Machine Learning Crypto Ransomware Detector

Travis Counihan
Physics Department
California State University, Fullerton
Fullerton, United States of America
tcounihan@fullerton.edu

David Nguyen
Dept. of Computer Science
California State University, Fullerton
Fullerton, United States of America
davidvn6@csu.fullerton.edu

Richard Perez
Dept. of Computer Science
California State University, Fullerton
Fullerton, United States of America
r.perez4@csu.fullerton.edu

***Abstract*—As the use of cryptocurrency continues to experience rapid growth among institutional investors and the general public, the threat of ransomware attacks also continues to rise. In order for cryptocurrency adoption to experience continued growth, the safety of these transactions need to be comparable to traditional forms of money. Ransomware attacks pose a significant security risk to everyday users and must be addressed if cryptocurrency is to become the dominant currency. For this project, we will utilize machine learning algorithms to properly detect unusual transactions and use it to identify ransomware payments. There are 2,916,697 data points with 10 features. The features include the address of the transaction, income, looped, length, label white/black (either ransom or regular transaction).**

***Index Terms*—cryptocurrency, supervised learning, ransomware detection, classification**

I. Introduction

As more public services accept cryptocurrency as a valid form of payment, it is becoming clear that digital currencies are here to stay. Cryptocurrency has existed for nearly two decades and will continue to grow and adapt. However, security concerns arise especially with the most valued coin, Bitcoin, reaching an all time high of over 120,000. With this valuation, the need for strong transaction identification is required to properly classify legitimate transactions apart from ransomware related threat transactions. Cryptocurrency does not require a third party intermediary for transactions and is secured through cryptographic protocols [1]. These transactions are irreversible and are recorded on a blockchain, which serves as a decentralized ledger. Bitcoin's blockchain technology authenticates and verifies each block of transactions, ensuring the integrity and continuity of the entire network. Since these systems process anonymous transfers, ransomware attackers aim to exploit blockchain transactions for their own gain. Within a two year span, 2023 and 2025 cryptocurrency thefts and transactions have contributed to a staggering ten billion dollar sum of financial cyber crime [2]. Using the Bitcoin Heist transaction data set, classification of transaction types would greatly improve security within the cryptocurrency environment.

II. Background

Ransomware has emerged as a prominent cyber threat and has been increasingly associated with cryptocurrency in recent years as a means of ransom collection. Cryptocurrency is being used instead of traditional means of financial payments because the use of blockchain based transactions provides users with pseudonymity that makes transaction tracing difficult [5]. Bitcoin is currently the largest cryptocurrency by market capitalization and is commonly associated with these illicit activities due to its widespread adoption. Ransomware attacks typically involve malicious software encrypting a victim's files and the release of those files in exchange for payment like Bitcoin [6]. Once these payments are sent, they are usually split multiple times and routed through numerous different blockchain addresses to prevent authorities from tracking their destinations [7]. Recent research, employing machine learning techniques, have shown that these blockchain transactions can be carefully examined to distinguish normal and legitimate cryptocurrency transactions from malicious ones [8]. This is because these transactions still contain many features that can be analyzed for patterns such as transaction volume and coin splitting intervals [9]. As ransomware operations continue evolving, employing the use of machine learning based detection methods provides an approach for improving blockchain security.

III. Data Set

The dataset used to train the different machine learning models, Bitcoin Heist Ransomware Address, is obtained from the UCI Machine Learning Repository [10]. The dataset contains 2,916,697 bitcoin transactions that were recorded between January 2009 to December 2018. The authors of this dataset formed a Bitcoin graph by aggregating daily transactions from the Bitcoin network using 24-hour intervals. The ransomware addresses were mainly adopted from the Montreal, Princeton, and Padua studies. Each instance in the dataset consists of ten features which are described as follows:

- Address - Represented as a string value indicating the bitcoin address
- Year - Represented as an integer value indicating the transaction year
- Day - Represented as an integer value indicating the day (1 to 365) of the transaction year
- Length - Represented as an integer value indicating how many times coins were split and moved through new addresses to hide their origins
- Weight - Represented as a float value indicating merge behavior

- Count - Represented as an integer value indicating the number of transactions sent to an address
- Looped - Represented as an integer value indicating the amount of transactions that split and eventually merge into a single address
- Neighbors - Represented as an integer value
- Income - Represented as an integer value indicating the total Bitcoin amount received by the address
- Label - Represented as a string value indicating the ransomware family or "white" if is not known to be ransomware

