling approach helps to improve security at both the infrastructure and application levels.

The paper includes a detailed analysis of a dataset of insurance claims hosted on the ClaimChain testbed, which reveals that the majority of fraudulent claims identified have no police report or witnesses. The authors conclude by highlighting the potential benefits of ClaimChain for the insurance industry and outlining future research directions. Overall, the PDF file provides a comprehensive overview of the benefits and challenges of using blockchain technology for insurance claims processing and introduces a promising new platform for improving the efficiency and security of this process.

[4] R. Roy and K. T. George, "Detecting insurance claims fraud using machine learning techniques," 2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT), Kollam, India, 2017, pp. 1-6, doi: 10.1109/ICCPCT.2017.8074258. <https://ieeexplore.ieee.org/document/8074258>

This paper focuses on detecting auto/vehicle insurance fraud using machine learning techniques. The authors focus on creating a set of rules and anomalies for creating raw data, which is dependent on a set of attributes. They then compare the performance of decision trees, random forests, and Naïve Bayes in detecting insurance fraud using a confusion matrix. The methodology adopted involves dividing the data into training and testing sets, and comparing the accuracy, precision, and recall of each method. The authors found that decision trees and random forests outperformed Naïve Bayes in detecting insurance fraud. The evaluation of the methodology involved comparing the performance of each method using a confusion matrix, and calculating accuracy, precision, and recall.

[5] Urunkar, Abhijeet & Khot, Amruta & Bhat, Rashmi & Mudegol, Nandini. (2022). Fraud Detection and Analysis for Insurance Claims using Machine Learning. 406-411. 10.1109/SPICES52834.2022.9774071.

<https://www.researchgate.net/publication/360646224_Fraud_Detection_and_Analysis_for_Insurance_Claim_using_Machine_Learning>

The traditional method for detecting fraud relies on manual intervention and heuristics, which can be limited in their ability to detect complex fraud schemes. The authors propose a machine learning approach that can analyze large amounts of data and identify patterns indicative of fraud. They use a combination of supervised and unsupervised learning algorithms, including decision trees, random forests, and clustering. The authors evaluate their approach using a dataset of insurance claims and compare their results to those of traditional fraud detection methods. They find that their machine learning approach outperforms traditional methods in terms of accuracy and efficiency. However, they also note that there are challenges to implementing these techniques in real-world insurance settings, such as data privacy concerns and the need for ongoing model maintenance.The authors suggest that it may not be practical to define optimum algorithmic techniques or use feature engineering processes for higher performance due to the inherent characteristics of various datasets. Instead, they propose that the models be used for specific business contexts and user priorities, allowing loss management units to focus on new fraud situations and ensure that the models are adapting to detect the

[6] Inayatulloh, Siti Elda Hiererra, Prasetya Cahaya S, Rozil Toyob, Nico Djundharto Djajasinga , Sawqi Saad El Hasan, Rofiq Noorman Haryadi, Rivaldhy N. Muhammad,“ Blockchain technology of fraud Detection and Risk Prevention in Insurance Industry “. Proceedings of the 3rd South American International Industrial Engineering and Operations Management Conference, Asuncion, Paraguay, July 19-21, 2022.

LINK : [47.docx (googleusercontent.com)](https://doc-10-8g-apps-viewer.googleusercontent.com/viewer/secure/pdf/77gfp996nr6b4qqh2qgl1g8e4t0lv9ej/2hvusaf9arr9domf1324i9sh39rgtnn0/1692286725000/gmail/06715741469926710199/ACFrOgBrcrsOqvV5LN7lBDorohY40T4RHb6Pmfi68FHikCPu2-EMrPf6mT8SWdxhBckrhWlX9brmAXk1f64M_NUAoHF405fIy0X4JjtAQK7WYAurqidDhI5O2iu2SBw=?print=true&nonce=2d7vgp0r0fbdq&user=06715741469926710199&hash=kupj6g9lt95ouaek0nsggtu7sgf38bmt)

Implementing Insurance Fraud Claim that customers register the insurance as a smart contract between customer and insurance participant that is stored as a new block/ledger broadcasting into the blockchain network . Validating process will be taken by each participant/node in the blockchain network which takes the advantage for difficulty in manipulation . After the validation from all blockchains network a new block will be part of the blockchain network and submission of the customer as an Insurance Company Customer is completed , the customer can access the approval via Smartphone or PC . The Submission claim from customer creates a new block as form of a ledger and broadcasted to the blockchain network . Each participant/node in the blockchain network will validate the claim submission from the customer and the process of potential fraud when submitting a claim will be lost. After the successful validation process the customer gets the result via Smartphone or PC finding out the submitted claims aren’t fake that revels in their application.

