iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52147

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Analysing Cryptocurrency Pump and DumpScams through Social Media

Soumya[1], Shreyasi Patra[2], Arham Saba[3], Pranshu Priyam[4], Prof. Roopa Banakar[5]
[1, 2, 3, 4]*Student, Dept. Of CSE, Sapthagiri College of Engineering Bangalore, Karnataka-560057*
[5]*Professor, Sapthagiri College of Engineering Bangalore, Karnataka-560057*

Abstract: ***Over the past few years, despite their vulnerability to market manipulation owing to limited liquidity, cryptocurrencies have attracted a lot of public attention and investment. More than $100 billion is currently traded on cryptocurrency exchanges each month, and individuals who are not necessarily experts in the field have begun purchasing these digital assets. A multi-modal technique is being utilised to identify and measure coordinated pump and dump scams that are carried out on bitcoin exchanges using social mediaplatforms like Telegram. This strategy attempts to anticipate if a pump effort will be effective by evaluating a variety of parameters.***

Index Terms: Pump-and-Dump, Cryptocurrency, Market Manipulation, Social Media, (Convolution neural network) CNN, LSTM (Long Short Term Memory), deep learning, Candidate Elimination.

I. INTRODUCTION

The popularity of cryptocurrencies as a form of investment has increased lately but the absence of regulation and transparency in the sector has made it susceptible to market manipulation, particularly in the form of pump-and-dump schemes. A group of traders artificially inflate the price of a certain cryptocurrency by coordinating their purchases and disseminating false information about the asset's potential value. The traders sell their shares when the price hits a certain point, which causes the price to fall and leaves other investors with huge losses. The establishment of the well-known cryptocurrency Bitcoin (BTC) was a crucial milestone enabled by the advent of blockchain technology. Following the first debut of various new cryptocurrencies, trading platforms suffered significant price volatility. These price movements increased the attraction of cryptocurrencies by allowing some investors to earn significantly. Despitethe fact that the bulk of investments are made in more well-known cryptocurrencies such as Bitcoin (BTC) and Ethereum (ETH), there are hundreds of additional smaller cryptocurrencies. The fraudulent "pump and dump" method of market manipulation is increasing the price of a particular investment that belongs to the manipulator and then selling it to other investors at a much higher price.

A. Problem Statement

The issues with bitcoin pump-and-dump schemes could significantly influence investors' financial well-being, particularly those who are new to the market or lack experience in discerning such fraudulent operations.
A few crucial questions are raised:

1) Pump-and-dump schemes can artificially inflate a cryptocurrency's price, leaving investors who invested at the peak with significant losses whenvalues fall.
2) Because the Bitcoin market is mostly unregulated, dishonest individuals can take part in pump-and-dump schemes without worrying about facing penalties.
3) Pump-and-dump scams may rely on coordinatedefforts by large groups of traders who use socialmedia and other online forums to spread false information and inflate the value of a particular currency, making it hard for investors to spot them.

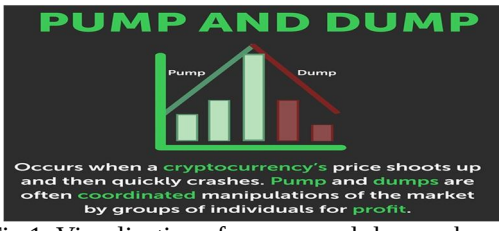


Fig 1: Visualization of a pump-and-dump scheme

Traditional methods of detecting pump-and-dump schemes rely on the manual analysis of trading data and identifying patterns that indicate manipulation. This approach can be time-consuming and prone to errors.

The goal of this survey paper is to explore the use of deep learning models for detecting cryptopump-and-dump scams. The survey will analyze existingresearch on the topic and identify gaps in the current literature.

B. Objectives

1) Detection of fake cryptocurrency inflation usingdata gathered from Twitter and Telegram, as well as simultaneous monitoring of cryptocurrency values on price tracking websites such as Coin market cap.

2) Checking to see whether a pump operation will be successful, that is, if it will achieve the goal price set in the Telegram and Twitter bot calls. To uncover these pump-and-dump frauds, we'll make predictions using machine learning techniques.

