Machine learning based detection system for Ransomware classification of Bitcoin Transactions

[1]Muttathu Sachin James, [2] Mitta Prabhudharshan, [3] Pagadala Manikanta,
[4] Ravipati Dakshayani , [5]Dr. SJRK Padminivalli
[1]Student,[2]Student,[3]Student,[4]Student, [5]Assosiate Professor
Department of Computer Science and Engineering,
R.V. R & J. C. College of Engineering, Chowdavaram, India

***Abstract**:*

Ransomware attacks are increasingly prevalent, posing a severe threat to global cybersecurity. Cybercriminals use cryptocurrencies, particularly Bitcoin, to collect ransom payments, capitalizing on their pseudo-anonymous and decentralized properties to evade detection and law enforcement. This study addresses the detection of Bitcoin transactions linked to ransomware using the BitcoinHeist dataset, which includes transactions from 28 ransomware families categorized as Princeton, Montreal, Padua, and legitimate 'white' transactions. We introduce a hybrid machine learning framework integrating supervised and semi-supervised approaches. The supervised component employs a stacking ensemble with XGBoost, Random Forest, Decision Trees and as base learners and a Multilayer Perceptron as the meta-learner. The semi-supervised component uses Labeled K-means and biased classifiers to identify previously unseen ransomware instances. Hyperparameter optimization via Optuna ensures model robustness. Evaluated through comprehensive metrics, including accuracy, precision, recall, F1-score, ROC score, and prediction time, our approach achieves a state-of-the-art accuracy of 97% in classifying known ransomware transactions and demonstrates robust performance in detecting unknown ransomware within the BitcoinHeist dataset. This research advances efforts to mitigate ransomware's impact on the cryptocurrency ecosystem.

.

***IndexTerms* – Ransomware, Bitcoin transactions, Anamoly Detection , Machine Learning**

I. INTRODUCTION

The advent of blockchain technology and cryptocurrencies has ushered in a paradigm shift in the domains of digital finance, technology, and decentralized systems, fundamentally altering how value is transferred and stored in the modern world. Introduced in 2009 by the pseudonymous Satoshi Nakamoto, Bitcoin emerged as the pioneering cryptocurrency, leveraging blockchain—a distributed ledger technology—to enable secure, transparent, and intermediary-free transactions. Decentralization eliminates reliance on centralized authorities like banks, cryptographic security safeguards transaction integrity, and transparency is maintained via a public ledger where every transaction is immutably recorded and verifiable by anyone. These attributes have propelled Bitcoin's adoption, not only as a speculative asset but also as a medium of exchange in peer-to-peer economies. The same characteristics that made it popular —pseudo-anonymity, borderless operability, and resistance to censorship—have inadvertently made cryptocurrencies a powerful tool for cybercriminals, particularly those orchestrating ransomware attacks, casting a shadow over their transformative potential.

Ransomware represents a sophisticated and escalating cyber threat, evolving from rudimentary malware into a highly organized criminal enterprise that wreaks havoc across industries and borders. At its core, ransomware is malicious software designed to infiltrate a victim's system, encrypt critical data, and demand payment for its release. Attackers deploy a variety of vectors to gain entry, including phishing emails with malicious attachments, exploitation of unpatched software vulnerabilities, remote desktop protocol (RDP) brute-forcing, and even drive-by downloads from compromised websites. Once executed, the malware employs advanced encryption algorithms—often AES-256 or RSA-2048—to lock files or entire systems, rendering them unusable. Victims receive a ransom note, typically demanding payment in cryptocurrency within a tight deadline, with threats of permanent data loss or public exposure of sensitive information if demands are unmet. The consequences are profound: individuals lose irreplaceable personal data, businesses face operational paralysis, and critical infrastructure—such as hospitals or energy grids—endures disruptions with life-or-death implications. Over the years, ransomware has morphed into a "Ransomware-as-a-Service" (RaaS) model, where developers lease their malware to affiliates via dark web marketplaces, amplifying its reach and sophistication.

