[1]Swapna Siddamsetti

[2]Chirandas Tejaswi

[3]Pallavi Maddula

Anomaly Detection in Blockchain Using Machine Learning

Journal of Electrical Systems

Abstract: - Blockchain technology has gained significant attention as a secure and decentralized platform for various applications. However, the immutable and distributed nature of blockchain also presents unique challenges for detecting anomalies and suspicious activities within the network. This research paper proposes a novel approach to anomaly detection in blockchain using machine learning techniques.

The goal of this study is to develop an effective and scalable anomaly detection framework that can analyze the vast amount of data generated within a blockchain network and identify irregularities or potential security threats. The proposed framework leverages the power of machine learning algorithms to learn patterns, relationships, and behaviours from historical blockchain data, enabling the detection of anomalous activities in real time.The research paper first focuses on feature extraction techniques tailored specifically for blockchain data. These techniques consider key characteristics of blockchain transactions, such as transaction size, timestamp, and involved addresses, to construct meaningful features that capture the underlying patterns and trends. Various dimensionality reduction techniques are also explored to handle the high-dimensional nature of blockchain data.Subsequently, several machine learning algorithms, including clustering, classification, and anomaly detection methods, are employed to train models using the extracted features. The performance of different algorithms is evaluated using benchmark datasets and real-world blockchain data to assess their accuracy, precision, and recall in detecting anomalies. Additionally, the scalability of the proposed framework is investigated to ensure its effectiveness in large-scale blockchain networks.Furthermore, the research paper investigates the integration of domain-specific knowledge, such as known attack patterns and regulatory compliance rules, into the anomaly detection framework. This hybrid approach combines the strengths of machine learning algorithms with expert knowledge to enhance the accuracy and interpretability of anomaly detection results.The experimental results demonstrate that the proposed anomaly detection framework achieves promising performance in identifying various types of anomalies in blockchain data. It exhibits high detection rates while minimizing false positives, thereby providing valuable insights for blockchain network administrators and regulators to mitigate security risks and safeguard the integrity of blockchain systems.

In conclusion, this research paper presents an innovative approach to anomaly detection in blockchain using machine learning. The proposed framework addresses the unique challenges posed by blockchain's decentralized and immutable nature, offering an effective solution for detecting suspicious activities and ensuring the security of blockchain networks. The findings of this study contribute to the growing field of blockchain analytics and have significant implications for real-world blockchain applications in domains such as finance, supply chain management, and healthcare.

Keywords: Blockchain,Brain Tumour,Machine Learning.

I.INTRODUCTION

Blockchain technology:Blockchain technology is a decentralized and distributed digital tally system that enables the secure recording, verification, and storehouse of deals across multiple computers or bumps. It was originally introduced as the underpinning technology behind cryptocurrencies like Bitcoin but has since set up operations beyond digital currencies. At its core, a blockchain is a continuously growing chain of blocks, where each block contains a set of deals. These deals are grouped, vindicated, and added to the blockchain in chronological order, forming a transparent and inflexible record of all deals. crucial characteristics of blockchain technology include Decentralization Unlike traditional centralized systems where a central authority controls the data, blockchain operates in a decentralized manner. The tally is distributed across multiple actors or bumps, each maintaining a dupe of the entire blockchain. This decentralization ensures translucency, adaptability, and resistance to single

[1]Research scholar, GITAM Institute of Sciences, GITAM University, Vishakapatnam, Assistant Professor, Neil Gogte Institute of Technology, Kachavanisingaram, Telangana, India

swapnangit2021@gmail.com

[2]Student, Department of Computer Science, Neil Gogte Institute of Technology, Kachavanisingaram, Telangana, India

chirandastejaswi22@gmail.com

[3]Student, Department of Computer Science & Engineering(AIML), Neil Gogte Institute of Technology, Kachavanisingaram, Telangana, India

pallavi9379@gmail.com

points of failure. translucency The blockchain tally is transparent, meaning that all actors can view and corroborate the deals recorded on the blockchain. This translucency enhances trust and responsibility as any changes or tampering attempts can be fluently linked. Security Blockchain employs cryptographic ways to secure deals and help unauthorized differences. Each block contains a cryptographic hash, a unique identifier generated grounded on the data within the block. Any change to the data in a block would affect a different hash value, making it computationally infeasible to modify once deals without discovery. invariability Once a block is added to the blockchain, it becomes virtually inflexible, meaning it's extremely delicate to alter or cancel the recorded deals. This property ensures the integrity and permanence of the data stored on the blockchain. Consensus Mechanisms Blockchain networks use agreement mechanisms to agree on the validity of deals and achieve agreement among the sharing bumps. Different agreement algorithms, similar to Proof of Work(PoW) or evidence of Stake(PoS), are employed to ensure agreement and help vicious conditioning. Blockchain technology has operations beyond cryptocurrencies. It can be used for colourful purposes, including but not limited to force chain operation, advancing systems, decentralized finance(DeFi), smart contracts, healthcare records, and identity operation. By barring the need for interposers, blockchain technology has the implicit to increase effectiveness, reduce costs, and enhance trust in colourful diligence.

Anomaly and anomaly detection in Blockchain:Anomaly refers to any deviation, irregularity, or outlier from an expected or normal pattern or behaviour. In the context of blockchain, an anomaly can indicate unusual activities, suspicious behaviour, or potential security threats within the blockchain network.Anomaly detection in blockchain involves the process of identifying and flagging such anomalies within the blockchain system. It aims to distinguish legitimate and expected transactions or behaviours from abnormal or fraudulent ones. Anomaly detection plays a crucial role in maintaining the security, integrity, and trustworthiness of blockchain networks.

Detecting anomalies in a blockchain setting can be challenging due to several reasons:

1. Distributed and Immutable Nature: Blockchain is a decentralized and distributed ledger, where each participant maintains a copy of the entire blockchain. Once a transaction is recorded and added to a block, it becomes practically immutable. This means that any anomalous or fraudulent transaction once added to the blockchain, cannot be easily removed. Anomaly detection techniques must consider this unique characteristic of blockchain while identifying and mitigating anomalies.

2. Vast Amount of Data: Blockchain networks generate a vast amount of data due to the continuous recording of transactions. Analyzing and processing this large volume of data in real-time can be computationally intensive. Anomaly detection methods should be scalable and efficient enough to handle the high throughput of blockchain transactions.

