ANALYSIS OF BEHAVIORAL PATTERNS IN BITCOIN ADDRESSES USING MACHINE LEARNING

V. Pravalika[1] J.Sushmitha[2], B.Bhavya Reddy[3], G.Siva Shankar[4], B.Vivek[5],

[1] Assistant Professor, *Department of CSE (ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING) TKR COLLEGE OF ENGINEERING & TECHNOLOGY*

[2,3,4,5] *UG Scholars in Department of CSE (ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING) TKR COLLEGE OF ENGINEERING & TECHNOLOGY*

Abstract: This study explores the application of machine learning algorithms to analyze behavioral patterns in Bitcoin addresses. Utilizing a dataset that includes various transaction details, we employ K-Nearest Neighbors (KNN), XGBoost, Random Forest (RF), and K-best feature selection as our base algorithms for initial pattern recognition. To enhance the predictive accuracy, we extend our analysis with advanced techniques including Stacking Classifier, Convolutional Neural Network (CNN), and Gradient Boosting Classifier. The primary objective is to accurately categorize Bitcoin addresses into specific categories such as blackmail, cyber-security service, darknet market, centralized exchange, P2P financial infrastructure service, gambling, money laundering, and more. This research aims to provide a robust framework for identifying suspicious activities and improving the security and transparency of Bitcoin transactions. The outcomes could significantly aid in the detection of illicit activities and the enforcement of cybersecurity measures in the cryptocurrency domain.

Keywords: Stacking Classifier, Convolutional Neural Network (CNN), and Gradient Boosting Classifier, classification algorithms, Kaggle dataset.

1. INTRODUCTION

Bitcoin, as a decentralized digital currency, has gained significant prominence since its inception. While its advantages include financial anonymity and a lack of centralized control, these same features have attracted illicit activities, making it a target for criminal enterprises. Addressing these concerns requires sophisticated methods to analyze and categorize Bitcoin transactions effectively. This study introduces a novel approach to understanding Bitcoin address behavior by leveraging advanced machine learning techniques to detect patterns indicative of various activities. In this research, we focus on a comprehensive dataset that encompasses a range of transactional details associated with Bitcoin addresses. To identify behavioral patterns, we apply a series of machine learning algorithms, starting with K-Nearest Neighbors (KNN), XGBoost, Random Forest (RF), and K-best feature selection. These algorithms serve as a foundation for initial pattern recognition, helping us to distinguish between different categories of Bitcoin addresses. Building upon this initial analysis, we employ more advanced techniques to refine our classification accuracy. The Stacking Classifier, Convolutional Neural Network (CNN), and Gradient Boosting Classifier are integrated into our framework to enhance predictive performance. Each of these methods brings unique strengths to the table: Stacking Classifier combines the strengths of multiple algorithms, CNNs excel in detecting intricate patterns through deep learning, and Gradient Boosting Classifiers improve model accuracy by focusing on difficult-to-predict instances. The primary goal of this research is to categorize Bitcoin addresses into specific

types such as blackmail, cybersecurity services, darknet markets, centralized exchanges, P2P financial services, gambling, and money laundering. By accurately identifying these categories, our work aims to bolster the security and transparency of Bitcoin transactions, providing a valuable tool for detecting illicit activities and strengthening cybersecurity measures in the cryptocurrency realm.

2. Motivation:

The growing complexity and usage of Bitcoin in various sectors underscore the need for advanced analytical methods to distinguish between legitimate and illicit activities. Traditional analysis techniques fall short in handling the vast and nuanced data associated with Bitcoin transactions. This study is motivated by the need to enhance transaction security and transparency through machine learning. By applying a diverse set of algorithms, including advanced methods like Stacking Classifier, CNN, and Gradient Boosting, we aim to create a sophisticated framework for identifying and categorizing Bitcoin addresses. This approach seeks to improve the detection of suspicious activities and bolster cybersecurity in the cryptocurrency realm.

