Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 9, September : 2023

CRYPTOCURENCY PRICE PREDICTION USING SENTIMENT ANALYSIS

P.V.R. Vishnu Prasad DEPARTMENT OF COMPUTER SCIENCE AND SYSTEMS ENGINEERING (A), ANDHRA UNIVERSITY, VISAKHAPATNAM-530 003.
Dr. G. LAVANYA DEVI Assistant Professor in DEPARTMENT OF COMPUTER SCIENCE AND SYSTEMS ENGINEERING (A), ANDHRA UNIVERSITY, VISAKHAPATNAM-530 003.

Abstract: The world's population produces 2.5 million terabytes of data every day in 2017 Every every day, there are 500 million tweets and 1.8 billion Facebook posts. These tidbits of information can be about anything, from the user's lunch to their displeasure with a referee in a football game. Particularly Twitter has established itself as a venue for the rapid and succinct dissemination of news. The level of public trust in a specific financial commodity is a major determinant of that commodity's value. Since social media's beginnings, it has given a forum for the expression of thoughts; The frequency of cryptocurrencies-related news stories and social media posts, notably tweets, is increasing quickly along with the economic and societal impact of cryptocurrencies. Similar to conventional financial markets, there seems to be a connection between media coverage and cryptocurrency coin prices. While there are many factors that influence cryptocurrency prices, it is important to investigate whether sentiment analysis of publicly accessible web media may help forecast whether a coin's price (or perceived value) will increase or decrease. In this project, our team specifically tried to address the following two questions: (1) Can sentiment analysis of headlines from news articles and/or social media posts make reliable forecasts about how much bitcoin, litecoin, and ethereum will fluctuate in price in the future? (2) If both sets of forecasts come true Text data from news headlines and tweets that have been aggregated by day and preserved in chronological order to maintain the time-series character of the data are the system's input. Then, each news item and tweet was given a label of 0 or 1 for each coin, denoting predictions of a price reduction or increase one day in the future, respectively, using conventional supervised learning binary classification algorithms. The final daily prediction was then determined using the majority label for each coin, on each day. Our long-term objective is to improve this price prediction algorithm and integrate it into a bigger system that automatically and wisely manages a cryptocurrency portfolio.

I Introduction

1.1 DIGITAL MONEY: EARLY ATTEMPTS

One of society's essential forces is money, an economic resource that is present everywhere but varies widely from society to society. It has a close relationship to quantification and is frequently used to denote one's ability to make purchases and to measure wealth and well-being.

Money has existed from the beginning of society. With economic, technical, and sophisticated commercial needs, money continued to change. Money has existed in many forms over the years, from bartering to cryptocurrencies, but its fundamental nature has not altered. Whatever its form, all agree that money serves three fundamental purposes: as a store of value, a means of trade, and a unit of account. While every type of currency Bartering, commodity money, metallic money, paper money, credit money, and electronic money are the six main stages of money's evolution. The development of digital money, which revolutionised the way people pay for goods and services and laid the groundwork for online commerce, is the most fascinating, important, and revolutionary phase. This begs the questions of how technology revolutionised money and what the underlying technology is. Even before bitcoin, the concept of digital currency was present. In the 1990s, efforts were made to develop digital currency by putting various computer programmes and cryptographic methods into practise. In order to streamline payments between petrol stations and customers, the Netherlands was the first country to create digital money.

1.2 THE DOUBLE-SPENDING PROBLEM, BYZANTINE GENERAL'S PROBLEM, AND THE ORIGIN OF BITCOIN

For a variety of reasons, including scalability, limited merchant adoption, power constraints, network congestion, network security, legal and regulatory challenges, digital money ventures before bitcoin have failed despite efforts. Something needs to be impossible to counterfeit in order to be used as money digitally, and the system that generates, distributes, and manages money needs to be robust enough to handle it. How to transmit value peer-to-peer without a middleman was the main problem that inspired the development of bitcoin. As a result, digital currency had to overcome:

Double-spending problem

The double-spending issue arises when a transaction's replicated copy is utilised to transact more than once, which means that a digital currency's transactional information, which is stored as bits, can be duplicated to spend again. Comparable to safeguarding conventional currency from counterfeiting, the double-spending issue with cryptocurrencies.

