Analysis of On-Chain Whale Signals

Jackie Chen

Advisor: Pramod Viswanath

Abstract

Within the last few years, blockchain and cryptocurrencies have expanded to a trillion-dollar market cap, and Ethereum is one of the largest, with a total supply worth \$229 billion. However, like most cryptocurrencies, the distribution of Ethereum is heavily skewed, with over 52% of the coins controlled by the top 300 wallets. In this paper, we perform an in-depth analysis of whales. Then we detail the design, development, and evaluation of a trading strategy that relies on on-chain signals created by Ethereum whales. We then compared our strategy to traditional trading methods to determine its feasibility and limitations. Our results showed that, even with a data sample of just one year, our strategy was able to forecast significant price changes.

1. Introduction

Cryptocurrencies are digital or virtual currencies that operate independently of a central bank and are not controlled by a single entity such as a government or financial institution. Instead, cryptocurrency relies on blockchain technology, which is a decentralized ledger that records all transactions and ensures their security. Historically, the first and most popular cryptocurrency was created by following a manuscript published in 2008 by an unknown author under the

1

pseudonym 'Nakamoto'.¹ Since then numerous other cryptocurrencies have appeared, the most successful of which is Ethereum, which has built-in programmability features. Within the last few years the crypto industry has expanded to a trillion dollar market cap. Even after numerous crashes, collapses, and supposed "death of crypto" forecasts, the current value of a single Bitcoin still holds at over \$20,000, which suggests that the field of blockchain and cryptocurrency trading will continue to endure and grow.

For the most part, floating cryptocurrencies such as Bitcoin and Ethereum tend to be treated as assets akin to investments—people buy them in hopes that they will appreciate in the future which allows them to be sold for profit. Within any free moving market, the price of an asset is determined by supply and demand, and therefore influenced by the largest shareholders. In the crypto space these large shareholders are called whales, akin to insiders in traditional financial services. Insiders, which also includes influential individuals such as politicians and magnates, are known to cause significant movement in a commodity's price. In order to increase investment transparency and confidence, when an insider makes a substantial transaction the SEC requires them to make a filing and disclose the trade to the public.² However, for large shareholders, they can wait up to two days before making this filing, which provides ample time for them to leave the market and suffer reduced consequences. Additionally, members of Congress are given a further 45 days to publicly disclose any financial transaction of securities,³ making it infeasible to trade on this information.

Like insiders, whales are known to have significant influence on the price of crypto. As mentioned in the abstract, of the 230 billion unique addresses in existence, over 52% of

-

¹ https://bitcoin.org/bitcoin.pdf

² https://www.sec.gov/files/forms-3-4-5.pdf

³ https://www.govinfo.gov/content/pkg/PLAW-112publ105/pdf/PLAW-112publ105.pdf

Ethereum is controlled by just 300 wallets,⁴ meaning that a small number of people have a disproportionate effect in the market. However, unlike traditional stocks, information about the trade of cryptocurrencies is always public due to the transparent nature of blockchain ledgers and wallet accounts. New blocks and transactions constantly update the decentralized ledgers of cryptocurrencies. Services such as blockchain explorers and RPC nodes read developments on the ledger, which effectively allows anyone with internet access to see all incoming and past transactions. As a result, when compared to that of traditional finance, information about the trade of cryptocurrencies is accessible with relative ease and sidesteps the need to buy expensive equipment such as Bloomberg terminals. Most importantly however, due to the liveness of blockchains, it is possible to immediately identify when substantial buy or sell orders occur. Thus, because these on-chain signals are always publicly available, we can make an analysis of insider trading on cryptocurrencies not possible with traditional stocks.

2. Problem Background and Related Work

Price prediction is a crucial aspect of the crypto industry. As the value of cryptocurrencies is highly volatile and can fluctuate rapidly, it's essential to have an accurate understanding of how prices might change in the future to minimize risks and maximize profits. This information is crucial for traders, investors, and other market participants who rely on accurate predictions to make informed decisions about buying, selling, or holding cryptocurrencies. With accurate predictions, traders can identify potential opportunities to buy or hold onto their assets when prices are expected to rise. Conversely, if prices are expected to fall, investors can sell their assets or take other steps to minimize their losses.