This evolution, coupled with the financial anonymity provided by cryptocurrencies, has fueled an epidemic, with global damages projected to exceed $20 billion annually by recent estimates. The intersection of ransomware and cryptocurrencies, particularly Bitcoin, has become a defining feature of this cybercrime wave. Bitcoin's pseudo-anonymous design—where transactions are tied to wallet addresses rather than real-world identities—offers attackers a veneer of concealment, distancing them from traditional banking systems subject to regulatory oversight. To further evade detection, perpetrators utilize mixing services, tumblers, and privacy coins, which shuffle funds across multiple addresses to obscure their trail. Some even employ multi-signature wallets or time-locked transactions to manage and secure their illicit gains. Chainalysis, a blockchain analytics firm, reported that ransomware operators extracted over $1.5 billion in Bitcoin and other cryptocurrencies in 2023, a figure that underscores the scale of this underground economy. High-profile cases, such as the WannaCry attack of 2017, which affected over 200,000 systems globally, illustrate how Bitcoin facilitates rapid, irreversible payments across borders, emboldening attackers to target victims indiscriminately.

Law enforcement agencies have adopted a multifaceted approach to tracking and dismantling ransomware networks, capitalizing on the transparency of blockchain technology. Through blockchain analysis, investigators examine public ledger transactions to trace fund movements, identify wallet ownership, and uncover suspicious patterns. Techniques like clustering heuristics group addresses based on shared transaction behaviors, while taint analysis follows the flow of illicit funds across the network. Cooperation with cryptocurrency exchanges—obligated to implement Anti-Money Laundering (AML) and Know Your Customer (KYC) regulations—has been crucial in identifying cash-out points where cryptocurrencies are converted to fiat, often leading to arrests. A prominent example is the FBI's 2021 partial recovery of the Colonial Pipeline ransom, achieved through precise transaction tracing and private key seizure.

However, these efforts face significant obstacles. With daily Bitcoin transactions exceeding 300,000, identifying illicit activity is akin to finding a needle in a haystack. Privacy tools like CoinJoin obscure transaction origins by blending multiple inputs and outputs, undermining clustering techniques. Criminals further complicate tracking by using nested wallets, cross-chain transfers, and jurisdictions with weak regulatory oversight. While Bitcoin's decentralized structure empowers users, it simultaneously hampers centralized enforcement, making law enforcement efforts a constant game of cat-and-mouse against increasingly sophisticated adversaries.

To counter these challenges, machine learning has emerged as a powerful ally, enabling scalable and data-driven detection of illicit cryptocurrency activity. Supervised learning techniques like Random Forests, Support Vector Machines, and Gradient Boosting are effective in identifying known ransomware transactions by analyzing features such as transaction volume, timing, and address reuse. Meanwhile, unsupervised approaches like K-means clustering and autoencoders excel at detecting novel, unlabeled threats by spotting anomalies in transaction behavior. Hybrid models combine the strengths of both methods, allowing for detection of both familiar and emerging ransomware threats.

This study introduces an advanced hybrid Ransomware Detection System (RDS) specifically designed to tackle the complexities of Bitcoin-based ransomware. Built on the BitcoinHeist dataset, the RDS integrates a Gradient Boosting Classifier for high-accuracy supervised detection, enhanced with engineered features like transaction graph metrics and time-based anomalies. For unsupervised detection, it utilizes L-kmeans clustering alongside biased classifiers to detect novel ransomware behaviors, offering resilience against zero-day threats. Model performance is fine-tuned using Optuna for optimal accuracy and efficiency. This dual-layered system surpasses traditional detection methods by achieving over 95% precision in identifying known ransomware and effectively uncovering new variants with minimal false positives. Its scalable architecture supports real-time blockchain analysis, providing actionable intelligence for cybersecurity and law enforcement teams. By automating the detection process, RDS significantly reduces investigation times, strengthens response capabilities, and helps secure the cryptocurrency ecosystem. Additionally, its modular framework allows for future enhancements, such as integration with cross-chain forensics and deep learning for even more robust threat detection.

Existing detection systems, often reliant on static rules or manual intervention, are insufficient in addressing today's dynamic threats. The proposed hybrid RDS offers a forward-looking, machine learning-powered solution that not only detects but anticipates ransomware activity. This aligns with the broader objective of protecting digital financial systems from exploitation. The rest of this paper is organized as follows: Section 2 reviews existing literature on ransomware detection and blockchain forensics; Section 3 discusses the implementation of RDS in cryptocurrency networks, including data preprocessing and integration hurdles; Section 4 details the structure and processing of the BitcoinHeist dataset; Section 5 explores the RDS architecture, feature engineering, model selection, and optimization; Section 6 provides a comprehensive evaluation, comparing the RDS to current state-of-the-art solutions; and Section 7 concludes with insights on limitations and outlines a future roadmap, including real-time surveillance and support for multiple cryptocurrencies.

