



## **Bitcoin Price Prediction Based on Convolutional Neural Network (CNN) and Long-Short-Term Memory (LSTM)**

*<sup>1</sup>M. Lakshmi Prabha, <sup>2</sup>Sankari D, <sup>3</sup>Sowgenya R, <sup>4</sup>Sujitha E*

<sup>1</sup>Assistant Professor, Sri Manakula Vinayagar Engineering College, Puducherry, India

<sup>2,3,4</sup>Student, Sri Manakula Vinayagar Engineering College, Puducherry, India

---

### **ABSTRACT**

Nowadays, the issue of fluctuations in the price of digital Bitcoin currency has a striking impact on the profit or loss of people, international relations, and trade. Hence, this article presents Bitcoin price prediction based on Convolutional Neural Network (CNN) and Long-Short-Term Memory (LSTM) using market sentiment and multiple feature extraction. In this proposed model, several parameters, including Twitter data, news headlines, news content, Google Trends, Bitcoin-based stock, and finance, are employed based on deep learning to make a more accurate prediction. Besides, the proposed model analyses the Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiments to examine the latest news of the market and crypto-currencies. According to the various inputs and analyses of this study, several effective feature selection methods, including mutual information regression, Linear Regression, correlation-based, and a combination of the feature selection models, are exploited to predict the price of Bitcoin.

---

### **I. INTRODUCTION**

The last decades have witnessed remarkable growth in the use of digital currencies by people and organizations. Nowadays, the issue of crypto-currencies has received much attention. In the modern world, crypto-currency has been introduced as a novel and emerging topic which is governed by the cryptographic protocol using Blockchain. Considering the concept of this crypto-currency, the way people think about money has been revolutionized. Also, the value of the crypto-currency has been significantly raised due to the continuous rise in adoption and widespread usage of that in the real world. According to the striking value of crypto-currencies, some people consider them are equal to real currencies or Fiat currencies.

Since the price of crypto-currencies does not behave as in the past, it is significantly difficult to predict the price of crypto-currencies. Additionally, the large fluctuations in the price of crypto-currency, random effects in the market, and the influence of various factors on the price of Bitcoin have become a globally novel challenge. Hence, the issue of predicting the variations in the price of the crypto-currency Bitcoin is of great importance. On the other hand, there are opportunities for a better understanding of the drivers of the Bitcoin price.

Considering CNN and LSTM, our project aims to propose a model for forecasting the variations in the price of the crypto-currency Bitcoin. For this purpose, a variety of methods of textual sentiment analysis such as news headlines, news, and tweets are considered. Such methods consist of using the Twitter API, a Python library, namely 'Tweepy,' extracting text and news content from the Telegram channel, the reference site regarding crypto-currencies, namely Kevin Telegraph, and receiving and extracting Google Trends data. In the beginning, using the tweets in which Bitcoin is mentioned, the data are collected from the storage. Then, the tweets are analyzed to calculate the sentiment score and compare it to other days. After that, that day's price is examined. to determine if there is a relationship between tweets and variations. As a result, variations in the price of crypto-currencies can be determined using sentiment.

The major contributions of this project are summarized as follows:

Presenting Bitcoin price prediction models based on Convolutional Neural Network (CNN) and Long-Short-Term Memory (LSTM) using market sentiment and multiple feature extraction. Analyzing the VADER sentiments to examine the latest news of the market and cryptocurrencies. Pricing Bitcoin price considering feature selection methods, including mutual information regression, Linear Regression, correlation-based, and a combination of the feature selection models.

---

### **II. LITERATURE SURVEY**

[1] In the modern world, crypto-currency has been introduced as a novel and emerging topic which is governed by the cryptographic protocol using Blockchain (Chohan, 2017). Considering the concept of this crypto-currency, the way people think about money has been revolutionized (Pant et al., 2018) Also, the value of the crypto-currency has been significantly raised due to the continuous rise in adoption and widespread usage of that in the real

world. According to the striking value of crypto-currencies, some people consider them are equal to real currencies or Fiat currencies. In comparison, others regard them as a good opportunity to invest. On January 9, 2017, the value of a Bitcoin had increased from \$ 863 by 2000% and reached its highest price level, i.e., \$ 17,550, on December 11, 2017. Eight weeks later, on February 5, 2018, the price of Bitcoin became less than half of the price mentioned earlier, i.e., about \$ 7900.