3) This would help protect retail investors from the media's pump-and-dump scams and make the crypto market a safer environment for retail investors who know nothing about cryptocurrencies but want to invest in them as an alternative to traditional financial instruments.

II. PRIOR WORKS

1) This paper's major goal is to thoroughly investigate the pump and dump ecosystem by examining the connections between various groups, exchanges, and targeted cryptocurrencies. Furthermore, it offers two case studies and develops a real-time detection tool that can spot pump and dump operations. Importantly, the article discovers a certain class of orders that work exceptionally well at spotting these fraudulent operations.

2) The motive of this paper is to categorize scams according to a proposed taxonomy that differentiates between pure and hybrid scams, as well as the distribution and correlation of different scam types. The classification is based on distinguishing between address-reported and URL-reported scams, and it begins with an examination of existing scientific literature on scams. A database of scammers is compiled using data from various sources, and the scams are then categorised using a free application in accordance with the specified classification scheme.

3) A notion that was previously believed to be exceedingly improbable or impossible is offered in this study, claiming that a dishonest actor utilising a strategy of price manipulation can build a price bubble. The model put forward in the research sheds light on quantitative phenomena such unusual variations in market prices and trade volumes and provides insights into how exchange manipulation might impact the entire market. The relevance of the relationship between order book liquidity, price formation, and the success of a pricemanipulation technique is also emphasised in the article. These findings are important for regulators and politicians trying to come up with effective ways to stop market exploitation.

4) This study assesses the spread of invite links to cryptocurrency-related channels by cross-checking over 50 million messages across the Twitter, Telegram, and Discord platforms. The results verified the controlling notion, which is based on the premise that the invite link exchange is a typical pattern associated with deceptive schemes, as well as a proxy for homophily and shared aims among the participating agents. It discovered a dense cluster of Ponzi scam channels. This work gives actionable knowledge that may be used to enact more effective responses.

5) This article investigates addresses in order to detect Bitcoin market manipulations using machine learning and statistical forecasting methodologies. As a result, the most generally utilized price estimating approaches, which included time series forecasting, and machine and deep learning techniques, were employed to identify the weekly/monthly increasing and decreasing patterns in Bitcoin pricing. This article employs SVM to detect manipulation periods as well as the Iterative Semi-Supervised Feature Selection (ISSFS) approach, which we created to more successfully analyzed text data. Furthermore, when combined with sentiment analysis findings and anomaly detection, SVM obtains the greatest performance in identifying manipulation zones and possible manipulators.

6) This paper's methodology focuses on the detrimental impact of cryptocurrencies' rising popularity on the cash tag mechanism via an experiment on LSE-100 businesses. The goal is to emphasie the contradicting problem in cash-tags while also offering classifiers that can accurately differentiate cryptocurrency and business cash-tags. The study effort was applied to real data gathered from Twitter and related to the London Stock Exchange to demonstrate the difficulty and present the classifiers (LSE). This study provides classifying methods that use Word-based Heuristic Filters, SVM (Support Vector Machine) Classifiers, and Logistic-regression-based Classifiers to handle the problem of conflicting cash-tags.

7) The purpose of this research is to undertake a systematic literature review (SLR) on current cryptocurrency-related cybercriminal activity. This research presented a defense strategy for detecting cybercrime. These assaults have been used to steal millions of dollars, exploit millions of connected devices, and cause much greater losses in service disruption and productivity losses. Random forest (RF) and support vector machine (SVM) were employed in this work to discriminate between benign and cybercrime datasets. The scope of these attacks is estimated in this research solely via the prism of technical scholarly literature.
8) This article provides an enhanced apriori approach fordetecting user groups that may be involved in "pump anddump" tactics. Using the leaked transaction history of thepopular Bitcoin exchange Mt. Gox, researchers discovered multiple user groups that purchase or sell at the same time. This study employs a technique that uncovered many anomalous trade records, i.e., abnormal trading behaviours and trading prices, to further investigate the detected groups

III. DATA ACQUISITION ANDPREPARATION

A. *Data Acquisition*

1) Identifies social media sites or platforms, including Twitter, Reddit, Telegram groups,or specialised forums, that are frequently used for cryptocurrency discussion.
2) Uses web scraping or APIs to gather information from these platforms. While web scraping involves directly obtaining information from websites, APIs offer structured data.
3) Concentrate on compiling information about conversations, mentions, and actions involving particular cryptocurrencies vulnerable to pump and dump scams.