3. Complex Transaction Patterns: Blockchain transactions can exhibit complex patterns and dependencies. Anomalies can manifest in various forms, such as abnormal transaction volumes, unusual transaction types, sudden changes in network behaviour, or malicious activities. Anomaly detection algorithms must be capable of capturing and understanding these intricate transaction patterns to effectively detect anomalies.

To address these challenges, anomaly detection in blockchain often involves the use of machine learning (ML) and data analytics techniques. These methods leverage historical blockchain data to learn patterns, relationships, and behaviours, allowing them to identify deviations from the expected norms.

The process of detecting anomalies in blockchain involves the following steps:

1. Data Collection: Information about participants, network metadata, and transaction records are all gathered and ready for study in blockchain data. You can get this information directly from the blockchain or from other sources.

2. Feature Extraction: To depict different facets of the transactions and network behaviour, pertinent characteristics or attributes are taken out of the data that has been gathered. Transaction sizes, timestamps,

involved addresses, transaction kinds, and network traffic patterns are a few examples of these aspects. To create meaningful representations of the data that capture significant qualities, feature engineering techniques are used.

3. Model Training: The extracted features are used to train machine learning models, such as clustering, classification, and anomaly detection methods. These models recognise typical patterns and behaviours by learning from historical data.

4. Anomaly Detection: The trained models are then applied to real-time or new incoming data to detect anomalies. Any deviation from the learned patterns or behaviours is flagged as a potential anomaly or suspicious activity. The models can provide anomaly scores or labels to indicate the degree of abnormality for further investigation.

5. Evaluation and Refinement: Appropriate measures including accuracy, precision, recall, and false positive rate are used to assess the anomaly detection models' performance. Based on the evaluation results and input from subject matter experts, the models can be improved and adjusted.

To put it briefly, anomaly detection in blockchain refers to the process of identifying and reporting unusual or suspicious activity within the blockchain network through the application of machine learning and data analytics tools. Blockchain technologies can improve security, stop fraud, and preserve the integrity of the network and its transactions by identifying abnormalities.

Types of anomalies in blockchain:

Several types of anomalies can occur in blockchain, which can be indicative of fraudulent or malicious activity. Here are some common types of anomalies in the blockchain:

1. Double-spending attacks: A double-spending attack occurs when an individual tries to spend the same cryptocurrency twice. This can be done by creating a fake transaction, which is then broadcasted to the network. If the fake transaction is accepted and added to the blockchain before the legitimate transaction, the individual can effectively spend the same cryptocurrency twice.

2. Network congestion: Network congestion occurs when the number of transactions being broadcasted to the network exceeds its processing capacity. This can lead to delays in transaction processing, increased transaction fees, and potentially, the rejection of valid transactions.

3. Fraudulent transactions: Fraudulent transactions occur when an individual tries to manipulate the blockchain by creating fake transactions or modifying existing ones. This can be done in an attempt to steal cryptocurrency or gain unauthorized access to the network.

4. Bugs in smart contracts: Smart contracts are self-executing contracts that run on the blockchain. If there are bugs in the code of a smart contract, this can result in unexpected behaviour, which can be exploited by malicious actors.

5. Sybil attacks: A Sybil attack occurs when an individual creates multiple fake identities, or "Sybils," on the network. This can be done in an attempt to manipulate the consensus mechanism and gain control over the network.

Overall, it is important to be aware of these different types of anomalies in blockchain and to implement effective measures for detecting and addressing them. This includes monitoring network activity, implementing security protocols, and ensuring the integrity of smart contract code.

II.LITERATURE REVIEW

Anomaly detection in blockchain has been an active area of research, with numerous studies focusing on developing effective techniques and frameworks to detect and mitigate anomalies within blockchain networks. Here are some key themes and existing research approaches in this field:

1. Machine Learning-Based Approaches: Many researchers have explored the application of machine learning algorithms for anomaly detection in the blockchain. These approaches involve training models using historical blockchain data and leveraging techniques such as clustering, classification, and anomaly detection algorithms to identify abnormal patterns or behaviours. Various algorithms, including k-means clustering, support vector machines (SVM), and deep learning models, have been employed for anomaly detection in the blockchain.

2. Network Analysis and Graph-based Techniques: Blockchain networks can be represented as graphs, where nodes represent entities (e.g., addresses, participants) and edges represent transaction flows or relationships. Network analysis and graph-based techniques have been utilized to detect anomalies in blockchain networks. These methods analyze the structural properties, connectivity patterns, and node attributes to identify suspicious or abnormal network behaviours.

3. Transaction-based Anomaly Detection: Anomalies can manifest at the transaction level in a blockchain network. Researchers have focused on developing techniques that analyze transaction-specific features, such as transaction size, transaction type, timestamp, input/output addresses, and transaction patterns. These features are used to identify abnormal transaction behaviours, including double spending, transaction spamming, or unusual transaction volumes.

4. Consensus Protocol Analysis: Anomaly detection approaches have been proposed to monitor and analyze the behaviour of blockchain participants concerning the consensus protocol. These methods examine deviations from expected behaviour, such as miners violating the consensus rules, selfish mining strategies, or abnormal block creation rates, to identify anomalies in the consensus process.

5. Hybrid Approaches: Some studies have explored the combination of machine learning techniques with domain-specific knowledge and heuristics for anomaly detection in the blockchain. These hybrid approaches integrate expert knowledge, known attack patterns, or regulatory compliance rules into the anomaly detection framework, enhancing the accuracy and interpretability of anomaly detection results.

6. Privacy-Preserving Anomaly Detection: Given the sensitive nature of blockchain data, preserving privacy while detecting anomalies is crucial. Research efforts have been made to develop privacy-preserving anomaly detection techniques that leverage cryptographic protocols, such as secure multi-party computation (MPC) or homomorphic encryption, to detect anomalies without compromising the privacy of the underlying blockchain data.

7. Benchmark Datasets and Evaluation: To facilitate research in anomaly detection in blockchain, researchers have developed benchmark datasets that simulate various types of anomalies in blockchain networks. These datasets enable the evaluation and comparison of different anomaly detection algorithms and techniques, fostering the advancement of the field.

Overall, existing research in the area of anomaly detection in blockchain has focused on leveraging machine learning, network analysis, graph-based techniques, and domain-specific knowledge to detect various types of anomalies, ensuring the security, integrity, and trustworthiness of blockchain networks. Further advancements in this field are expected to address the evolving challenges and emerging anomalies in blockchain technology.

