# Pump and Dump Cryptocurrency Detection Using Social Media

Domenico Alfano[1,2], Roberto Abbruzzese[1,2] and Domenico Parente[1]
[1]*Department of Management & Innovation Systems, University of Salerno, Fisciano, Italy*
[2]*Eustema S.p.A., Research and Development Centre, Napoli, Italy*

Keywords: Cryptocurrency, Pump and Dump, Anomaly Detection, Natural Language Processing, Explainable AI.

Abstract: The economic implications behind the fluctuation of cryptocurrencies prices, and, more importantly, the complexity of the variables involved in the process, have made price forecasting a very popular topic among researchers. Especially around detecting Pump & Dump events, where investors try to manipulate cryptocurrency owners to either buy or sell making a profit from them. Over the last decade, research has progressed by proposing new metrics (financial and non financial) capable of influencing and tracking the reasons for price fluctuations. Thanks to the advent of social media, major investment communities can be analysed through social channels to create new metrics. With developments in the field of Natural Language Processing, these social channels are used to extract opinions and mood of expert investors and cryptocurrencies owners. We propose to apply those innovative ways of creating metrics and to demonstrate that, taking these generated metrics into account, can significantly outperform other existing Pump & Dump detection methods. Moreover, to measure how each created metric contributes to the detection, a game theory approach called SHapley Additive exPlanations and a method that explains each prediction using a local, interpretable model to approach any black box machine learning model called Lime will be used.

## 1 INTRODUCTION

Cryptocurrencies are steadily gaining popularity, and more people are using them as platforms for investing. Cryptocurrencies are untested and generally uncontrolled, despite the significant sums of money invested in and traded in them. Its technical complexity and lack of regulation make them a desirable target for scammers aiming to prey on the uninformed (Kyle and Viswanathan, 2008). Pump-and-Dump (P&D) is a type of scam when speculators try to increase the value of a particular cryptocurrency by disseminating false information about it.

A P&D scheme is a sort of fraud in which perpetrators accumulate tokens over time, artificially raise the market by disseminating false information (pumping), and then sell what they have acquired to subsequent customers at the artificially inflated price (dumping). When the price has been boosted artificially, it typically drops, leaving the purchasers who made their purchase based on the misinformation at a loss (Kamps and Kleinberg, 2018).

Making specific public P&D groups is the method used to spread misinformation in the context of cryptocurrencies in order to drive up the price.

These organizations have developed as online chat rooms on social media platforms like Telegram with the specific aim of organizing pump-and-dump schemes on particular cryptocurrencies.

Studies suggest that these P&D groups almost mainly target less well-known currencies, especially those with low market capitalization and limited circulation since they are thought to be simpler to manipulate, in order to get the greatest outcomes (La Morgia, 2020; Mac and J., 2018). In a typical pump-and-dump scenario, group leaders advertise that the pump will occur at a specific time on a specific exchange, and that the currency will only be announced after that time. The group chat participants attempt to be among the first to purchase the currency after it is introduced in order to maximize their gains. They might even end up buying at the peak and not be able to sell for a profit if they move too slowly. Users are frequently urged to broadcast false information about the coin during the pump phase in an effort to get others to buy it so they may sell it more readily. Although there are many different types of misinformation, some frequent strategies include fake news, nonexistent ventures, phony alliances, or false celebrity endorsements.

P&D manipulation is currently not always un-



Alfano, D., Abbruzzese, R. and Parente, D.
Pump and Dump Cryptocurrency Detection Using Social Media.
DOI: 10.5220/0012059300003541
In *Proceedings of the 12th International Conference on Data Science, Technology and Applications (DATA 2023)*, pages 235-240
ISBN: 978-989-758-664-4; ISSN: 2184-285X

lawful due to the fact that the technology underpinning cryptocurrencies is still relatively new (Kramer, 2005).

This paper presents a novel application to detect P&D schemes in crypto. As most previous research in this area have only used financial data to address the issue, our work focuses on utilizing the vast majority of freely available data by using Telegram APIs to generate performance advantages (La Morgia, 2020; Kamps and Kleinberg, 2018).

## 2 RELATED WORK

The literature review led to the identification and study of scientific publications for anomaly detection in social media. The researcher Neda Soltani (Neda Soltani, 2018) identified anomalies and suspicious users through network analysis techniques, and through the study of the structure of nodes within communities. Another interesting work (Ryan G. Chacon, 2022) proposed the Fama French model to measure and demonstrate the influence of social media as a source of inspiration for new investments. In this case, an attempt was made to copy the investment strategy suggested by the Wallstreetbets group to generate profits. Then, the analysis from the social channel suggested the appropriate market to invest in. Other researcher (Firat Akba and Askerzade, 2022) attempted to identify Bitcoin price manipulation activities with the use of Machine Learning techniques. Specifically, the aim of the study was to investigate periods of manipulation studying emotions and user sentiments in social media.