B. *Data Preprocessing*

1) Cleaning up data to remove unwanted or annoying content, such as spam, pointless postings, and advertisements.
2) Extraction of important data points that can be used to recognise pump and dump actions. Cryptocurrency symbols, trade volumes, post sentiment analysis, mention counts, timestamps, and user profiles are a few examples of these functionalities.
3) For optimal continuity and complianceacross many sources and formats, normaliseand standardise the data.
4) Remove redundant or duplicate data points to preserve data integrity and eliminate bias.

IV. METHODOLOGY

A. *System Architecture*

Several crucial elements constitute the system structure for identifying and anticipating pump-and-dump scams in bitcoin via social media. Social media platforms are used by scammers to promote their fraudulent schemes and frequently to trick individuals into buying particular cryptocurrencies. The incorrect information in thesefraudulent posts on social networks might convince people to invest in the aforementioned cryptocurrencies in the hopes of making quick money. The system then gets pertinent information through a programming interface, or API that gives users access to historical or real-time bitcoin data. Analysis and forecasting are based on this data.

Candidate elimination, a preprocessing method, is used to remove irrelevant or unneeded information from the received data. This improves the data so it can be used for more analysis. Next, a training-set and a testing-set are created from the preprocessed data. The predictive model is trained using the training set, and its performance is assessed using the testing set. To guarantee uniformity and comparability across various features or variables, normalisation and standardizarion are employed. This procedure aids in eliminating any biases or contradictions that might present in the data. A certain amount of prior data points are utilised as inputs to forecast the outcome since the standardised data is segmented into window-sized parts. Using this windowing method, the model can detect temporal patterns and dependencies. In order to feed the preprocessed and windowed data into the predictive model, more processing and preparation must be done. It entails rearranging the input into a sequence that the model can recognise and efficiently process. The produced data is utilised to assess the prediction model's effectiveness. This assessment determines how effectively the model can identifypump and dump schemes using the supplied data. The model is developed using training data, where it attempts to capture the fundamental dynamics of pump and dump schemes by learning from previous trends. The performance of the model is then evaluated using the evaluation data. The model's performance is subsequently determined using the evaluation dataset. To analyse the sequential and temporal components of the data, the system uses a special kind of deep learning