[7] Rui Roriz, Jose Luis Pereria “ Avoiding Insurance Fraud : A Blockchain-based Solution for the Vehicle Sector “ <https://doi.org/10.1016/j.procs.2019.12.174>

In the vehicle sector the blockchain network was implemented with developer tools that were used according to the vehicle insurance claim and policy management for avoiding insurance fraud . The tools used to implement the plan are Solidity, Node.js, Truffle Framework, Ganache and MetaMask. Further, it uses

[8] Sun C., Li Q., Li H., Shi Y., Zhang S., Guo W., “Patient Cluster Divergence Based Healthcare Insurance Fraudster Detection”, IEEE Access, Vol. 7, pp. 14162–14170, 2019. <https://ieeexplore.ieee.org/document/8576507>

# Sun et al presented a novel approach for detecting frauds, called Patient Cluster Divergence-based Healthcare Insurance Fraudster Detection (PCDHIFD) in presence of camouflage responses. For the experimental purpose, the health care dataset was chosen and the dataset consisted of around 40M admission records of 10000 patients of the previous five years. The proposed technique worked in 3 steps for three basic records: Life history of patients, diagnosis record, and medical practitioners attended. Steps were in this sequence: first of all, a patient graph was constructed based on most similar info for the patient level hospital admission. Then a clustering-based graph algorithm was used for finding the peak and real meaning for individual clusters. Lastly, the difference in the patient cluster was found and the probability of fraud for each patient was calculated. The comparison was made with other state of the art algorithms i.e. Decision

# CHAPTER 5

# SYSTEM REQUIREMENTS

## 5.1 Software Requirements

**Blockchain Platform:** Choose a suitable blockchain platform for implementing your system. Eg: Ethereum

**Smart Contracts:** Develop smart contracts that handle the validation of insurance claims and their addition to the blockchain. These contracts should include logic for verifying claim details and authenticity.

**Machine Learning Models:** Integrate machine learning models for fraud detection. These models can analyze historical data to identify patterns of fraudulent claims.

## 5.2 Hardware Requirements

**Server Infrastructure:** Depending on the scale of your system, you might need server infrastructure to host the blockchain nodes, the DApp backend, and any other supporting services.

**Cloud Services:** Consider using cloud platforms like AWS, Azure, or Google Cloud for scalability, reliability, and ease of management.

**Security Measures:** Implement hardware security modules (HSMs) to enhance the security of private keys and cryptographic operations.

**Data Storage:** Set up storage solutions for storing blockchain data, historical data, and backups.

**Networking:** Ensure a reliable and fast internet connection to support the communication between nodes and users

# 

# CHAPTER 6

# SYSTEM ARCHITECTURE

**CHAPTER 7**

# TIMELINE

1. Problem Definition - Jul 27, 2023
2. Data Collection - Aug 16, 2023
3. Data Preprocessing - Aug 30, 2023
4. Model Selection and Training - Sep 20, 2023
5. Model Evaluation and Testing - Nov 5, 2023