3. Problem Statement:

The rise of Bitcoin as a digital currency has been accompanied by an increase in illicit activities, such as money laundering, blackmail, and transactions on darknet markets. Traditional methods for identifying suspicious behavior in Bitcoin addresses are often inadequate due to the pseudonymous nature of transactions and the vast scale of the Bitcoin network. This study aims to address this challenge by employing machine learning algorithms to analyze and categorize Bitcoin address behaviors. By using algorithms like K-Nearest Neighbors, XGBoost, Random Forest, and advanced techniques such as Stacking Classifier, CNN, and Gradient Boosting Classifier, this research seeks to develop a reliable framework for detecting and preventing illegal activities in the cryptocurrency ecosystem.

4. Objective of the Project:

The primary objective of this project is to leverage machine learning algorithms to analyze and categorize behavioral patterns in Bitcoin addresses, aiming to identify and classify suspicious activities. By utilizing K-Nearest Neighbors (KNN), XGBoost, Random Forest (RF), and K-best feature selection as foundational methods, the project seeks to recognize initial patterns in transaction data. To further improve predictive accuracy, advanced techniques like Stacking Classifier, Convolutional Neural Network (CNN), and Gradient Boosting Classifier will be employed. The ultimate goal is to accurately categorize Bitcoin addresses into specific types, such as those linked to blackmail, darknet markets, or money laundering, thereby enhancing security measures and transparency in Bitcoin transactions.

5. Scope of the project

This study focuses on analyzing Bitcoin address behaviors to identify patterns that correlate with specific activities such as blackmail, darknet market operations, and money laundering. By leveraging a combination of machine learning algorithms, including K-Nearest Neighbors, XGBoost, Random Forest, and advanced techniques like Stacking Classifier, Convolutional Neural Networks (CNN), and Gradient Boosting Classifier, the research aims to categorize Bitcoin addresses accurately. The study will utilize a Kaggle dataset containing detailed transaction records to achieve this goal. The scope encompasses developing a robust classification framework to enhance the detection of suspicious activities within the Bitcoin network, thereby contributing to the broader objective of bolstering cybersecurity and transparency in cryptocurrency transactions.

6. LITERATURE SURVEY

1. L. Serena, S. Ferretti, and G. D'Angelo, "Cryptocurrencies activity as a complex network: Analysis of transactions graphs," Peer-Peer Netw. Appl., vol. 15, no. 2, pp. 839–853, Mar. 2022.

The paper by L. Serena, S. Ferretti, and G. D'Angelo explores cryptocurrency transaction networks using complex network theory. Analyzing Bitcoin, DogeCoin, Ethereum, and Ripple, the study focuses on transaction patterns over time to uncover insights into user behavior on these Distributed Ledger Technologies (DLTs). The authors introduce the Distributed Ledger Network Analyzer (DiLeNA), a tool designed to investigate transaction networks. Their findings reveal that transaction graphs across all studied DLTs exhibit small-world properties, highlighting the importance of network analysis for understanding user interactions and the dynamics of cryptocurrency ecosystems.

2. B. Tao, I. W. Ho, and H.-N. Dai, "Complex network analysis of the Bitcoin blockchain network," in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), May 2021, pp. 1–5.

The paper by Tao, Ho, and Dai presents a comprehensive analysis of the Bitcoin blockchain network through a complex network approach. They introduce the BABD-13 dataset, which includes detailed Bitcoin transaction data from July 2019 to May 2021, featuring 13 types of addresses and 148 attributes. Using machine learning models like k-nearest neighbors, decision tree, and XGBoost, they achieve classification accuracies between 93.24% and 97.13%. Additionally, the study proposes a k-hop subgraph generation algorithm for in-depth network analysis and examines Bitcoin address behavior patterns, contributing to better understanding and tracking of blockchain transactions.

3. B. Tao, H.-N. Dai, J. Wu, I. W. Ho, Z. Zheng, and C. F. Cheang, "Complex network analysis of the Bitcoin transaction network," IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 69, no. 3, pp. 1009–1013, Mar. 2022.