3

⁴ https://etherscan.io/accounts

Whales are widely speculated to be significant market movers in the crypto industry. In a research paper examining cryptocurrency liquidity during extreme price movements, Viktor Manahov noted that whales use a variety of methods to manipulate the price of crypto [1]. He describes the 'rinse and repeat' strategy, which involves whales selling at a low price to induce panic from the market, waiting for prices to dip, then buying back the assets at a discount. He also discusses how whales can make 'buy and sell walls' with their disproportionate influence through creating large trade orders on exchanges, which allow them to closely control trade prices until they hit a desirable point [1].

Due to their impacts, whale behavior has been tracked across the crypto industry for years. Numerous news journalists publish articles about whale behaviors, and some directly attribute market downturns to whale selloffs. The analysis company Cryptoquant even has an entire section in their documentation describing "whale dumping" and that "significant Bitcoin deposits" into exchanges directly leads to price drops. Additionally, traders have developed techniques using "cryptocurrency whale watching" applications in their strategies. The comprehensiveness, effectiveness, and significance of these techniques are not well known, but Manahov does speculate that "whales are probably the biggest single contributor to herding behavior in the cryptocurrency markets" [1]. Therefore, in addition to directly manipulating prices, whales can also trigger secondary effects from other traders. This means even if a whale controls a comparatively small fraction of the entire supply, they can have far reaching indirect impacts.

_

https://coingape.com/ethereum-eth-price-drops-under-2000-as-dormant-whales-dump-heres-why-it-can-recover-fast/

⁶ https://cointelegraph.com/news/ethereum-whales-dumping-eth-as-price-slides-below-4k-data-shows

⁷ https://dataguide.cryptoquant.com/top-10-presets/whale-dumping

⁸ https://www.cryptostache.com/2017/11/21/cryptocurrency-whale-watching/

Bouri *et al.* in a 2019 research paper analyzing cryptocurrency herding behaviors does a closer study on herding and finds that it tends to occur as uncertainty increases [2]. This suggests, under the right settings, if a whale transaction triggers secondary reactions from other traders, those secondary trades can then cause more uncertainty and more reactive trades, creating a powerful and impulsive cascade that disrupts market efficiency. Bouri *et al.* also notes that there is a high degree of co-movement across different cryptocurrency markets [2], which means a single whale interaction can have profound impacts on the whole of the blockchain industry through herding and co-movement behavior.

Although the paper by Bouri *et al.* does not make any direct references, there exists indisputable evidence of whales triggering large cascading effects. The collapse of Terra-Luna is an infamous example. On May 7th, 2022, a whale traded 85 million UST for 84.5 million USDC.9 Although 85 million was only a fraction of the 18 billion UST in circulation, the transaction was immediately detected and triggered a series of panic sellings on the market. Within a day, the UST stablecoin had fallen to \$0.985, and by May 12 it was less than 10 cents. In total, the collapse and fallout lost billions in the entire crypto industry. While there were numerous underlying problems that contributed to Terra-Luna's downfall, sources point to the initial whale transaction as the starting point for the collapse. It's unknown if the whale intentionally triggered the crash, but the fact that a single transaction caused the implosion of a multi-billion dollar blockchain protocol emphasizes the massive effects a single whale signal can cause in the crypto industry.

-

⁹ https://etherscan.io/tx/0xaa23df48c53f221d0e8ac60ffc9e69340f3e8948fcdc936f3aee9c887d802abb

¹⁰ https://www.coindesk.com/learn/the-fall-of-terra-a-timeline-of-the-meteoric-rise-and-crash-of-ust-and-luna/

¹¹ https://www.forbes.com/sites/lawrencewintermeyer/2022/05/25/from-hero-to-zero-how-terra-was-toppled-in-crypt os-darkest-hour/?sh=22b46476389e