[2] Nevertheless, the promising technology behind crypto-currencies, namely the Chinese blockchain, is about to raise the use of crypto-currencies. Kristoufek stated that Bitcoin is a unique asset, and the price of Bitcoin crypto-currencies acts like a standard financial asset (Kristoufek, 2015). Bitcoin is regarded as the first decentralized digital currency in which transactions are conducted directly between the users and no intermediary (Matta et al., 2015; Naimy and Hayek, 2018). This type of currency is fundamentally different from what is typically employed in a prevalent monetary system. Based on mining, the crypto-currency is created, which has led to considerable variations in the online economic activities of users worldwide (Jain et al., 2018). Since the price of crypto-currencies does not behave as in the past, it is significantly difficult to predict the price of crypto-currencies.

Additionally, the large fluctuations in the price of crypto-currency, random effects in the market, and the influence of various factors on the price of Bitcoin have become a globally novel challenge. Hence, the issue of predicting the variations in the price of crypto-currency Bitcoin is of great importance. On the other hand, there are many opportunities for a better understanding of the drivers of the Bitcoin price (Karalevicius et al., 2018).

[3] since no central governing authority controls the digital currency and is mainly affected by the general public, Bitcoin is regarded as a volatile currency that changes based on socially constructed ideas. Therefore, the issue of sentiment analysis in the prediction of Bitcoin is of great importance, and many authors have studied it in this regard. The idea of some economists such as Daniel Kahneman and Amos Tversky has proved that the decisions made in this field are influenced by sentiments (KAI-INEMAN and Tversky, 1979). The study of R. J. Dolan regarding "Emotions, Cognition, and Behavior" further confirms that decision-making is extremely affected by sentiments (Dolan, 2002).

[4] the sentiment analysis indicates that demand for a good product, and consequently price, may be influenced more than its economic basics. In recent years, researchers have specifically found that purchasing decisions are made by people and are under the effect of online data collection (Mittal et al., 2019). Galen Thomas Panger stated that Twitter sentiments are related to people's overall sentimental state. In addition, it was revealed that social media such as Twitter has a calming effect rather than reinforcing the user's sentimental state (Panger, 2017). Based on a textual analysis conducted in a social context with the aim of investors called "Search Alpha", Chen et al. stated that the comments outlined in the submitted articles of "Seeking Alpha" were highly effective and even could predict the astonishments of profitability (Chen et al., 2013). In a similar study, Tetlock demonstrated that high levels of media passim in the stock market directly affect trading volume (Tetlock, 2007). Finally, in another study, Gartner pointed out that most users use social media to make their final decisions for purchasing (Petty, 2010).

Over time, an extensive literature has developed on the effectiveness of tweet sentiments. Kouloumpis et al. showed that standard methods of natural language processing like sentence scoring were ineffective due to the short nature of tweets and the uniqueness of this writing style (Kouloumpis et al., 2011). Pak and Patrick divided the individual tweets into positive, negative, or neutral categories that could better understand sentiments by the computer (Pak and Paroubek, 2010). O'Connor et al. indicated that the sentiments in tweets reflect public opinion on various topics in public opinion surveys (O'Connor et al., 2010). This study identified sentiment analysis as a more cost-effective option versus public opinion surveys. Nevertheless, according to this concept, the sentiments generated by tweets more accurately reflect the sentiments of the majority of people on the topic. Hence, it can be considered for predicting demand and the results of variations in the products' price. In another study, the researchers found that employment-related searches were related to the unemployment rate (Ettredge et al., 2005). A relationship between the volume of inquiries and the volume of stock trading on NASDAQ was observed in the study of Bordino et al. (Bordino et al., 2012). Choi and Varian have also conducted specific studies on Google Trends and presented remarkable results (Choi and Varian, 2012). According to the results of this study, it can be concluded that simple seasonal models of trend data are considered input data that outperform models that did not use Google Trends. Also, Asur et al. found that the extent to how much a keyword is a trend in newly released films accurately predicted their revenue at the box office (Asur and Huberman, 2010).