The analysis was carried out with the use of algorithms such as SVM and SARIMAX. Studies similar to the ones mentioned above are abundant, making it clear that social media are often used as a medium for fraudulent activities. In recent years, many of these have focused on the cryptocurrency market and the Pump & Dump phenomenon. However, this phenomenon was already used in the financial field before the creation of cryptocurrencies. Reviewing similar works in the financial field can help us understand more on the techniques used by malicious investors, making it easier to apply same or new approaches in the cryptocurrency market. One of the main means used by fraudsters to pump a currency is spreading misinformation, often with the help of bots to automate the process (Mehrnoosh Mirtaheri et al., 2021). On the other hand, from a financial perspective, market manipulation schemes are classified into three main categories (Allen and Gale, 1992): information-based, action-based and trade-based.

Analysing this classification, we can define pump and dump schemes as a combination of information-based and trade-based manipulation. The work of Kamps (Kamps and Kleinberg, 2018) shows a first attempt to detect pump and dump using an adaptive threshold. They emphasise the fact that there is no reliable dataset of the confirmed pump and dump pattern, so they cannot fully validate their results.

## 3 PUMP AND DUMP DETECTION

### 3.1 Data Set

The financial data collection utilized in this study is composed of manually labelled unprocessed transaction data from the cryptocurrency exchange Binance, which was first made public by (La Morgia, 2020). Known instances of P&D were found in transactions involving a variety of cryptocurrencies.

The authors initially joined various cryptocurrency P&D Telegram channels known for developing and carrying out P&D schemes in order to generate the data set. Following that, over a two-year period, the researchers gathered timestamps of the official pump signals that were announced in each of these groups by the group administrators. Depending on what was accessible for access, the authors were able to gather all bitcoin transactions pumped up to a week before and after pumping using these timestamps and the Binance API. This method led to the collection of 104 P&D data occurrences. After obtaining this raw data from the Binance API, the authors pre-processed it further by grouping the transactions into chunks of five seconds, fifteen seconds, and twenty-five seconds, resulting in three distinct aggregated data sets. The following features are present in each of these aggregated data sets:

- **PumpIndex, Symbol:** The pump's base 0 index, identifying it as one of the 104 pumps that were available, and the coin's symbol that the pump took place on;
- **StdRushOrder, AvgRushOrder:** The average percentage change in the number of rush orders and the moving standard deviation;
- **StdTrades:** The number of buy- and sell-side trades' moving standard deviation;
- **StdVolume, AvgVolume:** The average percentage change in order volume and the moving standard deviation;
- **StdPrice, AvgPrice, AvgPriceMax:** The asset price's average percentage change, standard devi-

ation, and maximum percentage change, respectively.

On the social media side, it has been proven that actual, organized price rigging occurs there (Kamps and Kleinberg, 2018). The least active groups only carry out one P&D operation per week, whilst the most active groups carry out roughly one operation every day. Generally speaking, the following steps are taken during the procedure (Martineau, 2018):

- A few days or hours prior to the operation, the administrators make public the fact that P&D will occur, the exchange to be used, the precise time the operation will begin, and whether the operation will be Free for All, in which case everyone receives the message simultaneously, or Ranked, in which case VIPs and members at higher levels in the hierarchy receive the initial message before other members;
- The notice is made more frequently as the operation's execution time draws near;
- The organizers give these straightforward advice just before the event begins: as you wait for an outside investor, check your Internet connection, buy low and sell high, and hold currency as much as you can;
- The free chat rooms are currently closed to prevent "Fear, Uncertainty, and Doubt" (FUD) as a result of deception efforts put out by those looking to disrupt the operation and cause panic within the group;
- Depending on where you are in the group hierarchy, you will know exactly when the designated time comes for the targeted cryptocurrency to be exposed. The name of the cryptocurrency is typically written in a fuzzy image that can only be read accurately by humans. The purpose of the obfuscation is to hinder bots' ability to analyze the message using OCR methods and begin the process more quickly than humans;
- Admins release a news item shortly after the operation begins and ask everyone in the group to spread the message that the price of the cryptocurrency is increasing. This is done on Twitter, in forums, and in special chats. With the use of a special investment opportunity, this activity hopes to draw in outside investors;
- Ultimately, the admins reopen the free chat rooms after the operation and give the users some P&D statistics.