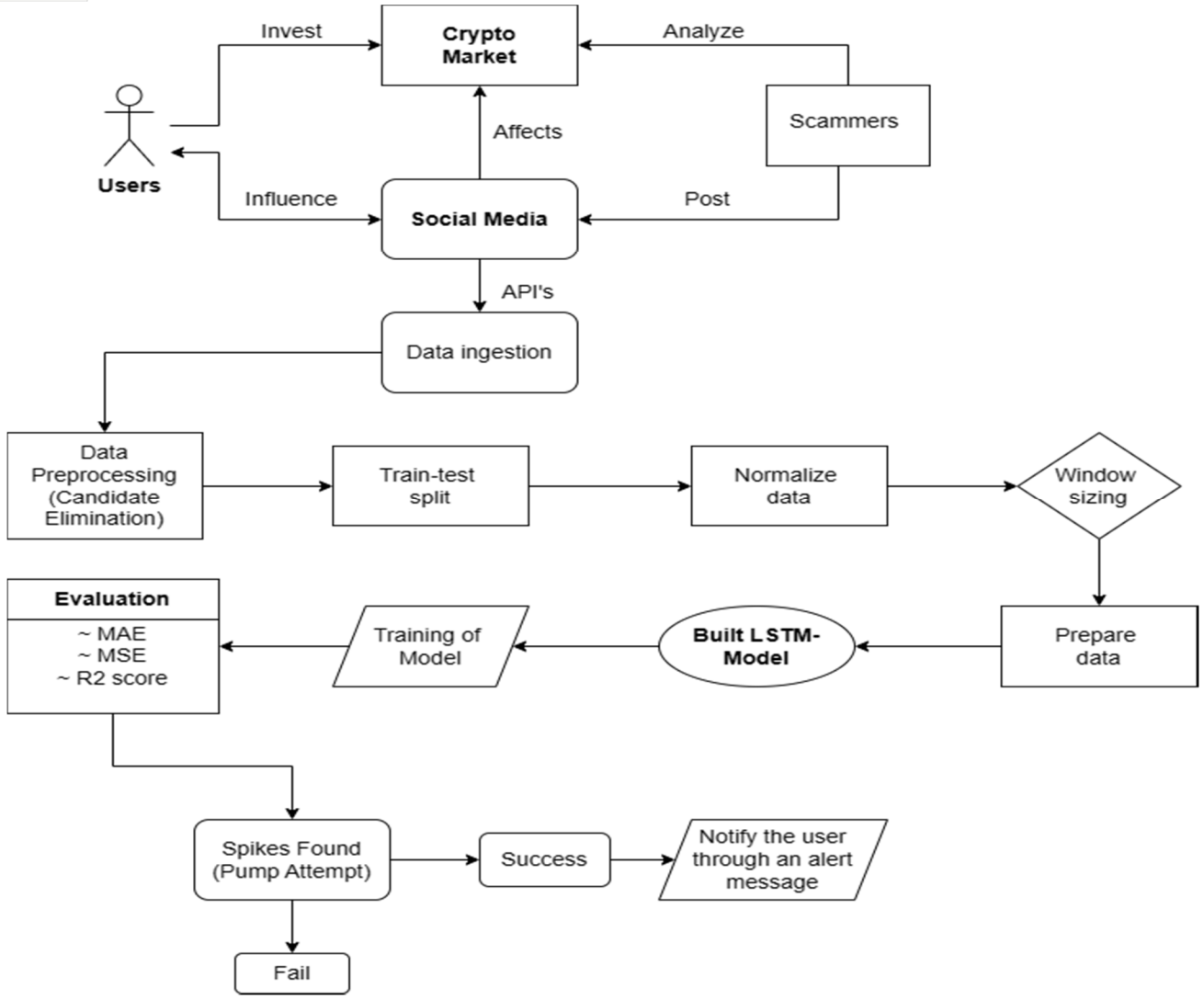


Fig 2: System Architecture for analyzing pump-and-dump scams

Model called Long Short-Term Memory (LSTM). Because LSTM models are effective at time-series analytic tasks, they may be used to anticipate pump and dump schemes. The LSTM model is tested and trained to look for potential spikes or other irregular patterns in the data that might be signs of a pump and dump scheme.

The model examines the input data to find any notable departures from predicted behaviour. An alert or warning is delivered to the appropriate people to advise them about the possible fraudulent conduct if the model is successful in identifying a pump or dump scam. If the model is unable to detect a fraud, more research or model improvement may be required. To forecast upcoming frauds, the system also uses the K-Nearest Neighbours (KNN) algorithm. KNN is a supervised machine learning method that classifies data points based on similarity criteria. The technology can forecast the possibility of upcoming pump and dump frauds by examining previous trends and parallels.

In general, this system architecture incorporates methodologies for data collection, preprocessing, model training, and predictive analysis to identify and anticipate pump and dump schemes in cryptocurrencies using social media.

B. LSTM Model Architecture

Recurrent neural networks (RNNs) of the LSTM variety are created to address the shortcomings of conventional RNNs in capturing long-term dependencies in sequential input. This is accomplished through the use of memory cells as well as the input gate, forget gate, and output gate—three gating mechanisms. Its unique structure makes it possible to identify long-term dependencies in sequential data. It is made up of LSTM units or cells that repeat and each of whichhas a specific set of parts and connections.

1) *Input Gate:* The input gate determines whether components of the input signal should be preserved in the cell state (Ct-1). Throughout the whole sequence, the cell state serves as the LSTM's memory and has the capacity to store and propagate information for extended periods of time.