[5] The data of sentiment can be used to predict variation in macroeconomic statistics, and many studies have been performed in this field. Several researchers, including Choi and Varian (Choi and Varian, 2012) and Ettredge et al. (Ettredge et al., 2005), have claimed that web-based search data, which is the same as Google Trends data, can be particularly utilized to predict the price of Bitcoin. Dennis and Yuan collected capacity scores in tweets associated with 500 S&P companies and realized a correlation between them and stock prices (Sul et al., 2014). De Jong et

al. analyzed minute-by-minute stock prices and tweet data of 30 stocks at the Dow Jones industrial average (de Jong et al., 2017).

[6] Traditional time series prediction techniques like Holt-Winters exponential smoothing models are fundamentally related to linear assumptions and need data with the capability of breaking down into trend, seasonal, and noise to be effective (Chatfield and Yar, 1988). Since the Bitcoin market mainly lacks seasonality and high volatility, the traditional methods are not useful. To tackle this drawback, deep learning (DL) technology has been introduced as a novel technique that reduces the costs and complexity of the calculations (McNally et al., 2018). Unlike the traditional linear statistical models, the artificial intelligence (AI) method can consider the nonlinear property. Notably, artificial neural networks (ANNs) with deep learning (DL) algorithms are regarded as the most thriving methods due to their remarkable predictive capabilities

[7] A hybrid Bitcoin price prediction method based on ANN (Nakano et al., 2018). and using Bi-LSTM and Bi-RNN was also presented by Das et al. in 2021, and the benefits of the proposed method were revealed (Das et al., 2021). Despite this interest, no one as far as we know has studied the issue of Bitcoin price prediction considering Twitter data, news headlines, news content, Google Trends, Bitcoin-based stock, and finance using CNN and LSTM.

### III. Data Collection

The analysis presented in this paper needs an understanding of why and how crypto-currencies are different from valid currencies or stocks in the companies of the traditional stock market. Since crypto-currencies are part of the more extensive technology (China Blockchain), Twitter activities can be considered very effective. It should be noted that Google Trends data and the volume of tweets represent an overall tendency to have crypto-currencies

#### *Blockchain and crypto-currencies*

The data of the first crypto-currency in the world are analyzed in this paper. Bitcoin is the largest crypto-currency in terms of market size, followed by Atrium. Bitcoin was the first crypto-currency to be created. In contrast to cash, this system outlines a peer-to-peer payment method using an electronic system. The crypto-currency can be sent directly from one party to another without the use of a third party to verify the transaction between them. This innovation is presented by employing the "blockchain," which is like a common ledger in the whole transaction. This is a peer-to-peer network in which the network verifies the whole transaction to prevent forging them.

Since the applications of blockchain technology go far beyond peer-to-peer payments, this technology provides security, privacy, and decentralization. A decentralized office exploits the blockchain for IoT applications, isolated storage systems, healthcare, and more. The range of blockchain applications has led to the creation of more blockchains and crypto-currencies. Furthermore, using the blockchain increases the usage of crypto-currencies and gives them intrinsic value whose amount depends on many factors. The main reason is this it is a new technological debate. Notably, the information regarding the type of currency and how it stores its new value is useful to improve understanding of what can lead to price changes.

