Journal of Ecohumanism
2025
Volume: 4, No: 2, pp. 2091 – 2111
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i2.6604

Detecting Cryptocurrency Scams in the USA: A Machine Learning-Based Analysis of Scam Patterns and Behaviors

Bimol Chandra Das[1], Babul Sarker[2], Amit Saha[3], Kanchon Kumar Bishnu[4], Biswajit Chandra das[5], M Saif Sartaz[6], Muhammad Hasanuzzaman[7], Md Majedur Rahman[8], Medhat Mohiuddin Khan[9]

Abstract

The exponential growth of cryptocurrency implementation in the USA has brought with it a surge in correlated risks, particularly in the form of scams that exploit the relative novelty and complexity of digital currencies. The primary objective of this study was to develop machine algorithms for identifying fraud trends in cryptocurrency transactions. By employing complex analysis, this research project attempted to identify certain trends and behaviors that fall under a variety of scams, providing a platform for effective detection and counter-strategies. This study will have a definite objective in terms of Bitcoin, Ethereum, and other high-profile cryptocurrencies in America when it comes to scam analysis. The scam-related transaction dataset comprised in-depth information regarding suspicious fraud activity in the cryptocurrency environment, such as a specific ID for a transaction, timestamps, values for transactions, and labels distinguishing between suspicious and legitimate activity. A variety of proven models were selected such as Logistic Regression, Random as well Multinomial Naive Bayes, where each model had its respective weaknesses and strengths. The Random Forest algorithm attained the highest accuracy, nearing perfection which underscores its robustness and reliability in classifying both legitimate and fraudulent reports. To effectively counter fraud in cryptocurrencies, U.S. policies must be strengthened with a merger of machine intelligence in them. Regulatory agencies have to work towards developing a system that encourages exchanges to utilize complex analysis for fraud detection, perhaps in terms of reduced compliance burden for entities with effective anti-fraud controls in position. Leveraging AI insights can go a long way in supporting investigations into scams in cryptocurrencies conducted by governments. By utilizing machine algorithms trained with datasets of past scams, governments can monitor and follow illicit fund flows through the blockchain with ease.

Keywords: *Cryptocurrency, Machine Learning, Fraud Detection, Scams, Bitcoin, Ethereum, U.S. Market, Scam Patterns, Predictive Analytics.*

Introduction

Background and Context

In recent years, America experienced an unprecedented boom in cryptocurrency use, with companies and millions of investors dealing in cryptocurrencies. All these have been spurred by a combination of factors, including hope for decentralized finance, the attraction of speculative investments, and the growing acceptance of cryptocurrencies as a payment system [6]. Nevertheless, with its high growth, the sector has not been exempted, and with it, rogue entities have flocked in, and with them, a sharp rise in scams in cryptocurrencies. All these scams have taken numerous forms, including Ponzi schemes, scams in exchanges, and phishing, all designed to collect innocent investors with no expertise in working with new technology in cryptocurrencies [9].

[1] Master of Science in Business Analytics, Trine University, Indiana, USA, Email: bdas23@my.trine.edu, (Corresponding Author)
[2] Master of Science in Business Analytics, Trine University, Indiana, USA.
[3] Computer and Information Science, University of Michigan Dearborn.
[4] MS in Computer Science, California State University Los Angeles.
[5] BS in Computer Science, Los Angeles City College
[6] Electrical Engineering and Computer Science, Florida Atlantic University
[7] Master's in Strategic Communication, Gannon University, Erie, PA, USA
[8] Master of Business Administration (MBA), Trine University, Indiana
[9] Master of Business Administration, Westcliff University

According to Rahman et al. [16], the rise in scams in the cryptocurrency environment is most disorienting in consideration of such virtual assets' decentralized and anonymous nature. Unlike traditional financial networks, with a regulating environment and supervision, the cryptocurrency marketplace is relatively unregulated, and, as such, a fertile breeding ground for scams. Fraudsters apply sophisticated techniques in luring victims, including social engineering, counterfeit websites, and fraudsters' ads, whose unveiling can become a challenge for both regulators and victims alike. On top of that, anonymity through blockchain technology makes tracking illicit transactions and punishing offenders even more arduous [18].

Problem Statement

As per Sumsuzoha et al. [19], the dynamic and ever-evolving nature of scams in cryptocurrencies creates a significant challenge in terms of detection and prevention. Cryptocurrency scammers regularly update and adapt their modus operandi in a way that exploits new trends and technology, and regulators and law enforcers have difficulty keeping pace with them. Traditional fraud detection tools, whose premise for analysis tends to depend on older information and predetermined algorithms, cannot detect new and emerging forms of scam activity that don't follow traditional trends. That vulnerability underlines the necessity for a fraud detection mechanism that can effectively scan through massive volumes of transaction information produced in the environment of cryptocurrencies. Furthermore, both its high level and velocity of transactions make them even more challenging to detect. Buiya et al [5], reported that cryptocurrency trading is 24/7, with a constant tide of millions of transactions taking place daily through many platforms. All of that constant information can overwhelm current tools for detection, creating a lag in discovering and acting concerning suspected scams. In addition, no one controls it, and thus, victims have no recourse, heightening the need.

Research Objective

The primary objective of this study is to develop machine algorithms for identifying fraud trends in cryptocurrency transactions. By employing complex analysis, this research project will try to identify certain trends and behaviors that fall under a variety of scams, providing a platform for effective detection and counter-strategies. In concrete terms, the study will focus on an analysis of information in leading cryptocurrencies Bitcoin Ethereum, and other important virtual assets, to find trends that can represent fraud activity. In addition to developing detection algorithms, this work seeks to produce actionable intelligence that can inform security in America's cryptocurrency marketplace. With an awareness of fraudsters' root behavior and techniques, investors, exchanges, and regulators can implement wiser security implementations in an attempt to protect against scams. Ultimately, it seeks to contribute towards a safer and more secure environment for cryptocurrencies, with heightened trust and confidence for participants.

Scope and Relevance

This study will have a definite objective in terms of Bitcoin, Ethereum, and other high-profile cryptocurrencies in America when it comes to scam analysis. With a definite target in terms of such high-profile virtual assets, the study will have a deep analysis of the most applicable scam trends for U.S.-based investors and trading participants. The use of fraud analysis through machine learning algorithms will enable early fraud detection of sophisticated and evolving scam trends that can go undetected in any other manner. The significance of such a study is in its potential to make the overall security and integrity of the marketplace for cryptocurrencies even enhanced. With the growing industry, information derived from such a study can be utilized in creating a fraud detection mechanism more efficiently, and subsequently, less fraud can be perpetrated and investors can save financial loss. Besides, such a study can serve as a useful tool for regulators and policymakers in creating effective frameworks and guidelines for regulating the marketplace for cryptocurrencies in a better manner. By resolving a key issue of fraud detection, such a study helps contribute to a larger discussion about the responsible and sustained development of the marketplace for cryptocurrencies in America.

ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i2.6604

Literature Review

Cryptocurrency Deceptive Practices and Fraud

Argarwal et al. [1], asserted that the emergence of cryptocurrencies not only revolutionized the financial sphere but also witnessed an outburst of scams and frauds in its aftermath. As cryptocurrencies gained widespread acceptance, scammers have exploited a lack of governing entities and technological complexity in blockchain technology to introduce a range of scams and frauds. Some of the most common types of scams include investment scams, rug pulls, and pump-and-dump scams. Investment scams most frequently entail high returns for investments in ventures in cryptocurrencies that sometimes don't even have any existence, and sometimes ventures poorly constructed. Investment scams most frequently utilize social networks and web forums for marketing their scams, recruiting investors with convincing marketing strategies. For instance, in 2020, "Bit Connect" gained widespread prominence, with marketers providing investors with tremendous returns through a trading bot that, in theory, earned its owners a profit through algorithmic trading. Bit Connect, in reality, was a Ponzi scheme that scammed investors out of approximately $1 billion before its failure [2].

Ali et al. [4], argued that rug pulls are yet another underhanded form of fraud in DeFi. In a rug pull, a collection of a cryptocurrency project's developers simply vanishes when a lot of investment flows in, and investors can sometimes even lose everything. There was a big one with the "Squid Game Token" in 2021 when its developers simply vanished when its price soared through the roof, and investors were left with nothing but valueless assets. These scams speak volumes about investors in unregulated crypto and about how easy it is for villains to manipulate markets. Kabila et al. [7] contended that pump-and-dump schemes in the virtual marketplace have not been new for long. In a scheme, a low-value-trade cryptocurrency is artificially increased in value through coordinated buying and manipulative advertisement, and then at its peak, its owners sell out, with innocent investors holding valueless assets in their hands. Cryptocurrency "Dogecoin" is a perfect example in which social personalities have been a part of a case of pump-and-dump, and its price saw a sharp rise purely out of hype and not value.

