

# Cryptocurrency Price Prediction using Sentiment Analysis

Abdul Rehman Khurshid

akhurs18@earlham.edu

Computer Science Department at Earlham College

Richmond, Indiana, USA

## 1 INTRODUCTION

As the economic and social impact of cryptocurrencies continues to grow rapidly, the importance of related news articles and social media posts also increase, particularly tweets as Twitter is growing popular in the financial world as fresh news and advice are from the top players in the world of finance appear on Twitter [8]. As seen previously with traditional financial markets, there appears to be a relationship between media sentiment and the prices of cryptocurrency coins. The world is more connected than ever, and the traditional ways of communication, such as newspapers and word of mouth are declining, and the use of social media is rising. This is why it has an impact on the consumer's decisions on buying and selling. While many causes of cryptocurrency price fluctuation can be seen, it is important to explore whether sentiment analysis on available online media can inform predictions on the prices of coins [5].

This project analyzes the influence of news and social media data to predict price fluctuations for two cryptocurrencies: Bitcoin and Cardano. In addition to its growing popularity, the low barrier of entry and high data availability of the cryptocurrency market makes it a subject of study. Social media is increasingly used as a news source that influences purchase decisions by informing users of the currency and its increasing popularity [10].

## 2 DATASET

To predict cryptocurrency price changes, several different data sources are considered as possible inputs to the model. The first input taken is sentiment analysis of collected tweets about Bitcoin and Cardano. The second is Google Trends data, and the third is tweet volume. This section details how each of these data sources are going to be gathered.

### 2.1 Tweets

Stenqvist [9], Abraham [5] and Valencia [6] all used data extracted from Twitter as their data set for sentiment analysis. Twitter is a popular social network where messages are shared between users. Data can be mined through Twitter API. Keys are required which I can get by applying on Twitter's developer site. A lot of the researchers use Twitter as a source because these researchers believe that public sentiment affects the cryptocurrency market.

### 2.2 Twitter Tweet Volume

As the Python library Tweepy can not be used to find the count of the Tweets, the Tweet volume can be gathered from bitinfocharts [1] which provides the number of Tweets by day about both Bitcoin and Cardano.

### 2.3 Google trends

Abraham [5] gathered data from Google Trends. The internet is navigated is through search engines, and Google is by far the world's most popular search engine accounting for 74.52 percent of all internet searches [5]. This means that Google search data can provide a greater insight into what the world is interested in, and how interested in any given topic it is. Google Trends data provides information on how popular given search terms are relative to other search terms at any given time. Furthermore, these search term popularity values can be compared over a period of time. By using Google trends, the popularity of cryptocurrency over the last five years could be found and, this data compared with the price charts of Bitcoin and Cardano is useful for my analysis [2].

## 3 METHODS OF PRICE PRIDITION

### 3.1 Sentiment analysis

Sentiment Analysis is a text classification tool that analyses a sentiment, to calculate if the sentiment is positive, negative or neutral. There are researchers who have solely used the data on the sentiment of the public to predict the stock market and cryptocurrency prices, such as Abraham [5] and Stenqvist and Lönnö [9].

Abraham [5] conducted a Twitter sentiment analysis to predict cryptocurrency prices. The data was analyzed to determine if it would be a valuable input to the final model. VADER sentiment analysis determined tweets to be more neutral, which would make the results less efficient as if the public sentiment is neutral because neutral sentiment usually does not indicate a pattern for buying or selling. Both Google Trends and tweet volume were highly correlated with price. A linear regression algorithm was used to predict Bitcoin closing daily price. However, Twitter sentiment on cryptocurrencies tends to be positive regardless of future price changes. A more complex model than linear regression could be used in future work to improve the results as the results from this paper were taken when the prices were only going up.

Lamon [8] used news and social media sentiment to predict cryptocurrency prices. The model uses a classifier to learn feature weights that are used for labelling data. Linear support vector classification, multinomial Naive Bayes, and Bernoulli Naive Bayes were tried. However, logistic regression produced the best results. This model was able to predict the largest price increases and decreases correctly. As this research paper analyzed news and Twitter data separately, a more efficient result can be obtained if the model can work with a combination of news and social media data.

Valencia, Gómez-Espinosa, and Valdés-Aguirre [10] proposed a model with the use of machine learning tools and social media data predicted the prices of a specific few cryptocurrencies. This model utilized the use of neural networks (NN), support vector machines (SVM) and random forest (RF). The results for this model

show that predicting cryptocurrency is possible through sentiment analysis and by using machine learning tools. Furthermore, some cryptocurrencies prices can be predicted through Twitter data only, usually the ones which have a large following.

Stenqvist and Lönn [9] investigated public sentiment from a posts on Twitter to predict the price of Bitcoin. 2.27 million Bitcoin-related tweets were gathered for sentiment analysis to indicate a price change for the near future. This is done by a method of solely attributing fall or rise based on the severity of aggregated Twitter sentiment change over time periods ranging between 5 minutes and 4 hours, and then shifting these predictions forward in time 1,2, 3 or 4 time periods to indicate the corresponding BTC interval time. This method yielded an 83 per cent accuracy. Further, a prediction was only made when the mean of sentiment was limited by a minimum 2.2 percent change. This analysis can be improved in the future by adding a domain-specific lexicon which would yield a more representative sentiment.