III.METHODOLOGY

The methodology used (isolation forest, k means): When it comes to anomaly detection in the Bitcoin blockchain dataset, two commonly used methodologies are Isolation Forest and K-means clustering.

Isolation Forest: Specifically created for anomaly identification, Isolation Forest is an unsupervised machine learning system. Instead of characterising typical data points, it is predicated on the idea of isolating anomalies. The technique builds a binary tree structure and determines an instance's anomaly score based on how many random splits are needed to isolate it. Isolation Forest can be used to find anomalous transaction volumes, double spending attempts, and abnormal transaction patterns in the context of the Bitcoin blockchain dataset.

K-means Clustering: K-means clustering is a popular unsupervised learning algorithm used for data clustering. In the context of anomaly detection in the Bitcoin blockchain dataset, K-means clustering can be applied to

group similar transactions or addresses together based on features such as transaction size, timestamp, or transaction type. Anomalies can then be identified as instances that do not belong to any well-defined cluster or are significantly distant from the centroid of their assigned cluster.

Both Isolation Forest and K-means clustering is effective in detecting anomalies in the Bitcoin blockchain dataset, but they approach the problem from different angles. Isolation Forest focuses on isolating anomalies by constructing binary trees, while K-means clustering aims to identify abnormal instances based on their distance from cluster centroids.

It's crucial to remember that the methodology selected will rely on the particular goals, the dataset's properties, and the kinds of anomalies anticipated in the Bitcoin blockchain. Depending on the needs of the anomaly detection task, additional approaches like support vector machines (SVM), neural networks, or density-based clustering methods (like DBSCAN) may also be taken into consideration. Furthermore, a hybrid strategy or mix of several methodologies might be investigated to improve the efficacy and precision of anomaly detection in the Bitcoin blockchain dataset.

Using ML to Find Anomalies in Blockchain Technology:

The essential components of the blockchain architecture are:

A person or a machine is called a node in the blockchain.A transaction is the smallest unit in a blockchain system.A group of transactions that are shared by every network node is kept in one form of a data structure called a block.Blocks arranged in a specific order form a chain.Certain nodes known as miners are responsible for performing the block verification process.A set of rules and agreements called consensus is employed to control blockchain operations.

This kernel organization consists of:

1.Importing Essential Libraries.

2.Reading and Processing data.

3.Model Building & Evaluation.

4.Isolation Forest Algorithm.

5.K Means Algorithm.

6.Conclusion.

The organization of the kernel can be described as follows:

1. Importing Essential Libraries: This section involves importing the necessary libraries and modules required for data processing, modelling, and evaluation. This typically includes libraries such as pandas, numpy, scikit-learn, and any other specific libraries needed for anomaly detection algorithms like Isolation Forest and K-means.

2. Reading and Processing Data: In this section, the dataset is read into the kernel, and any necessary preprocessing steps are performed. This may involve cleaning the data, handling missing values, transforming variables, and scaling the data if required.

3. Model Building & Evaluation: This section focuses on constructing anomaly detection models. It involves splitting the data into training and testing sets, training the models using the training data, and evaluating their performance using appropriate metrics. The performance evaluation can include metrics like accuracy, precision, recall, and false positive rate.

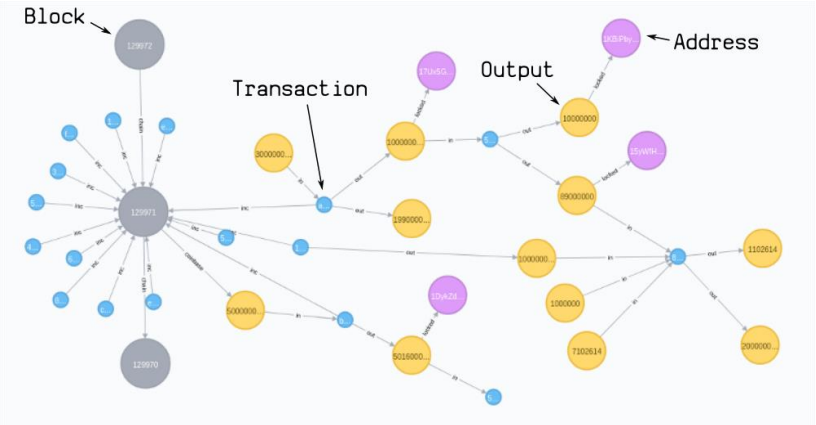
4. Isolation Forest Algorithm: This section specifically focuses on implementing the Isolation Forest algorithm for anomaly detection. It may include explaining the algorithm's principles, discussing parameter settings, fitting

the model to the training data, and predicting anomalies using the trained model. The performance of the Isolation Forest algorithm is typically evaluated and compared with other methods or baselines.

5. K-means Algorithm: This section is dedicated to implementing the K-means algorithm for anomaly detection. It may involve explaining the K-means algorithm, discussing the number of clusters and other parameter settings, fitting the model to the training data, and assigning anomalies based on their distance from cluster centroids. The performance of the K-means algorithm is evaluated and compared with other methods or baselines.

6. Conclusion: The conclusion section provides a summary of the findings and results obtained from the anomaly detection models. It may discuss the performance of the Isolation Forest and K-means algorithms, their effectiveness in detecting anomalies in the given dataset, and any insights gained from the analysis.

Overall, the organization of the kernel follows a logical flow, starting with importing libraries, processing the data, building and evaluating models, implementing the Isolation Forest and K-means algorithms, and concluding with a summary of the findings. This framework makes it possible to comprehend the use and assessment of anomaly detection techniques in the provided dataset.



IV.FRAMEWORK

Step 1: Data Collection and Preprocessing

Gather the blockchain data, including transaction records, block information, and relevant attributes.Preprocess the data by cleaning any noise or inconsistencies, handling missing values, and normalizing the data if necessary.

Step 2: Feature Engineering and Selection

Identify and extract relevant features from the blockchain data. This may include transaction size, transaction type, timestamp, input/output addresses, network traffic patterns, or any other domain-specific features.

Perform feature selection techniques to choose the most informative and relevant features for anomaly detection. This step helps reduce dimensionality and improve model efficiency.

Step 3: Model Training

Choose appropriate machine learning algorithms for detecting anomalies, like K-means clustering, isolation forest, or other methods that are tailored to the unique needs and features of the blockchain data.Divide the preprocessed data into sets for testing and training.Utilising the training data, train the chosen machine learning models by applying the characteristics that were extracted to understand typical patterns and behaviours in the blockchain network.