Case studies of high-profile scams in cryptocurrencies in America go a long way in defining emerging trends in fraud in such a field. One of them is the "One Coin" fraud, in which investors were guaranteed education and investment in a then-emerging, revolutionary form of cryptocurrency, with an estimated loss of $4.4 billion [10]. It operated worldwide but with a significant impact in America, with its founders utilizing a multi-level model that exploited investors' trust, culminating in a general financial loss for many investors. "Centra Tech" fraud, in which $25 million in an initial coin offering (ICO) was reaped through a claim of collaborations with big financial entities, then sued for fraud by the U.S. Securities and Exchange Commission (SEC) and culminating in criminal indictments, is yet another high-profile case. These examples serve to present a picture of diversity in scams in the cryptocurrency sphere and illustrate the necessity for efficient preventive and detection tools. Cryptocurrency transactions' anonymity and complexity make them most resistant to controls, and such an environment encourages fraud and scams with ease [8].

Traditional Scam Detection Methods

According to Manful & Hashford [12], in conventional financial networks, fraud discovery has long been achieved through rule-based methodologies that rely on past information and predefined thresholds for discovering suspicious activity. In such methodologies, thresholds for numerous statistics, such as values and volumes of transactions, are predefined, and when such thresholds are attained, an alarm is triggered for investigation. For example, credit card companies utilize algorithms that monitor transaction behavior and identify suspicious activity, such as purchases in geographically disparate locations over a short timeframe [13].

However, these traditional approaches have several weaknesses, most prominently in fickle cryptocurrency environments. The high-speed and high-volatility trading environment in cryptocurrencies makes it unfeasible for rule-based frameworks to react to new trends in scams emerging promptly [15]. Since scams in cryptocurrencies entail new methodologies that evolve over a period, rule-based frameworks become outdated, and new ones have to be formulated in a loop, with little opportunity for them to respond and react to new scams promptly. Besides, cryptocurrencies' anonymity and decentralized feature make it unfeasible to reverse-engineer transactions to a specific entity, and many traditional countermeasures become ineffectual [14].

Moreover, traditional fraud detection tools have a strong bias towards producing high volumes of high-false positive, innocent transactions being inaccurately marked suspicious. That can result in customer dissatisfaction and loss of confidence in financial institutions. In financial cryptocurrency markets, with seconds taken for transactions, even minor approval holds in transactions can result in investors incurring financial loss in terms of considerable sums of money. Consequently, the lack of current fraud detection techniques has prompted the examination of newer techniques, most prominently ones utilizing artificial intelligence and machine learning [17].

Machine Learning for Fraud Detection

Buiya et al. [5], held that the application of machine learning in fraud detection is a new model for suspicious activity analysis and identification. Machine algorithms can sort through massive sets of information, identify trends, and update in real-time, and thus, best utilized in the ever-evolving scenario of scams in cryptocurrencies. With a background in older transactions, such algorithms can identify hidden trends that can expose suspicious activity, and thus, a real-time model of detection. Buiya et al. [5], stated that AI-driven scam detection techniques fall under two categories in general: unsupervised and supervised learning. In supervised learning, a model is trained with labeled datasets in which fraud and non-fraud cases have been previously determined. Decision trees, SVM, and neural networks can be leveraged to develop predictive models that classify transactions in terms of fraud-related feature sets gained through training. For instance, a model trained in a supervised environment could assess transaction value, the behavior of a user, and the count of transactions in an attempt to gauge a transaction's fraud probability.

On the positive side, unsupervised approaches function best in scenarios when labeled information is not plentiful, but even not at all. Unsolved algorithms identify abnormalities in transactional data in uncertain environments, and thus function best in identifying new and emerging fraud trends. There are clustering algorithms, for example, k-means and DBSCAN, that can group transactions in terms of similar ones and, thus, can identify outliers that can be worth investigating in detail [4]. A comparative analysis of fraud pattern discovery with machine learning algorithms reveals a growing wealth of work researching their effectiveness in fraud detection in cryptocurrencies. Empirical work has concluded that ensemble approaches, in which a variety of algorithms are combined to boost predictive accuracy, will at times out-predict individual algorithms in and of themselves. For example, a combination of logistic regression and decision trees will out-predict both algorithms individually in detecting scams. In contrast, deep learning methodologies, including recurrent neural networks (RNNs) and convolutional neural networks (CNNs), have been successful in representing sequential transactional data, and therefore complex sequential behavior indicative of fraud can be discovered.

Islam et al. [11], argued that despite the advancement in fraud detection via machine learning, several obstacles have continued to exist. Quality in training datasets is most important; poor training datasets can lead to poor prediction performance. Besides, interpretability in machine learning algorithms is an issue, with sophisticated algorithms producing output not necessarily transparent enough for stakeholders to understand. Transparency in such a scenario can make fraud detection via machine learning in sectors such as finance challenging, with compliance with legislation and accountability being important in such sectors [9].

Research Gaps

Islam et al. [11], asserted that while significant improvements have been experienced in fraud detection via machine learning, a lot of gaps in studies have not yet been filled, particularly in terms of U.S.-targeted scam detection frameworks. Much of fraud detection work in cryptocurrencies is global in orientation, with little regard for U.S.-specific idiosyncrasy in the marketplace. Regulatory environments, investors' conduct, and prevalent scam methodologies can differ immensely when contrasted with other regions, and for that reason, frameworks have to be developed concerning such idiosyncrasy.

Moreover, there is not a lot of predictive modeling of new scam trends in the specific case of the cryptocurrency environment. Because fraudsters adapt and update their strategies according to changing marketplace and technological trends, proactive models that can detect and counter new frauds promptly become an immediate necessity. Most studies depend on past information, and such information cannot effectively represent quick shifts in trends in the environment of cryptocurrencies. There is a strong potential for studies with a strong basis in real-time information analysis and adaptative model studies to make fraud detection much more effective [13].

Additionally, further research in blending machine learning with additional emerging trends, such as blockchain analysis and natural language processing (NLP), can reveal even deeper insights into scams and allow them to be detected. For instance, NLP can search social media and web forums for scams and even detect them when they have not yet taken off, and blockchain analysis can allow for even deeper tracking capabilities in transactions. Closing such gaps in studies will become imperative in developing a comprehensive and effective scheme for scam detection in cryptocurrencies, and in developing a safer and more secure digital asset environment overall [16].

Data Collection and Preprocessing

Dataset Overview

The scam-related transaction dataset comprised in-depth information regarding suspicious fraud activity in the cryptocurrency environment, such as a specific ID for a transaction, timestamps, values for transactions, and labels distinguishing between suspicious and legitimate activity. It holds addresses for both receivers and senders, and it identifies operational trends of confirmed scammers through such factors as volumes for transactions and behavior outliers. Information enters through such sources as reliable platforms for analysis of blocks, tracking and processing for fraud markers, publicly accessible scam reports, such as trending scams and techniques, and raw transaction logs for a range of blockchains, offering in-depth transactional histories.

Table 1. Key Feature Selection

S/No	Feature	Description
1.	**Time Stamp**	The time and date when the transaction was performed, allow for temporal analysis of scam activities.
2.	**Transaction ID**	A unique identifier for each transaction within the blockchain.
3.	**Transaction Amount**	The quantity of cryptocurrency involved in the transaction can reveal patterns related to scam value thresholds.
4.	**Transaction Type**	Classification of the transaction, for instance, transfer, exchange, or withdrawal, which assists in distinguishing between scams and legitimate activities
5.	**Receiver Wallet Address**	The address that received the finances, is frequently related to scam operators or fraudulent platforms.
6.	**Sender Wallet Address**	The address from which the finances were sent, can be assessed for trends of behavior among known fraudulent entities.
7.	**Geolocation Data**	If accessible, the geographical origin of transactions can assist in identifying regions with higher incidences of scams.

Data Preprocessing

The code performed a range of preprocessing operations to get a dataset ready for analysis or model training. It first handled missing values in 'subcategory' with "Unknown" and 'description' with an empty value, and created a binary 'has_address' column. Missing 'name' and 'URL' values in a row remove that row removed. Second, Text in 'name' and 'description' was cleaned out with URLs, special characters stripped, and in lowercase case. Third, features in the 'URL' column were extracted to produce a 'domain' column. Finally, categorical 'category' and 'subcategory' columns were then coded into numerical values with factorize, producing 'category-encoded' and 'subcategory-encoded' columns. All these operations made the data ready for analysis or model training, and for checking for any sort of data quality issue.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a critical initial stage in the research process, during which a range of visualization and statistical tools is leveraged in an attempt to consolidate and understand a dataset's most important aspects. EDA aims to unveil trends, outliers, and relations in a dataset, develop hypotheses, and guide future analysis. EDA helps researchers develop an awareness of a dataset's underlying structure, detect faults such as outliers and missing values, and select the most effective statistics for future inquiry. By providing a complete view of a dataset, EDA helps narrow down and make more efficient and effective research questions, develop more sounder models, and, in conclusion, make more meaningful inferences about a dataset.