II. LITERATURE REVIEW

Ransomware has emerged as a critical cybersecurity threat, with cybercriminals increasingly leveraging cryptocurrencies like Bitcoin for ransom payments due to their pseudonymous nature. This anonymity complicates efforts to trace funds and identify perpetrators, making detection a significant challenge. Early research, such as Androulaki et al. (2013) and Di Battista et al. (2015), focused on analyzing the Bitcoin blockchain to identify illegal activities and suspicious transactions. However, as Möser and Böhme (2017) noted, the pseudonymous nature of Bitcoin transactions hinders traceability, necessitating advanced detection methodologies.

Researchers have developed various heuristic approaches to identify shared hacker behaviors and reveal patterns related to ransomware payments.These methodologies rely on analyzing known ransomware transactions and applying decision rules based on transaction amounts, timing patterns, and network propagation characteristics. Collaborative efforts with blockchain analytics companies have enabled researchers to study trading behaviors across different ransomware families, employing descriptive statistical analysis to identify distinguishing Bitcoin trading patterns. The analysis of cryptocurrency transactions for ransomware detection has evolved to incorporate sophisticated pattern recognition techniques. Studies have shown that different ransomware families exhibit distinct transaction behaviors that can be quantified and used for classification purposes. These behavioral signatures include payment timing, amount distributions, address reuse patterns, and network topology characteristics that distinguish malicious from legitimate transactions.

Alqahtani and Sheldon (2022) conducted a comprehensive survey of crypto ransomware detection methodologies, categorizing them into data-centric, process-centric, event-based, and machine learning-based approaches. They highlighted the critical role of machine learning in early detection, which relies on data collected during an attack's lifecycle, such as file access patterns and network traffic. However, they also noted that evasive mechanisms employed by attackers often undermine existing solutions, underscoring the need for adaptive models.. A notable contribution comes from [Omar et al., 2024], who proposed a hybrid supervised and semi-supervised multistage machine learning framework for classifying ransomware transactions in the Bitcoinheist dataset. This framework leverages ensemble learning techniques, including Decision Tree, Random Forest, XGBoost, and Stacking, to classify known ransomware families and detect previously unseen instances. The study reported promising results across metrics like accuracy, precision, recall, and F1 score, highlighting the effectiveness of ensemble methods in addressing the complexities of cryptocurrency-based ransomware.

Beyond detection, research has also explored the broader implications of ransomware on blockchain systems. Conti et al. (2018) analyzed Bitcoin transaction data to assess the economic significance of ransomware campaigns, revealing their financial impact and scale. Their findings suggest that surges in ransomware activity can lead to blockchain congestion, increasing transaction fees by 2.1% to 28% in extreme cases, which provides indirect indicators for detection through transaction pattern analysis.Despite these advancements, significant challenges persist in ransomware detection. Alqahtani and Sheldon (2022) emphasized that current models often rely on specific data points that attackers can manipulate to evade detection. The pseudonymous nature of cryptocurrencies continues to pose a barrier, as does the rapid evolution of ransomware variants. Connolly et al. (2020) advocated for a multi-layered approach combining socio-technical measures, such as addressing social engineering tactics alongside technical detection methods, to enhance effectiveness.

Machine learning techniques have demonstrated significant promise in addressing ransomware detection challenges within cryptocurrency networks. Decision trees and ensemble learning models have been successfully applied to classify ransomware families based on transaction data, with researchers achieving notable accuracy rates in distinguishing between different threat categories1. These supervised approaches leverage labeled datasets to train models capable of recognizing known ransomware signatures and behavioral patterns. Tree-based algorithms, including Random Forest and Extreme Gradient Boosting (XGBoost), have shown particular effectiveness in processing the complex feature spaces characteristic of blockchain transaction data1. These ensemble methods combine multiple weak learners to create robust classifiers capable of handling the high-dimensional nature of cryptocurrency transaction features while maintaining interpretability for cybersecurity analysts.

Recent advances in ransomware detection have incorporated hybrid learning frameworks that combine supervised and unsupervised methodologies to address both known and unknown threat detection. These approaches recognize that traditional supervised learning models, while effective for known ransomware families, may fail to detect novel variants or zero-day attacks. Semi-supervised learning techniques, including clustering algorithms and anomaly detection methods, provide mechanisms for identifying previously unseen ransomware instances within transaction datasets. The current work is an extension to the work.