Whales undeniably cause major market movement. Moreover, the analyzes and comments from Manahov and Bouri et al. also seem to indicate that there is a way of utilizing whale behavior to predict prices meaningfully [1, 2]. Despite that, few academic studies have examined data on whale signals. While Manahov and Bouri et al. describe the strategies used by whales and their significance, neither study actually directly analyzes whales [1, 2]. Instead their research focuses largely on the cryptocurrency herding behavior and its impacts on investors as a whole. They attribute whales as the source of market influence and a possible leading contributor for herding, but refrain from providing statistics, or in depth analysis to shed further light on this subject. In fact, neither of the papers utilize data collected directly from the blockchain at all. Manahov uses information scraped from centralized exchanges and Bouri et al. uses data collected from online price-tracking tools [1, 2]. In other words, their data relies entirely on services that are in some measure removed from the blockchain. Despite this, the works of Manahov and Bouri et al. are one of the few papers that even mention the influence of whales in cryptocurrencies. As a result, from a research perspective, it's still largely unknown whether whale signals can be comprehensively utilized to create meaningful price predictions.

3. Approach

In this research paper, we detail the development of a machine learning price prediction model trained with data collected from whales on the Ethereum blockchain. Through building a viable trading strategy that solely utilizes on-chain whale signals and past price values, we demonstrate that whale behavior can be interpreted in a useful way to forecast prices. Our paper specifically monitors whales on the Ethereum blockchain because it is the largest, oldest, and most well-known programmable blockchain. In addition, the Ethereum blockchain also has a number of well established data collection tools such as the blockchain explorer Etherscan and the RPC

node provider Alchemy. We will be using these services extensively to monitor and track whale signals.

Simply following Ethereum whale trades is not efficient. Manahov comments that copying whales can be dangerous in part because whales understand how much impact they have [1]. He warns that, for example, they can create false signals of a price increase by creating a large buy order on an exchange, duping small traders to invest only to immediately cancel the order and sell their assets at a higher price [1]. In fact, Bouri *et al.* concluded in their research that herding behavior is actually indicative of relative market inefficiencies [2]. Therefore it is not a good idea to just mimic whale trades on exchanges or blockchains. Whales are also known to make mistakes. For example, the crypto hedge fund companies 3AC, Voyager, and Celsius were one of the biggest and most influential whales before they went bankrupt due to poor management of funds. 12

Thus, in order to adapt around these issues, we opt to record multiple signals from a variety of whales to capture a more complete picture of the entire whale sentiment. For this project, we identify whale wallets as any non-project related address with an Ethereum balance in the top 5% of active wallets. Ideally, we'd like to collect data from every whale over the course of multiple years to make a robust analysis. Unfortunately, due to the amount of time it takes to collect this data from the blockchain, such an approach would be infeasible for the scope of this paper. Thus for the purposes of this study we will solely focus on data collected daily from 92 wallet addresses during the year 2022. These wallets were picked randomly in order to get a better spread of data. For comparison, our largest whale held 12,035.15 ETH worth \$14.7 million and

¹² https://amycastor.com/2022/07/09/crypto-collapse-3ac-voyager-celsius-and-other-defi-casualties/

our smallest whale held 26.6 ETH worth \$31,900. Section 4.1 details more about our collection methods and results.

The advantages of collecting data directly from the blockchain is that it ensures accuracy with granularity. The blockchain is also the most direct source for whale signals. As mentioned in the introduction, the public nature of the cryptocurrency's ledger allows anyone with internet access to explore the entire ledger of a blockchain. Therefore every aspect of the whales we select can be closely examined down to the block, allowing us to be prudent with our data selection. We chose to record three distinct features as our whale signals: Ethereum balance changes, Ethereum to stablecoin ratio, and finally interactions with known trading exchanges. At a high level, all of these signals can potentially indicate a whale's intent to move the market. An increase in Ethereum stores across the whales likely points to a bullish market attitude. A general shift to buying stablecoins could indicate a coming market downturn. Similarly whales sending Ethereum onto exchanges could also indicate an incoming selloff. Section 4.2 details more about our data collection methods and results.