Step 4: Model assessment:

Evaluate how well the trained models perform using pertinent assessment metrics, such as accuracy, precision, recall, F1 score, or area under the receiver operating characteristic curve (AUC-ROC).Use the testing dataset to assess how well the ML models generalise to detect problems.

Step 5: Anomaly Detection and Interpretation

Apply the trained ML models to detect anomalies in new, unseen blockchain data.

Analyze the output of the ML models to identify instances flagged as anomalies. This may involve setting appropriate anomaly detection thresholds or using anomaly scores provided by the models.

Interpret the detected anomalies by analyzing the characteristics and patterns that distinguish them from normal blockchain activities.

Step 6: Post-processing and Visualization

Perform any necessary post-processing steps on the detected anomalies, such as filtering out false positives or aggregating anomalies based on their characteristics.

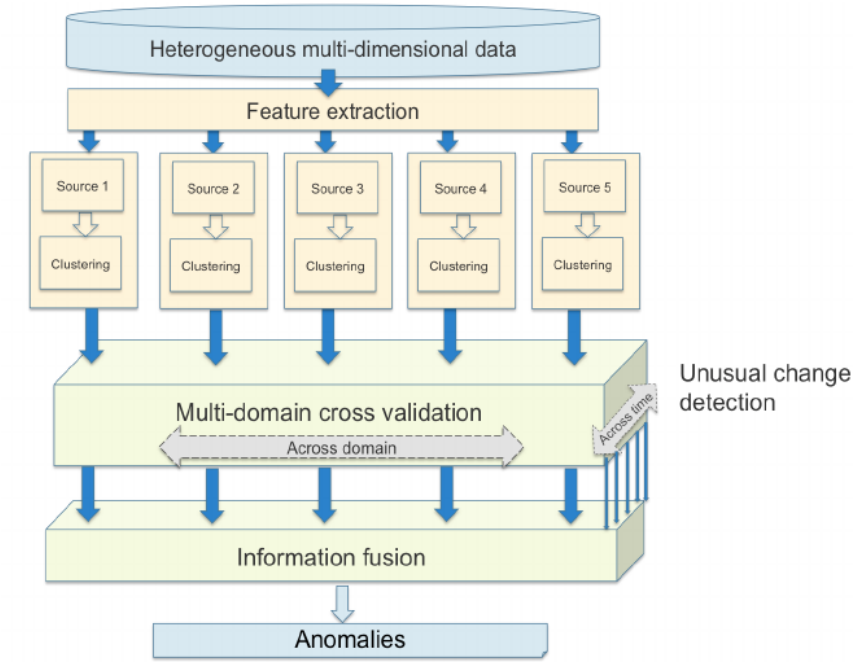
Visualize the detected anomalies and their associated attributes, patterns, or network relationships to facilitate interpretation and decision-making.

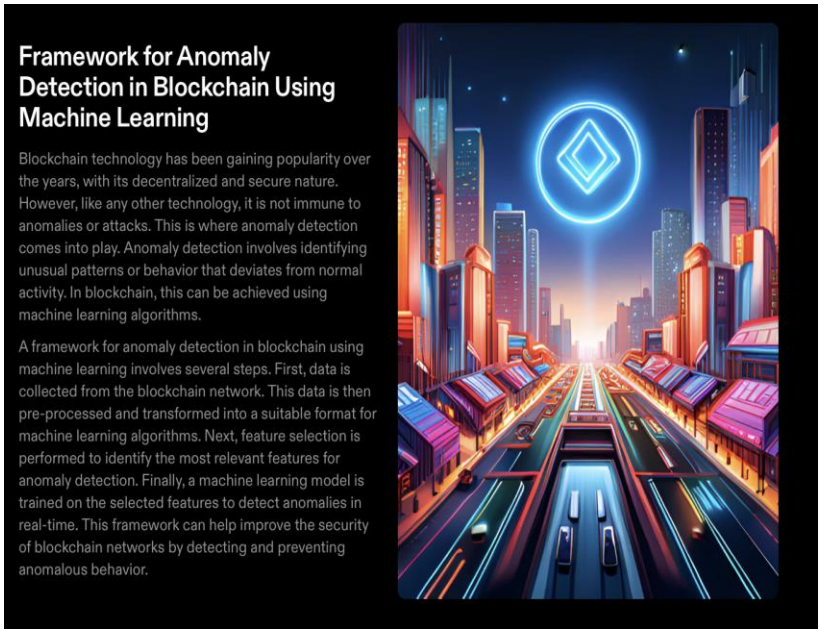
Step 7: Iterative Improvement

Assess the performance and effectiveness of the anomaly detection framework.

Explore and experiment with different ML algorithms, feature engineering techniques, or parameter settings to improve the accuracy and efficiency of the framework.

Continuously update and refine the framework based on new data, emerging anomalies, and evolving blockchain characteristics.





V.RESULTS

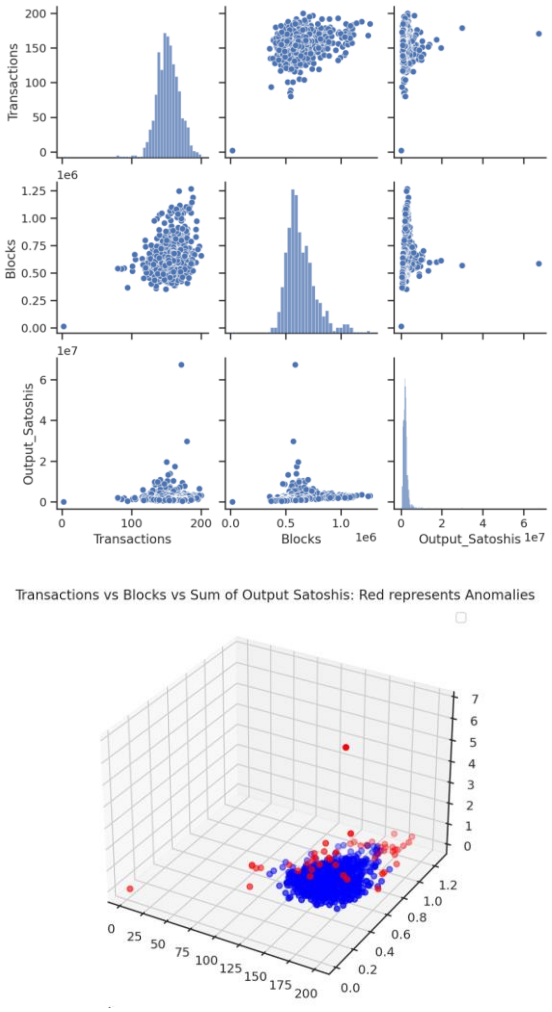
In the context of blockchain, transactions, blocks, and output_satoshis are fundamental components that contribute to the functioning and structure of the blockchain system.