This paper presents a comprehensive framework for analyzing Bitcoin transactions to identify illicit activities within the cryptocurrency ecosystem. The study introduces the BABD-13 dataset, the largest publicly available labeled dataset of Bitcoin addresses, covering transactions from July 2019 to May 2021. It features 13 address types, 5 indicator categories, and 148 features, with 544,462 labeled entries. A novel subgraph generation algorithm, BTC-SubGen, extracts k-hop subgraphs from the Bitcoin transaction network. Classification using various machine learning models achieved accuracy rates between 93.24% and 97.13%. The study also explores feature importance and behavior patterns of Bitcoin addresses.

4. N. Tovanich, N. Soulié, N. Heulot, and P. Isenberg, "An empirical analysis of pool hopping behavior in the Bitcoin blockchain," in Proc. IEEE Int. Conf. Blockchain Cryptocurrency, May 2021, pp. 1–9.

In their 2021 paper, N. Tovanich et al. investigate pool-hopping behavior in Bitcoin mining, where miners switch pools to maximize rewards. They propose a new detection methodology using time window analysis of mining rewards. Their approach includes algorithms for miner identification and revenue tracking, tested on the top five mining pools over two periods in 2020 and 2021. The study finds that while pool-hopping remains beneficial, the newer reward systems have improved fairness, reducing the disparity between pool-hoppers and static miners. Despite this, pool-hoppers still achieve a 33% higher median cumulative gain compared to static miners.

5. I. Alqassem, I. Rahwan, and D. Svetinovic, "The anti-social system properties: Bitcoin network data analysis,"

IEEE Trans. Syst., Man, Cybern., Syst., vol. 50, no. 1, pp. 21–31, Jan. 2020.

The paper by Alqassem, Rahwan, and Svetinovic, titled "The Anti-Social System Properties: Bitcoin Network Data Analysis," published in IEEE Transactions on Systems, Man, and Cybernetics: Systems (2020), examines the Bitcoin network through a lens of anti-social system properties. This study is part of a broader index covering technical items published in 2020, including papers, correspondence, and reviews. The Author Index details the primary entries for each item, organized by the first author's name, and includes coauthors, paper titles, and publication specifics. The Subject Index categorizes entries by relevant subject headings, providing comprehensive bibliographic information. The focus of the paper is on analyzing Bitcoin's network dynamics and its implications within an anti-social system framework.

7. Block Diagram

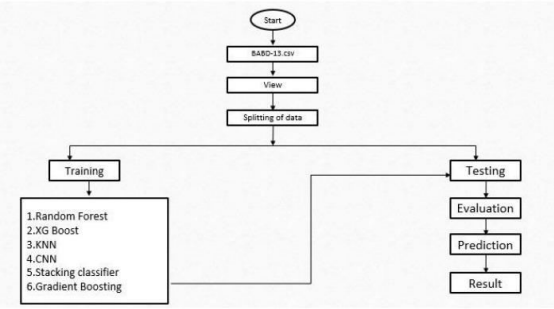


Fig.1: block diagram

8. System Architecture

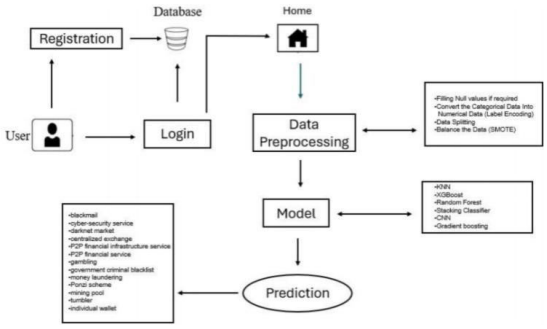


Fig.2: work flow diagram

9. Methodology and Algorithms:

A. K-Nearest Neighbours (KNN):

K-Nearest Neighbours (KNN) is a fundamental machine learning algorithm utilized for pattern recognition and classification tasks. For the BABD (Bitcoin Address Behaviour Dataset), KNN can be effectively applied to analyse and categorize Bitcoin addresses based on their transaction behaviours. This algorithm works by measuring the distance between a given address and its neighbouring addresses in the feature space. By identifying the 'k' nearest addresses to a target address, KNN classifies the target based on the majority class among its nearest neighbours. This method is particularly useful for pattern analysis in the BABD as it does not require a predefined model and can adapt to the intricacies of the dataset. By leveraging KNN, one can uncover patterns related to address behaviour and transaction characteristics, which are crucial for understanding and predicting Bitcoin address activities.