A payee cannot confirm double-spending unless there is a centralised, trusted middleman in place to do so. In a conventional banking system, where only money issued by a central authority is trusted, central banks are tasked with avoiding duplicate spending. The objective was to create a system that would enable value to be transferred without the use of a middleman and to prevent double spending.

Byzantine general's problem

Byzantine general's problem involves generals of a large army deciding with a common course of action either to attack or retreat with an assumption that some of the generals are traitors. To achieve consensus, the generals must have a mechanism in place which ensures that:

- All loyal generals decide on the same course of action
- A small number of traitors cannot influence loyal generals to adopt a bad plan

Since there is no centralised authority to stop the spread of incorrect information throughout the network, a Byzantine failure might happen in a decentralised system. Decentralised systems may experience byzantine failures for several causes, including malicious communication, network issues, and hardware malfunctions. The ability of the network to continue operating in the presence of Byzantine faults is known as byzantine fault tolerance in the context of blockchain. The proof-of-work consensus method used by the Bitcoin network, which requires network members to reach consensus before adding a block to the existing chain, provides Byzantine fault tolerance. This stops rogue nodes from submitting fraudulent transactions with the aim of double spending.

In order to solve the double-spending and Byzantine general's dilemma and avoid the necessity for a third party, bitcoin is innovative.

1.3 THE TECHNOLOGY BEHIND CRYPTOCURRENCIES: BLOCKCHAIN

Following the success of bitcoin, which securely kept transactions, blockchain technology became extremely popular. Stuart Haber and colleagues suggested a method to date digital documents such that the document cannot be tampered with, which resulted in the introduction of the first and longest-running Blockchain.

Blockchain is a sort of distributed ledger technology that uses ledgers or blocks to progressively store transactions on each system connected to the network.

Blockchains are intricate systems that utilise layers of the network, infrastructure, data, consensus, and application. The data layer, which is also the basic architecture, includes cryptographic techniques, chain structure, and data blocks. In addition to block propagation verification, blockchain transactions and implementation happen at the network layer. In distributed processes that demand agreement on a single data value or a network state, the consensus layer has a fault-tolerant mechanism. For instance, Ethereum and Bitcoin use Proof of Stake (PoS) and Proof of Work (PoW), respectively. The application layer focuses on various blockchain solutions from different industries, including money and other financial products.

2 Literature survey

The most well-known and well-established cryptocurrency is Bitcoin. In contrast to "normal" currencies, Bitcoin's value was derived from computer complexity rather than a physical good. Bitcoin is essentially a piece of open source software that runs on networked computers called nodes. Together, these nodes form a distributed database known as the blockchain. The blockchain acts as the single source of truth for all network transactions and enables Bitcoin to operate as intended, touching on the fields of cryptography, software engineering, and economics

The blockchain can be used for any system in which one would exchange value, even if the Bitcoin currency is its most well-known application This is because it forbids Fiat currencies are governed, hence there are flaws in how the controlling body chooses to affect a currency. By employing quick fixes to address issues or crises, such as printing money, which increases the quantity of money while decreasing its value, irresponsible monetary policies can create an artificial long-term deflation

Contrarily, Bitcoin lacks a central authority and any means of directly influencing its price or supply By design, this eliminates the intermediary that most monetary systems—the central bank and the banking system—are built around [4]. Participating in transaction calculations is the sole way to increase the supply of Bitcoins, which causes it to expand predictably over time and The same factors that affect a fiat currency's value also affect cryptocurrency's value concurrently The architecture of the Bitcoin network is also indicative of the decentralised approach. In order for Bitcoin to function as the decentralised peer-to-peer network of nodes it is designed to be, at least half of the peers must approve any changes to the architecture or technical implementation details The shared database, often known as the ledger or blockchain and of which all nodes have a copy, is a component of the decentralised design. Both historical transactions and present Bitcoin owners are listed in this ledger . The database is built up in chunks of time-related transactions. Previous attempts to use sentiment from tweets to forecast changes in the price of bitcoin have failed. Using comparable supervised learning methods, Coliannni et al. [1] achieved 90% accuracy in price fluctuation prediction; however, their data was labelled using an online text sentiment API. Therefore, rather than accuracy in terms of forecasting price changes, their accuracy assessment related to how well their model matched the online text sentiment API. Similar to this, Stenqvist and Lonno [2] used deep learning algorithms to forecast changes in the price of bitcoin with 79% accuracy using 2.27 million tweets, but at a considerably greater frequency of every 30 minutes. Neither of these strategies examines the average size of price variations or uses data that is directly labelled based on price movements.