Transactions: Transactions represent the fundamental unit of activity in a blockchain network. They are records of value transfers or other operations that are broadcasted to the network by participants. Transactions typically include information such as the sender's address, the recipient's address, the amount of cryptocurrency being transferred, and any additional data or instructions associated with the transaction. Each transaction is digitally signed by the sender to provide authenticity and ensure that only the rightful owner can initiate the transfer.

Blocks: Blocks are containers or collections of transactions that are bundled together and added to the blockchain in a sequential and immutable manner. In a blockchain network, transactions are grouped into blocks to create a chronological chain of blocks, hence the term "blockchain." Blocks contain transaction data, a reference to the previous block (creating a link between blocks), a timestamp, and a unique identifier called a block hash. The block hash serves as a cryptographic signature that ensures the integrity of the block and verifies its position in the blockchain.

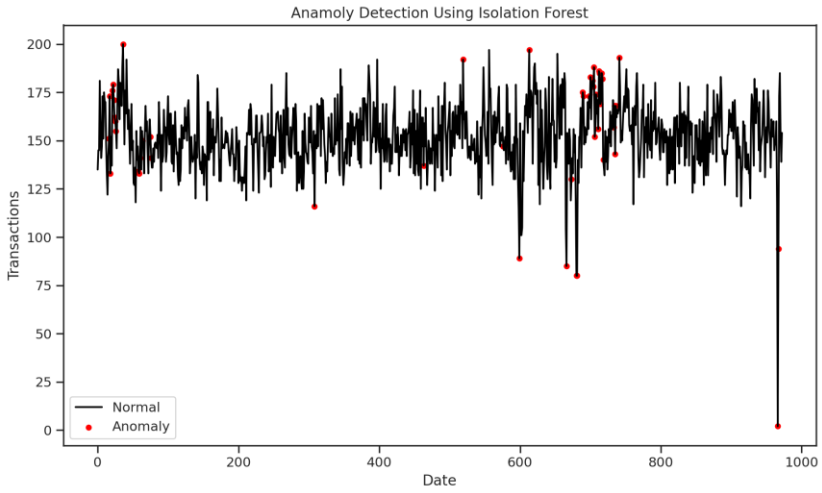
Output_satoshis: Output_satoshis refers to the amount of cryptocurrency associated with a particular output of a transaction. In most blockchain networks, cryptocurrency amounts are measured in smaller units, such as satoshis in Bitcoin. Satoshis are the smallest divisible unit of Bitcoin, named after the pseudonymous creator of Bitcoin, Satoshi Nakamoto. Output_satoshis indicates the specific amount of cryptocurrency being transferred or assigned to a particular recipient's address in a transaction. It represents the value that is stored and can be further spent or transferred in subsequent transactions.

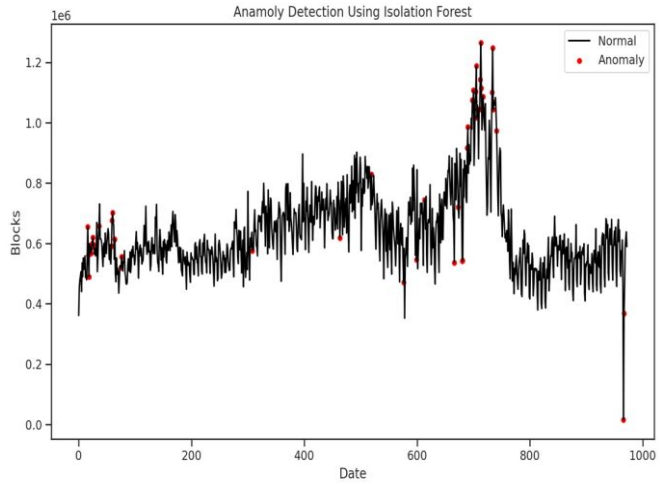
Together, transactions, blocks, and output_satoshis form the foundation of a blockchain system. Transactions represent the actions performed by participants, blocks provide a structure to organize and validate transactions, and output_satoshis determines the amount of cryptocurrency associated with each transaction output. These components work in harmony to ensure the secure and transparent transfer of value within the blockchain network.

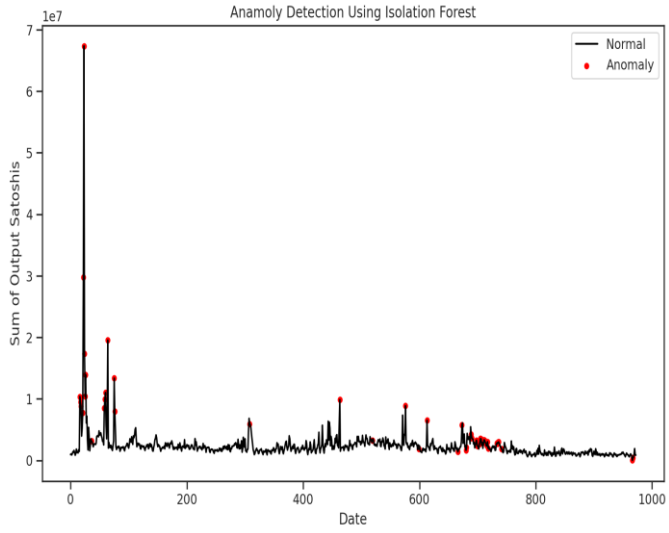


The above plot cites Transactions, Features, and Sum of Output Satoshis features to represents anomalies in the data using the Isolation Forest method. Now, summarize the predictions just performed by assigning them binary values i.e., 0 or 1 (0 for normal, 1 for anomaly).