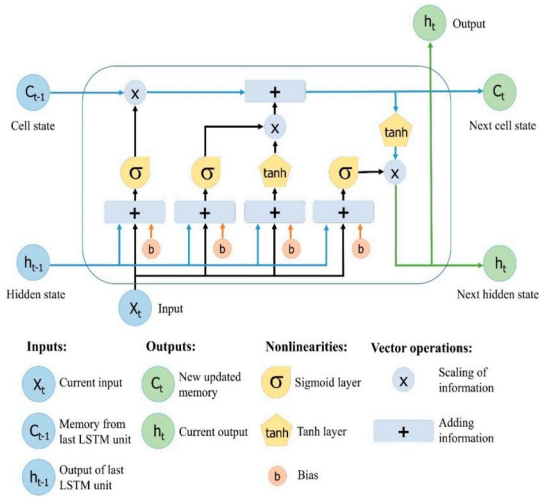


Fig. 3: The structure of the Long Short-TermMemory (LSTM) neural network

Current input (Xt) and the preceding hidden state (Ht-1) are combined with a sigmoid(σ) activation function. The sigmoid function's output, which has a range from 0 to 1, represents the importance or significance of each input element. The candidate values derived from the activation function of the tan hyperbolic (tanh) transformation applied to the current input (Xt) are multiplied by the outcome. The candidate values can be updated due to the non-linearity generated by the tanh function. Which data ought to be included in the cell's status is decided duringthis process.

2) *Forget Gate:* The forget gate determines which components of the prior state of a cell should be forgotten. A sigmoid() activation function is used after combining the previous hidden state with thepresent intake. The recall factor that represents every component in the previous state of a cell is represented by the sigmoid function's output, which ranges from 0 to 1. The result is element-wise multiplied by the preceding cell state, which aids in eliminating extraneous data.
3) *Update Cell State (Ct):* The outputs of the inputs from the input gate and forget gate are combined to update the state of the cell. The candidate state of the cell (C~t) and the previous cell state (Ct-1) are combined to form the updated cell state. The potential addition of fresh information to the cell state is represented by the candidate cell state. Based on the judgements made by the input gate, the latest values entered get added to the cells state. The decisions made by the forget gate determine whether components of the previouscell state are retained.
4) *Output Gate:* The hidden state (Ht) of the cell should be outputted based on the decisions made by the output gate. The result of an LSTM cell ata particular time step is called the hidden state,and it contains the pertinent data from the present input and previous inputs. Through the output gate, the cell state has an impact on it. It acceptsas inputs the present intake (Xt) and the prior hidden state (Ht-1). The output gate linearly combines Xt and Ht-1 with a sigmoid activation function. It accepts as inputs the present intake (xt) and the prior hidden state (Ht-1). The output gate linearly combines xt and Ht-1 with a sigmoid activation function. To compress the values between -1 and 1, the updated cell state is run via the tanh activation function. The last hidden state (Ht) is produced by multiplying the output of the output gate by the tanh output.

C. Pump and dump Activities using LSTMclassifier

To identify and categorise pump-and-dump operations in the bitcoin market, an LSTM classifier can be used. Pump-and-dump strategies include orchestrated purchasing and selling to artificially inflate the price of a cryptocurrency. The LSTM classifier makes use of historical pricedata, trade volume, order book data, and sentiment from social media to spot patterns related to pump-and-dump actions. To find patterns linked to pump-and-dump schemes, it makes use of historical price data, trade volumes, order book data, and social media sentiment. Collection of data, preprocessing, feature extraction, dataset preparation, creation of the LSTM model architecture, training, assessment, and prediction with an alarm system are all steps in the process.

By analyzing market data and employing machinelearning techniques, the LSTM classifier can identify suspicious trading activities indicative of pump-and-dump schemes, thereby enabling informed decision-making and risk mitigation. Continuous monitoring and updates are essential to adapt the classifier to new market trends and evolving scam patterns.

D. Model Evaluation and Selection

Model evaluation is a crucial step in assessing the performance of an LSTM classifier for detecting pump-and-dump activities in the cryptocurrency market. The collected dataset is typically split intotraining and testing sets, with the former used to train the model and the latter used to evaluate its performance on unseen data. Various performancemetrics, such as RMSE (Root Mean Square Error), R-squared, Mean Square Error (MSE), and Mean Absolute Error (MAE), are employed to measure the classifier's effectiveness. These metrics are commonly used for regression tasks, where the goal is to predict continuous numerical values.