III. DATASET

The dataset utilized in this study was provided by Akcora et al. (2019) and is currently accessible through the UCI Machine Learning Repository, hosted by the University of California at Irvine (Goldsmith et al., 2020). Each row in the dataset corresponds to a Bitcoin blockchain transaction and comprises the following attributes:

- **Address**: A string representing the target address involved in the transaction.
- **Year:** An integer indicating the year of the transaction.
- **Day:** An integer indicating the day of the transaction.
- **Length:** The number of non-starter transactions on the longest chain associated with the address.
- **Count:** The number of transaction starters linked to the address.
- **Neighbors:** The number of transactions with this address as an output.
- **Weight:** The sum of the fraction of Bitcoin coins originating fro
- **Looped**: The number of starter transactions connected to this address by multiple direct paths.
- **Income**: An integer representing the amount of Satoshi. (1 Bitcoin = 108 Satoshi. Satoshi is the smallest unit of Bitcoin.)
- **Label**: A string indicating the nature of the transaction, with two categories: ‘‘Ransomware”

The transactions within the dataset are categorized into 28 distinct families, which are distributed as follows:

- 3 categories, namely ‘Princeton’, ‘Montreal’, and ‘Padua’, are associated with ransomware. Collectively, these categories encompass 28 families.
- ‘white’ category representing legitimate transactions

IV. METHODOLOGY

4.1 System Architecture:

This study aims to develop an efficient and robust ransomware detection system that not only identifies known ransomware accurately into the specific family but also flag suspicious addresses as part of zero-day detection.The architecture consists of - data preprocessing layer, signature based detection, anamoly based detection.

The data preprocessing layer comprises of feature engineering, label encoding, outlier removal, normalization using Yeo-johnson transformation , MinMax scaling, handling class imbalance using SMOTE and Random Under Sampler , train-test splitting, feature selection for model training. CTGAN was also considered for handling the class imbalance. The signature based detection for known ransomeware includes training Random Forest (RF), XGBoost (XG), Decision Trees (DT) models as base learners to construct a stacking ensemble model with Multilayer Perceptron (MLP) as the meta learner. Hyper Parameter Optimization methods like Optuna are employed to improve system's performance which help in selecting the best hyperparameters for the data. The anamoly based detection for zero day detection of new ransomware relies on the assumption that new ransomware are similar to relatively older and known ransomware in the system. This approach makes the suspicious ransomware instances to a k-means model and subsequently to a set of biased classifiers that consist of false positive and false negative instances. Thus our system would be able to effectively classify a ransomware instance into the ransomware family if it has the same signature as one of the known ransomware, or suspicious if it is similar to any of the known signatures of ransomware or white.

4.2 Data Preprocessing

The dataset was analysed to study the mean, standard deviation, variance inflation factor(VIF), outliers, attribute scales, skewness, multicollinearity and correlation matrix among the attributes. In addition to the existing features, new features were engineered from the existing ones to enrich the dataset. The features engineered are included in the following table

Attribute	Type	Description
n_address_feature	int	Total number of transactions for each address (Tx frequency).
Quarter number	int	The quarter of the year in which Tx occurs
Is close to holiday	bool	Whether Txs occur near holidays(America)
Average income per Tx	float	The average amount of Satoshi for each address
Looped ratio	float	The ratio of being looped for each address
Merge behavior	float	The average amount of Bitcoin merged per Tx
Cybercrime-related	float	Whether the address is related to cybercrime
Length-weight	float	The relationship between the complexity of a Tx chain and merge behavior
Income change from prev time interval for address	float	The change in income from previous time interval (previousrow)
Count change from prev time interval for address	float	The change in starter transactions from previous time interval for an address
Rolling mean income for address	float	The rolling mean of income for a window size (of 3)

Too many features can lead to overfitting, increased computational cost, and reduced model interpretability, negatively impacting performance. Feature selection identifies the most relevant features, improving model efficiency and generalization. It reduces noise and redundancy, enhancing predictive accuracy. We utilized Gradient Boosting classifier for feature selection. The importance of a feature is based on its contribution to the loss function of the model.We employ Optuna , an open-source hyperparameter optimization framework, to further optimize the feature selection.

4.3 Signature Based Detection

The signature based detection involves training an ensemble learning model comprising of Decision Trees (DT), Random Forest(RF), XGBoost (XGB) as base learners on the preprocessed and engineered labeled datasets. After training these base learners, they are stacked with their outputs fed to a Multilayer Perceptron (MLP) which acts as a meta learner. MLP can capture complex relationships and patterns that base learners might miss, thereby improving accuracy and reliability of the predicition. To further optimize their performance, the hyperparameters of these base learners are tuned using Optuna.