Studying this process and starting from the timestamps of the manually labelled pumps and the telegram group link from which the information was retrieved by the authors, we joined into these groups and through the Telegram APIs we downloaded all the messages exchanged in the chats from two days before and two days after the pump.

However, only some of these telegram groups allowed access to the message history, so we went from a data set of 104 pumps to one of 89 pumps.
Using Natural Language Processing techniques, a pre-processing of these messages was carried out, consisting of the following steps:

- **Removal of Emoji, Images, Stop-Words and External Links:** process of reducing non-informative text;
- **Lemmatization:** process of grouping together the inflected forms of a word so they can be analysed as a single item;
- **Pos Tagging:** process of marking up a word in a text as corresponding to a particular part of speech based on both its definition and its context by extracting only words referring to nouns (NOUN) and proper nouns (PROPN).

After gathering the messages from the relevant chat for each pump index, we proceeded to the stage of extracting the features.

In particular, for each of the three data sets the authors (La Morgia, 2020) created, for each of the financial transactions and consequently for each pump index defined within them, the new feature takes on a value of 1 if the currency symbol occurs in the messages exchanged 5 minutes before the transaction was made, otherwise 0. We thus produce the final data set, which consists of 89 pumps, by only adding a categorical feature to the financial ones.

## 3.2 Model

The authors' adopted classifier, (La Morgia, 2020), was employed in order to assess the scores and significance of utilizing a new Telegram feature in the most effective way. We are discussing the Random Forest classifier, which is a collection of decision tree classifiers that depend on the values of a random vector sampled independently, each of which casts a vote, with the prediction being the class with the most votes overall (Breiman, 2001). Since our data set consisted of 89 pump and dumps, we did not divide it into normal train and test sets. Instead, we used a 5 fold cross-validation to provide a more accurate assessment of the performance. For the Random Forest classifier we use a forest of 200 trees, each leaf node must have at least 6 samples, and a maximum depth of 4 for each tree.

# 4 RESULTS

In all the data sets, with the integration of the social feature, the classifier was able to beat the previous approach by a statistically significant margin. This demonstrates the effectiveness of using features from social data in this previously unexplored area.

State-of-the-art results across all data sets are showed in our table.

Table 1: 5-Folds Random Forest Performance on Financial Features.

| Chunk-Size | Precision | Recall | F1 |
|---|---|---|---|
| 5 Sec | 89.1% | 83.1% | 86.04% |
| 15 Sec | 97.5% | 89.8% | 93.5% |
| 25 Sec | 97.5% | 91.1% | 94.1% |

Table 2: 5-Folds Random Forest Performance on Financial and Social Features.

| Chunk-Size | Precision | Recall | F1 |
|---|---|---|---|
| 5 Sec | 94.2% | 91% | 92.5% |
| 15 Sec | 98.8% | 94.3% | 96.5% |
| 25 Sec | 97.6% | 94.3% | 95.9% |

We found that predictions using the 5-second chunked data set are much less accurate than those on the 15-second and 25-second chunked data set, which suggests that predicting anomalies using smaller chunk sizes corresponds to a harder problem in general. This confirms findings from previous works (La Morgia, 2020).

# 5 EXPLAINABILITY

## 5.1 Features Contribution Analysis

SHAP (SHapely Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model (Strumbelj and Kononenko, 2014). SHAP is used to solve an attribution problem, distributing the prediction score of a model for a specific input to its base set of features, showing how influential a feature is to make a decision. One of the leading approaches for contribution analysis is based on the Shapley value (Lundberg and Lee, 2017), a construct from cooperative game theory. In cooperative game theory, a group of players come together to consume a service, and this incurs some cost. The Shapley value distributes this cost among the players.

The algorithm computes the prediction assigning different values to the feature in order to calculate its contribution to the predicted outcome. Multiple Shap methods are available to explain different machine learning models. In the Random Forest case the TreeExplainer should be used. The Python SHAP library has multiple methods to estimate the feature contribution, based on the model used. In this case the model selected is the Random Forest and the method of the library that can be used with random forest is TreeExplainer, which explains the output of ensemble tree models.