#### *Twitter*

Twitter was created in July 2006 as an application that consists of other applications, websites (such as Instagram, Facebook, LinkedIn, etc.), and microblogging. A microblog is a medium that allows smaller and more frequent updates compared to blogging to be performed. Twitter allows users to send messages publicly (called "tweets") up to 140 characters long, which was doubled on November 6, 2017, to 280 characters per tweet. Users can add a "hashtag" to the tweet, denoted by the symbol of "#." This symbol follows a sequence of characters employed to identify the subject of a tweet and search for that. Hashtags are considered later when collecting tweets in the data section. Because 83% of the world's leaders have Twitter accounts, Twitter earns nearly \$ 330 million a month with 1.3 billion users. Due to such considerable statistics, it should be noted that the Twitter database can be significantly rich and efficient. It is considered a great source of information showing how people almost feel about anything you want. Also, you can observe how these feelings change over time since it can inform you when a tweet has been sent. Hence, Twitter is regarded as a remarkable resource for collecting textual data on a topic such as crypto-currencies to explore possible relationships between them and their prices.

#### *Sentiment analysis*

It can be estimated that 90% of the global data has been generated in the last two years. Most of this data is in the form of textual data without structure. This data can also be in the form of tweets, articles posted on the Internet, text messages, emails, or others that create such a wide amount of unstructured data. "Natural language processing" (NLP) is considered a novel discussion that is being studied or developed. There is a set of methods for computers to analyze and understand the

text. In this paper, a set of natural language processing tools called "emotion analysis" is employed. Sentiment analysis is conducted for extracting and measuring the sentiments or mental opinions outlined in the text. There are several methods to do this, but the "VADER" (Valence Aware) method is selected in this study. The aim here is to use sentiment analysis in the collected tweets for determining what tweets have positive or negative comments regarding crypto-currencies.

#### *Google Trends*

In many parts of the world, almost the whole aspect of daily life includes the Internet. Browsing the Internet is conducted through search engines and Google. Nowadays, the most popular search engine in the world is Google with 74.52% of searches. Therefore, Google search data can provide credible insights into what the world is interested in and the extent of this interest in anything. Google makes this data available through Google Trends. The data provides information concerning the popularity of the searched words compared to other words. There is a variation in

the ranking of Google Trends data at different times in crypto-currencies, which can be related to increasing and reduction of public profit and the price of crypto-currencies.

#### *The headline and the main text of the day's news*

Since the price of crypto-currencies significantly depends on positive and negative news and the crypto-currency market follows more fundamental analysis, we decided to extract news from the most globally reputable news site in the field of crypto-currency, i.e., Kevin Telegraph, for increasing the accuracy. In this period, the extraction and analysis of sentiments based on Twitter data have also been conducted based on the news to see how the news is effective for determining the Bitcoin price.

---

## IV. ALGORITHMS

As for different neural network models, Recurrent Neural Networks (RNN), especially Long Short-Term Memory (LSTM) is a prevalent algorithm to perform time-series prediction in financial markets and the Bitcoin market. In this project, a price forecasting assistant or a predictor model based on CNN and long-short-term memory (LSTM) is analyzed using market sentiment and multiple feature extraction. LSTM networks are a special type of recurrent neural network that can learn long-term dependencies. The convolutional neural network is similar to other neural networks and is composed of neural layers with bias and weights and the ability to learn. The following items occur in each neuron: The neuron receives a set of inputs. Internal multiplication is conducted between the weights of the neurons and the inputs. The result is added to bias.

Finally, a nonlinear function (the same as the activation function) is passed.

The above process is conducted layer by layer and reaches the output layer, creating the network forecast.

The proposed model consists of different parts, and each part has information and details. Besides, the flowchart of the proposed method is demonstrated in Figure 1 for better understanding. Additionally, in the proposed models, VADER sentiment analysis is exploited to examine the latest market news of crypto-currencies. In the proposed model, the Twitter data analysis, the news headlines, news content, Google Trends, Bitcoin stocks, and financials based on deep learning are employed to forecast the Bitcoin price better and more accurately.

This section consists of several main subsections, including data collection and data set, text pre-processing and text feature extraction, data normalization, VADER -based sentiment analysis, feature selection, proposed models based on deep learning, and evaluating the performance criteria.