# CHAPTER 8

# BIBLIOGRAPHY

1. **“** A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology Anokye Acheampong Amponsah ∗ , Adebayo Felix Adekoya, Benjamin Asubam Weyori **“**
2. Sathya, M. ., & Balakumar , B. . (2022). Insurance Fraud Detection Using Novel Machine Learning Technique. *International Journal of Intelligent Systems and Applications in Engineering*, *10*(3), 374–381. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2178
3. “ Naga Ramya Bhamidipati\*, Varsha Vakkavanthula\*, George Stafford\*, Masrik Dahir\*, Roshan Neupane\*, Ernest Bonnah, Songjie Wang, J. V. R. Murthy, Khaza Anuarul Hoque, Prasad Calyam University of Missouri-Columbia, USA; Jawaharlal Nehru Technological University, India. “
4. R. Roy and K. T. George, "Detecting insurance claims fraud using machine learning techniques," 2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT), Kollam, India, 2017, pp. 1-6, doi: 10.1109/ICCPCT.2017.8074258.
5. Urunkar, Abhijeet & Khot, Amruta & Bhat, Rashmi & Mudegol, Nandini. (2022). Fraud Detection and Analysis for Insurance Claims using Machine Learning. 406-411. 10.1109/SPICES52834.2022.9774071.
6. Sun C., Li Q., Li H., Shi Y., Zhang S., Guo W., “Patient Cluster Divergence Based Healthcare Insurance Fraudster Detection”, IEEE Access, Vol. 7, pp. 14162–14170, 2019. <https://ieeexplore.ieee.org/document/8576507>
7. <https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis>
8. <https://www.kaggle.com/datasets/arashnic/imbalanced-data-practice>
9. [https://labelyourdata.com/articles/data-collection-methods-AI#:~:text=Simply%20 put%2C%20data%20collection%20 is,fed%20into%20an%20 ML%20mod](https://labelyourdata.com/articles/data-collection-methods-AI#:~:text=Simply%20put%2C%20data%20collection%20is,fed%20into%20an%20ML%20model)