Finally, we use machine learning to interpret our data. Theoretically a machine learning model would be able to interpret a whale's data and identify patterns not immediately apparent to human traders, thereby providing us with an edge. For this reason, ML models are also commonly utilized in price predictions for financial institutions. We use the recurrent neural network Long Short-Term Memory (LSTM) for our design. Research [3] has shown LSTM to be good for time series analysis, such as price prediction. One of the main advantages of LSTM is its ability to capture complex temporal dependencies in sequential data. In price prediction, LSTM can learn to recognize complex patterns and trends in historical price data, including cyclicity and trends, and use this information to make predictions about future prices [3]. This is

particularly important in the cryptocurrency market, which is highly volatile and subject to sudden price movements. LSTM also operates well with noisy and incomplete data. The cryptocurrency market is highly complex and dynamic, as a result the whales we track are constantly trading, buying, or moving around their assets. LSTM's ability to filter out irrelevant information and focus on the most significant features in the data improves its accuracy in predicting future price movements. Neural networks like LSTM are also known to operate well with small sets of data, which is perfect because we only have a year's worth of data.

Our data will be split 60%, 20%, and 20% for training, validating, and testing. This means the first 219 days will be used as training data, the next 73 as validating, and the last 73 days used for testing data. We evaluate our model through its returns and risks. Additionally, we examine its strengths, weaknesses, and other noteworthy characteristics. For additional insight, we will compare our model results from momentum trading, a simple but well known traditional technique. Figure 1 demonstrates that momentum strategies have a remarkable capability to generate profits even in Ethereum's volatility. If our on-chain analysis model is able to make predictions on par with momentum or prove useful in a different way, it would be well within reason to say that we have developed a good trading strategy.



Figure 1: Momentum returns & Eth price growth from 2018

Momentum can be calculated easily using pandas. Here is a snippet of our code:

```
df = pd.read_csv('eth_price.csv',header=None, names=['date', 'open_price'])

df['momentum_returns'] = df['open_price'].pct_change(periods=n)

df['signal'] = 0 # 0 = hold, 1 = buy, -1 = sell

df.loc[df['momentum_returns'] > 0, 'signal'] = 1 # buy signal

df.loc[df['momentum_returns'] < 0, 'signal'] = -1 # sell signal</pre>
```

4. Implementation

The next few sections go into detail on how we collected whale addresses, their signals, and finally the design of our price prediction model. Afterwards, we will also make a comparison with our strategy and momentum to put our methods in perspective.

4.1 Whale address collection

By our definition of whale wallets, we searched for active addresses with an Ethereum balance in the top 5%. Through Alchemy's RPC endpoints, we randomly selected 10 blocks mined during 2022, collected their lists of transactions, then used those transactions to create a database of 2186 unique active wallets. We then picked the top 5% of the richest wallets to designate as our whales, resulting in 109 unique whale addresses.

However, because we scrapped our whale addresses directly from the blockchain, our list also contained protocol vaults, exchange deposits, bridge portals, and even mining pool addresses. These corporate and organization owned wallets don't tend to behave like ordinary whales, so we opted to prune them from our list. Unfortunately, due to the decentralized nature of blockchains, there is no easy way of identifying if an address is owned by a business or

protocol. As a result each address on our list had to be individually examined. By the end we had a total of 92 clean whale addresses.

The largest whale we tracked was 0xD7e... with 12,035.15 ETH worth \$14.7 million. The smallest whale we tracked was 0x8c4... with 26.6 ETH worth \$31,900. A complete list of our wallets can be found here: *Data removed for privacy reasons*

4.2 Signal collection

Using the list of whales, we collect daily data signals from each address about their Ethereum balances, Ethereum to stablecoin holding ratio, and interactions with known centralized exchanges during 2022. Each new day begins at midnight UCT.

4.2.1 Historical Ethereum Balances

Collecting the historical Ethereum balance of a wallet proved surprisingly difficult due to the structure of the blockchain. The web3.eth Python library and Etherscan's API both contained methods which returned an address's balance at a given block. However, querying 365 times for each of our 92 whales proved to be infeasible due to the frequency of crashes. As a work around, through Etherscan we instead collected the transaction history of each of our whales. Using a wallet's balance on January 1st, we then build their past Ethereum balance using their transaction history during 2022, thus completing our first set of signals.