V. RESULTS AND DISCUSSION

Insightful information on market manipulation was gleaned through an analysis of pump-and-dump behaviours in the Bitcoin market. Potential pump-and-dump events were visually identified using a graph showing the variations in the price of bitcoin over a certain time period. The price of bitcoin showed periods of abrupt and substantial price jumps on the graph, followed by periods of rapid falls, suggesting possible pump-and-dump actions. On the graph, the identified pump-and-dump occurrences were highlighted to show the precise time windows during which these manipulative behaviours took place. There were noticeable variations in trade volumes and liquidity during the detected pump-and-dump episodes, which pointed to unusual market behaviour. These anomalies provide additional evidence of planned pricing manipulation. The graph's analysis revealed that Bitcoin pump-and-dump actions had place over very brief periods of time. The sharp rise and subsequent decline in prices over a short period of time were signs of a pump phase, followed by a dump phase in which market manipulators dumped their assetsat inflated prices, resulting in substantial losses for unwary investors. A better knowledge of the dynamics of market manipulation in Bitcoin is possible thanks to the ability to visually watch the pump-and-dump activities and pinpoint the precise time periods. These findings highlight the significance of diligent surveillance and detection of such operations in order to safeguard investors and preserve the market's integrity.

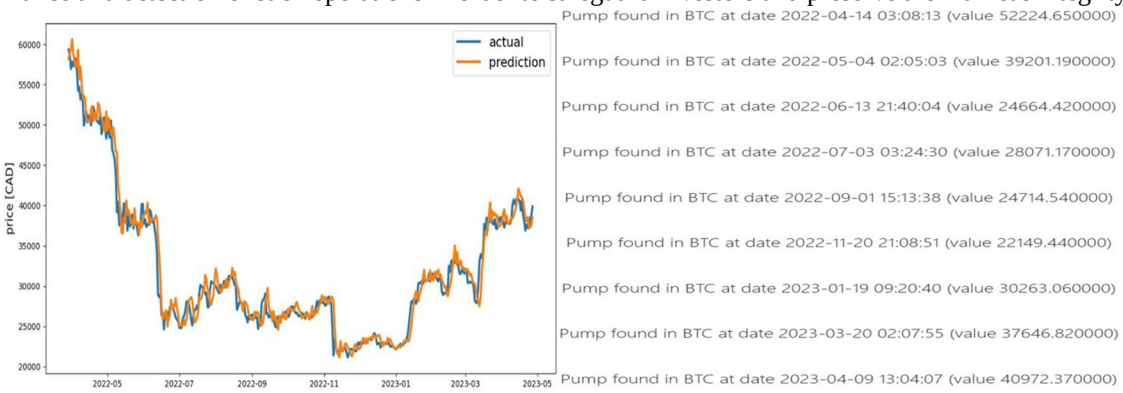


Fig. 4: Bitcoin : Model Evaluation and pump detection

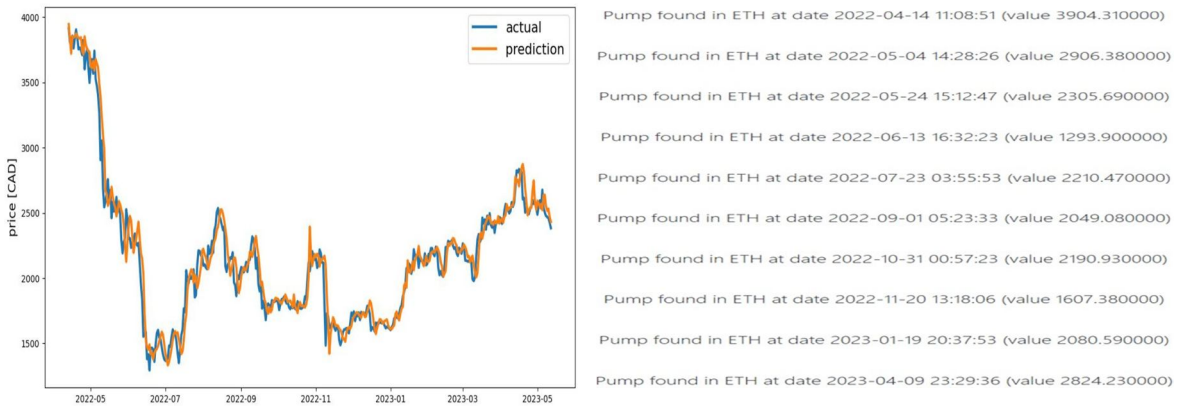


Fig. 5: Bitcoin : Model Evaluation and pump detection

VI. CONCLUSION

In conclusion, assessing bitcoin pump and dump schemes is essential for avoiding losses and safeguarding investors. These frauds may be efficiently identified and avoided by utilising cutting-edge technology like data analytics, machine learning, and blockchain analysis. Investors may choose which currencies to investin and which ones to avoid by using LSTM models to analyse a variety of coins. To remain ontop of the most recent scams and safeguard investors from possible losses, cooperation between investors, regulators, and technological specialists is essential. To preserve the integrity of the cryptocurrency market and make sure it continues to be a dependable and trustworthy investment alternative, the examination of pump and dump frauds in cryptocurrencies is crucial.