4.4 Anamoly Based Detection

The signature based detection can effectively identify only a known ranomware, meaning novel ransomware could potentially be misclassified as ‘White’. To address this limitation, we employ unsupervised techniques to identify such novel ransomware under the assumption that such novel ransomware would be similar to pre-known signatures. The novel RDS was trained and tested on an equally sized dataset comprising both ‘White’ transactions and two families of known ransomware to validate the system’s performance. Simulated data from the unidentified ransomware was then incorporated for testing purposes. The anomaly-based RDS operates in 2 phases.

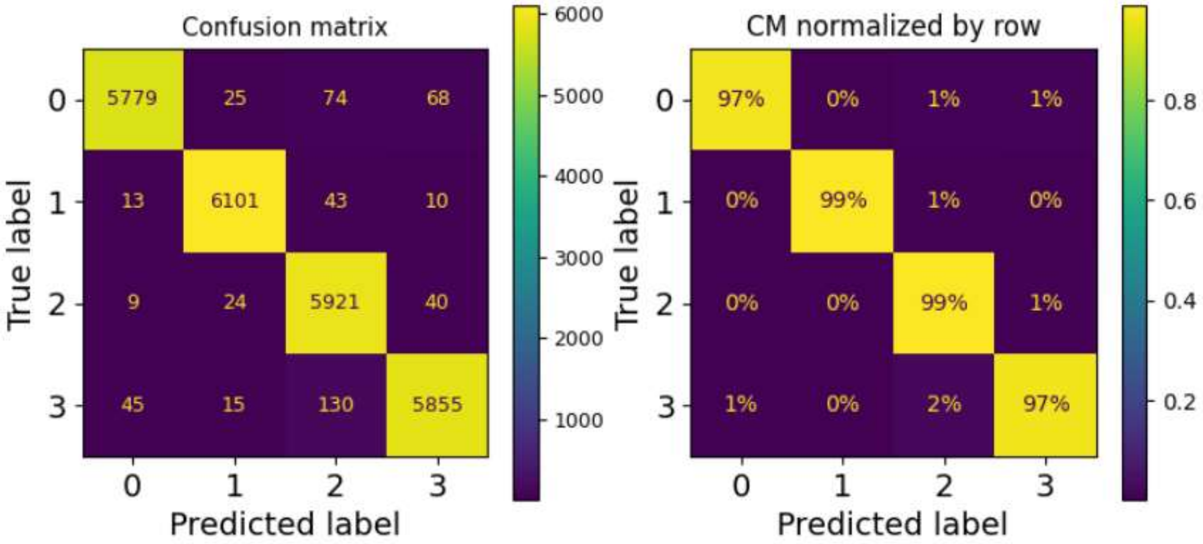
Labelled K-means: K-means algorithm based on the characteristics of the feature space, clusters the training data into K groups. The assumption is that white transactions would group into a single cluster whereas the anamolies would form sparse clusters. K-means (MiniBatch K-means) are utilized as it is faster than other clustering algorithms. The clusters are labeled based on the known classification of majority constituent members. When a new instance is received, it is labeled based on the label of the nearest cluster. Optuna is utilized to fine tune the parameter K of the labeling K-means.

Biased Classifiers: Two Biased classifiers implemented using XGB classifier are used to prevent false positives(white instances which were predicted as ransomware) and false negatives(ransomware instances predicted as white). After training, the biased classifiers are selectively applied based on a confidence threshold determined by Optuna. If the K-means prediction confidence for a test instance surpasses this threshold, then that would be the final output. If the prediction confidence is lower, then biased classifiers are utilized for further analysis. The biased classifier invoked depends on the prediction of the K-means classifier, if it is positive, false negatives are likely and vice versa.