Figure 2 displays the combined Ethereum of our whale wallets. At the peak, the whales held a total of 155363 Ethereum, about 0.125% of the entire supply.

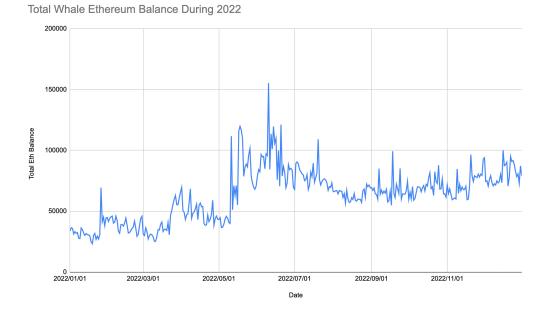


Figure 2: Total Whale Ethereum Balance During 2022

Shifting Ethereum balances of whales can also provide insights into market manipulation.

After all, market manipulation is something whales are known to do. However the cryptocurrency market is highly complex and dynamic. As a result, a wallet could have multiple motives for changing their ethereum balance. But with a diverse spread of whales and data, we reduce our exposure to irregularities while also increasing our exposure to general trends indicative of underlying market sentiments.

4.2.2 Ethereum to Stablecoin Ratio

Moralis has a convenient API endpoint for tracking ERC 20s. This allowed us to quickly retrieve data about our whale's stablecoins. We focused on the three most popular stablecoins: USDC, USDT, and DAI. For each whale, we recorded the amount of stablecoins they owned each day with a simple querying loop. Using our previously collected data on Ethereum balances, we are able to create a new signal presented below in figure 3.

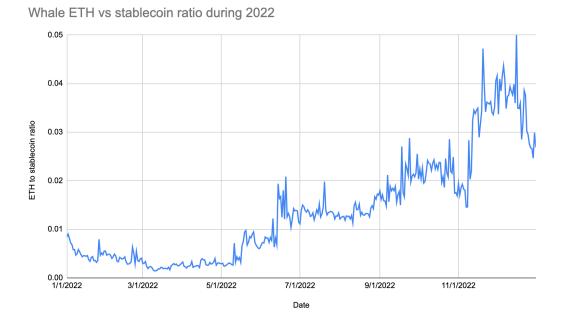


Figure 3: Combined Whale Ethereum to Stablecoin ratio During 2022

The stability of stablecoins makes them attractive to traders who want to avoid the risk of price volatility. Moreover, stablecoins are commonly used as a hedging tool. Traders can protect their investments against price fluctuations by converting their cryptocurrencies into stablecoins during times of market volatility. This can help to reduce risk and protect gains, particularly in a market where the prices can fluctuate wildly.

As such, it should come as no surprise to see a close inverse relationship between Figure 3 and the price of Ethereum. 2022 as a whole was a massive bear market, and on June 18th Ethereum even dipped below \$1,000. Curiously, there is a sharp increase in our ETH to stablecoin ratio right on June 14th just before the price dipped, suggesting that our whales could have foreseen or even caused the downturn. Evidently this signal in particular appears very strong.

4.2.3 Interactions with known Centralized Exchanges

During the process of pruning our whale wallets described in section 4.1, we also performed a number of searches on Etherscan to collect a list of corporate wallets. Etherscan provides public name tags to any wallet an organization publicly declares ownership of. This feature allowed us to collect a total of 540 wallets belonging to centralized exchanges, which was essential for this part of data collection. Using our whale transaction histories again, we can record every interaction our whales made with a known centralized exchange.¹³

Figure 4 shows that almost all of these transactions brought more Ethereum from our list of exchanges to whales. Here is the list of centralized exchange wallets we used:

☐ Centralized Exchange Wallet List

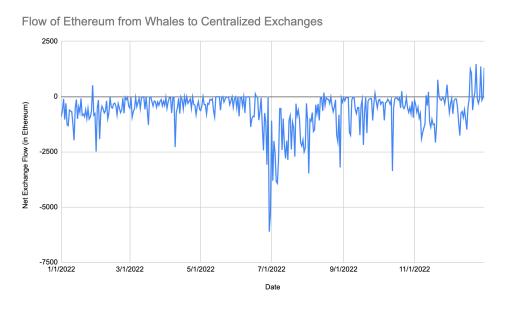


Figure 4: Net Flow of Ethereum From Whales to Centralized Exchanges During 2022

-

¹³ Annoyingly Etherscan does not actually have a list of these named organization wallets available for easy reference. They more or less have to be found manually through the explorer. As a result, there are certainly more than just 540 wallets belonging to centralized exchanges. Unfortunately, I simply cannot find them all.

Centralized exchanges are one of the few ways crypto currency traders can get their money off-chain or put their money on-chain. While there are many reasons why a whale may change in their Ethereum balance, an interaction with a centralized exchange is usually a direct indicator of underlying market sentiments. If someone is moving thousands of Ethereum onto an exchange, it probably doesn't mean that they plan to hold onto that Eth for much longer. Similarly, the fact that so much Ethereum moved off of exchanges and into whale wallets in the middle of the year suggests that Eth isn't going to be sold anytime soon.

4.3 Price Prediction Model

As mentioned before, LSTM neural networks are well-suited for price prediction because they are capable of modeling complex temporal dependencies in the data [3]. In finance and cryptocurrency markets, prices are influenced by a wide range of factors that can be difficult for humans to pick up. LSTM models, on the other hand, are able to capture these complex interactions and changes by incorporating the sequence of historical price and other relevant data as inputs. The model's hidden state, which is updated at each time step, allows it to remember important features of the data and update its predictions based on new information. This makes LSTM well-suited for analyzing time-series data, such as price data. They are able to identify patterns and trends in the data that can be used to make price forecasts.

We begin by preprocessing our whale signals through normalizing and scaling it to ensure that the model is better able to capture temporal patterns. We also divided the data into training, validating, and testing sets by 60%, 20%, and 20% of the total data, with the training set comprising 219 days, the validation set 73 days, and the remaining 73 days used for testing the model's performance.

Validation data sets are an essential part of machine learning models, especially ones as delicate as price prediction models. While LSTM is extremely powerful in their abilities to analyze complex interactions, this also increases the likelihood of overfitting and creating false correlations. Through the validation data set, we determine that approximately 80 epochs of training with one hidden layer appears to be optimal as our validation MSE no longer decreases as much beyond that point. We also noticed that our model was very sensitive and frequently made small predictions of inconsequential price shifts that could open it up to unnecessary risk. As a result we also added a threshold of 1.5% to control the sensitivity of a model and improve its reliability. Our results are shown on Figure 5.

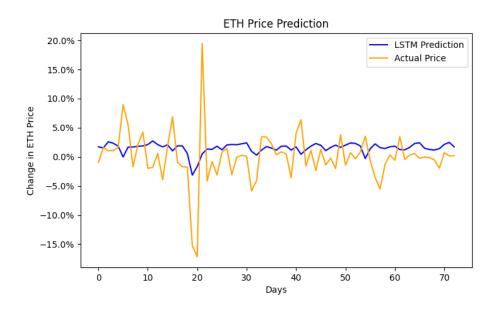


Figure 5: LSTM Price Predictor on the last 73 days of 2022

Although the model is not perfect and fails to capitalize on the many ups and downs of the market during this two month period, it appears adept at sensing significant signals. In fact, its inability to capture every market movement is a good indicator that our model doesn't suffer

from overfitting. Additionally, while the model generally predicts small positive prices every day, it is remarkably accurate at predicting the crash at day 20.

In a more comprehensive light, Figure 6 displays what a portfolio starting with 100 dollars would look like following our strategy, assuming that shorting is allowed.



Figure 6: LSTM Price Predictor compared to Relative Eth Price

4.4 Comparisons

Momentum trading is a popular strategy in financial markets that involves buying assets that have shown an upward trend in price over a given period, and selling assets that have shown a downward trend in price. This strategy is based on the idea that assets that have performed well in the past are likely to continue performing well in the future, and vice versa.