Future efforts to combat cryptocurrency pump-and-dump scams will focus on improving data analysis methods, creating real-time monitoring systems, bolstering regulatory frameworks and enforcement, promoting investor education and awareness, putting market surveillance measures in place, and investigating blockchain technology options. By concentrating on these areas, the goal is to enhance pump-and-dump activity detection, prevention, and mitigation, providing investors and stakeholders with a more safe and reliable cryptocurrency market.

VII. ACKNOWLEDGEMENT

We would like to convey our heartfelt appreciation to everyone who has contributed important support and assistance during the development of this project. We would like to express our deepest gratitude to Dr. Kamalakshi Naganna, Professor and Head, Departmentof Computer Science and Engineering, Sapthagiri College of Engineering, and our project guide, Prof. Roopa Banakar, Professor, Department of Computer Science and Engineering, Sapthagiri College of Engineering, for their continual advice, support, and encouragement during the project. Their important opinions and suggestions have helped shape our project. Finally, we are grateful to our friends and family for their continual encouragement and support.

REFERENCES

[1] M. La Morgia, A. Mei, F. Sassi and J. Stefa, "Pumpand Dumps in the Bitcoin Era: Real Time Detection of Cryptocurrency Market Manipulations," 2020 29th International Conference on Computer Communications and Networks (ICCCN), Honolulu, HI, USA, 2020, pp. 1-9, doi: 10.1109/ICCCN49398.2020.9209660.

[2] M. Bartoletti, S. Lande, A. Loddo, L. Pompianu and S. Serusi, "Cryptocurrency Scams: Analysis and Perspectives," in IEEE Access, vol. 9, pp. 148353-148373, 2021, doi:10.1109/ACCESS.2021.3123894.

[3] Fratrič, P., Sileno, G., Klous, S. et al. Manipulationof the Bitcoin market: an agent-based study. Financ Innov **8**, 60 (2022). doi.org/10.1186/40854-022-00364-3.