# - - Additionals - - -

# 

# 

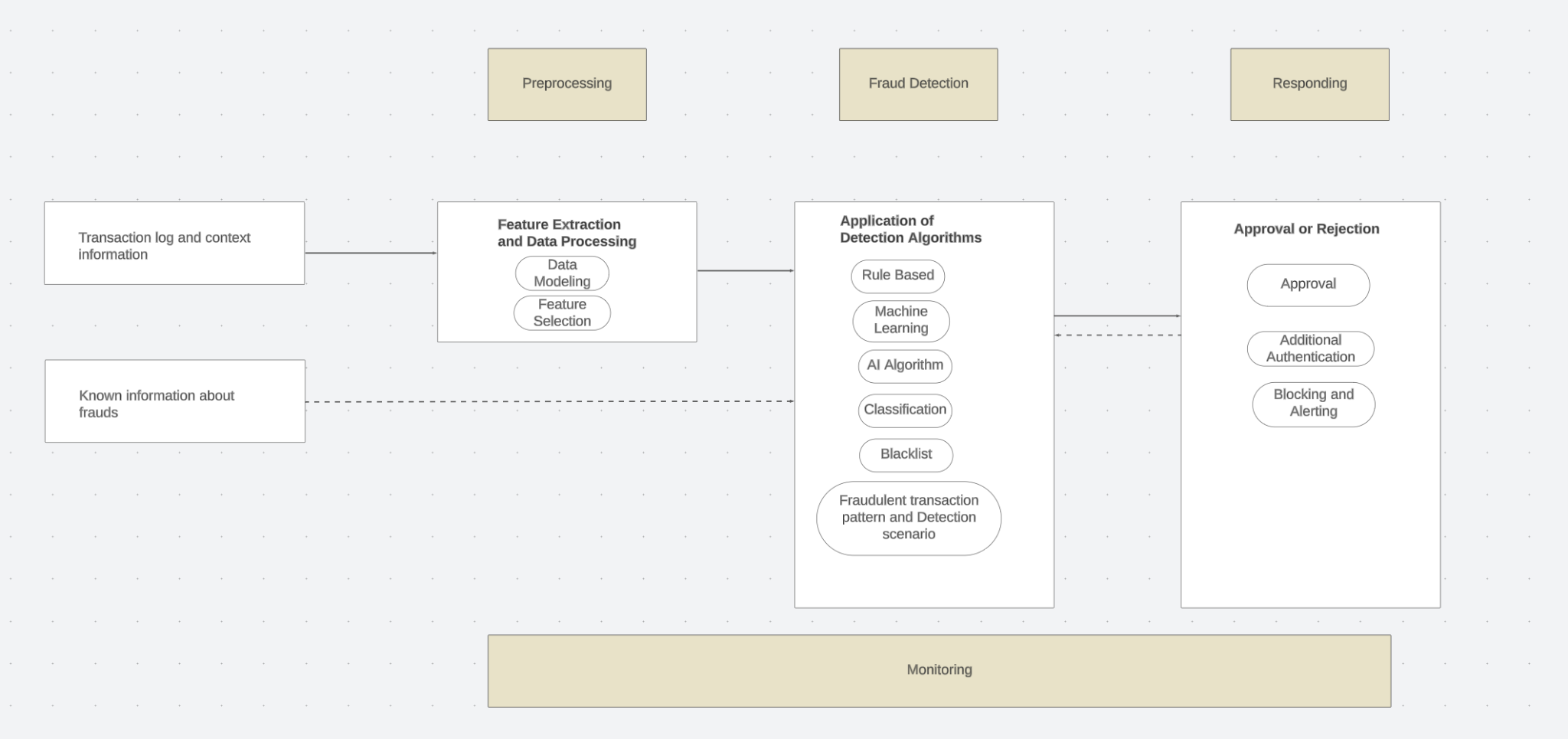