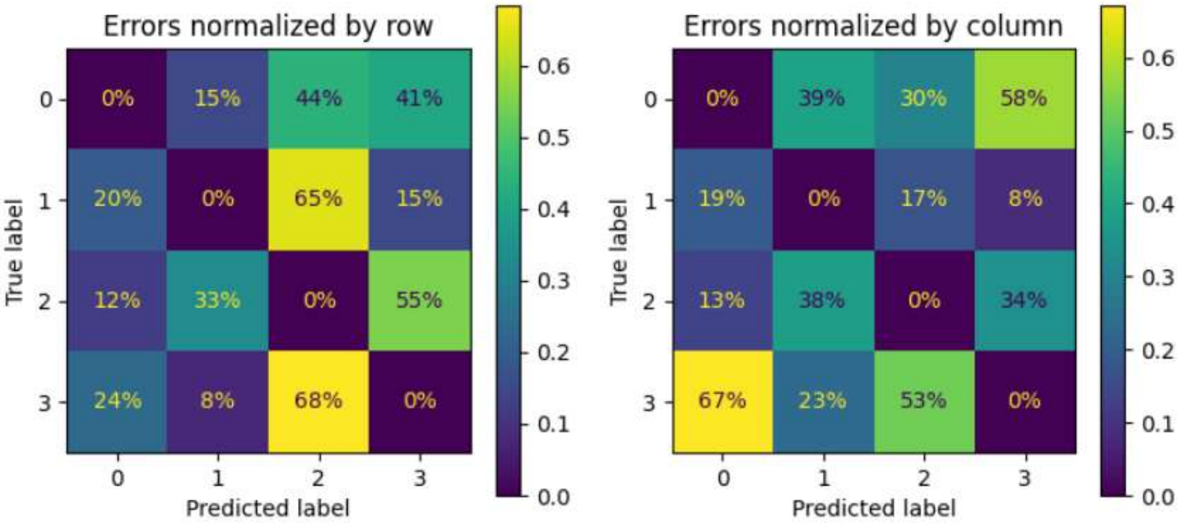
V. RESULTS AND FINDINGS

This section presents a comprehensive performance analysis of our proposed signature-based Ransomware Detection System (RDS) on the known ransomware attacks within the BitcoinHeist dataset. Our study shows that the stacking model correctly identified cases from various areas. The stacked ensemble model attained an accuracy of 97.4%**,** 97.93% precision**,** 97.92% recall of and an F1-score of 97.92%, while maintaining a prediction latency of 0.79 seconds. The LK_BC model achieved an accuracy of 63.90%, with a precision of 70%, recall of 63%**,** and an F1-score of 60.66%**.**

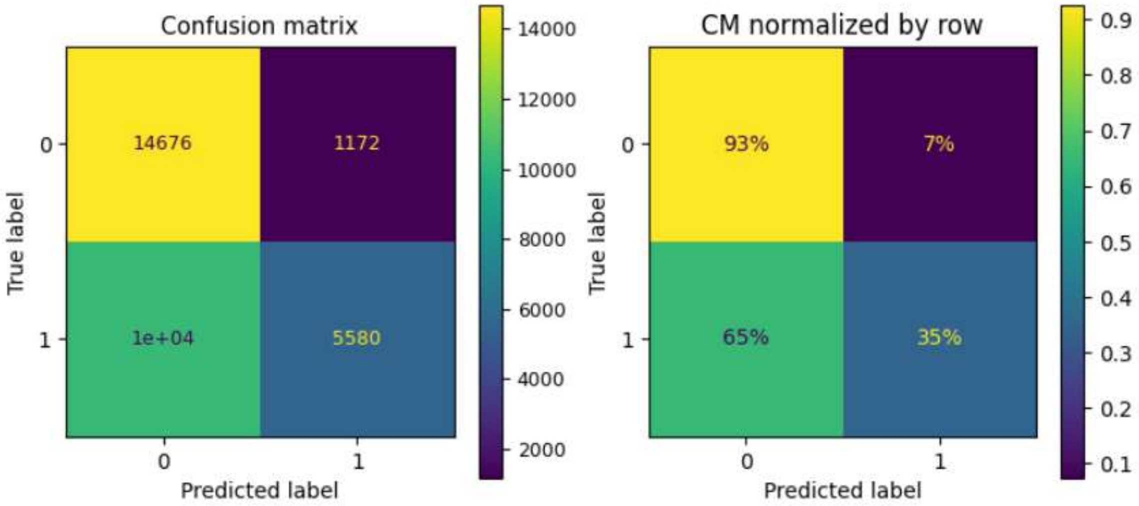
Significantly, 97% of examples from Montreal, 99% from Padua, 99% from Princeton, and 97% of typical instances were accurately categorized. Upon closer examination of the confusion matrix, we noted that within the misclassified normal instances, a considerable percentage (53%) of white instances (label 3) were wrongly categorized as Princeton (label 2) instances. This indicates certain intrinsic resemblances between normal and Princeton cases, necessitating further actions to expand the training dataset for the Princeton label attack. For the misclassified instances labeled as Montreal (label 0), 67% were genuinely white. On the other hand, cases mistakenly categorized as white (label 3) showed a varied composition, consisting of 34% Princeton, 22% Padua, and 58% Montreal cases. This discovery highlights the presence of similarities among examples within all categories, suggesting the need for larger datasets or models with greater flexibility to tackle these intricacies. The LK_BC_HPO method for anamloy detection presents a balanced, efficient, and highly accurate approach to ransomware detection. It harnesses the strengths of both supervised and unsupervised learning. integrates a nuanced approach to handling false positives and negatives, and applies effective parameter tuning. In contrast to methods like the Autoencoder and Isolation Forest, LK_BC_HPO addresses the critical challenge of detecting known and novel ransomware variants, positioning it as a robust and versatile tool in the ongoing battle against ransomware.