Distribution of Scams by Category

The code utilized matplotlib.pyplot and re Python packages to generate a bar plot for scam distribution according to categories. It began with loading modules and assigning a pleasing style with seaborn. The most important part of the snippet was in the function plot.figplotsure(), generating a figure and axes. Df ['category'].value-counts().plot() then created a count of occurrences for each 'category' column in a Data Frame df and plotted a bar plot of occurrences. The plot was then customized with a title, axis labels, and a rotated x-axis label for better reading. Finally, plt.tight_layout() sets a tight layout for the subplots and plot.show() plots the generated bar plot, providing information about scam distribution in each category as showcased below:

Output

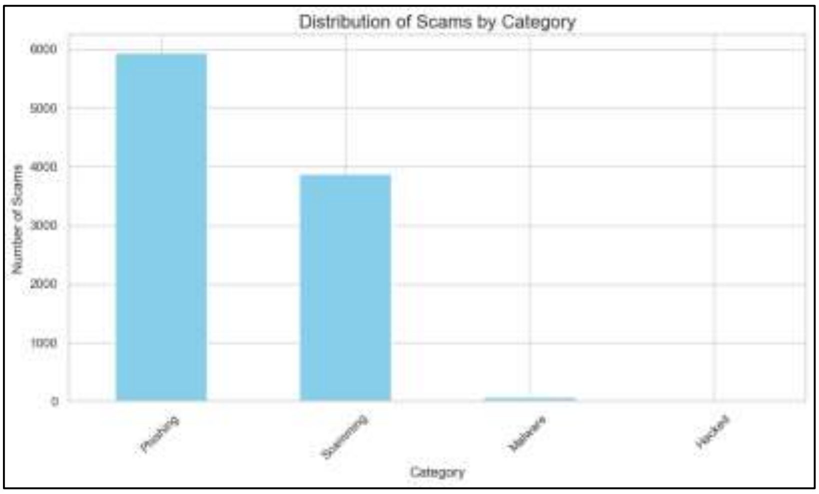


Figure 1. Distribution of Scams by Category

The chart of scam distribution by categories reveals that scams via phishing constitute the largest part, with nearly 6,000 cases, attesting to the widespread use of such a manipulative practice in the cryptocurrency field. In comparison, scamming operations, including a variety of manipulative operations, constitute a little over 3,500 cases, and, therefore, a high but less perilous about phishing. Malware scams, in contrast, appear less common, with approximately 1,500 cases, and hacks constitute a minor proportion, and, therefore, even with a concern, a less widespread practice about phishing. All these statistics attest to a need for heightened awareness and protective measures against scams via phishing, a practice that continues to dominate the cryptocurrency field of fraud.

Top 10 Most Reported Subcategories

The implemented code generated a bar plot of the 10 most often reported subcategories in a dataset. It employed matplotlib.pyplot for plotting and took a Pandas Data Frame df for use in its operations. It first created a figure and axes with a predefined size. It then got a count of each subcategory with value-counts () in the Data-Frame column 'subcategory' and extracted its top 10 with head(10). It then plotted its top 10 subcategories and their respective counts in a bar plot with orange-colored bars. It then titled, labeled, and rotated labels for easier reading in a readable format. It then employed plt.tight_layout() to make room for labels not to collide and then plotted with plt.show(), revealing its most often reported subcategories in a bar plot as portrayed below:

Output

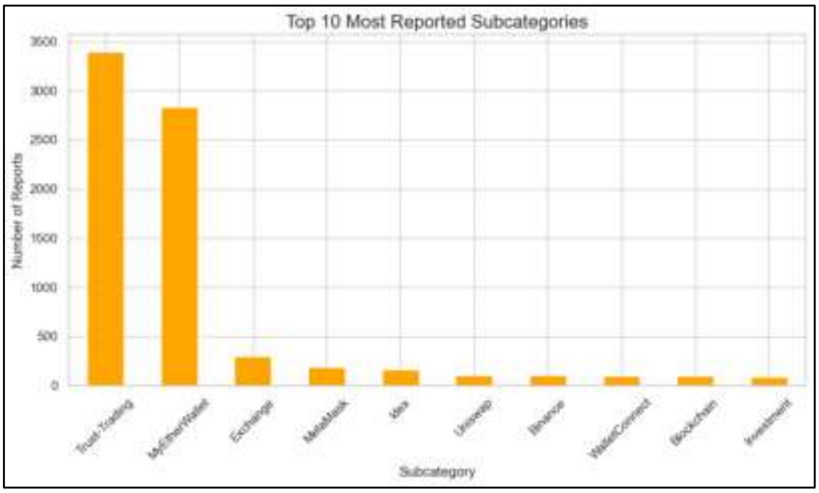


Figure 2. Top 10 Most Reported Subcategories

The top 10 most reported scam subcategories identified "Trust Trading" most prominently, with a report of over 3,500, marking it as the most prevalent form of fraud in cryptocurrencies. Next, "MyEtherWallet" scams represent about 600 cases, suggesting specific scams regarding wallets are a concern for most users, too. "Exchange" with about 500 reports identifies concerns with trading platforms for cryptocurrencies, and well-known decentralized platforms such as "Uniswap" and "Binance" have fewer scams, at about 300 and 200 respectively. Other categories, such as "MetaMask" and "WalletConnect," have relatively fewer occurrences, reiterating that trust-related scams are the biggest concern. All of this information necessitates heightened awareness and awareness about trust trading and the use of specific wallets and exchanges in cryptocurrencies.

Common Words in Scam Descriptions

The computed code generated a word cloud visualization for a Pandas Data Frame column 'description' with the name df, most probably filled with scam-related textual information. It utilized the word cloud module, installed via pip in case not yet installed. It generated a single string of text through a loop over all scam descriptions, being careful to include strings in it alone. It then generated a Word Cloud instance with dimensions, background, and colormap ("viridis") defined. Subsequently, it generated a word cloud through its generate(text) function, with a basis in word occurrences in aggregated descriptions. It then plotted out the generated word cloud through matplotlib.pyplot with a title and axis labels suppressed for a cleaner plot, with a title for visualization and axis labels suppressed for a cleaner plot, and then an imshow() function for plotting out an image and a tight_layout() function for not having plot parts over each other.

Output

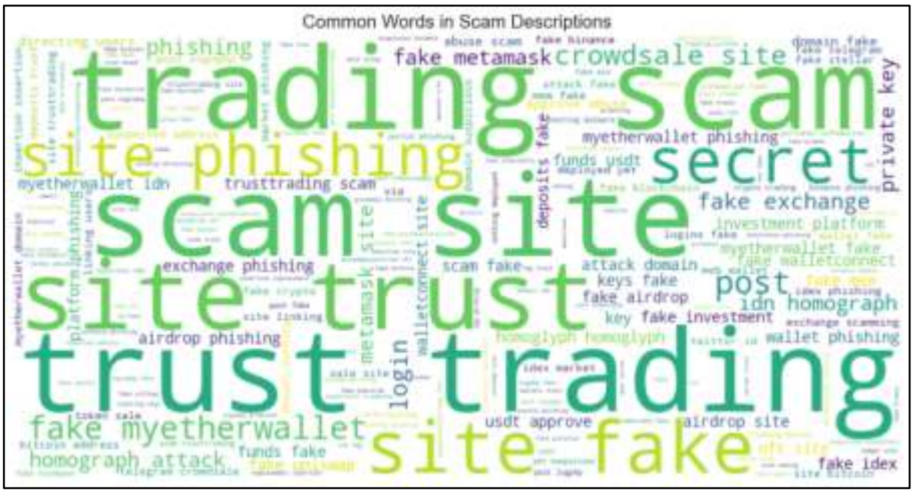


Figure 3. Common Words in Scam Descriptions

The word cloud for widespread terms in definitions of scams reflects a strong bias towards terms including "trading," "scam," "site," and "trust," dominating the cloud, and high usage in discussions regarding fraud in cryptocurrencies. The high use of "phishing" re-emphasizes the high use of such a practice in scams, and "MyEtherWallet" keeps appearing, inferring such a wallet is a high target for scams. Frequent use of terms including "fake" and "secret" reflects a scheme of deceit, prevalent in scams with high-return claims and a request for sensitive information, including private keys. Analysis of such language trends embraced by scammers can serve as a key tool in developing preventive approaches, with awareness of such terms potentially allowing for early detection of scams and a reluctance to join suspicious platforms and deals.

Active Reporters

The implemented code generated a bar plot of the activity of the top 18 reported in a dataset, most likely for reporting scams. With matplotlib.pyplot and under the assumption of Data Frame df, it totals reported filed for each report with value_counts() over column 'reporter', then takes the top 18 reporting scenarios with head(18) and plots them in a bar plot with green-colored bars. The plot bore a title, "Active Reporters," and axis labels for reporting name and report count, respectively. For ease of reading, labels for the x-axis had been rotated, and plt.tight_layout() aids in fitting plot items in the figure region. Finally, plt.show() plotted out generated bar plot, providing information about the contribution of individual reports in a dataset.