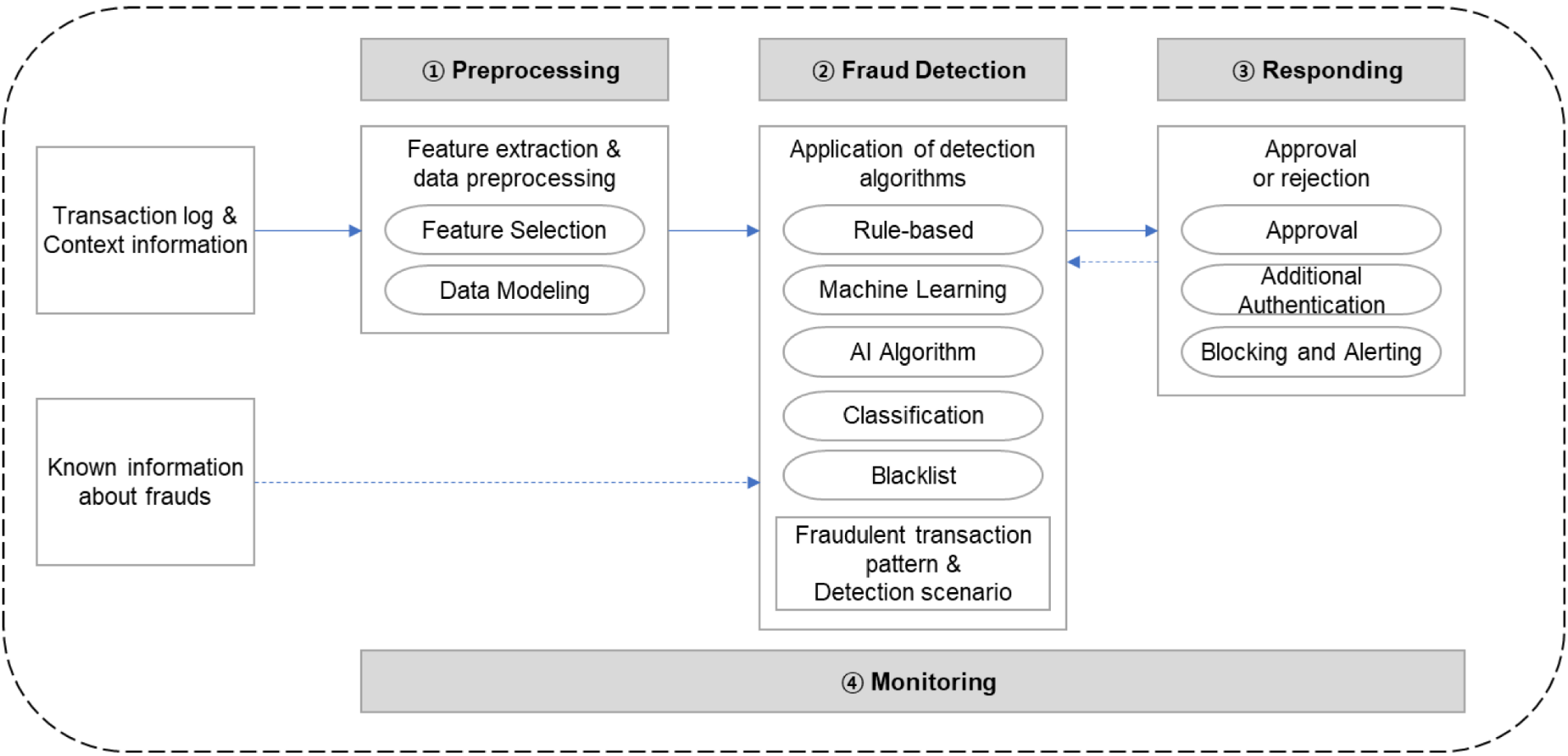
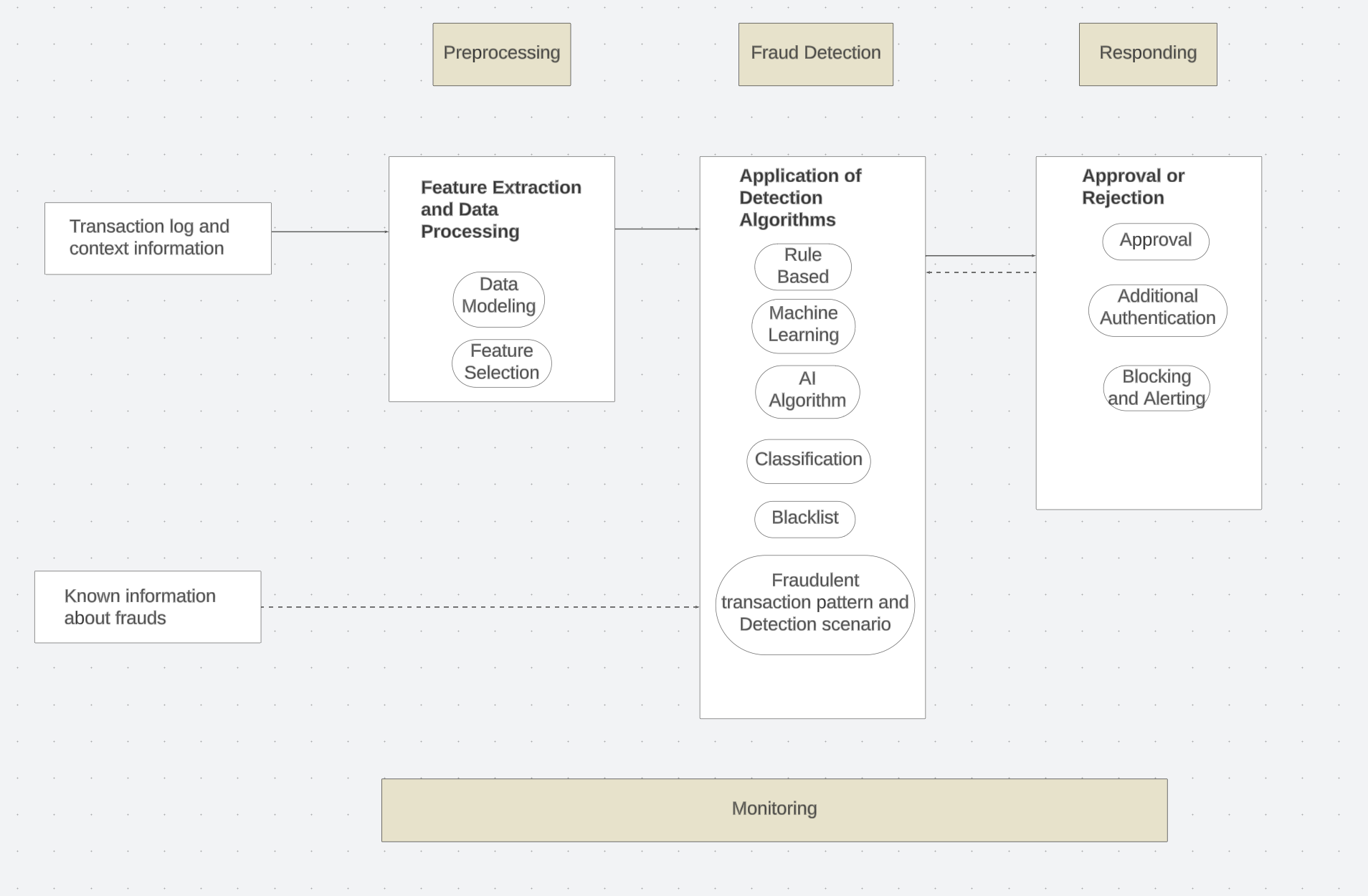
# Original abstract