ANOMALY DETECTION USING ISOLATION FOREST

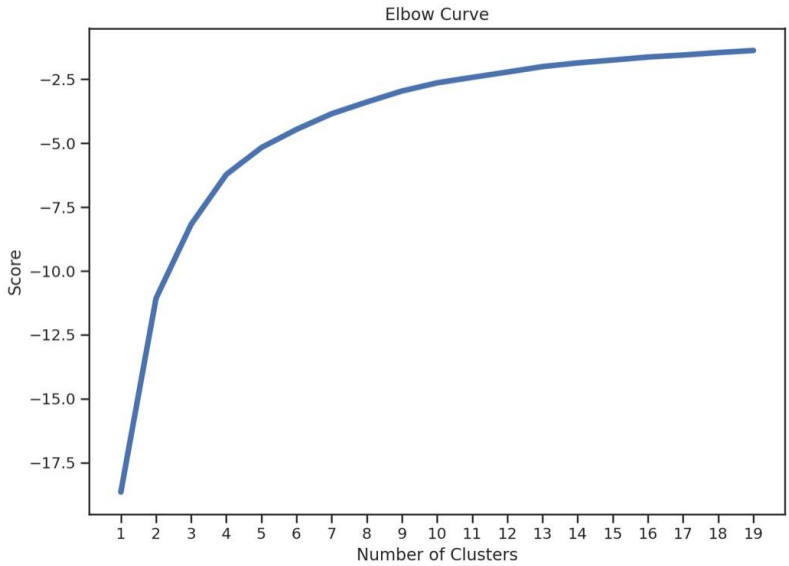




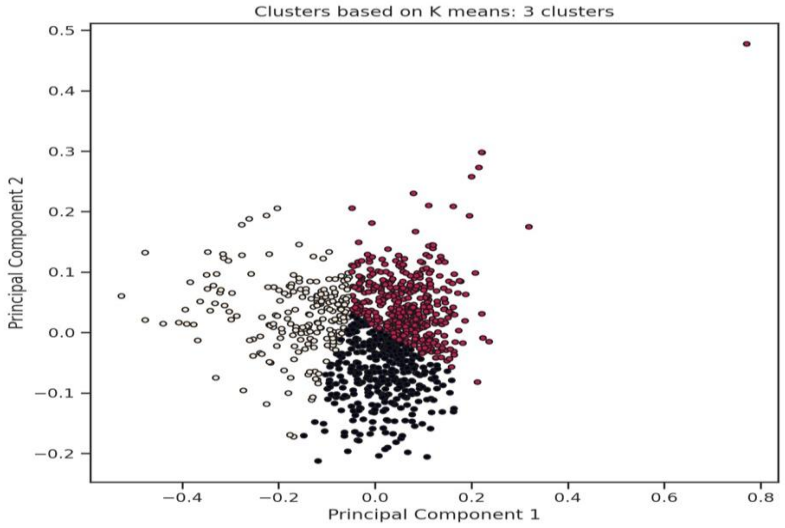


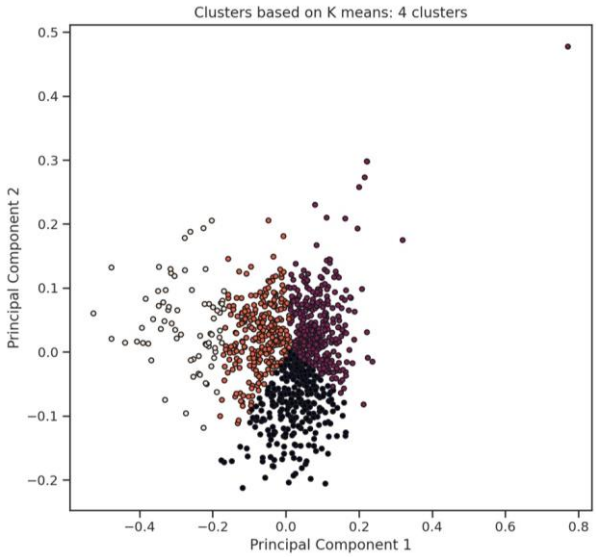
Inherited the use of those three previous plots (Transactions, Blocks, Output Satoshis vs. date) for a better understanding of the Isolation Forest method predicted anomalies by visualizing them with outliers anomaly.