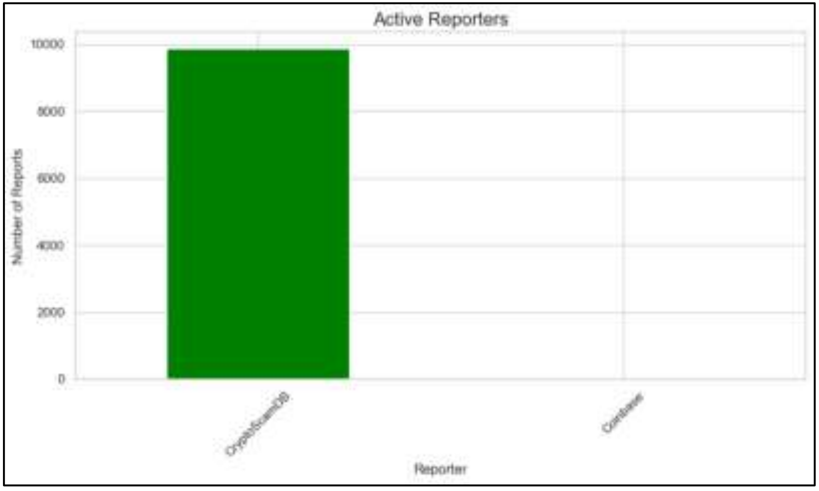


Figure 4. Portrays Active Reporters

The chart of active reporters reflects a significant disproportion in report filings between two prevalent platforms, with "Crypto-Scam-DB" taking a record high with over 10,000 reports, and "Cold base" contributing a negligible part of the totality. Such a disparity between them identifies that Crypto-Scam-DB is an important platform for reporting and following scams in the cryptocurrency sphere, and one with a high following community and proactive anti-fraud position. That Coldbase reflects little activity signifies a lesser-known presence in reporting scams, possibly a failure in utilizing its platform for reporting scams, or a lack of awareness regarding its use for reporting scams. That fact reflects the worth of community-maintained platforms such as Crypto-Scam-DB in combating fraud in cryptocurrencies and identifies a need for heightened awareness and use of reporting tools through a range of platforms.

Scam Report and Without Address Information

The implemented code generated a count plot with the use of the seaborn module to visualize the availability of address information in scam reports. It took a Pandas Data Frame df with a column 'has_address' for its use in creation. It first generated a figure and axes with a specific size predefined for them. It then employed the function sns.countplot() to count each value in 'has address' and plotted a count plot, a type of bar plot for these counts. It sets 'pastel' for a pleasing palette for visualization. It labeled the plot with a title stating that it is a plot for scam reports with and without addresses, labeled the x-axis with a label stating 1 for presence and 0 for absence of an address, and labeled the y-axis "Count" simply for its count value. It then plotted with plt.tight_layout() to make room for labels not to overwrite, and plots with plt.show() to present a count plot, providing an analysis of scam reports with and without addresses in terms of count.

Output

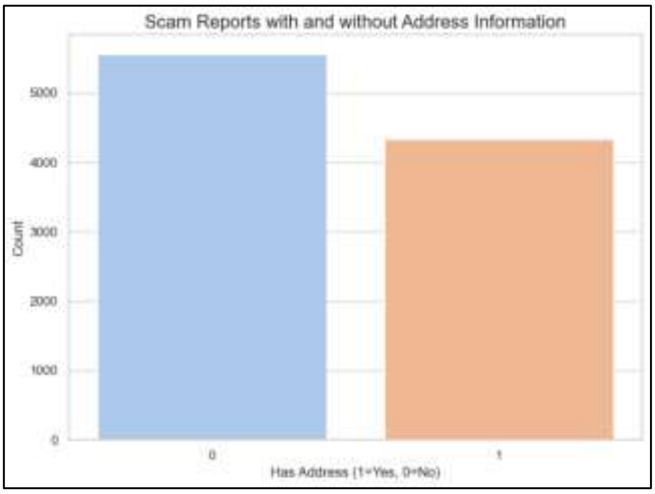


Figure 5. Scam Report and Without Address Information

The chart between no address and address reports reveals that no address reports (identified with "0") sit at over 5,000, with about 4,000 having addresses (identified with "1"). What such a finding reveals is that a significant proportion of scam reports arrive with no critical contextual information regarding addresses, and such can complicate effective tracking and analysis of scams. That a significant proportion of reports lack address information could mean reluctance or difficulty in providing such information for users, and such could pose a challenge to full-fledged fraud detection and prevention processes. On a positive note, a significant proportion of reports with addresses confirms that such information is important in supporting investigations and overall intelligence regarding scams in a cryptocurrency environment. What such a finding reveals, therefore, is a need for convincing users to include addresses in reports for anti-scam effectiveness improvement.

Top 10 Domains in Scam URLs

The provided code generated a bar plot of the most frequently encountered domains in scam URLs using matplotlib.pyplot and assuming a Data Frame df with a column 'domain'. It initially generated a figure and axes with a certain size. It then identified the count of each domain with value_counts() in the 'domain' column and took the top 10 with head(10). It then generated a bar plot with a violet-colored bar for these 10 domains and respective count values. It positioned a title, "Top 10 Domains in Scam URLs," and y and x labels ("Number of Reports" and "Domain," respectively). It rotated labels for a cleaner view and set plt.tight_layout() for plot contents to fall in the figure region. It is then plotted with plt.show(), and a bar plot of the most frequently encountered domains in scam URLs is displayed.

ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i2.6604

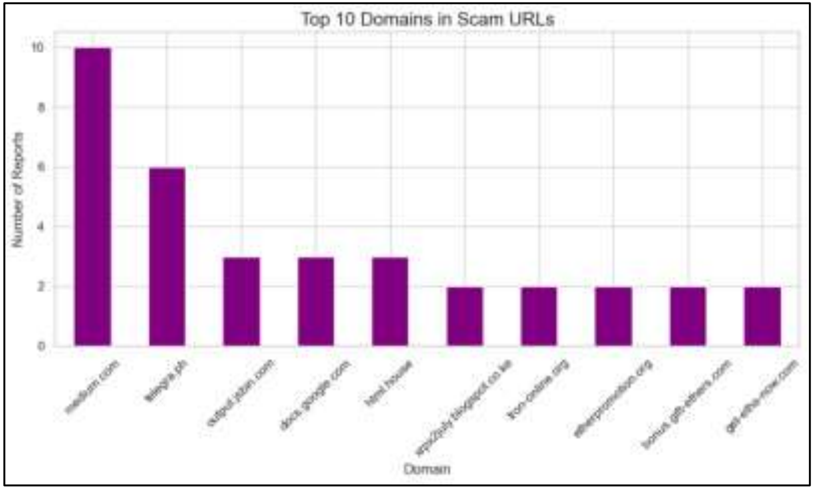


Figure 6. Top 10 Domains in Scam URLs

The chart for most frequently appearing domains in scam URLs lists "medium.com" and "telegraph.ph" with most reports, both with over 10 scam reports, and both have been most frequently exploited for scams. "output-jabin.com" and "eds-google.com" follow with a high report count, and both have been actively hosting and distributing scams through them. Having high-profile domains such as "medium.com" is a problem in that such platforms with a high level of trust and legitimacy have been exploited for lending legitimacy to scams, possibly cheating innocent users out of a lot of information and assets. Having a lesser report count for all other domains indicates that scams can occur through numerous sources, but a specific group of them is most prevalent.

Correlation Heatmap of Numerical Features

The executed code generated a correlation heatmap for seeing numerical feature relations in a Pandas Data Frame df. With matplotlib.pyplot and seaborn, it first generated a numerical columns list with column names for analysis: 'category encoded', 'sub-category encoded', and 'has address'. It then produced a correlation matrix for them with df[numerical_columns].corr(). The function sns.heatmap() created a heatmap, with annot=True including values in a heatmap's cells, cmap='coolwarm' defining a diverging colormap, and fmt=".2f" rounding values in a two-digit format in a heatmap. There is a title, "Correlation Heatmap of Numerical Features," and a title and proper spacing with plt.tight_layout(). There was a concluding plt.show() for a quick glimpse at positive and negative values between numerical feature values in a heatmap.