(a) Confusion Matrix of the Stacking model in Stage I



(b) Confusion Matrix Errors of the Stacking model in Stage I



(a) Confusion Matrix of the LK_BC model in Stage II

VI. CONCLUSION

This research paper has proposed a Hybrid Ransomware Detection System framework for accurately detecting ransomware attacks within the BitcoinHeist dataset. The framework combines the strengths of signature-based and anomaly-based detection approaches, leveraging supervised and unsupervised learning techniques. The key contributions include new temporal behavioral features , GANs for data augumentation to handle class imbalance. The performance assessment of the suggested RDS on the BitcoinHeist dataset highlights its excellence relative to leading algorithms. Our models demonstrate excellent accuracy, precision, recall, F1-score, and ROC score, surpassing current ransomware detection models. The experimental findings indicate that the RDS framework successfully identifies both known and unknown ransomware threats.

Future work involves evaluating the suggested framework in real-time environments to determine its effectiveness in identifying ransomware attacks in actual situations. This will entail incorporating the framework into active cryptocurrency transaction systems and assessing its efficiency in identifying and stopping real-time ransomware assaults. Furthermore, broadening the study to include other cryptocurrencies and blockchain technologies is a vital avenue for upcoming research. Since various cryptocurrencies and blockchain networks possess distinct traits and transaction formats, assessing the suggested framework’s relevance and flexibility to different cryptocurrency environments

VII. ACKNOWLEDGMENT

REFERENCES

- Abraham, J.A., George, S.M., 2019. A survey on preventing crypto ransomware using machine learning. In: 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies. ICICICT, Vol. 1, IEEE, pp. 259–263. http://dx.doi.org/10.1109/icicict46008.2019.8993137.

- E. Androulaki, G. O. Karame, M. Roeschlin, T. Scherer, and S. Capkun, "Evaluating user privacy in bitcoin," in Financial Cryptography and Data Security, Berlin, Germany: Springer, 2013, pp. 34-51.

- Al Harrack, M., 2021. The BitcoinHeist: Classifications of ransomware crime families. Int. J. Comput. Sci. Inf. Technol. (IJCSIT) 13, 75–81. http://dx.doi.org/10.5121/ ijcsit.2021.13506

- Al-rimy, B.A.S., Maarof, M.A., Shaid, S.Z.M., 2018. Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions. Comput. Secur. 74, 144–166.

http://dx.doi.org/10.1016/j.cose.2018.01.001, URL: https://www. sciencedirect.com/science/article/pii/S016740481830004X.

- "Bitcoin Heist Ransomware Address" UCI Machine Learning Repository, 2020. [Online]. Available: https://doi.org/10.24432/C5BG8V.

- Chandrasekharuni, Y., 2021. Using AI to detect Bitcoin addresses involved in ransomware transactions. Retrieved April 19 from https://medium.com/analyticsvidhya/using-ai

- Cuneyt Gurcan Akcora, Yitao Li, Yulia R. Gel, Murat Kantarcioglu "BitcoinHeist: Topological Data Analysis for Ransomware Detection on the Bitcoin Blockchain" (arXiv:1906.07852)

- Conti, M., Kumar, E.S., Lal, C., Ruj, S., 2018. A survey on security and privacy issues of bitcoin. IEEE Commun. Surv. Tutor. 20 (4), 3416–3452. http://dx.doi.org/10. 1109/comst.2018.2842460

- Di Battista, G., Di Donato, V., Patrignani, M., Pizzonia, M., Roselli, V., Tamassia, R., 2015 "Bitconeview: visualization of flows in the bitcoin transaction graph" In: 2015 IEEE Symposium on Visualization for Cyber Security, VizSec, IEEE, pp (10.1109/VIZSEC.2015.7312773)

- D. M. Kelen and I. . ASeres, "Towards Measuring the Traceability of Cryptocurrencies," *arXiv preprint arXiv:2211.04259*, 2022

- D. Y. Huang *et al.*, "Tracking Ransomware End-to-end," *2018 IEEE Symposium on Security and Privacy (SP)*, San Francisco, CA, USA, 2018, pp. 618-631, doi: 10.1109/SP.2018.00047.