These are the results of the SHAP analysis:

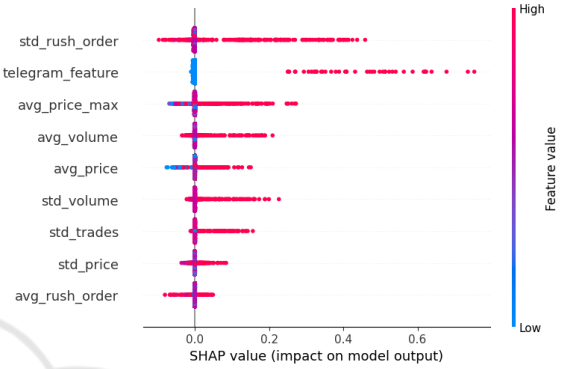



Figure 1: Impact of Financial and Social features on RF (5 Folds) for Chunk-Size 5 Sec.

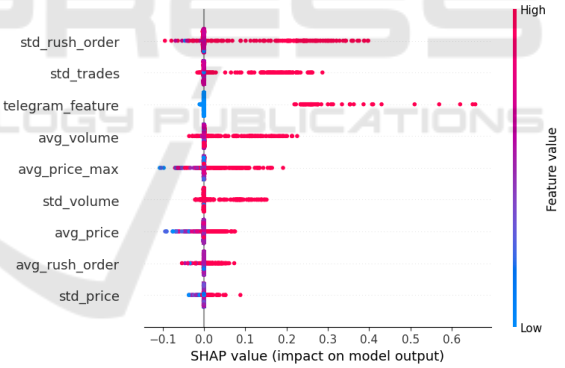



Figure 2: Impact of Financial and Social features on RF (5 Folds) for Chunk-Size 15 Sec.

As shown in these figures, the social feature has a considerable impact on all data sets. Its importance, however, wanes as the chunk-size window increases. This is unsurprising and the motivation lies in the fact that for 15-second and 25-second chunk sizes, fewer data are required to capture the same amount of information on average when compared to the more data contained in 5-second chunk. Consequently, this implies a drop in the amount of social information present in the input for larger chunk sizes.

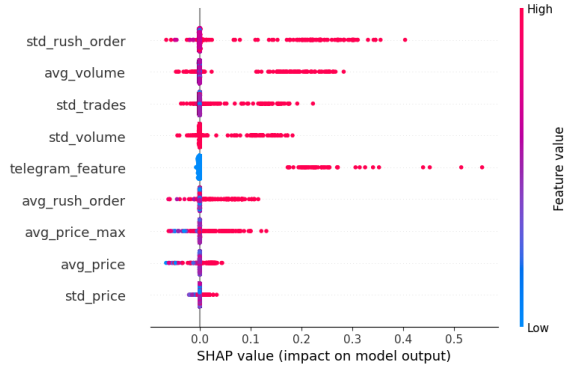
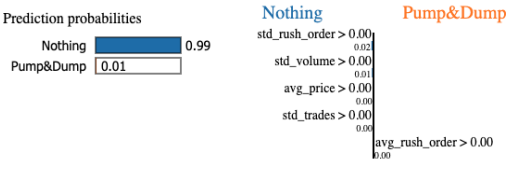



Figure 3: Impact of Financial and Social features on RF (5 Folds) for Chunk-Size 25 Sec.

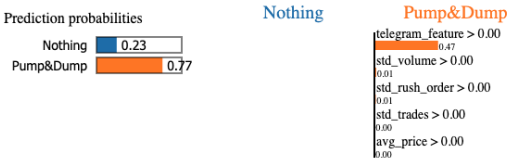
## 5.2 Single Instance Explanation

LIME (Local Interpretable Model-agnostic Explanations) is an explanation technique that learns an interpretable model locally around the prediction to explain any classifier's prediction in a way that is understandable and accurate (Marco Tulio Ribeiro, 2016). Each component of the name reflects an explanation we seek. *Local* fidelity means that we want the explanation to accurately capture how the classifier behaved "around" the instance that is being predicted. This explanation is pointless unless it can be understood by a person, or else it cannot be *interpretable*. LIME is *model-agnostic* because it can describe any model without having to "peak" into it and concentrates on training local models to explain specific predictions rather than developing a global model. The objective is to comprehend why a particular prediction was made by the machine learning model. LIME investigates what happens to predictions when different sets of data are fed into a machine learning model and creates a brand-new data set made up of altered samples and the related black box model predictions. These are the results coming from Lime analysis:

In particular, for each data set, instances that the (La Morgia, 2020) model predict as class 0 (Nothing) and the proposed model incorporating social features predict as class 1 (Pump&Dump) are taken into account. The analysis demonstrates the significant weight of social characteristics for all types of data sets. In many instances, especially when the other features' values are absent, it is able to fully overturn the local forecast (all equal to 0).