B. XGBoost:

XGBoost, or eXtreme Gradient Boosting, is an efficient and scalable machine learning algorithm renowned for its performance in classification and regression tasks. In the context of the Bitcoin Address Behavior Dataset (BABD), XGBoost can be leveraged to analyze and model patterns in Bitcoin address transactions. This method excels in handling

large datasets and complex interactions between features, making it ideal for detecting subtle patterns and anomalies in Bitcoin address behaviors. XGBoost operates by constructing a series of decision trees in a sequential manner, where each new tree corrects the errors of its predecessors. The model employs gradient boosting techniques to minimize prediction errors and enhance accuracy. Its built-in regularization further prevents overfitting, ensuring robust performance on unseen data. By applying XGBoost to the BABD, researchers can uncover intricate behavior patterns and derive actionable insights from Bitcoin transaction data.

C. Random Forest:

In the context of analyzing Bitcoin address behavior, the Random Forest algorithm serves as a powerful tool for pattern recognition and classification. This ensemble learning method operates by constructing a multitude of decision trees during training and outputting the mode of the classes or mean prediction from individual trees for classification or regression tasks, respectively. Each tree is built using a subset of the training data and a random subset of features, which helps in reducing overfitting and improving model generalization. For the Bitcoin address behavior dataset, Random Forest can effectively identify and differentiate between various transaction patterns, clustering addresses with similar activity traits and revealing anomalies or emerging trends. Its robustness to noisy data and ability to handle large datasets with high-dimensional features make it particularly suited for complex analyses in the cryptocurrency domain.

D. Stacking Classifier:

A stacking classifier is an ensemble learning technique that combines multiple base classifiers to enhance predictive performance. In the context of the BABD (Bitcoin Address Behavior Dataset), a stacking classifier can be particularly useful for identifying and analyzing patterns in Bitcoin address usage. This method involves training several different base models, such as logistic regression, decision trees, and support vector machines, each capturing unique aspects of the data. The predictions from these base models are then used as input features for a final meta-model, which learns to make the most accurate predictions by aggregating the strengths of the base classifiers. By leveraging diverse algorithms and combining their outputs, a stacking classifier can effectively address the complexities of Bitcoin address behavior, offering a more robust analysis of patterns and improving overall classification accuracy. This approach helps in uncovering nuanced insights and trends within the dataset.

E. Gradient Boosting Classifier:

In the context of the BABD (Bitcoin Address Behavior Dataset), the Gradient Boosting classifier serves as a robust machine learning model for analyzing and classifying Bitcoin transaction patterns. This classifier builds an ensemble of decision trees in a sequential manner, where each subsequent tree corrects the errors of the previous ones, improving the model's accuracy over time. By focusing on the residuals of the previous models, Gradient Boosting effectively captures complex patterns in Bitcoin address behaviors, which can be crucial for identifying fraudulent activities or behavioral anomalies. Its adaptability to various data complexities and its capacity to handle imbalanced datasets make it particularly suitable for the nuanced analysis required in cryptocurrency transactions. The model's performance can be further enhanced by tuning hyperparameters and incorporating feature selection techniques to better understand and interpret the underlying behavioral patterns within the dataset.