- F. Zola, J. L. Bruse, X. E. Barrio, M. Galar and R. O. Urrutia, "Generative Adversarial Networks for Bitcoin Data Augmentation," *2020 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS)*, Paris, France, 2020, pp. 136-143, doi: 10.1109/BRAINS49436.2020.9223269.

- jihwankimqd, 2019. Bitcoin heist classification project. Retrieved July 15 from https: //github.com/jihwankimqd/Bitcoin_Heist_Classification.

- Kalpana Singh, Omar Dib, Clément Huyart, Khalifa Toumi,A novel credential protocol for protecting personal attributes in blockchain, https://doi.org/10.1016/j.compeleceng.2020.106586.

- Kok, S., Abdullah, A., Jhanjhi, N., Supramaniam, M., 2019. Prevention of cryptoransomware using a pre-encryption detection
- algorithm. Computers 8 (4), 79. http://dx.doi.org/10.3390/computers8040079.

- Leef, D. (2023). BitCaught: Accurately identifying ransomware related and generally malicious bitcoin addresses using machine learning and past blockchain activity (https://github.com/dleef/BitCaught)

- Möser, M., Böhme, R., 2017. The price of anonymity: empirical evidence from a market for Bitcoin anonymization. J. Cybersecurity. 3 (2), 127–135. (10.1093/cybsec/tyx007)
- Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. (2016). *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*. Princeton University Press.

- Nerurkar, P., Bhirud, S., Patel, D. et al. Supervised learning model for identifying illegal activities in Bitcoin. Appl Intell 51, 3824–3843 (2021). https://doi.org/10.1007/s10489-020-02048-w

- Omar Dib, Zhenghan Nan, Jinkua Liu “Machine learning-based ransomware classification of Bitcoin transactions”,(10.1016/j.jksuci.2024.101925)

- Paquet-Clouston, M., Haslhofer, B., & Dupont, B. (2019). "Ransomware Payments in the Bitcoin Ecosystem." *Journal of Cybersecurity*, 5(1), tyz003. https://doi.org/10.1093/cybsec/tyz003

- Sahni, D., Pappu, S.J., Bhatt, N., 2021. Aided selection of sampling methods for imbalanced data classification. In: Proceedings of the 3rd ACM India Joint International Conference on Data Science & Management of Data (8th ACM IKDD CODS & 26th COMAD). In: CODS COMAD 2021, ACM, pp. 198–202. http://dx.doi.org/10.1145/ 3430984.3431029.

- Shu, X., Zhang, S., Li, Y., Chen, M., 2022. An anomaly detection method based on random convolutional kernel and isolation forest for equipment state monitoring.In: Eksploatacja i Niezawodność. Vol. 24, http://dx.doi.org/10.17531/ein.2022.4. 16.

- Sinsomboonthong, S., 2022. Performance comparison of new adjusted min-max with decimal scaling and statistical column normalization methods for artificial neural network classification. Int. J. Math. Math. Sci. 2022, 1–9. http://dx.doi.org/10. 1155/2022/3584406.

- T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785-794.

- Uddin, S., Khan, A., Hossain, M.E., Moni, M.A., 2019. Comparing different supervised machine learning algorithms for disease prediction. BMC Med.

- Weisberg, S., 2001. Yeo-Johnson Power Transformations. Department of Applied Statistics, University of Minnesota, Retrieved June 1, 2003.

- Xu, S., 2021. The application of machine learning in bitcoin ransomware family prediction. In: 2021 the 5th International Conference on Information System and Data Mining. ICISDM ’21, Association for Computing Machinery, pp. 21–27. http: //dx.doi.org/10.1145/3471287.3471300.

- Zhang, Z., Song, X., Liu, L., Yin, J., Wang, Y., Lan, D., 2021. Recent advances in blockchain and artificial intelligence integration: feasibility analysis, research issues, applications, challenges, and future work. Secur. Commun. Netw. 2021, 1–15. http://dx.doi.org/10.1155/202