3 Implementation Study

- ML models based on the Twitter dataset. The researchers aim to find an association between user sentiment and BTC price. However, they use a variety of algorithms, such as Support Vector Re gression, Decision Tree Regression DTR, and Linear Regres sion LR. As a result of the experiment, there is a discernible relationship between sentiment on Twitter and price change, based on the highest accuracy obtained from the decision tree algorithm compared with other algorithms is 75%. Similarly,

3.1proposed methodology

- The best performing models from each of the aforementioned predictive tasks, more specifically, the Direction-BiLSTM and Magnitude-CNN models, were merged together to create a voting classifier model which takes into consideration the outputs from the two models. As Fig. 7 shows, the voting classifier works by first predicting the next day's closing price direction and then, the magnitude of the next day's closing price using the second model. Ten, it checks whether the next day's closing price direction matches the direction of the predicted change magnitude. In other words, a match happens: (i) if the first model outputs a 0, which means a decrease in price, and the second model outputs a class from 1 to 5 (negative magnitude of price change); or (ii) if the first model outputs a 1, which means an increase in price, and the second model outputs a class from 6 to 10 (positive magnitude of price change). The prediction of the next day's closing price direction is kept if there is a match in the output of the two classifiers. Moreover, the voting classifier is evaluated on 50 different runs with 50 differently shuffled datasets.

4. Methodology

MODULES:

4.1 Data Preprocessing and Cleaning

Sentences pre-processing, which is the first step in our method, converts the Arabic sentences to a form that is suitable for a sentiment analysis system. These pre-processing tasks include Punctuations removal, Latin characters removal, Stop word removal, Digits removal, Tokenization, Normalization, and Light Stemming. These linguistic are used to reduce the ambiguity of words in order to increase the accuracy and the effectiveness of our approach.

4.2 Tokenization

Tokenization is a method for dividing texts into tokens; Words are often separated from each other by blanks (white space, semicolons, commas, quotes, and periods). These tokens could be individual words (noun, verb, pronoun, and article, conjunction, preposition, punctuation, numbers, and alphanumeric) that are converted without understanding their meaning or relationships. The list of tokens becomes an input for further processing. In this work, we use "Tokenizer" from *Keras* [1], which is the Python Deep Learning library.

4.3 LSTM (long Shorth Term Memory Network)

It's a unique type of recurrent neural network that can learn long-term data relationships. This is possible because the model's recurring module is made up of four layers that interact with one another. An LSTM module has a cell state and three gates, giving it the ability to learn, unlearn, or retain information from each of the units selectively. By permitting only a few linear interactions, the cell state in LSTM allows information to travel across the units without being altered. Each unit contains an input, output, and a forget gate that adds or removes data from the cell state. The forget gate utilizes a sigmoid function to determine which information from the previous cell state should be ignored. The input gate uses a point-wise multiplication operation of 'sigmoid' and 'tanh' to control the information flow to the current cell state. Finally, the output gate determines which data should be transmitted on the next hidden state

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 9, September : 2023

LSTM can be used in many applications such as for weather forecasting, NLP, speech recognition, hand writing recognition, time-series prediction, etc .The cell state is represented by the horizontal line that runs across the top of the figure. The condition of the cell is similar to a conveyor belt. This flows straight down the chain with just minimal linear interactions. The ability of LSTM to add or delete information from the cell state is controlled by gates. Gates are used to allow information to pass through if desired. A sigmoid neural net layer plus a point wise multiplication operation make up gates. The sigmoid layer produces values ranging from 0 to 1, indicating how much of each component should be allowed to pass. Let nothing through with a value of 0, and everything through with a value of 1! To safeguard and govern the cell state, an LSTM contains three of these gate

4.4 Algorithm(BI_LSTM)

Two unidirectional LSTMs that process the sequence in both forward and backward directions make up the bidirectional LSTM architecture. This design can be thought of as having two independent LSTM networks, one of which receives the token sequence in its original order and the other of which receives it in reverse. The output from each of these LSTM networks is a probability vector, and the combined output is the result of these probabilities.