---

## V. CONCLUSION

Several Bitcoin price prediction models based on CNN and LSTM were designed. The sentiment analysis using the VADER tool and feature extraction of the Bitcoin news is employed in the proposed models. Twitter data analysis, news headlines, news content, Google Trends, Bitcoin stocks, and financials based on deep learning were considered in the proposed model to better and more accurately predict the Bitcoin price. Due to the high extraction features of different input data, three methods of mutual information regression, Linear Regression, and correlation-based feature selection A combination of three feature selection methods were presented in a separate model to take advantage of such feature selection methods. The proposed models are compared with each other in terms of the performance criteria such as mean square error (MSE), root means square error (RMSE), mean absolute error (MAE), median absolute error (MedAE), and coefficient of determination (R<sup>2</sup>). The results of various implementations and experiments proved the remarkable performance of the proposed hybrid model based on sentiment analysis and combined feature selection with MSE value of 0.02 and R<sup>2</sup> value of 0.1 in obtaining better results and less error in predicting the Bitcoin price.

---

## REFERENCES

- [1] ALWESHAH, M., ALKHALAILEH, S., ALBASHISH, D., MAFARJA, M., BSOU, Q. and DORGHAM, O., 2021. A hybrid mine blast algorithm for feature selection problems. *Soft Computing*, 25, 517- 534.
- [2] AWOKE, T., ROUT, M., MOHANTY, L. and SATAPATHY, S.C., 2021. Bitcoin Price Prediction and Analysis Using Deep Learning Models, 631-640. Singapore.
- [3] BUI, D.-K., NGUYEN, T., CHOU, J.-S., NGUYEN-XUAN, H. and NGO, T.D., 2018. A modified firefly algorithm-artificial neural network expert system for predicting compressive and tensile strength of high-performance concrete. *Construction and Building Materials*, 180, 320-333.
- [4] CAMBRIA, E., PORIA, S., HAZARIKA, D. and KWOK, K., 2018. Sentic Net 5: Discovering conceptual primitives for sentiment analysis utilizing context embeddings. *Proceedings of the AAAI conference on artificial intelligence*.
- [5] CHAUDHARI, H. and CRANE, M., 2020. Cross-correlation dynamics and community structures of cryptocurrencies. *Journal of Computational Science*, 44, 101130.
- [6] DAI, B., JIANG, S., LI, C., ZHU, M. and WANG, S., 2021. A multi-hop cross-blockchain transaction model based on improved hash-locking. *International Journal of Computational Science and Engineering*, 24, 610-620.
- [7] DAS, S., BILLAH, M. and MUMU, S.A., 2021. A Hybrid Approach for Predicting Bitcoin Price Using Bi- LSTM and Bi-RNN Based Neural Network, 223-233. Cham.
- [8] DE JONG, P., ELFAYOUMY, S. and SCHNUSENBERG, O., 2017. From returns to tweets and back: an investigation of the stocks in the Dow Jones Industrial Average. *Journal of Behavioral Finance*, 18, 54-64.
- [9] DUTTA, A., KUMAR, S. and BASU, M., 2020. A Gated Recurrent Unit Approach to Bitcoin Price Prediction. *Journal of Risk and Financial Management*, 13, 23.
- ELRAHMAN, S.A. and ALLUHAIIDAN, A.S., 2021. Blockchain technology and IoT-edge framework for sharing healthcare services. *Soft Computing*, 25, 13753-13777.

- 
- [10] HOTA, H.S., SHARMA, D.K. and VERMA, N., 2021. 14 - Lexicon-based sentiment analysis using Twitter data: a case of COVID-19 outbreak in India and abroad.
- [11] KARALEVICIUS, V., DEGRANDE, N. and DE WEERDT, J., 2018. Using sentiment analysis to predict interday Bitcoin price movements. *The Journal of Risk Finance*.
- [12] LI, D., HAN, D., WENG, T.-H., ZHENG, Z., LI, H., LIU, H., CASTIGLIONE, A. and LI, K.-C., 2021.