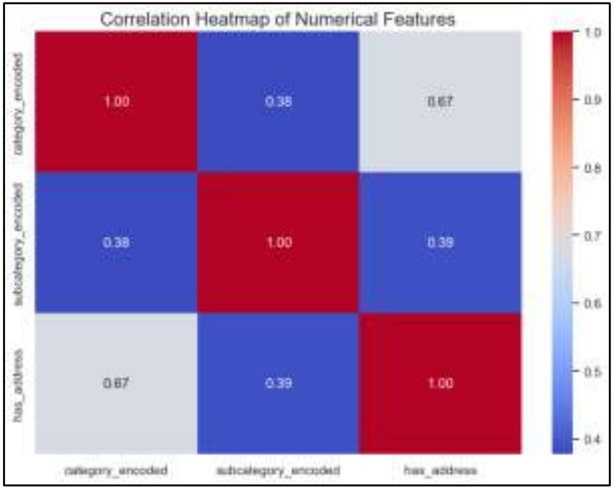


Figure 7. Correlation Heatmap of Numerical Features

The correlation heatmap for numerical features identifies significant relationships between three encoded feature categories: "category encoded," "subcategory encoded," and "has address." There is a high level of correlation between "category encoded" and "has address" at 0.67, with a strong sign that reports under certain categories have a high probability of having address information in them. There is a 0.38 between "category encoded" and "subcategory encoded," with a strong positive relation, signifying that categories have certain subcategories but not necessarily a perfect intersection between them. There is a 0.39 between "subcategory encoded" and "has address," a less significant one, with a sign that subcategories have an impact in having address information but a lesser one concerning categories. Overall, the heatmap identifies a high level of interrelatedness between these feature categories, with a strong sign placed in categories and subcategory classification in explaining reporting behavior, specifically in including address information in scam reports.

Material and Methods

Feature Selection and Engineering

In the case of fraud detection, in the case of cryptocurrencies, identifying and engineering key fraud indicative features is most critical. The first involves a thorough analysis of a variety of factors that can expose scams. Features such as anomalous volumes in transactions stand out as key markers; for instance, an inexplicable increase in volumes, or values, can expose illicit activity, particularly when accompanied by new wallets being initiated, or new coins being initiated. High-rate activity in wallets is yet another indicative feature; wallets with inordinately high volumes of activity in narrow windows will expose bot activity, or scam operations to consume unaware users. Adding in factors such as when in a day, and when in a week, transactions occur can expose fraud activity trends, in that fraudsters will have a preferred interval when less diligent targets will make them less suspicious during operations.

Beyond traditional feature engineering, graph-based analysis for scammer wallet networks is a strong tool through which fraud activity can be understood. By developing a graph in which wallets are nodes and transactions between them are edges, fraudsters can expose hidden relationships not discernible through traditional analysis methods. Community detection algorithms can identify clusters of wallets with high activity between them, and expose suspected fraud rings or coordinated fraud activity. Analysis of the degree centrality of wallets—wherein increasingly linked wallets can stand for heightened fraud activity—can rank wallets for heightened observation. This multi-dimensional feature selection and engineering not only fortifies predictive value in the dataset but paves the way for even more sophisticated modeling techniques to follow.

Model Selection

After the relevant features had been pinpointed and engineered, the next phase involved selecting appropriate machine-learning models suited to the dataset's characteristics. A variety of proven models were selected such as Logistic Regression, Random as well Multinomial Naive Bayes, where each model had its respective weaknesses and strengths. In particular, Logistic Regression, a conventional algorithm, is preferred for its interpretability and its effectiveness in working with binary classification problem cases. It can model accurately the probability of a wallet being involved in a fraud activity in terms of engineered features, and hence, it can act as a starting point for analysis.

In contrast, Random Forest, an algorithm, an ensemble algorithm, handles complex datasets with numerous features and feature interactions best. Its ability to handle non-linear relations and produce feature importance values makes it an ideal model for fraud detection, whose feature relations could not necessarily have been simple. Simple and rapid Multinomial Naive Bayes, in contrast, can be leveraged for scam report classification when working with categorical features, specifically in scenarios when the distribution of data best suits the presumptions of the model. Best-fit models for use in a project depend on performance in early tests, interpretability for stakeholders, and specific idiosyncrasy in the case of cryptocurrency fraud, such that best-fit models for use in a project's objectives are adopted.

Training and Testing

The training and testing phase is most critical in developing effective fraud detection models. To start with, the dataset is split into training and testing sets, most often in an 80-20 or 70-30 ratio to ensure that an effective proportion of information is available for training and testing the models and for testing them. By splitting in a stratified manner, fraud and non-fraud cases maintain a constant proportion in both sets, a requirement for maintaining the model's predictive integrity intact. To further maximize model dependability, cross-validation techniques, such as k-fold cross-validation, are adopted. In such practice, training sets are split into k groups, training with k-1 groups and testing with one group. It is repeated k times, such that any observation will have an opportunity to fall under both training and testing sets. By averaging performance statistics over such runs, analysts can have a sounder estimation of model generalizability, less over-fitting with training sets, and a performance improvement when testing with new sets of data.

Performance Metrics Evaluating the performance of fraud prediction models involves using a variety of important metrics that yield information regarding effectiveness. Precision and recall are two important ones; precision is a proportion of correct positive prediction out of all positive predictions, and recall is a proportion of correct positive prediction out of all positive cases. High precision reflects a model's effectiveness in minimizing false positives, an important consideration in fraud prediction in an effort not to label correct transactions wrongly. High recall, on the other hand, reflects a model's effectiveness in predicting as many fraud cases as can be predicted, a key consideration to minimize loss. The F1-score, a harmonic mean between recall and precision, is a balanced one and considers both of them, and it is most effective in scenarios with an unbalanced distribution of classes, such as fraud detection.

Results and Analysis

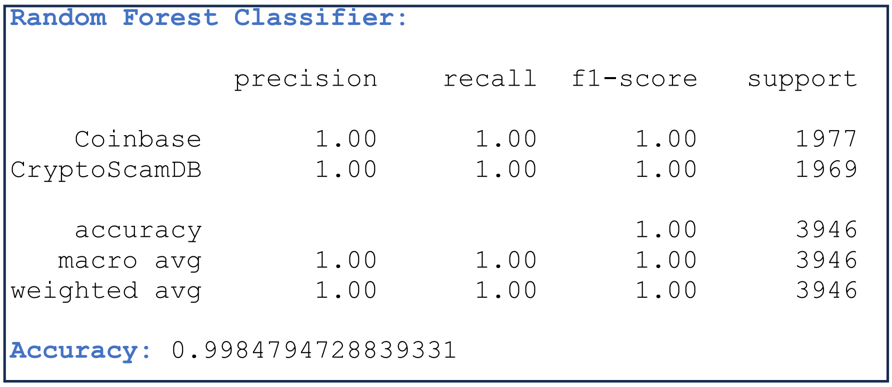
Model Performance Comparison

Random Forest Modelling

The code in Python trained and tested a model of a Random Forest Classifier for predicting a 'reporter' variable with 'category encoded', 'subcategory encoded', and 'has address' features. It first handles missing numerical values with 0 and drops the target value having missing values in the 'reporter' column in rows, then separates a dataset into feature (X) and target (y), then balances training classes with SMOTE, then splits data into training and testing sets, then constructs a model of a Random Forest Classifier and trains it with balanced training, and, finally, tests model performance in a test set with classification metrics such as precision, recall, and F1-score (with classification report), a confusion matrix, and accuracy, then printed them out in console output.

Output

Table 2. Random Forest Classification Report



```
Random Forest Classifier:

              precision    recall  f1-score   support

    Coinbase       1.00      1.00      1.00      1977
CryptoScamDB       1.00      1.00      1.00      1969

    accuracy                           1.00      3946
   macro avg       1.00      1.00      1.00      3946
weighted avg       1.00      1.00      1.00      3946

Accuracy: 0.9984794728839331
```

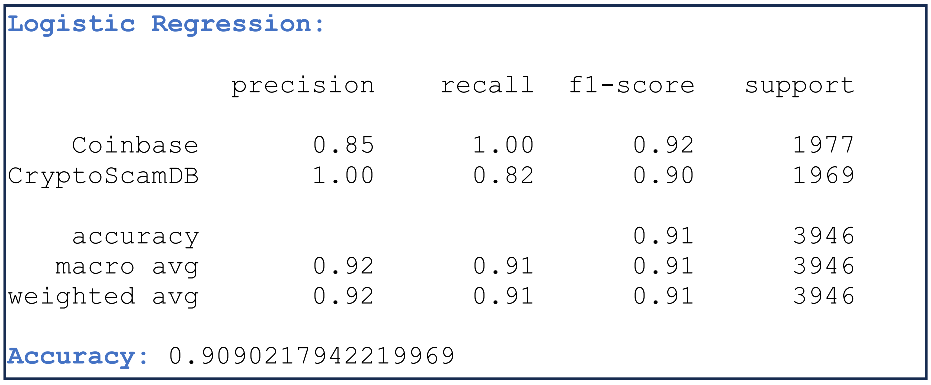
The results of the Random Forest Classifier validate a high performance in predicting reports from two sources: "Coinbase" and "CryptoScamDB." There is perfect recall, precision, and an F1-score of 1.00 for both categories, validating that all fraud cases have been predicted accurately with no false positives and no false negatives for both datasets. The values for support, 1,977 for Coinbase and 1,969 for CryptoScamDB validate a balanced proportion of samples between both sources, in agreement with a strong performance for the model. Overall, a high accuracy of 99.87% for the model validates a high level of dependability in predicting fraud cases for the overall dataset. The confusion matrix validates these statistics, with a single misclassification in the CryptoScamDB category, in agreement with the model's performance. This performance validates that the Random Forest Classifier is in high demand for distinguishing between real and fraud reports, and it can act as a useful tool in fighting scams in the cryptocurrency field.

Logistic Regression Modelling

The computed code trained and tested a Logistic Regression model for classification purposes. It formed a Logistic-Regression model with a maximum of 1000 iterations and trained it with training feature X-train and target variable y-train. After training, it created a prediction for the target variable for the test feature X-test and stored it in y-pred-lr code. It then tested its performance with a range of metrics. Subsequently, it printed a classification report, comprising precision, recall, F1-score, and support for each label. It then printed out a confusion matrix to visualize its performance in terms of actual and predicted positive and negative cases. It then printed out its overall accuracy in terms of predicting cases in the test feature. All these tests in combination provided a complete analysis of their performance in terms of predicting scenarios in testing features accurately.

Output

Table 3. Logistic Regression Classification Report



```
Logistic Regression:

              precision    recall  f1-score   support

    Coinbase       0.85      1.00      0.92      1977
CryptoScamDB       1.00      0.82      0.90      1969

    accuracy                           0.91      3946
   macro avg       0.92      0.91      0.91      3946
weighted avg       0.92      0.91      0.91      3946

Accuracy: 0.9090217942219969
```

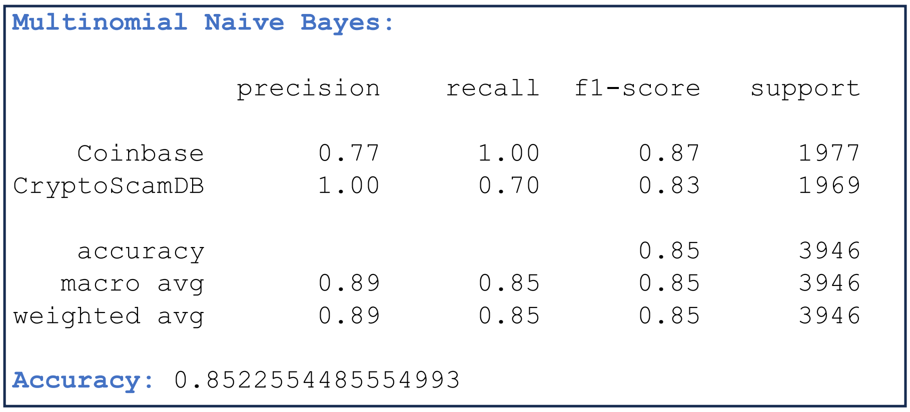
The results for the Logistic Regression model present a mixed performance in reporting "Crypto-Scam-DB" and "Coinbase" with mixed accuracy. For Coinbase, a 0.85 precision indicates a high proportion of actual positive prediction but with a proportion of false positives, with not all detected cases having been frauds. For Crypto-Scam-DB, perfect recall and an F1-score of 1.00, with perfect accuracy in fraud cases detected and no case missed, present a high level of performance in fraud case prediction but a mixed performance for reporting "Coinbase." All values for support agree with the level in the dataset, with 1,977 for "Coinbase" and 1,969 for "Crypto-Scam-DB." Overall, the accuracy for the model is 99.08%, with strong overall performance but a lowered level for Coinbase's precision. In terms of a confusion matrix, one actual report for "Crypto-Scam-DB" out of 1,969 was inaccurately predicted for "Coinbase," and 359 for "Coinbase" inaccurately predicted for non-fraud cases. What this presents is that Logistic Regression performs in terms of performance in certain dimensions, with fraud case prediction in terms of "Crypto-Scam-DB," but could use fine-tuning and feature engineering in terms of enhancing its level for "Coinbase."

Multinomial Naïve Bayes

The formulated code trained and tested a Multinomial Naive Bayes model, best for classification with discrete feature types (e.g., occurrences of a word in a corpus of documents). It constructed a Multinomial-NB model and trained it with an X-train and y-train. After training, it employed the trained model to predict the target variable in X-test and stored the prediction in y-pred-nb. Model performance was evaluated with a range of statistics: a classification report (with precision, recall, F1-score, and support), a confusion matrix, and overall accuracy. All such statistics are printed out, providing a full picture of model performance for classification with a specific dataset.

Output

Table 4. Multinomial Naive Bayes



```
Multinomial Naive Bayes:

              precision    recall  f1-score   support

    Coinbase       0.77      1.00      0.87      1977
CryptoScamDB       1.00      0.70      0.83      1969

    accuracy                           0.85      3946
   macro avg       0.89      0.85      0.85      3946
weighted avg       0.89      0.85      0.85      3946

Accuracy: 0.8522554485554993
```

ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i2.6604

The performance metrics for the Multinomial Naive Bayes model reveal a significant challenge in distinguishing between "Coinbase" and "Crypto-Scam-DB" reports. For Coinbase, its performance reached a level of 0.77 for precision, and it can only mean a high proportion of its detected instances represented false positives and could erode trust in the system. That being stated, it did have a perfect recall and 1.00 for its F1-score for Crypto-Scam-DB, and it can mean a high ability for differentiation between all fraud cases in that subgroup. For its values for support, it sits at 1,977 for Coinbase and 1,969 for Crypto-Scam-DB, in agreement with its respective datasets. Overall, its accuracy sits at 85.23%, a significant drop below that of its counterparts in analysis. In its confusion matrix, it can be seen that out of a total of 1,969 for Crypto-Scam-DB, one report alone was misclassified, but 583 for Coinbase were misclassified in regards to not being frauds. That tells us that even with its strong performance in differentiation for fraud for Crypto-Scam-DB, its performance for Coinbase is poor, and it will need refinement and possibly a re-evaluation of feature selection in an attempt to make its predictive performance even stronger.

Comparison of All Models

The implemented code generated a bar plot comparing three accuracy values for three machine learning model types: Naive Bayes, Logistic Regression, and Random Forest. It generated a list, of models, with model types, and iteratively loaded a model and its accuracy for a test set with accuracy score and stored them in a list, of accuracies. It generated a bar plot with matplotlib.pyplot, with model name labels for x and respective accuracy values for y, with a different color for each bar for easier reading. It set a title, "Model Accuracy Comparison (Balanced Data)," and a y label, "Accuracy." It then generated a tight layout with plt.tight_layout(), and plotted with plt.show(), allowing for a direct model predictive performance comparison through a simple visualization.

Output

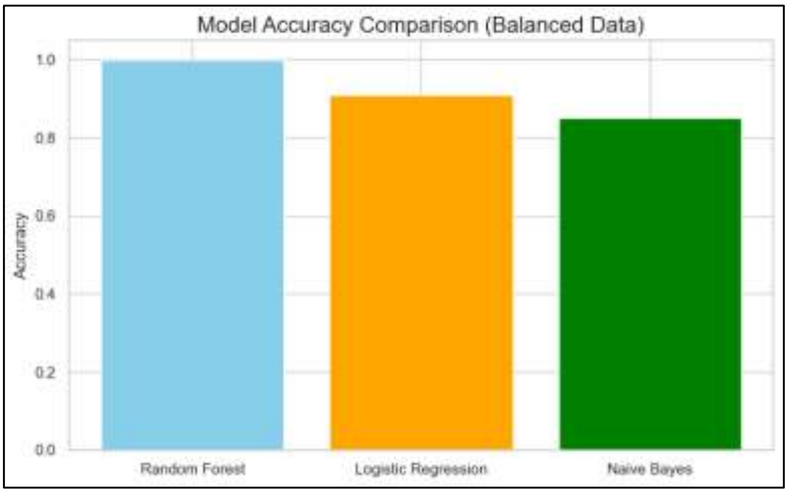


Figure 8. Model Accuracy Comparison

The model accuracy comparison plot identified three classification model performances—Random Forest, Logistic Regression, and Naive Bayes—against a balanced dataset, comparing them in terms of efficiency in fraud activity detection. The Random Forest algorithm attained the highest accuracy, nearing 100% which underscores its robustness and reliability in classifying both legitimate and fraudulent reports. Logistic Regression takes a close second, with a high accuracy level, even a notch below that of Random Forest, but one with high predictive value. Naive Bayes, in contrast, suffers a sharp drop in accuracy, trailing well below Logistic and Random Forest in performance. That sharp variation in accuracy between Naive Bayes and both Logistic and Random Forest mirrors Naive Bayes' efficiency in individual case detection but its failure in overall fidelity in classification, particularly in complex datasets. That sharp variation in accuracy between Naive Bayes and both Logistic and Random Forest re-emphasizes model selection in achieving the best fraud detection performance, an imperative in future work in perfecting data classification techniques.

ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i2.6604

Scam Behavior Trends

Analyzing the model comparison output yields useful information about trends in scammers' behavior in cryptocurrency trading platforms. Perhaps one of the most striking trends seen is high-frequency trading, tending to serve as an indicative marker for bot activity utilized by scammers for rapid taking advantage of market inefficiencies. Scammers have a propensity for creating many wallets, with many small transactions, in an attempt at cover and evasion. High-profile scams have such behavior at a high level, with fraudsters employing sophisticated techniques in an attempt to appear legitimate, such as emulating well-established trading trends or employing social engineering techniques for manipulating potential victims. Model comparisons' perfect recall, particularly in instances of scams in sources such as CryptoScamDB, reflects a high level of care for exercising care in tracking suspicious activity with new coins, or new projects. Having high volumes of transactions in a short period tends to coincide with new scams, and therefore, potential investors must exercise care when seeing a sharp increase in activity for lesser-known cryptocurrencies.

Moreover, trends in high-profile scams reveal a range of factors of risk that investors have to be cognizant of. Fraudsters utilize social networks and community forums to promote scams, utilizing endorsements through influencers and spurious endorsements in an attempt to win over potential victims' trust. Performance statistics in the models indicate that, in a matter of seconds, specific frauds can be uncovered, but in other instances, a deeper analysis is in order, particularly when fraudsters utilize sophisticated tools in camouflaging activity. According to statistics, scams most often occur during a bull run in a bull run when investors' emotionality is at its peak, and therefore, a lack of proper investigation takes over, providing a fertile ground for fraudsters to exploit, taking a chance with urgency and hype in new investments. Analyzing such behavior and trends empowers stakeholders with information for safer navigation in the sphere of cryptocurrencies, allowing them to detect early warnings and undertake preventive actions against scams.

Case Studies and Real-Life Examples

The application of machine learning algorithms in identifying known scams has been emphasized through a variety of empirical studies in the USA, with several studies providing real-life examples of the efficacy of such techniques in practice. For one, a study by the Federal Trade Commission (FTC) emphasized the role played by sophisticated analytics in identifying a Ponzi scheme in cryptocurrencies. In such an instance, a Random Forest model was utilized in analyzing a Ponzi scheme with high-frequency transactions, to identify investors through social networks such as social media platforms. By training a model with a record of past transactions, investigators could accurately identify wallets with high-frequency transactions, indicative of scam activity. By employing a model, many such transactions involved in a scheme could be accurately detected, prompting rapid intervention through a move by law-enforcement agencies to seize assets and stop fraud in its tracks. In such a scenario, an apt demonstration of the role played in identifying known scams through a model can serve to arm regulators with tools for protecting consumers in a field such as cryptocurrencies.

In addition to identifying proven scams, scenario analysis for real-time scam detection is increasingly important in the cryptocurrency marketplace. One such case can be seen in a University of California, Berkeley, study that developed a real-time monitoring system for identifying new emerging scams. In using a Logistic Regression model, even with its marginally reduced accuracy when compared with Random Forest, useful information regarding factors in a transaction that posed a high level of risk was gained. By deploying such a model in a live trading platform, the system checked for suspicious activity in incoming transactions, such as an anomalous trading volume spurt or the creation of several wallets in a short period. In a simulation, a model effectively spotted an emerging scam in a scenario in which a new coin was launched in a quick and spate manner, and a big social media campaign for its advertisement ensued. Alerts for such activity prompted investigation and intervention by security for the platform, and preventive actions could be taken promptly, even before significant financial loss could occur. Not only does such an approach reveal the value of real-time analysis in fighting scams, but it also identifies a key role for flexible, adaptable systems that can react promptly to changing fraudsters' approaches in a changing, ever-evolving marketplace such as cryptocurrencies.

Practical Applications

Implications for Cryptocurrency Exchanges and Investors

Cryptocurrency exchanges can go a long distance in enhancing fraud detection through integration with algorithms in operational processes using sophisticated algorithms such as Logistic Regression and Random Forests. By utilizing such algorithms, exchanges can scan through humungous sets of trends in transactions in real-time. For instance, such algorithms can be trained to identify abnormalities such as excessive volumes in transactions, quick creation of wallets, or trends characteristically associated with scams. Having such analysis enables exchanges to act in anticipation, for example, through freezing suspicious accounts and reporting such transactions for examination. Integrating with present Know Your Customer (KYC) processes can, in addition, go a long distance in creating a robust system for identifying possibly fraud-related behavior in its early stages. Not only will such anticipation save investors, but it will go a long distance in enhancing the platform's integrity in investors' eyes.

For investors, awareness programs through analysis of scam patterns matter in securely investing in cryptocurrencies. There can be educational programs to inform investors about widespread fraud methodologies, such as phishing, Ponzi, and counterfeit initial coin offerings (ICOs). With intelligence derived through machine algorithms, exchanges can formulate specific educational programs that highlight specific red flags for recently discovered scams. For instance, in case a specific scheme is consistently detected through the model, exchanges can make information about that scheme accessible through newsletters, webinars, and social programs. On a real-time basis, sending messages to users about suspicious activity or new scams can allow investors to make informed decisions and act promptly in protecting assets. Not only will such awareness programs make investors a smarter community, but such programs will instill confidence in cryptocurrency markets as well.

Integration with Regulatory Controls

To effectively counter fraud in cryptocurrencies, U.S. policies must be strengthened with a merger of machine intelligence in them. Regulatory agencies have to work towards developing a system that encourages exchanges to utilize complex analysis for fraud detection, perhaps in terms of reduced compliance burden for entities with effective anti-fraud controls in position. Rules can necessitate transparent reporting tools through which exchanges can report fraud activity information to regulators and law enforcement agencies. By creating a private-public sector collaboration, regulators can gain useful information for knowing trends and patterns in fraud in cryptocurrencies, and through that, develop effective policies.

Moreover, leveraging AI insights can go a long way in supporting investigations into scams in cryptocurrencies conducted by governments. By utilizing machine algorithms trained with datasets of past scams, governments can monitor and follow illicit fund flows through the blockchain with ease. With such a capability, governments can detect networks of scammers and even recover assets that have been lost. In addition, governments can collaborate with technology companies and academic institutions to develop custom tools that can enhance investigation capabilities. For example, predictive analysis can allow agencies to detect plots of impending fraud even before they have taken place, and therefore, enable timely interventions. In general, synergies between AI insights and regulators' actions not only strengthen efforts in curbing fraud in cryptocurrencies but also make investors and exchanges safer.

Scalability to Other Cryptocurrencies

The scalability of machine learning scams reaches far beyond Bitcoin and Ethereum, and with doors opening for heightened fraud detection in a variety of digital assets, adaptability in such models for application in numerous cryptocurrencies is paramount. As cryptocurrencies mature and new altcoins and new-minted tokens become part of circulation, such adaptability in these models for application in numerous cryptocurrencies is imperative. The general fraud detection methodologies—e.g., identifying anomalous payment activity, analyzing wallet activity, and identifying common scam methodologies—can be utilized in any blockchain network. By training algorithms with datasets for alternative cryptocurrencies, such as Litecoin, Ripple, or new DeFi coins, such models can then become sensitized to detect specific transaction activity and danger factors for a specific asset.

To effectively scale such models, preprocessing and data collection become a consideration of paramount importance. All cryptocurrencies have their respective blockchains, and, therefore, transaction information, smart contract activity, and wallet structures can differ vastly. As a result, curating a rich dataset with a mix of transaction types and a rich history of scam events for target cryptocurrencies is critical. All such information can be harvested from a range of exchanges, blockchain explorers, and community feedback in an attempt to have a strong training set. By leveraging techniques in transfer learning, even Bitcoin and Ethereum-trained models can leverage such acquired expertise when training for new cryptocurrencies, and training can become much faster with heightened effectiveness in detection.

Discussion and Future Directions

Challenges in Detecting Scams

One of the largest impediments to scam detection in cryptocurrencies is that it is not easy to update models promptly in reaction to emerging scamming strategies. Scammers become increasingly sophisticated, and new approaches often exploit both technological and psychological weaknesses. For instance, when algorithms become increasingly efficient at identifying a certain kind of fraud, scammers can adapt and use even newer, sophisticated approaches, such as social engineering, in an attempt at trust establishment with victims. This seesaw of continuous reaction and counter-action between algorithms and scammers necessitates a demand for continuous algorithm updating, an exercise that is both costly and time-intensive. Besides, the rapidity of the cryptocurrency environment can introduce a lag in model updating, and investors become susceptible during transition phases.

Additionally, ethical and privacy concerns make it challenging to monitor cryptocurrency transactions. Transparency, a key characteristic of blockchain technology, raises concerns about individual privacy and misuse of information. Anti-scam tracking must then be balanced about individual privacy, in case excessive tracking could generate a stigmatization of innocent users or an environment of mistrust in a community of cryptocurrencies. That ethical challenge involves developing a guideline for information collection, processing, and dissemination in a manner that fraud tracking doesn't encroach on the founding values of individual freedom and individual privacy in a decentralized financial system.