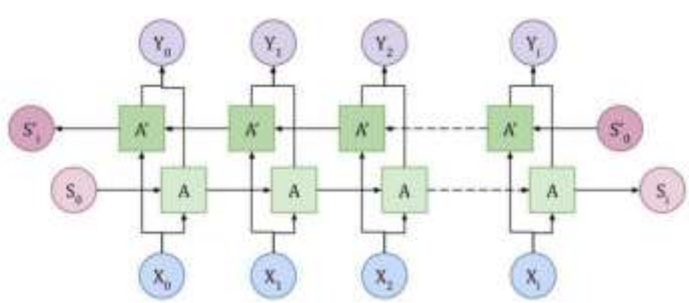

Figure1:_ proposed BI-LSTM Model

review system using BiLSTM layers in Python using the Tensorflow library. We would be performing sentiment analysis on the IMDB movie review dataset. We would implement the network from scratch and train it to identify if the review is positive or negative.

5 Results and Evolution Metrics

	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verif
0	[illegible]	Atlanta, GA	[illegible] real estate, fintech, startups...	[illegible]	8534.0	[illegible]	[illegible]	False
1	[illegible]	NaN	BITCOINLIVE is a Dutch platform aimed at bit...	2019-10-17 20:12:10	[illegible]	[illegible]	[illegible]	False
2	[illegible]	London, England	[illegible] The best #forex, [illegible]	2014-11-10 10:50:37	[illegible]	[illegible]	[illegible]	False
3	Crypto is the future	NaN	I will post a lot of buying signals for BTC tr...	[illegible]	[illegible]	[illegible]	[illegible]	False
4	[illegible]	Europe	Co-founder [illegible] | Forbes 30Under30 | I...	[illegible]	[illegible]	[illegible]	[illegible]	False

Fig. No. 2: Sample Diagram from Dataset

ends	user_favourites	user_verified	date	text	hashtags	source	is_retweet	compound	score
	1712.0	False	2021-06-22 05:22:52	nice project \n\n\n\n\nortcoin ort okratech ...	['ortcoin', 'ort', 'okratech', 'bitcoin', 'air...	Twitter for Android	False	0.5994	1025.108064
	45077.0	False	2021-08-25 07:06:40	Long Bitcoin short the banks	['Bitcoin']	Twitter Web App	False	0.0000	0.000000

Fig. No.3: Score of Each Tweet

	Date	Open	High	Low	Close	Adj Close	Volume
0	2014-09-17	465.864014	468.174011	452.421997	457.334015	457.334015	21056800
1	2014-09-18	456.859985	456.859985	413.104004	424.440002	424.440002	34483200

Fig. No. 4: Fit rate of Bitcoin

Fig. No. 5: Cryptocurrency compared with Sentiment

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 9, September : 2023

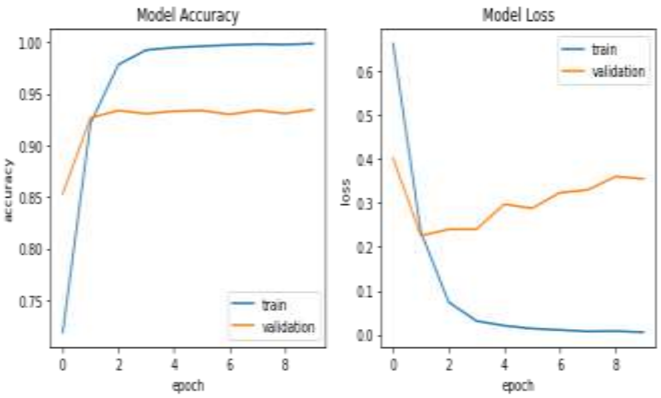

Fig. No. 6: Training and Validation Loss

Accuracy: 93.5%

	precision	recall	f1-score	support
0	0.84	0.78	0.81	478
1	0.94	0.95	0.95	1782
2	0.95	0.95	0.95	2224
accuracy			0.93	4484
macro avg	0.91	0.90	0.90	4484
weighted avg	0.93	0.93	0.93	4484