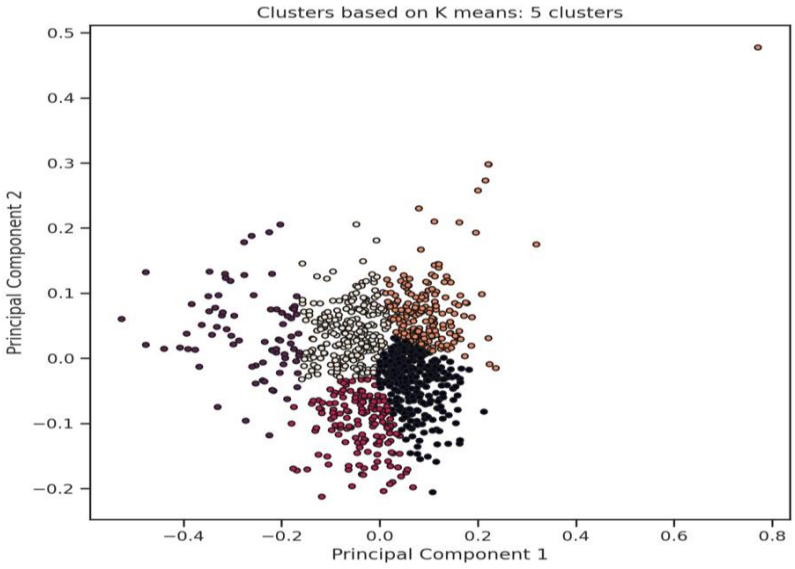
ANOMALY DETECTION USING K-MEANS

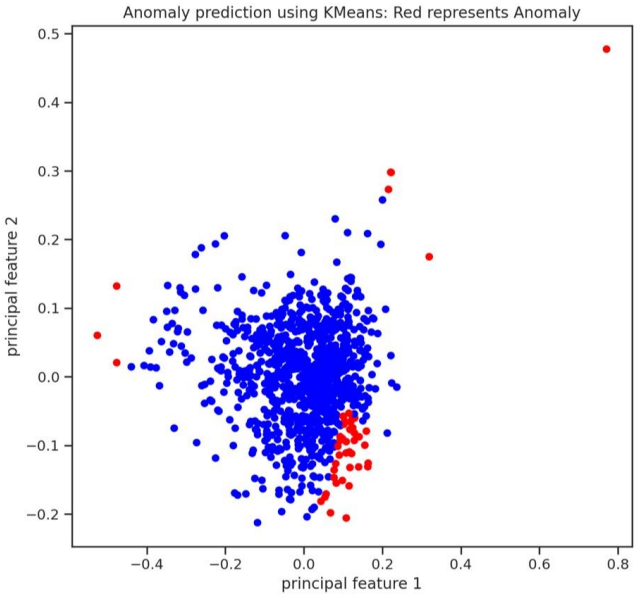


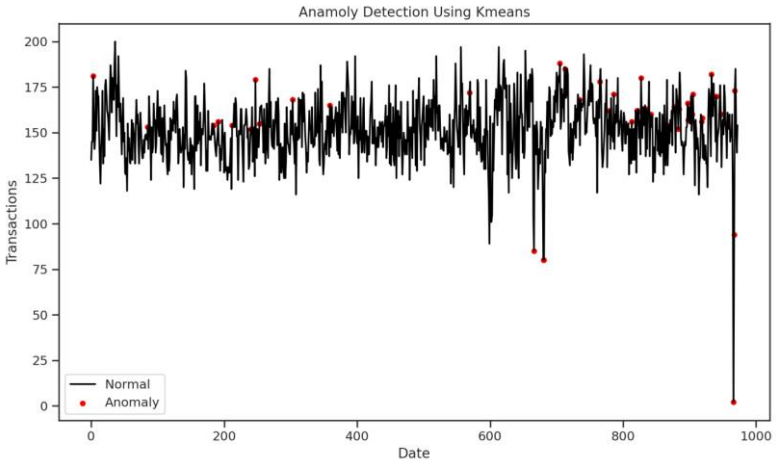
The above elbow curve determines the number of clusters for the K-means algorithm. The most drastic change was recorded in the elbow curve at 3 or 4 or 5 clusters. So let's which one of them is better for our case.

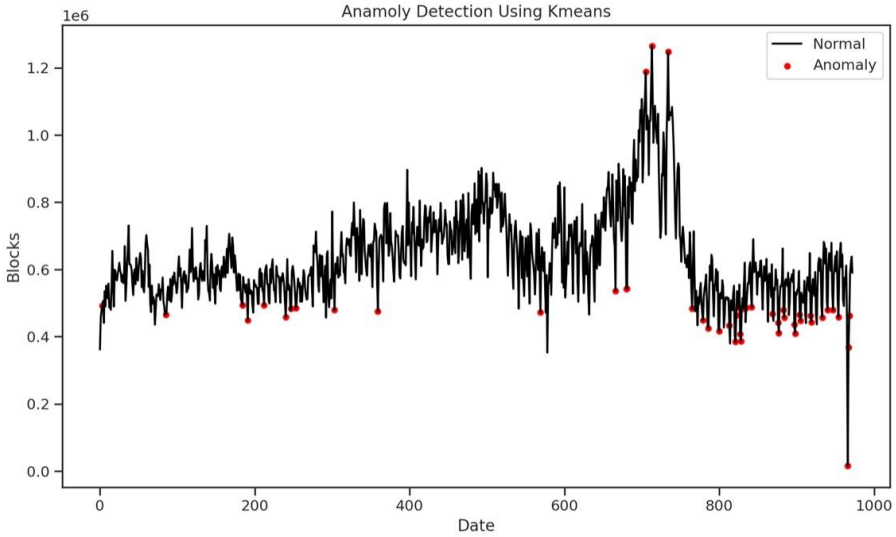


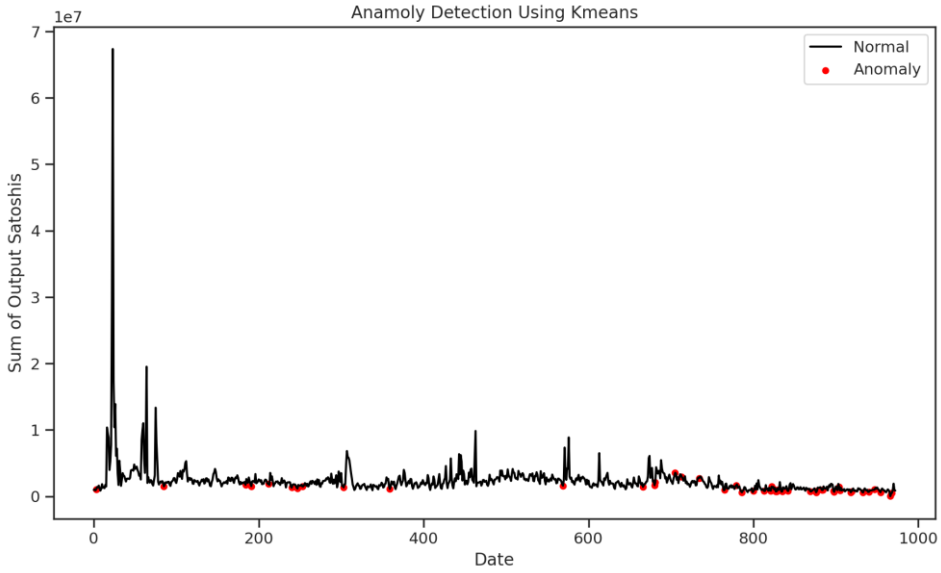





Anomaly prediction using KMeans: Red represents Anomaly
principal feature 2
principal feature 1


Anamoly Detection Using Kmeans
Transactions
Date
Normal
Anomaly


Anamoly Detection Using Kmeans
1e6
Blocks
Date
Normal
Anomaly



FINAL RESULTS

Date	Transactions	Blocks	Output_Satoshis	anomaly_IsolationForest	anomaly_kmeans	
0	2016-01-01	135	361519	9.574813e+05	1	0
1	2016-01-02	145	455120	1.037920e+06	0	0
2	2016-01-03	145	478708	8.985480e+05	0	0
3	2016-01-04	181	492865	1.067068e+06	0	1
4	2016-01-05	157	506371	1.392599e+06	0	0
...	...	...	...	...	...	...
968	2018-08-26	173	463319	5.155083e+05	0	1
969	2018-08-27	185	585010	9.744078e+05	0	0
970	2018-08-28	153	616972	1.898013e+06	0	0
971	2018-08-29	139	638060	9.013310e+05	0	0
972	2018-08-30	154	589764	8.546115e+05	0	0

Select the cases for final anomaly in which both the algorithms predicted anomaly:

anomaly_kmeans == 1 & anomaly_IsolationForest == 1:

	Date	Transactions	Blocks	Output_Satoshis	anomaly_IsolationForest	anomaly_kmeans
666	2017-10-28	85	536785	1.417696e+06	1	1

	Date	Transactions	Blocks	Output_Satoshis	anomaly_IsolationForest	anomaly_kmeans
680	2017-11-11	80	542026	1.664880e+06	1	1
681	2017-11-12	80	542990	2.087931e+06	1	1
705	2017-12-06	188	1187481	3.531511e+06	1	1
713	2017-12-14	185	1264802	2.943543e+06	1	1

