Cryptocurrencies

Cryptocurrencies and scams: An analysis of the most common frauds

Sara Abreu Hernández

AI and Data Analytics Specialist

INTRODUCTION

Cryptocurrencies have revolutionized the financial world, offering new investment opportunities, decentralization, and anonymity. However, this same ecosystem has led to a significant increase in fraud and scams. The lack of regulation in many countries and the ignorance of users have allowed cybercriminals to develop increasingly sophisticated strategies to deceive investors and enthusiasts of the sector.

This report aims to identify common patterns in cryptocurrency scams by analyzing the methods most used by scammers, the terms employed to attract victims, and the platforms most affected. Through the collection and analysis of data, we seek to detect geographical and temporal trends, as well as explore possible predictive models that can help prevent future fraud.

To do this, data sources such as Crypto Scam DB, Blockchain Explorer, and datasets available on Kaggle will be used. Through data analysis and visualization tools, the main characteristics of these scams and their impact on the crypto community will be examined. Finally, conclusions and recommendations will be presented to mitigate risks and improve security in the cryptocurrency ecosystem.

HYPOTHESIS

Cryptocurrency frauds, particularly scams on platforms such as Twitter, Telegram, and phishing sites, are more frequent in certain categories and subcategories (such as "trust trading" in the "scamming" category or "fake ICO" frauds). Additionally, the addresses associated with these scams are generally linked to specific platforms used to receive funds from victims.

MATERIAL

1. **Dataset**: I used the "Cryptocurrency Scam Dataset" available on Kaggle (https://www.kaggle.com/datasets/zongaobian/cryptocurrency-scam-dataset), which includes information on scam URLs, descriptions of frauds, the platforms involved, and cryptocurrency wallet addresses.

2. **Analysis Tools**:

a. **Programming Language**: Python
b. **Libraries Used**: Pandas, Matplotlib, Seaborn, NetworkX, Numpy, and urllib for data processing and analysis.
c. **Platform**: Kaggle for performing the analysis and creating visualizations.

PROCEDURE

1. Define the project goal
2. Load and explore the data: The `urls.csv` and `uris.csv` files from the dataset were loaded, and the columns and data types were inspected.
3. Data cleaning: Several cleaning operations were performed, such as removing null rows, transforming Ethereum addresses, and normalizing descriptions.
4. Scam URLs Analysis: The most used platforms (e.g., popular domains in scams) were extracted and the distributions of fraudulent URLs were visualized.
5. Categories and subcategories analysis: The scam categories, such as "scamming," "phishing," and "fake ICO," were analyzed, and the associated subcategories were explored.
6. Platforms and reporters analysis: The platforms contributing the most to scams and the reporters identifying them were evaluated.
7. Scammers' connections network analysis: A network of connections between platforms, reporters, and scam sites was created using NetworkX.
8. Visualization: Bar graphs, histograms, and networks were generated to present the results.

DATOS

Nombre	Descripción
Scam URLs::	Includes domains associated with cryptocurrency frauds.
Categories and Subcategories:	The data contains category labels such as "Phishing," "Scamming," "Fake ICO," among others.
Descriptions:	Detailed descriptions of the

	reported scams are included, which allowed for identifying patterns in the language used.
Cryptocurrency Addresses:	Some rows contain Ethereum and other cryptocurrency addresses that the scammers used to receive funds.
Platforms and Reporters:	The platforms involved in the scams and the entities that report these scams.

RESULTS

The most common platforms for scams were Twitter, Telegram, and sites with domains like medium.com.

1. The most common scam categories were Scamming, Phishing, and Fake ICO, with subcategories such as Trust Trading in "scamming" and MyEtherWallet in "phishing."
2. **CryptoScamDB was the most frequently cited source in scam reports.**
3. No valid Ethereum addresses associated with the scams were found within the dataset.
4. The network analysis showed that CryptoScamDB and Twitter are strongly connected, with Twitter being a popular channel for spreading scams.

CONCLUSION

The analysis showed that cryptocurrency scams are prevalent on platforms like Twitter and Telegram, with certain categories and subcategories dominating the landscape, such as "scamming" and "phishing.".

Although an attempt was made to analyze cryptocurrency addresses associated with scams, no valid information was found in the dataset. However, the research provided insights into the prevalence of these scams across different platforms and categories.

It is essential to continue monitoring platforms and scammers' tactics to improve security within the cryptocurrency ecosystem.

REFERENCES

1. Dataset: "Cryptocurrency Scam Dataset" by Zongao Bian, available on Kaggle (https://www.kaggle.com/datasets/zongaobian/cryptocurrency-scam-dataset).
2. Tools: Pandas (https://pandas.pydata.org/), Matplotlib (https://matplotlib.org/), Seaborn (https://seaborn.pydata.org/), NetworkX (https://networkx.github.io/).
3. **Further Research: Articles and studies on cryptocurrency frauds, such as CryptoScamDB reports.**