In today's insurance industry, the rise of fraudulent activities is causing a major headache for insurers and regular policyholders alike. To tackle this problem head-on, our project aims to create an innovative and advanced system that combines the power of Machine Learning with the security of blockchain technology. According to the 2015 report on the financial cost of healthcare fraud a total of about £303.8 million was lost to healthcare fraud. This amount can be broken down into £237 million for prescription charge fraud, £43.9 million for dental charge fraud, and £22.9 million for optical charge fraud. These amounts exclude the losses incurred by expenditure and income which will inflate the figure to a range of £3.73 – 5.74 billion.

Our system will be using a Machine Learning algorithm to analyze a massive amount of historical claims data. By learning from past cases of fraud, it will become an expert at spotting suspicious patterns and subtle indicators that often go unnoticed by human investigators. This means we can detect fake insurance claims faster and more accurately, saving time and money for everyone involved.

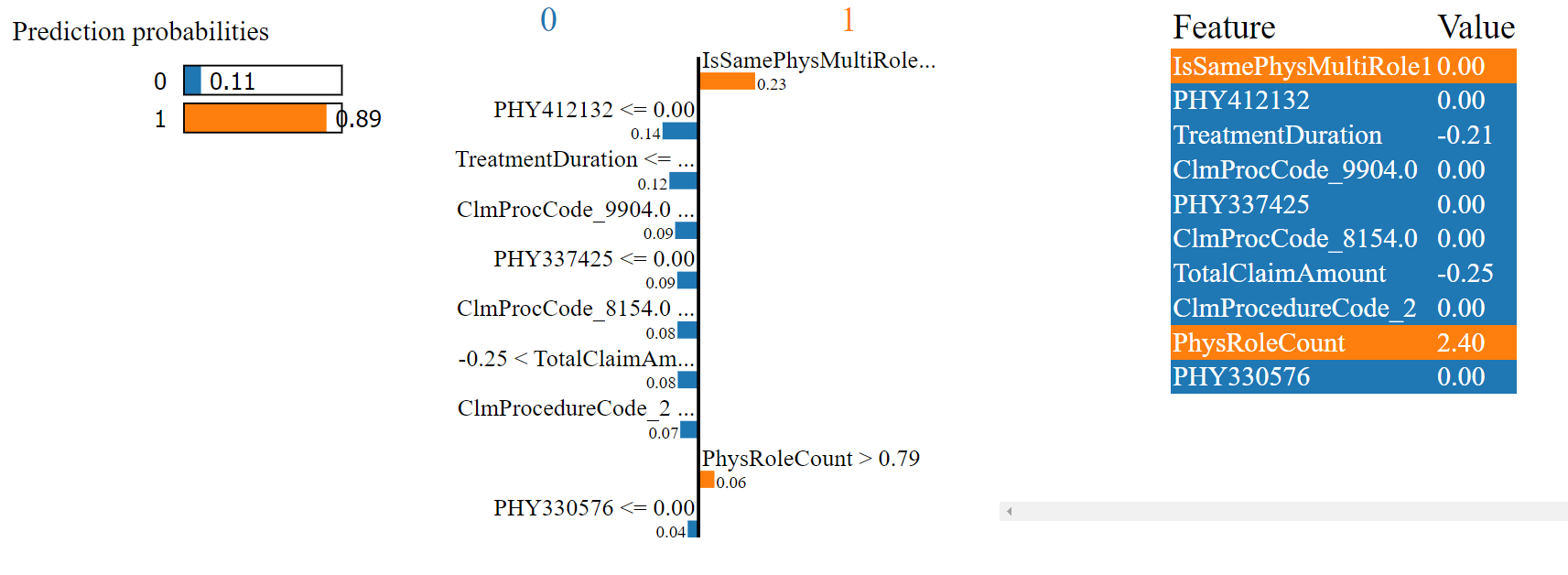
We're also taking data security to the next level by incorporating blockchain. This technology acts like an iron-clad vault, ensuring that all genuine insurance transactions are securely stored and impossible to tamper with. It's like having a transparent ledger accessible to all relevant parties, providing trust and confidence in the entire process.

Our goal is to revolutionize how the insurance industry fights fraud. By speeding up claims verification, reducing financial losses, and fostering trust, our system aims to create a safer and more efficient insurance ecosystem. So, rest assured, we've got your back – protecting both insurers and honest policyholders, while promoting integrity and credibility within the insurance world. Together, we're building a brighter, more secure future for everyone.



|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

Positive



Negative