F. CNN:

The paper "BABD: A Bitcoin Address Behavior Dataset for Pattern Analysis" presents a comprehensive dataset designed to enhance the analysis of Bitcoin address behaviors. This dataset, known as BABD, includes a diverse

collection of Bitcoin addresses along with detailed transaction histories and behavioral patterns. By aggregating data from various sources, the dataset facilitates the exploration of transactional trends and user behaviors within the Bitcoin network. Researchers and analysts can leverage BABD to identify emerging patterns, detect anomalies, and gain insights into the dynamics of cryptocurrency transactions. The dataset aims to support advancements in financial security, fraud detection, and blockchain analysis by providing a rich resource for pattern recognition and behavioral studies in the cryptocurrency domain. This work underscores the importance of data-driven approaches in understanding and improving the security and functionality of blockchain technologies.

10. IMPLEMENTATION

MODULES:

1. **User:**

1.1 Select the model:
User have to select the model.

1.2 View score:
Here user have ability to view the accuracy in %

1.3 Input Model:
The user must provide input values for the certain fields in order to get results.

1.4 View Results:
User view's the generated results from the model.

2. **System**

2.1 Working on dataset:
System checks for data whether it is available or not and load the data in csv files.

2.2 Pre-processing:
Data need to be pre-processed according the models it helps to increase the accuracy of the model and better information about the data.

2.3 Training the data:
After pre-processing the data will split into two parts as train and test data before training with the given algorithms.

2.4 Model Building
To create a model that predicts the personality with better accuracy, this module will help user.

2.5 Generated Score:
Here user view the score in %

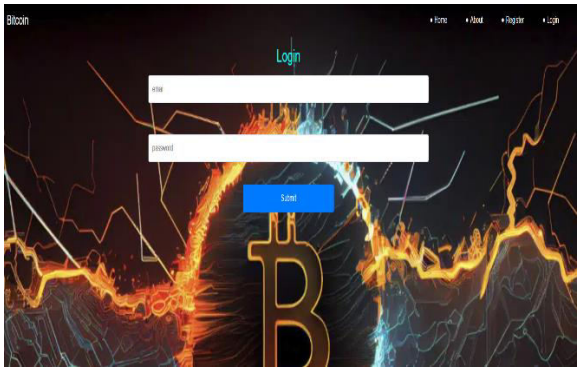
2.6 Generate the Result:
It will give the Prediction or Final Result.

11. OUTPUT SCREENS

Home: Welcome to BABD: A Bitcoin Address behavior Dataset For Pattern Analysis.

Register: The registration page enables new users to create an account so user have to provide name, email, password, confirm password to register.

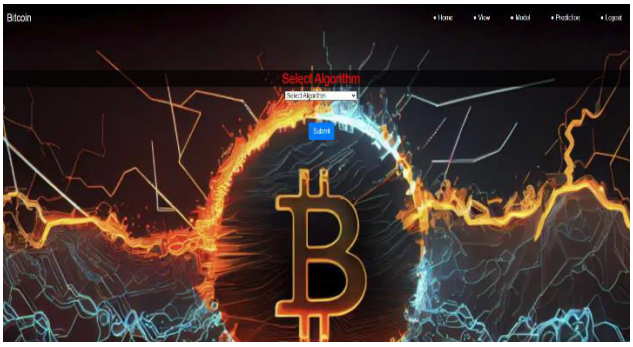
Login: The login page allows registered users to securely access the application. User have to provide register email and password.



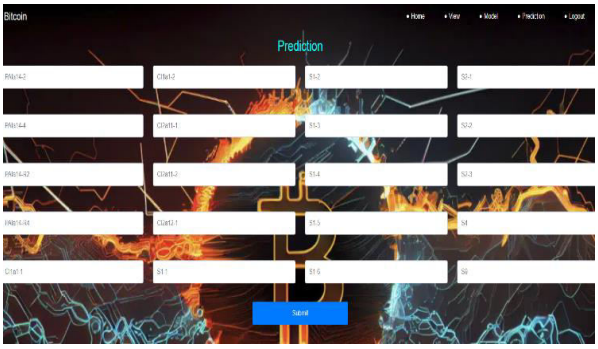
User home: After Login User Enter in User home page so user have another option relates the project.

View: The view data page provides users with access to the data used for Bitcoin Pattern Analysis.