Fig. No. 7: Accuracy and Evaluation Matrix Results

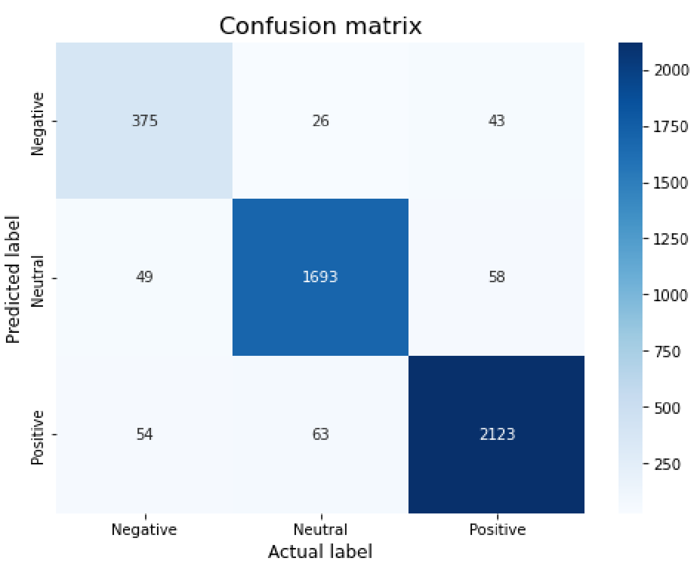

Fig. No. 8: Confusion Actual Matric vs Prediction

6 Conclusions

Bitcoin-related sentiment research on Twitter data can be used as a basis for forecasting whether the price of bitcoin will climb or decline. Based on the intensity of sentiment variations from one time interval to the next, a crude prediction model was presented. The algorithm revealed that 1 hour was the most precise aggregated time to make predictions over, indicating a change in the price of bitcoin 4 hours in the future. Furthermore, a forecast was only made when the change in the emotion mean was no more than 2.2%.The main finding is that even though the offered prediction model had an accuracy rate of 83%, there were so few forecasts that drawing any firm inferences from it would be foolish.

7 References

1) IBM Corporation. IBM - what is big data? URL https://www- 01.ibm.com/software/data/bigdata/what-is-big-data.html.

2) Lisa Lowe. 125 amazing social media statistics you should know in 2016. URL https://socialpilot.co/blog/125- amazing-social-media-statistics-know-2016/.

3) Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash sys- tem. 2008.

4) Pedro Franco. Understanding Bitcoin: Cryptography, engineering and economics. John Wiley & Sons, 2014.

5) Hong Kee Sul, Alan R Dennis, and Lingyao Ivy Yuan. Trading on twitter: Using social media sentiment to predict stock returns. Decision Sciences, 2016.

6) Carruthers, B.G. Money and Society: New Interdisciplinary Perspectives. Sociol Forum 20, 643–649 (2005). https://doi.org/10.1007/s11206-006-9062-4

7) König, S. (2001). THE EVOLUTION OF MONEY - from Commodity Money to E-Money. [online] Available at: http://www.ww.uni-magdeburg.de/fwwdeka/student/arbeiten/009.pdf [Accessed 18 Jan. 2022].

8) Weidai.com. (2019). [online] Available at: http://www.weidai.com/bmoney.txt.

9) www.hashcash.org. (n.d.). Hashcash.org. [online] Available at: http://www.hashcash.org/.

10) Editor, C.C. (n.d.). Double spend (problem) - Glossary | CSRC. [online] csrc.nist.gov. Available at: https://csrc.nist.gov/glossary/term/Double_spend_problem [Accessed 27 Jan. 2022].