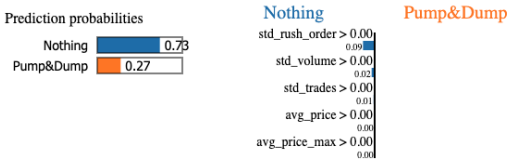



(a) Without Social feature.

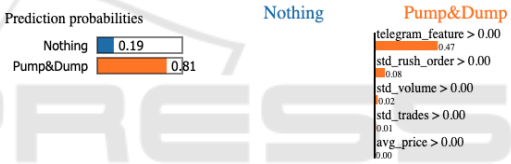



(b) With Social feature.

Figure 4: Single Instance Explanation Chunk-Size 5 Sec.

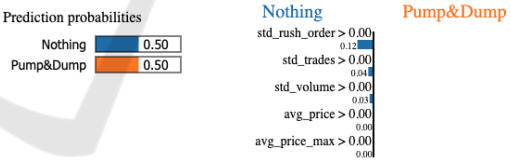



(a) Without Social feature.

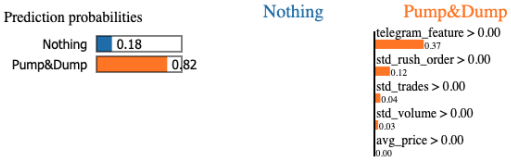



(b) With Social feature.

Figure 5: Single Instance Explanation Chunk-Size 15 Sec.




(a) Without Social feature.




(b) With Social feature.

Figure 6: Single Instance Explanation Chunk-Size 25 Sec.

# 6 CONCLUSIONS

This paper studies the application of social features in addition to the cryptocurrencies fraud detection problem space. We propose a novel method that can reach state-of-the-art performance on the data available.

Future work includes fine-tuning this model with new features to better account for the volatility generally found in cryptocurrencies, and exploring the potential for deep learning techniques (V. Chadalapaka and Vasil, 2022).

Moreover, having proved the impact of the social feature within the proposed models, future research will focus on improving the construction of new and more complex social features capable of capturing more information contained in the Telegram text messages.

## REFERENCES


Allen, F. and Gale, D. (1992). *Stock-price manipulation*. The Review of Financial Studies.

Breiman, L. (2001). *Random forests*. Machine learning, vol. 45, no. 1, pp. 5–32.

Firat Akba, Ihsan Tolga Medeni, M. S. G. and Askerzade, I. (2022). *Manipulator Detection in Cryptocurrency Markets Based on Forecasting Anomalies*. IEEE International Conference on Artificial Intelligence in Engineering and Technology.

Kamps, J. and Kleinberg, B. (2018). *To the moon: defining and detecting cryptocurrency pump-and-dumps*. Crime Science.

Kramer, D. B. (2005). *The way it is and the way it should be: Liability under § 10 (b) of the exchange act and rule 10b-5 thereunder for making false and misleading statements as part of a scheme to pump and dump a stock*. University of Miami Business Law Review.

Kyle, A. S. and Viswanathan, S. (2008). *How to define illegal price manipulation*. American Economic Review.

La Morgia, A. Mei, F. S. (2020). *Pump and Dumps in the Bitcoin Era: Real Time Detection of Cryptocurrency Market Manipulations*. IEEE.

Lundberg, S. M. and Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. Advances in neural information processing systems.

Mac, R. and J., L. (2018). *Here's how scammers are using fake news to screw with Bitcoin Investors*. BuzzFeedNews.

Marco Tulio Ribeiro, Sameer Singh, C. G. (2016). *"Why Should I Trust You?" Explaining the Predictions of Any Classifier*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016.

Martineau, P. (2018). *Inside the group chats where people pump and dump cryptocurrency*. The Outline.

Mehrnoosh Mirtaheri, Sami Abu-El-Haija, F. M., Steeg, G. V., and Galstyan, A. (2021). *Identifying and Analyzing Cryptocurrency Manipulations in Social Media*. IEEE Transactions on Computational Social Systems.

Neda Soltani, Elham Hormizi, S. A. H. G. (2018). *Anomaly Detection in Q&A Based Social Networks*. Advances in Intelligent Systems and Computing.

Ryan G. Chacon, Thibaut G. Morillon, R. W. (2022). *Will the reddit rebellion take you to the moon? Evidence from WallStreetBets*. Financial Markets and Portfolio Management.

Strumbelj, E. and Kononenko, I. (2014). *Explaining prediction models and individual predictions with feature contributions*. Knowledge and information systems.

V. Chadalapaka, Kyle Chang, G. M. and Vasil, A. (2022). *"Crypto Pump and Dump Detection via Deep Learning Techniques*. arXiv.