As evidenced by Figure 1, under the right situations momentum can be exceptionally profitable as well. For this section, in order to better understand the usefulness of our own strategies, we will compare our results in Figure 6 with momentum in Figure 7.

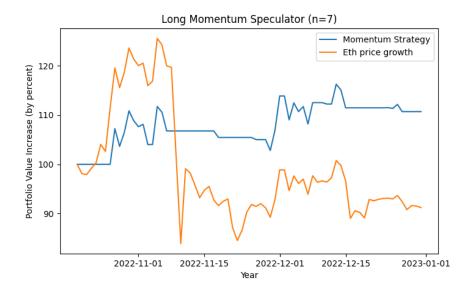


Figure 7: Momentum Trading & Relative Eth Price during the last 73 days of 2022

Interestingly, even though our results from Figure 6 solidly beat out the strategy in Figure 7, momentum trading has performed exceptionally well in these circumstances. Unlike our whale strategy, momentum lacks specific signals for its predictions, instead entirely relying on past price trends. Thus, the fact that it has been able to generate a profit in the bear economy is impressive.

On the other hand, our strategy was almost able to achieve a 50% return. While more statistics, analytics, and data collection needs to be done, it is evident with our findings that we have created a solid trading strategy.

5. Evaluation

While the returns of our algorithm in Figure 6 were impressive, a closer inspection on the performance of our algorithm reveals that, aside from the success around day 20, our returns are more or less identical to Ethereum's general fluctuations. That itself is not a significant issue.

After all, we trained our strategy solely based on whale signals under the impression that whales

are major market movers. Therefore it would make sense that our strategy was most effective during the moment of major price movement because such actions are likely directly influenced by a whale.

However, in closely following whale signals and behaviors, we can unintentionally fall into a trap. Back in our approach in section 3 we noted that Manahov [1] warned against simply mimicking whales because whales can create false signals, duping traders to invest only to pull out and dump their now appreciated assets. It's difficult to say if our approach can fall into a similar trap because we only monitor whales from a daily perspective and thus reduce impulsive behavior. But the fact that whales control such a disproportionate amount of power suggests some risk in using them as our main strategy. Unfortunately the data we collected only had one instance of a significant price shift. Therefore, it is inadequate to definitively prove how risky our approach was. Despite that, more generally, the fact that our model was able to solidly predict the sudden downturn of Ethereum prices on day 20 in Figure 6 indicates undeniable proof that whale signals can be used to great effect.

6. Conclusion

This paper has presented a model for cryptocurrency price prediction using on-chain whale data, and our results have shown promising performance despite only having a year's worth of data to train on. By collecting and analyzing on-chain data, we were able to capture important information about whale behavior and use it to make accurate predictions about future price movements.

As mentioned before, unfortunately there is a lack of academic research related to the entire field of cryptocurrencies and even fewer about whales in particular. But the significance of whales cannot be overstated. As the blockchain industry continues to expand, more and more

individuals will end up buying and trading cryptocurrencies. Therefore, for the most part,

whales, their signals, and influences will also continue to exist.

References

[1] Viktor Manahov (2021) Cryptocurrency liquidity during extreme price movements: is there a

problem with virtual money?, Quantitative Finance, 21:2, 341-360, DOI:

10.1080/14697688.2020.1788718

[2] Bouri, Elie, et al. "Herding Behaviour in Cryptocurrencies." Finance Research Letters, vol.

29, 2019, pp. 216–221., https://doi.org/10.1016/j.frl.2018.07.008.

[3] M. A. Istiake Sunny, M. M. S. Maswood and A. G. Alharbi, "Deep Learning-Based Stock

Price Prediction Using LSTM and Bi-Directional LSTM Model," 2020 2nd Novel Intelligent and

Leading Emerging Sciences Conference (NILES), Giza, Egypt, 2020, pp. 87-92, doi:

10.1109/NILES50944.2020.9257950.

Code:

https://github.com/pie575/IW-2023

20