Select the cases in which either of the two algorithms predicted anomaly:

anomaly_kmeans == 1 | anomaly_IsolationForest == 1:

	Date	Transactions	Blocks	Output_Satoshis	anomaly_IsolationForest	anomaly_kmeans
0	2016-01-01	135	361519	9.574813e+05	1	0
3	2016-01-04	181	492865	1.067068e+06	0	1
16	2016-01-17	151	655229	1.035818e+07	1	0
17	2016-01-18	173	561828	9.462153e+06	1	0
18	2016-01-19	133	489778	8.804629e+06	1	0

Select the cases where no algorithm predicted anomaly:

anomaly_kmeans == 0 & anomaly_IsolationForest == 0:

	Date	Transactions	Blocks	Output_Satoshis	anomaly_IsolationForest	anomaly_kmeans
1	2016-01-02	145	455120	1.037920e+06	0	0
2	2016-01-03	145	478708	8.985480e+05	0	0
4	2016-01-05	157	506371	1.392599e+06	0	0
5	2016-01-06	141	440544	1.337497e+06	0	0
6	2016-01-07	147	535250	1.606574e+06	0	0

Complete records: 973

Final anomaly number: 8.

There are 89 potential anomalies.

Total deviation: 97

9.97 percent of the overall anomaly in the data

VI.CONCLUSION

As a result, we have identified 76 potential anomalies in the latest three iterations of the Google BigQuery Bitcoin Blockchain dataset, or 10% of the total anomaly. Insulation wood and K-means are two well-liked anomaly discovery approaches that we've used. One noteworthy finding from the research is that 38 anomaly cases are predicted by both anomaly discovery approaches, which is quite harmonic. By using more data or adding more features, we can investigate anomaly finding in blockchain systems in more detail. It is possible to employ anomaly discovery techniques to automatically detect and remove anomalous conditions, thereby strengthening and safeguarding blockchain systems.

VI.FUTURE SCOPE

The field of anomaly detection in the blockchain is expected to continue evolving with future advancements and research efforts. Here are some potential areas of future scope in anomaly detection in the blockchain:

1. Advanced Machine Learning Techniques: As machine learning algorithms and techniques continue to advance, there is a scope for exploring more sophisticated approaches for anomaly detection in the blockchain. This includes the application of deep learning models, reinforcement learning, and ensemble methods to improve detection accuracy and handle the complexity of blockchain data.

2. Real-Time Anomaly Detection: Real-time anomaly detection is crucial for quickly identifying and mitigating potential security threats in blockchain networks. Future research can focus on developing efficient algorithms and frameworks that can handle the high throughput of blockchain transactions in real-time, enabling timely detection and response to anomalies.

3. Privacy-Preserving Anomaly Detection: Privacy is a critical concern in blockchain networks. Future research can focus on developing privacy-preserving anomaly detection techniques that can detect anomalies without exposing sensitive transaction details or compromising the privacy of participants' data. This can involve the use of advanced cryptographic techniques, such as zero-knowledge proofs or secure multi-party computation.

4. Adversarial Anomaly Detection: With the evolving nature of attacks and malicious activities in blockchain networks, there is a need for robust anomaly detection techniques that can detect adversarial behaviours and sophisticated attacks. Future research can explore the use of adversarial machine learning and game-theoretic approaches to detect and counteract such adversarial anomalies.

5. Explainable Anomaly Detection: Interpretability and explainability of anomaly detection results are crucial for understanding and validating the detected anomalies. Future research can focus on developing methods that provide clear explanations or visualizations of detected anomalies, enabling stakeholders to understand the underlying reasons for anomaly detections and take appropriate actions.

6. Blockchain-Specific Anomalies: Blockchain technology continues to evolve, and new types of anomalies specific to blockchain networks may emerge. Future research can investigate and identify these novel anomalies, developing tailored anomaly detection techniques to address them effectively.

7. Integration with Blockchain Forensics: Anomaly detection can be integrated with blockchain forensics to enhance the investigation and analysis of suspicious activities within blockchain networks. Future research can explore the integration of anomaly detection techniques with forensic tools and methodologies, enabling a comprehensive approach to detecting, tracing, and attributing anomalous behaviours in blockchain systems.

8. Industry-Specific Anomaly Detection: Different industries have unique requirements and challenges in blockchain adoption. Future research can focus on industry-specific anomaly detection techniques, catering to

sectors such as finance, supply chain, healthcare, and energy, where anomalies can have significant implications for security, compliance, and operational efficiency.

In conclusion, the future scope of anomaly detection in blockchain encompasses advancements in machine learning techniques, real-time detection, privacy preservation, handling adversarial anomalies, explainability, addressing emerging anomalies, integration with blockchain forensics, and industry-specific applications.

VII.REFERENCES

[1]. M. Ul Hassan, M. H. Rehmani and J. Chen, "Anomaly Detection in Blockchain Networks: A Comprehensive Survey," in *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 289-318, First quarter 2023, doi: 10.1109/COMST.2022.3205643.

[2]. M. Signorini, M. Pontecorvi, W. Kanoun and R. Di Pietro, "BAD: A Blockchain Anomaly Detection Solution," in *IEEE Access*, vol. 8, pp. 173481-173490, 2020, doi: 10.1109/ACCESS.2020.3025622.

[3]. Huang, D., Chen, B., Li, L., Ding, Y. (2020). Anomaly Detection for Consortium Blockchains Based on Machine Learning Classification Algorithm. In: Chellappan, S., Choo, KK.R., Phan, N. (eds) Computational Data and Social Networks. CSoNet 2020. Lecture Notes in Computer Science(), vol 12575. Springer, Cham. https://doi.org/10.1007/978-3-030-66046-8_25

[4]. Shafik, M., and K. Case. "Anomaly Detection System for Ethereum Blockchain Using Machine Learning." (2022): 311.

[5]. Singh, Priyanshi, Deepika Agrawal, and Sudhakar Pandey. "Anomaly detection and analysis in blockchain systems." (2023).

[6]. Martin, K., Rahouti, M., Ayyash, M., & Alsmadi, I. (2022). Anomaly detection in the blockchain using network representation and machine learning. *Security and Privacy*, *5*(2), e192.

[7].https://www.kaggle.com/code/mrsohelranapro/anomaly-detection-in-blockchain-system-via-ml/notebook