Xin Huang [7] proposed LSTM based sentiment analysis model. The data gathered to identify the sentiment was from the most popular Chinese social media platform Sina-Weibo. A long short-term memory (LSTM) based recurrent neural network was used along with the historical cryptocurrency prices to predict the price trend for the future. The results yielded an 87 per cent accuracy rate. This was 15.4 higher than the traditional autoregression method which is used.

Bharathi and Geetha [6] investigated the sentiment of the public using social media data to predict stock market prices. Using an algorithm for sentiment analysis, the correlation between the stock market values and sentiments in the social media data is established. Moving average method was used as an indicator to predict prices. The results show that the moving average method, in addition to sentiment analysis, yields a 14.43 per cent higher efficiency than when only the moving average method is used.

**3.1.1 VADER.** Although there are multiple methods available, I will use the "VADER" (Valence Aware Dictionary and sEntiment Reasoner) system in this analysis. The goal of this analysis is to apply sentiment analysis to tweets that are collected so that it can be determined if the tweets are generally positive or negative in their opinions of cryptocurrencies. VADER is a lexicon and rule-based sentiment analysis tool that is specifically made for sentiments expressed on social media. VADER calculates a sentiment score which can be used to find out how positive or negative a sentiment is [4]. VADER is used by some of the researchers who have worked on this topic, such as Evita Stenqvist and Jacob Lönnö [9]. In their performance review, VADER was compared against eleven other semantic analysis tools for determining its sentiment. VADER consistently performed and was among the top in all test cases and outperformed all the other techniques in the social media text-domain.

## 3.2 Machine learning algorithms

I will be using both logistic regression and linear regression in my model.

**3.2.1 Linear Regression.** Linear Regression is a machine learning algorithm that is based on a supervised regression algorithm. Regression models target prediction values based on independent variables. It is used for finding out the relationship between variables and forecasting. Different regression models are unique based on the relationship between the dependent and independent variables, they consider the number of independent variables being used [3].

**3.2.2 Logistic Regression.** Logistic regression is used to classify elements of a set into two groups (binary classification) by calculating the probability of every element of the set. In this classification problem, the target variable,  $y$ , can take only discrete values for a given set of features(or inputs),  $X$  [3].

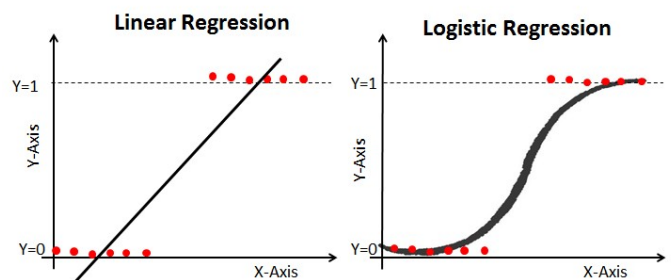


Figure 1: Logistic regression vs Linear Regression

## 4 CONCLUSION

This literature review discussed how sentiment analysis is used to predict cryptocurrency prices using multiple sources of social media such as Twitter and Sina-Weibo. Some researchers use social media to predict cryptocurrency prices as many studies prove that machine learning tools in addition to sentiment analysis produces significantly higher efficiency then when only machine learning tools are used [10]. In this literature review, I explained which method I will use to predict the price changes of coins depending on the sentiment of the public using social media data.

## REFERENCES

- [1] [n.d.]. bitinfocharts. [www.bitinfocharts.com](http://www.bitinfocharts.com).
- [2] [n.d.]. Google Trends. <https://trends.google.com/trends/?geo=US>.
- [3] 2020. ML | Linear Regression vs Logistic Regression. <https://www.geeksforgeeks.org/ml-linear-regression-vs-logistic-regression/>.
- [4] 2021. Python | Sentiment Analysis using VADER. <https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/>.
- [5] Jethin Abraham, Daniel Higdon, John Nelson, and Juan Ibarra. 2018. Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Science Review* 1, 3 (2018), 1.
- [6] Shri Bharathi and Angelina Geetha. 2017. Sentiment analysis for effective stock market prediction. *International Journal of Intelligent Engineering and Systems* 10, 3 (2017), 146–154.
- [7] Xin Huang, Wenbin Zhang, Yiyi Huang, Xuejiao Tang, Mingli Zhang, Jayachander Surbiryala, Vasileios Iosifidis, Zhen Liu, and Ji Zhang. 2021. LSTM Based Sentiment Analysis for Cryptocurrency Prediction. *arXiv preprint arXiv:2103.14804* (2021).
- [8] Connor Lamon, Eric Nielsen, and Eric Redondo. 2017. Cryptocurrency price prediction using news and social media sentiment. *SMU Data Sci. Rev* 1, 3 (2017), 1–22.
- [9] Evita Stenqvist and Jacob Lönnö. 2017. Predicting Bitcoin price fluctuation with Twitter sentiment analysis.

[10] Franco Valencia, Alfonso Gómez-Espinosa, and Benjamín Valdés-Aguirre. 2019. Price movement prediction of cryptocurrencies using sentiment analysis and

machine learning. *Entropy* 21, 6 (2019), 589.