Training: Here user have to select the model and and view the accuracy so that we know which algorithm is giving the highest accuracy.



Prediction: This is the prediction page here user have to provide values of a 20 features and our model is predict the final output.



12. CONCLUSION

In conclusion, our research demonstrates the effectiveness of combining traditional machine learning algorithms with advanced techniques like Stacking Classifier, CNN, and Gradient Boosting for categorizing Bitcoin addresses based on transactional behavior. The comprehensive analysis not only improves predictive accuracy but also provides valuable insights into identifying and mitigating suspicious activities in the Bitcoin ecosystem. By successfully classifying addresses into specific categories, this study contributes to enhancing the security and transparency of cryptocurrency transactions, potentially serving as a critical tool for combating illicit activities and strengthening cybersecurity in the digital financial landscape.

FUTURE ENHANCEMENTS:

Building on the current framework, future work could explore integrating additional machine learning models like Reinforcement Learning and AutoML for improved detection accuracy.

Expanding the dataset to include more diverse and recent Bitcoin addresses, especially those related to emerging threats, would enhance the model's robustness. Incorporating natural language processing (NLP) techniques to analyze transaction notes or metadata could provide further insights. Additionally, real-time detection systems could be developed to identify suspicious activities instantly, improving the security and transparency of Bitcoin transactions and potentially extending this approach to other cryptocurrencies.

13 .REFERENCES:

[1] I. Alqassem, I. Rahwan, and D. Svetinovic, "The anti-social system properties: Bitcoin network data analysis," IEEE Trans. Syst., Man, Cybern., Syst., vol. 50, no. 1, pp. 21–31, Jan. 2020. [11] J. Kim, J. Kim, H. Kim, M. Shim, and E. Choi, ''CNN-based network intrusion detection against denial-of-service attacks,'' Electronics, vol. 9, no. 6, p. 916, Jun. 2020.

[2] P. Nerurkar, D. Patel, Y. Busnel, R. Ludinard, S. Kumari, and M. K. Khan, "Dissecting Bitcoin blockchain: Empirical analysis of Bit- coin network (2009–2020)," J. Netw. Comput. Appl., vol. 177, Mar. 2021, Art. no. 102940.

[3] M. K. Popuri and M. H. Gunes, Empirical Analysis of Crypto Curren- cies. Berlin, Germany: Springer, 2016, pp. 281–292.

[4] D. Ron and A. Shamir, "Quantitative analysis of the full Bitcoin transaction graph," in Proc. Int. Conf. Financial Cryptography Data Secur. Cham, Switzerland: Springer, 2013, pp. 6–24.

[5] L. Serena, S. Ferretti, and G. D'Angelo, "Cryptocurrencies activity as a complex network: Analysis of transactions graphs," Peer-Peer Netw. Appl., vol. 15, no. 2, pp. 839–853, Mar. 2022.

[6] B. Tao, I. W. Ho, and H.-N. Dai, "Complex network analysis of the Bitcoin blockchain network," in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), May 2021, pp. 1–5.

[7] B. Tao, H.-N. Dai, J. Wu, I. W. Ho, Z. Zheng, and C. F. Cheang, "Complex network analysis of the Bitcoin transaction network," IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 69, no. 3, pp. 1009–1013, Mar. 2022.

[8] S. Ranshous et al., "Exchange pattern mining in the Bitcoin transaction directed hypergraph," in Financial Cryptography and Data Security. Sliema, Malta: Springer, 2017, pp. 248–263.

[9] M. Romiti, A. Judmayer, A. Zamyatin, and B. Haslhofer, "A deep dive into Bitcoin mining pools: An empirical analysis of mining shares," 2019, arXiv:1905.05999.

[10] N. Tovanich, N. Soulié, N. Heulot, and P. Isenberg, "An empirical analysis of pool hopping behavior in the Bitcoin blockchain," in Proc. IEEE Int. Conf. Blockchain Cryptocurrency, May 2021, pp. 1–9.