IV. Data Set Processing

The dataset is partially preprocessed. For example, the authors reported that no values are missing. As a precaution, we incorporated a step to remove any values that might be missing. To address outliers, we applied a clamp transformation that restricted each numerical feature to fall within two standard deviations of its mean. Feature engineering enhances the data relevance for model training. This includes applying log transformations to highly skewed variables and constructing new ratio-based features to capture more meaningful relationships among the original features. Feature extraction is additionally performed by removing features such as year and day, as they did not contribute any meaningful data. Since the target variable consists of categorical labels, we apply integer encoding to convert these labels into numerical form. Classes with fewer than two samples are also excluded, otherwise splitting of the data into train and test sets would not give a fair split. We use an 80/20 train-test split with random sampling, and stratification is applied to preserve the original class distribution across the two sets. The engineered features are defined as follows:

- log_income: log-transform of raw income to reduce right skew.
- log_count: log-transform of transaction count.
- log_weight: log-transform of merge intensity.
- income_per_trx: average income per transaction.
- merge_ratio: average merge intensity per transaction.
- looped_ratio: fraction of transactions that are looped.
- is_looped: (Binary) 1 if looped 0, else 0 — any looping activity?
- high_income: (Binary) 1 if income 100M satoshis (1 BTC), else 0 — large payments common in ransom.
- log_neighbors: reduces skew in neighbor count.
- encoded_label: integer encoded label for model handeling

V. Methodology

To evaluate ransomware detection performance on the Bitcoin transaction dataset, a combination of supervised and unsupervised machine learning techniques are employed. Supervised models predict whether a Bitcoin address is related to ransomware, while unsupervised clustering examine the structure of the dataset and identify potential groupings beyond the provided labels. All models are evaluated using the engineered features derived during preprocessing.

A. KNN

K-Nearest Neighbor is a supervised learning algorithm that labels new data points by comparing the data points feature values to the closest data point in the training set [3]. It can compute the distance in various ways, but euclidean distance is optimal because it measures feature relationship on proximity. The parameter k value decides the amount of neighboring data points to be evaluated with. A greater k value reduces variance but increases bias [3]. To determine the best value for k, with balance variance and bias an elbow plot is constructed to map the correlation between the error rate and different k values. This algorithm is selected because our data has low dimensionality so, computing the distance similarity has easy and meaningful interpretability.

B. Random Forest

Random forest is a supervised learning algorithm for classification and regression that aggregates the predictions of different decision trees during training for a single resulting value [4]. In classification the trees vote for the most frequent class, while as in regression the prediction is averaged. An advantage of random forests is that the combination of decision trees, each trained with random data subsets, will prevent overfitting in the results and improve generalization [4]. Feature parameters which impacted the algorithm performance are the number of trees, its maximum depth, the number of splits for each node, and feature weights. The weighted class is applied to address the imbalance in the dataset, as an estimated 99 percent of transaction data belonged to a single class. We are using this algorithm because of it's ability to capture complex relationships in the data, reduce overfitting, and performs well on large datasets.

C. K-Means Clustering

K-Means Clustering is an unsupervised learning algorithm that separates data into K clusters by assigning each data point to the nearest centroid using a distance metric [11]. The centroids are then updated using the mean of the points in each cluster, repeating iteratively until the algorithm converges or a stopping threshold is reached [11]. Since the algorithm utilizes Euclidean distance to calculate centroid proximity, all numerical features are standardized to have a mean of 0 and standard deviation of 1 to prevent a single feature from dominating the clustering scale. As seen in Figure 7 and Figure 8, the number of clusters for this model are chosen through the use of an elbow plot depicting inertia against various K values. The k value is 4 for the engineered full features model as it is the point where the rate of decrease in inertia began to plateau. This algorithm enables exploration of the data's underlying structure and identification of hidden patterns without relying on the original class labels. Previous work has applied K-Means Clustering to blockchain transactions to detect anomalous or potentially fraudulent activity [12]. Utilizing K-Means Clustering for this dataset will allow us to determine whether Bitcoin addresses associated with

ransomware will naturally form distinct clusters compared to legitimate transactions.

D. Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm which falls under classification and regression problems. It works by identifying or defining the optimal hyperplane that separates data points of one class from another and maximizes that hyperplane as much as possible while trying to minimize the classification error. SVMs can handle both binary and multi class classification problems. For multiclass problems, techniques such as one-versus-all or all-versus-all classify a data point into one of the multiple classes. Because of its ability to handle multi class classification situations, this makes an SVM model perfect for the Bitcoin Heist Data. Because the data is classified into different ransomware families, such as Cryptolocke, Wannacry, White, this makes predicting the class of the data from its features much more manageable and can better generalize once properly trained.

VI. Results

Looking at the results, deductions of the model performance can be made more clearly. Between the Random Forest and KNN models, it is evident from Figure 1 that the Random Forest model performs better than the KNN model, as it has a higher true positive rate. The same result can be seen from the precision and recall in Figure 2. The PCA graphs show how well the KMeans model performed for classifying the different ransomware classes. Due to the large overlap between cluster boundaries in the high density areas of the graphs, it yielded a silhouette score of -0.4, the KMeans model fails to successfully identify ransomware threat categories. There is a noticeable difference in the general shape of the PCA plots. The full features plot shows a linear positive trend while the reduced feature plot shows a negative linear trend. This difference is most likely due to noise or variance that the reduced features cannot see without the full features, therefore causing the alternate trend. For the SVM model, the confusion matrix seen in Figure 5 shows a poor true non-ransomware outcome, but a fairly accurate ransomware outcome, leading to a high rate of false positives.

References

[1] K. Bholane,"Pros and cons of cryptocurrency: A brief overview."SSRN,pp.2–8, 2025.

[2] A. Alizadeh,"Global Cyber Financial Crimes 2023-2025: Cryptocurrency Heists, Banking Breaches, and the Convergence of Digital Theft with Terrorist Financing."SSRN,pp.3,Oct.2025.

[3] Z. Zhang,"Introduction to machine learning: k-nearest neighbors." Ann Transl Med.,vol.4,no.11,p.2,Jun.2016,doi: 10.21037/atm.2016.03.37.

[4] A.A.J.Al-Abadi,M.B. Mohamed, and A.Fakhfakh,"Enhanced Random Forest Classifier with K-Means Clustering (ERF-KMC) for Detecting and Preventing Distributed-Denial-of-Service and Man-in-the-Middle Attacks in Internet-of-Medical-Things Networks."Computers,vol.12,no.12,p.262,2023,http://www.mdpi.com/2073-431X/12/12/262

[5] "What Is Crypto Ransomware?," Check Point Software, 2025. https://www.checkpoint.com/cyber-hub/ransomware/what-is-crypto-ransomware/

[6] "Ransomware," FBI, 2025. https://www.fbi.gov/how-we-can-help-you/scams-and-safety/common-frauds-and-scams/ransomware

[7] Financial Crimes Enforcement Network, "Financial Trend Analysis: Ransomware," FinCEN, Oct. 2021. https://www.fincen.gov/system/files/2021-10/Financial

[8] K. Wang, J. Pang, D. Chen, Y. Zhao, D. Huang, C. Chen, and W. Han, "A large-scale empirical analysis of ransomware activities in Bitcoin," ACM Trans. Web, vol. 16, no. 2, Art. 7, Dec. 2021, doi: 10.1145/3494557

[9] Y.-J. Lin, P.-W. Wu, C.-H. Hsu, I.-P. Tu, and S.-W. Liao, "An Evaluation of Bitcoin Address Classification Based on Transaction History Summarization," in Proc. IEEE Int. Conf. Blockchain and Cryptocurrency (ICBC), Seoul, Korea, 2019, pp. 302–310, doi: 10.1109/BLOC.2019.8751410.

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

[11] "K means Clustering – Introduction," GeeksforGeeks, Aug. 22, 2025. https://www.geeksforgeeks.org/k-means-clustering-introduction/

[12] S. E. Vadakkethil Somanathan Pillai and G. S. Nadella, "Blockchain fraud detection using unsupervised learning: Anomalous transaction patterns detection using K-means clustering," in Proceedings of the ACM, Aug. 2024, doi: 10.1145/3675888.3676080.

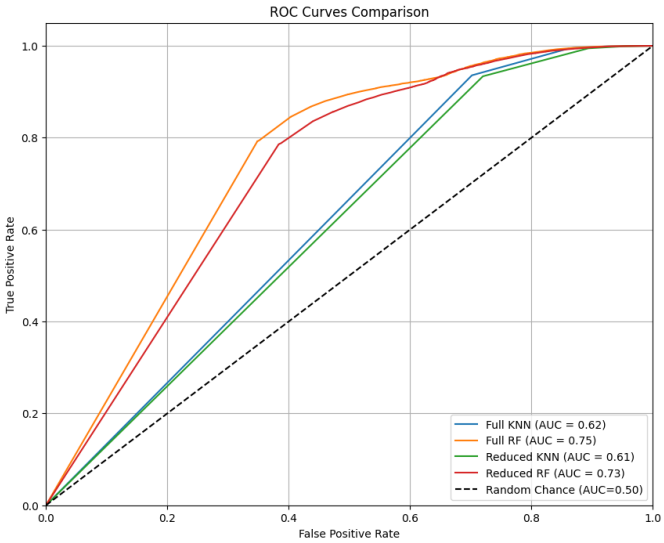

Fig. 1. ROC Curve of Full and Reduced features for KNN and Random Forest

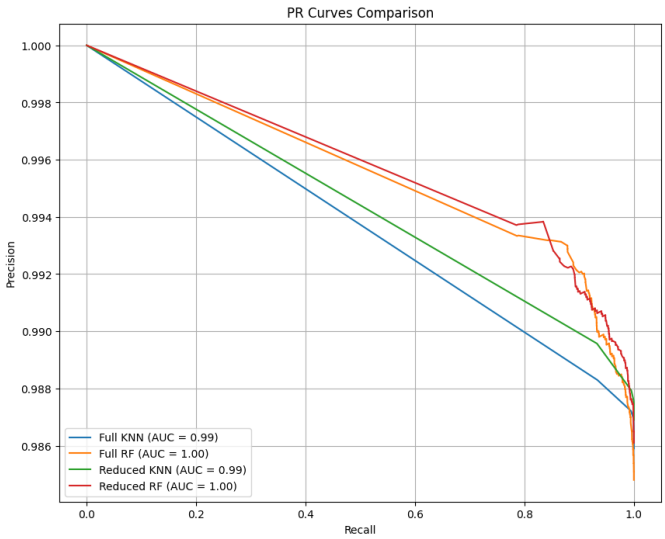

Fig. 2. PR Curve of Full and Reduced features for KNN and Random Forest

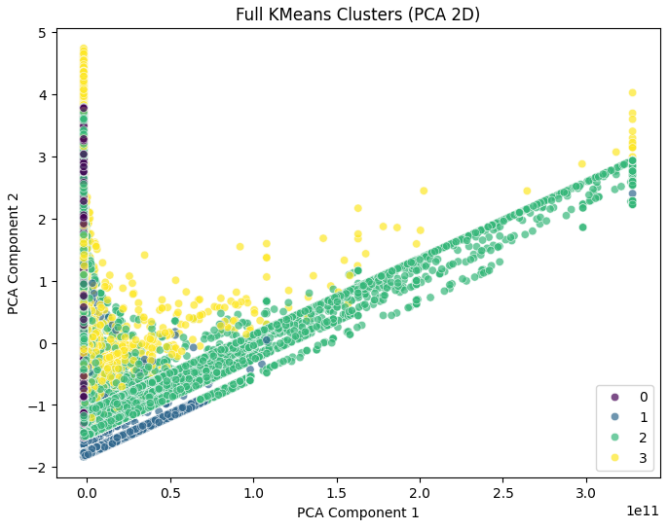

Fig. 3. PCA for full features showing clustering ability

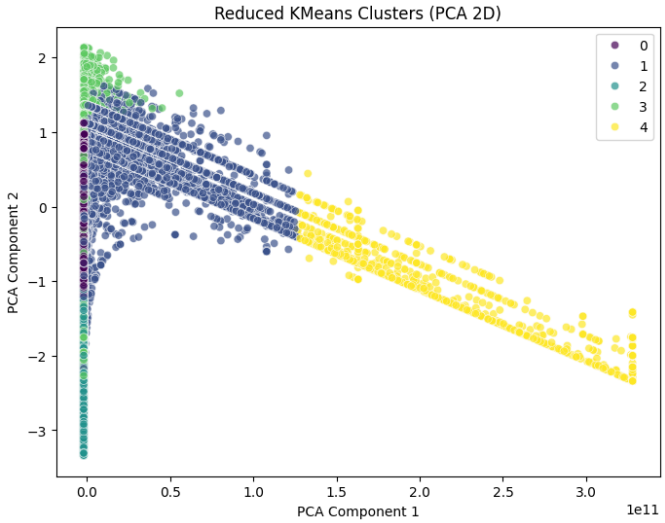

Fig. 4. PCA for reduced features showing clustering ability

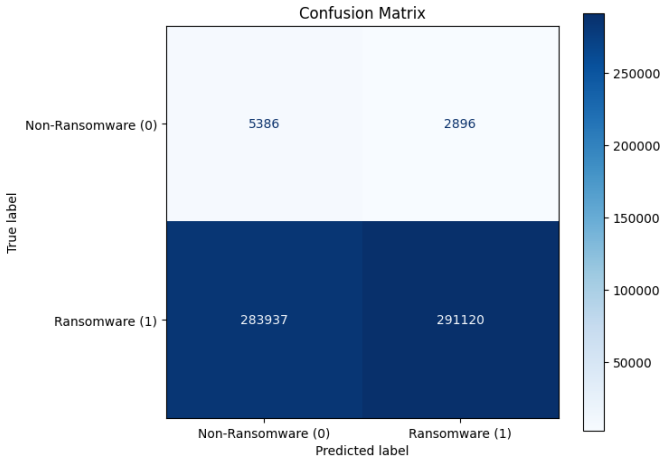

Fig. 5. Confusion Matrix of SVM model

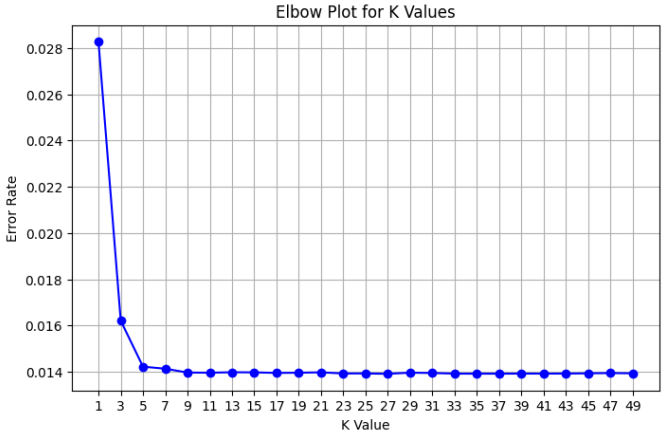

Fig. 6. KNN Elbow Plot

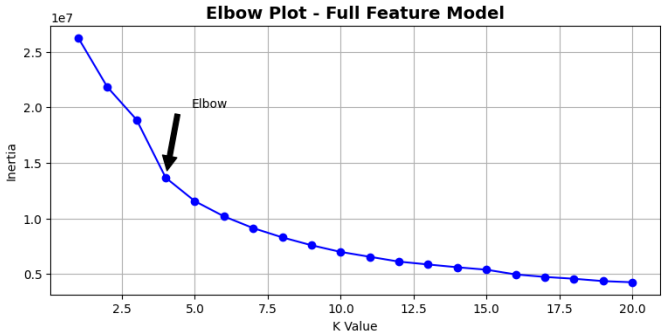

Fig. 7. K-Means Clustering Full Feature Elbow Plot

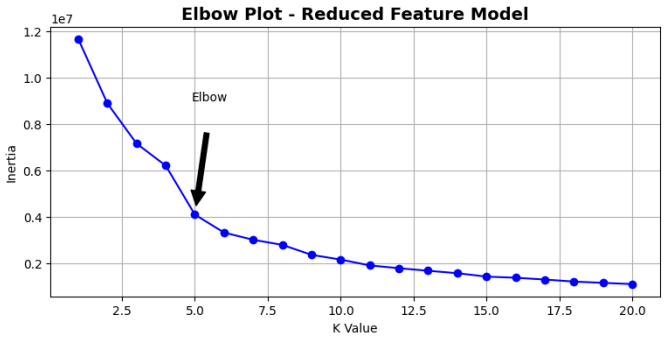

Fig. 8. K-Means Clustering Reduced Feature Elbow Plot