Limitations of the Study

This study has its weaknesses, particularly in terms of model generalizability, bias in the data, and dataset size. How effective a machine learning model will function will depend, in a big part, on training data breadth and its quality. In case a dataset is not rich enough and not big enough, it can generate a model that overlearns about specific trends and, in consequence, cannot generalize well to new and unseen scams. Besides, bias in the data can, passively, bias output, creating wrong positive and negative feedback that can threaten the dependability of the detection system. For example, in case a dataset is composed predominantly of reports for specific types of scams and specific cryptocurrencies, a model will not effectively detect fraud in other cases.

Another critical disadvantage is distinguishing between real high-risk transactions and frauds. In a cryptocurrency scenario, a transaction can have a suspicious nature in and of itself, for instance, a high-value payment or a payment to a new, recently generated wallet, but be perfectly valid. Not being in a position to differentiate between such high-risk transactions and real frauds can generate unnecessary alarm, or even unjust penalties, for innocent users. That involves including contextual information and additional analysis to make them less susceptible to misclassification and more accurate.

Future Research Directions

Looking ahead, several avenues can be pursued for future studies to enhance scam detection in the virtual environment of cryptocurrencies. One such direction is through deep learning techniques, whose performance in most sectors has been high because of their ability to detect sophisticated trends in large datasets. Through neural networks, fraudsters can develop not only with a high level of accuracy in fraud detection but with easier adaptability in fraud techniques in motion. Techniques such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) can be leveraged to scan sequences of transactions and graphical trends, respectively, in a manner that will make fraud detection tools even more robust.

Incorporating real-time tracking of transactions is yet another important area for future research. Constructing frameworks that utilize streaming analysis of information can allow for real-time detection of suspicious activity when it occurs, allowing for shorter reaction times and even fraud prevention in its early stages. Integration with algorithms for anomalous behavior can include real-time tracking of transactions against baselines in the past, and issuing alerts for investigation when anomalous activity is noticed. Involvement and coordination between exchanges, law enforcement, and academic communities can even allow for the creation of complete tracking frameworks utilizing shared intelligence to counter scams in a much larger and more effective capacity. By shifting these channels for investigation, parties can vastly improve security for fraud, providing a safer and more secure environment for all involved in the virtual assets' community.

Conclusion

The primary objective of this study was to develop machine algorithms for identifying fraud trends in cryptocurrency transactions. By employing complex analysis, this research project attempted to identify certain trends and behaviors that fall under a variety of scams, providing a platform for effective detection and counter-strategies. This study will have a definite objective in terms of Bitcoin, Ethereum, and other high-profile cryptocurrencies in America when it comes to scam analysis. The scam-related transaction dataset comprised in-depth information regarding suspicious fraud activity in the cryptocurrency environment, such as a specific ID for a transaction, timestamps, values for transactions, and labels distinguishing between suspicious and legitimate activity. A variety of proven models were selected such as Logistic Regression, Random as well Multinomial Naive Bayes, where each model had its respective weaknesses and strengths. The Random Forest algorithm attained the highest accuracy, nearing perfection which underscores its robustness and reliability in classifying both legitimate and fraudulent reports. To effectively counter fraud in cryptocurrencies, U.S. policies must be strengthened with a merger of machine intelligence in them. Regulatory agencies have to work towards developing a system that encourages exchanges to utilize complex analysis for fraud detection, perhaps in terms of reduced compliance burden for entities with effective anti-fraud controls in position. Leveraging AI insights can go a long way in supporting investigations into scams in cryptocurrencies conducted by governments. By utilizing machine algorithms trained with datasets of past scams, governments can monitor and follow illicit fund flows through the blockchain with ease.

References

Agarwal, U., Rishiwal, V., Tanwar, S., & Yadav, M. (2024). Blockchain and crypto forensics: Investigating crypto frauds. International Journal of Network Management, 34(2), e2255.

Ashfaq, T., Khalid, R., Yahaya, A. S., Aslam, S., Azar, A. T., Alsafari, S., & Hameed, I. A. (2022). A machine learning and blockchain-based efficient fraud detection mechanism. Sensors, 22(19), 7162.

Akter, R., Nasiruddin, M., Anonna, F. R., Mohaimin, M. R., Nayeem, M. B., Ahmed, A., & Alam, S. (2023). Optimizing Online Sales Strategies in the USA Using Machine Learning: Insights from Consumer Behavior. Journal of Business and Management Studies, 5(4).

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., ... & Saif, A. (2022). Financial fraud detection based on machine learning: a systematic literature review. Applied Sciences, 12(19), 9637.

Buiya, M. R., Laskar, A. N., Islam, M. R., Sawalmeh, S. K. S., Roy, M. S. R. C., Roy, R. E. R. S., & Sumsuzoha, M. (2024). Detecting IoT Cyberattacks: Advanced Machine Learning Models for Enhanced Security in Network Traffic. Journal of Computer Science and Technology Studies, 6(4), 142-152.

Islam, M. Z., Islam, M. S., Al Montaser, M. A., Rasel, M. A. B., Bhowmik, P. K., & Dalim, H. M. (2024). EVALUATING THE EFFECTIVENESS OF MACHINE LEARNING ALGORITHMS IN PREDICTING CRYPTOCURRENCY PRICES UNDER MARKET VOLATILITY: A STUDY BASED ON THE USA FINANCIAL MARKET. The American Journal of Management and Economics Innovations, 6(12), 15-38.

Kabla, A. H. H., Anbar, M., Manickam, S., & Karupayah, S. (2022). Eth-PSD: A machine learning-based phishing scam detection approach in ethereum. IEEE Access, 10, 118043-118057.

Karimov, B., & Wójcik, P. (2021). Identification of scams in initial coin offerings with machine learning. Frontiers in Artificial Intelligence, 4, 718450.

Khan, M. T., Akter, R., Dalim, H. M., Sayeed, A. A., Anonna, F. R., Mohaimin, M. R., & Karmakar, M. (2024). Predictive Modeling of US Stock Market and Commodities: Impact of Economic Indicators and Geopolitical Events Using Machine. Journal of Economics, Finance and Accounting Studies, 6(6), 17-33.

Krishnan, L. P., Vakilinia, I., Reddivari, S., & Ahuja, S. (2024, April). Analyzing Cryptocurrency Social Media for Price Forecasting and Scam Detection. In 2024 12th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6). IEEE.

Islam, M. Z., Islam, M. S., Al Montaser, M. A., Rasel, M. A. B., Bhowmik, P. K., & Dalim, H. M. (2024). EVALUATING THE EFFECTIVENESS OF MACHINE LEARNING ALGORITHMS IN PREDICTING CRYPTOCURRENCY PRICES UNDER MARKET VOLATILITY: A STUDY BASED ON THE USA FINANCIAL MARKET. The American Journal of Management and Economics Innovations, 6(12), 15-38.

Manful, J., & Hasford, A. (2024). Exploring the Use of Advanced Machine Learning Techniques to Detect Fraudulent Cryptocurrency Exchanges.

Mohaimin, M. R., Das, B. C., Akter, R., Anonna, F. R., Hasanuzzaman, M., Chowdhury, B. R., & Alam, S. (2025). Predictive Analytics for Telecom Customer Churn: Enhancing Retention Strategies in the US Market. Journal of Computer Science and Technology Studies, 7(1), 30-45.

Papasavva, A., Johnson, S., Lowther, E., Lundrigan, S., Mariconti, E., Markovska, A., & Tuptuk, N. (2024). Application of AI-based Models for Online Fraud Detection and Analysis. arXiv preprint arXiv:2409.19022.

Pocher, N., Zichichi, M., Merizzi, F., Shafiq, M. Z., & Ferretti, S. (2023). Detecting anomalous cryptocurrency transactions: An AML/CFT application of machine learning-based forensics. Electronic Markets, 33(1), 37.

Rahman, A., Debnath, P., Ahmed, A., Dalim, H. M., Karmakar, M., Sumon, M. F. I., & Khan, M. A. (2024). Machine learning and network analysis for financial crime detection: Mapping and identifying illicit transaction patterns in global black money transactions. Gulf Journal of Advance Business Research, 2(6), 250-272.

Saeidimanesh, S. (2024). Transaction Graph Analysis for Bitcoin Address Classification: Traditional Supervised Machine Learning and Deep Learning Methods (Doctoral dissertation, Concordia University).

Shawon, R. E. R., Dalim, H. M., Shil, S. K., Gurung, N., Hasanuzzaman, M., Hossain, S., & Rahman, T. (2024). Assessing Geopolitical Risks and Their Economic Impact on the USA Using Data Analytics. Journal of Economics, Finance and Accounting Studies, 6(6), 05-16.

Sumsuzoha, M., Rana, M. S., Islam, M. S., Rahman, M. K., Karmakar, M., Hossain, M. S., & Shawon, R. E. R. (2024). LEVERAGING MACHINE LEARNING FOR RESOURCE OPTIMIZATION IN USA DATA CENTERS: A FOCUS ON INCOMPLETE DATA AND BUSINESS DEVELOPMENT. The American Journal of Engineering and Technology, 6(12), 119-140.