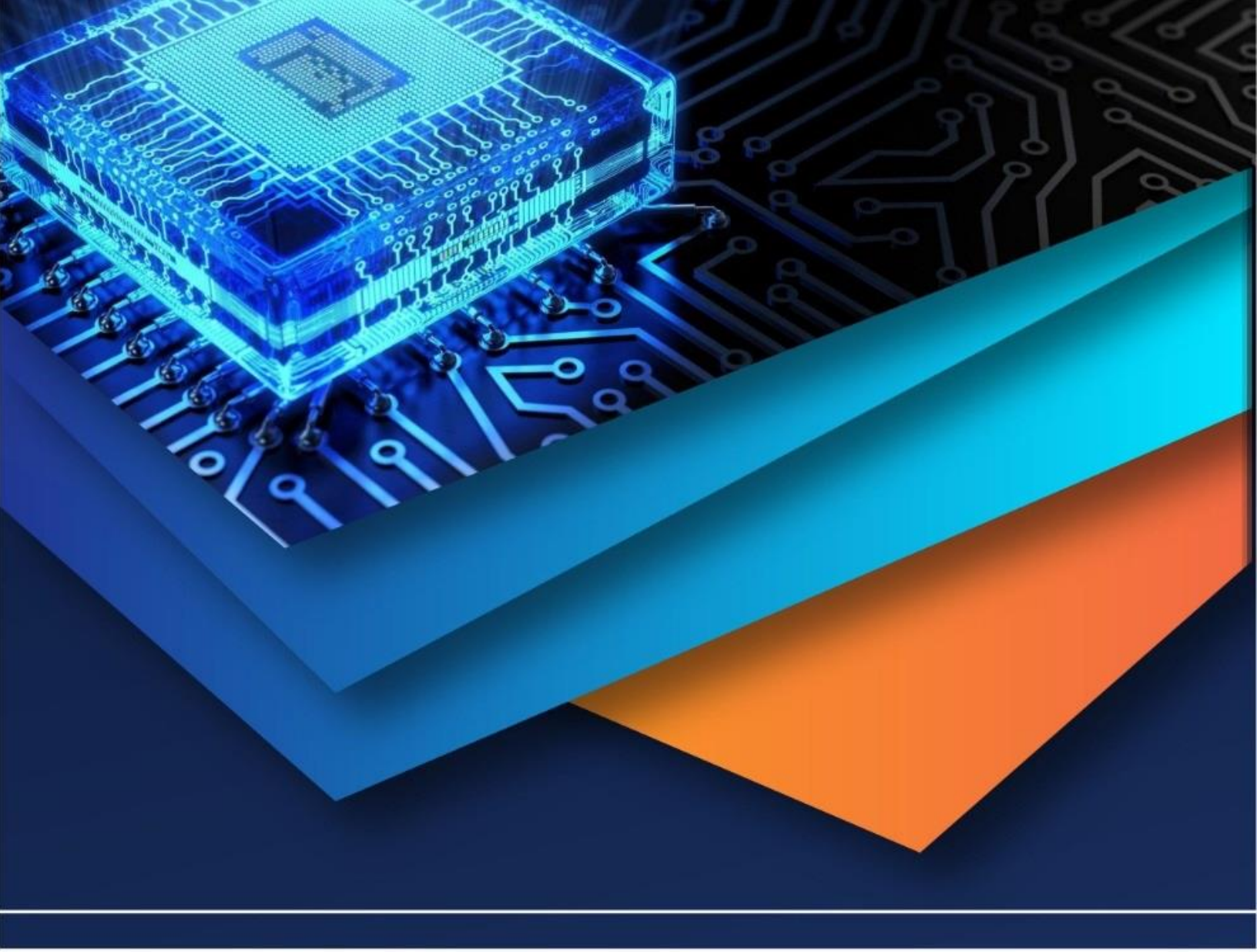
[4] Leonardo Nizzoli, Serena Tardelli, Marco Avvenuti(Member, IEEE), Stefano Cresci (Member, IEEE), Maurizio Tesconi and Emilio Ferrara (Senior Member, IEEE), "Charting the Landscape of Online Cryptocurrency Manipulation," IEEE Access, Vol. 8, pp. 113230 - 113245, Jun. 2020.

[5] Ihsan Tolga Medeni, Mehmet Serdar Guzel, Firat Akba (Member, IEEE) and Iman Askerzade , "Manipulator Detection in Cryptocurrency Markets Based on Forecasting Anomalies," IEEE Access, Vol. 9, pp. 108819 - 108831, Jul. 2021

[6] Antón Lorenzo García, Ana Fernández Vilas andRebeca P. Díaz Redondo, "The Irruption of Cryptocurrencies Into Twitter Cashtags: A ClassifyingSolution," IEEE Access, Vol. 9, pp. 32698 - 32713, Feb. 2020.

[7] Emad Badawi and Guy-Vincent Jourdan, "Cryptocurrencies Emerging Threats and Defensive Mechanisms: A Systematic Literature Review," IEEE Access, Vol. 8, pp. 200021 - 200037, Oct. 2020.

[8] Weili Chen, YueJin Xu, Zibin Zheng, Yuren Zhou,Jianxun Eileen Yang and Jing Bian, "Detecting Pump & Dump Schemes" on Cryptocurrency Market Using An Improved Apriori Algorithm," 2019 IEEE International Conference on Service-Oriented System Engineering, May. 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY