A novel Bitcoin price prediction method using Transformer-LSTM Ensemble model

Raj Madhu [1]

Ahmedabad, Gujarat, India.

raj.madhu0406@gmail.com;

Abstract

Bitcoin is the world's first distributed digital currency. It was proposed and established by Satoshi Nakamoto in 2009. Since then, it has become an essential decentralized financial asset. Bitcoin offers tremendous potential returns, but it comes with high risks. Predicting the prices of Bitcoin can help investors and individual earn huge amounts of profits but the high volatility of Bitcoin market makes forecasting its price very difficult. Many models are proposed for predicting bitcoin prices like ARIMA, LSTM, GRU, ANN, etc., but most of them have been trained on a small dataset, or they can not efficiently capture long-range dependencies. This paper proposes Transformer-LSTM ensemble model. The Transformer model is based on the paper 'Attention Is All You Need' and uses self-attention to capture long-range dependencies. The proposed model was trained on almost 9 years of historical bitcoin prices and can efficiently capture long-range dependencies. The proposed model performs better than most deep learning models and achieves a RMSE error of **181.1013**, which is one of the lowest error rates.

Keywords: Transformer, Bitcoin, Time-series, Cryptocurrency, prediction

1 Introduction

There are more than 18000 cryptocurrencies, but Bitcoin is the most popular. Since its launch in 2009, it has become a global phenomenon due to its high returns and decentralized nature. This has led many investors to invest in Bitcoin. In July 2021, the total number of crypto users stood close to 230 million. However, as August came and the market began to rally, the same figure shot up to 263 million. In April 2021, the Bitcoin market cap reached an all-time high and had grown by over 1,000 billion USD (YahooNews). Bitcoin is built on a technology called blockchain. Blockchain is a growing list of records, called blocks, that are securely linked together using cryptography. They are resistant to modification as the data in any given block cannot be altered without altering all the subsequent blocks (Wikipedia) . Also, each block is uniquely recognized. Blockchain offers very high security and features like Immutability and decentralization. In recent years, many application has been developed using blockchain technology, such as cryptocurrencies, digital signature, NFT, web3, etc.

Bitcoin is the first cryptocurrency based on blockchain technology. Similar to stocks, Bitcoin price prediction is a time-series prediction task. The significant difference between them is that stocks are traded only on weekdays, but the Bitcoin market operates around the clock, and investors can buy and sell Bitcoin any day. The stock market is volatile, but it has only went up over the years. While Bitcoin market, on the other hand, is more volatile, and the risk of it fluctuating suddenly is higher. The prediction of Bitcoin prices is considered a difficult task than predicting stock prices due to its high volatile nature and unpredictable factors that affect its price. With the rise of neural networks, many recent studies employed deep neural networks to predict Bitcoin prices due to their high performance. Models like Recurrent Neural Network (RNN), Long short-term memory (LSTM), Gated Recurrent Unit (GRU), etc., have been used for prediction. The most common method for predicting Bitcoin price is using the previous historical price data to predict the future price. In addition to this, some methods use external features that have some correlation with Bitcoin prices. Some have also used sentiment analysis scores along with historical prices to predict future prices. All the above mentioned regression methods are very difficult, so to reduce the complexity, many have converted the task to a classification problem, i.e., predicting whether the prices will go up or down.

Since its publication in 2017, the transformer model has been a state-of-the-art model for many NLP tasks. Recently, many different models based on transformer architecture have been proposed that give unprecedented accuracy in different NLP tasks. Models like T5, BERT, GPT-3, etc give very high accuracy in different NLP tasks like text translation, text summarization, Audio classification, etc. This paper, proposes a Transsformer-LSTM ensemble model to predict Bitcoin prices from historical data. Only recently, the Transformer model has been employed in predicting time-series data, but no one has used Transformer-LSTM ensemble model to predict bitcoin prices.

2 Related Work

The popularity of Bitcoin has attracted the attention of many researchers and investors in blockchain cryptocurrencies. A lot of research work is done in predicting Bitcoin prices and other time-series data such as stocks, influenza, weather, etc. Some of them are mentioned below.

2.1 Statistical and sentiment Analysis

A great deal of statistical analysis is done on Bitcoin prices and other cryptocurrencies. Many external factors affect Bitcoin prices, and it is necessary to understand their correlation. [12] studied the correlation between Bitcoin prices and search queries on google trends and Wikipedia; and found that there is a strong correlation between them. [11] analyzed user comments in online cryptocurrencies communities to predict the fluctuation in their prices and the number of transactions. They proved a strong correlation between user comments and replies in online communities and the number of transactions among the users. [8] concluded that media sentiment affects the prices of Bitcoin and that investors tend to overreact to news in a short period. [1] used ARIMA model for prediction. [5] conducted a comparison between statistical models and Machine Learning models, specifically the ARIMA model, Back-propagation Neural Network (BPNN), and Generic Algorithms (GA). The author of [22] used a Bayesian regression model along with expert correction to predict Bitcoin prices. [19] used google trends media and sentiment analysis scores from Bitcoin tweets to analyze and predict Bitcoin prices. [7] proposed a Bayesian neural network model that predicts the price of Bitcoin using blockchain information.

2.2 Deep Learning Models

With the increase in deep learning models and popularity, many different Deep Learning models have been employed for time-series forecasting. Many research papers employed LSTM as it can capture long-range dependencies and store temporal information. [14] was one of the first to use Deep learning models for the prediction of cryptocurrency. The authors used a LSTM model for prediction and concluded that they performed better than generalized regression neural architecture. [15] employed a hybrid model based on CNN and LSTM for predicting Bitcoin prices. [20] inputted historical prices of Bitcoin to LSTM, RNN, and ARIMA models for future forecasting. [21] used different external features that affected the Bitcoin price. These features were given as input to 4 models- artificial neural network (ANN), stacked artificial neural network (SANN), support vector machines (SVM), and long short-term memory (LSTM). [2] proposed a multimodal Adaboost-LSTM ensemble approach for forecasting Bitcoin price. They used various external factors and Twitter sentiments as input. [6] used LSTM and GRU for stock prediction. [10] employed a feed-forward neural network and RNN. [24] also used LSTM and RNN for stock prediction. [13] used historical Ethereum features like open, close, high, low, volume, etc., as input and Multi-layer perceptron and LSTM models for prediction. [23] evaluated the performance of advanced deep learning algorithms for predicting the price and movement of cryptocurrencies and concluded that deep learning techniques are not able to solve forecasting problems efficiently, and a new approach should be explored.

2.3 Transformer Models

Many research work is done using various deep learning models. However, with the rising popularity and accuracy of the transformer-based model, it is compelling to explore its potential in forecasting time-series data. Different time-series transformer models have been proposed for predicting various time-series data, such as [26] influenza-like illness (ILI), stocks, traffic, etc. However, very few have proposed a transformer model for predicting Bitcoin prices. [28] used the VADER sentiment score and transformer model for predicting Bitcoin and Ethereum prices. They used Twitter API to scrape tweets related to Bitcoin and Ethereum and Time2Vec [9] vector to capture periodic and non-periodic patterns. The authors of [25] analyzed the use of transformer for different time-series tasks like forecasting, anomaly detection, etc. The authors of [27] proposed a framework for multivariate Time-Series learning based on the transformer encoder architecture. They used this model for regression as well as classification tasks. [16] proposed a new transformer-based model called Temporal Fusion Transformer (TFT) for predicting time-series data. The model selects relevant features and reduces unnecessary components, achieving high accuracy. [17]presented the Gated Transformer Network, which is a simple extension of the current transformer network with gating. This model is employed in solving multivariate time-series classification tasks.

3 Data and preparation

3.1 Data

The data contains daily Bitcoin information collected from 2010-07-16 to 2019-04-20 and contains 3201 data-points. The dataset has been collected and combined from yahoo finance and kaggle. The data contains the following information about the Bitcoin:

- Open - This is the opening price of the time period
- High - This is the highest price of the time period
- Low - This is the lowest price of the time period
- Close - This is the closing price of the time period
- Volume - This is the number of Bitcoin bought and sold.

The Bitcoin close price data is plotted on a graph as shown in figure 1 for a better understanding of the data and visualization purposes. After the graph is plotted, we can clearly observe the features' fluctuation. The dataset is divided into training data (90%) and testing data (10%).

4 Preparation

The proposed model uses Bitcoin's historical **Close Price** as input to the model. To prepare the data for the input layer, Date and Close columns from the pandas data frame are extracted. The data is then sorted according to dates

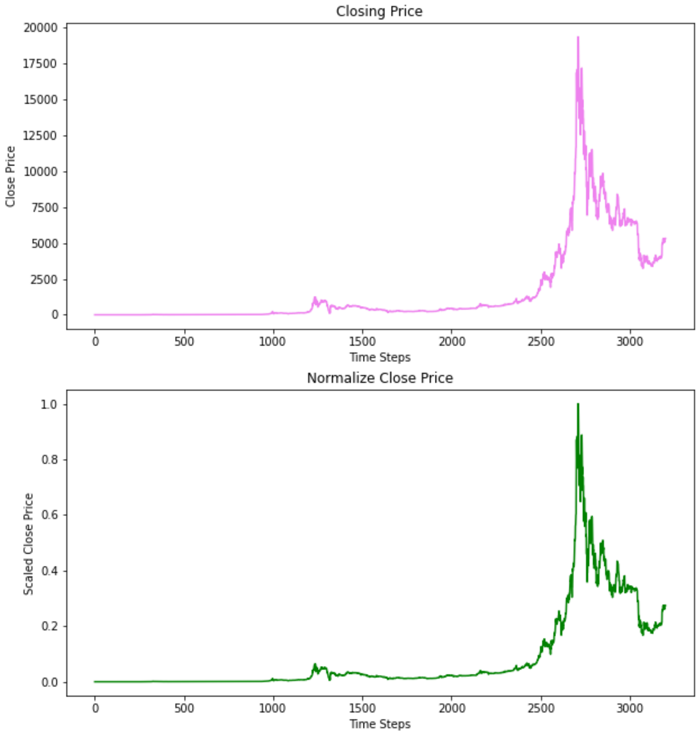

Figure 1: Bitcoin Close price graph (unscaled and scaled)

and checked for null values and missing dates. After confirming that the data contains no anomalies, it is converted to a NumPy array. The NumPy array contains Bitcoin close price. This array is normalized by passing it through the MinMaxScalar function from the sklearn library and given as input to the model.

5 Proposed Method

The proposed model is an ensemble of LSTM model and transformer model for predicting time-series data. The LSTM or the Long-Short Term Memory model is a kind of Recurrent Neural network (RNN) that is capable of learning long-term dependencies. The proposed transformer model has been modified for predicting time-series data. The final predicted price is the the average of the output of both the models. The different components and mechanisms of the proposed model are described below:

5.1 Transformer Model

The transformer model is an Encoder and Decoder model. The Encoder contains the multi-head self-attention mechanism and a fully connected feed-forward layer. The Decoder layer of the proposed model is a simple linear layer that outputs one day ahead price prediction. The different mechanism of the transformer model are described below:

5.1.1 Positional Encoding

Since the transformer model has no recurrence or convolution like in other models such as RNN, LSTM, or CNN, the model uses positional encoding to capture the sequence information of the input data. The positional encoding module injects information about the sequence's relative or absolute position.

$$PE_(pos, 2i) = sin\left(pos/10000^{2i/d_{model}}\right) \tag{1}$$

$$PE_(pos, 2i+1) = cos\left(pos/10000^{2i/d_{model}}\right) \tag{2}$$

5.1.2 Scaled Dot-product Attention

Scaled dot-product attention or self-attention is an attention mechanism where the dot products are scaled down by $\sqrt{d_k}$. The attention mechanism of the transformer model creates three vectors: Q(query) of d$_k$ dimension, K(key) also of d$_k$ dimension, and V(value) of d$_v$ dimension. The attention function computes the dot product of the query with all keys, divides each key by $\sqrt{d_k}$ and applies a softmax function to obtain weights on the value. In practice, the attention function computes the weights simultaneously on a set of queries.

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{3}$$

5.1.3 Multi-head Attention

Multi-head Attention computes 'h' number of single-head attention parallelly and concatenates the output weights. The Q (queries), K (keys), and V (values) are linearly projected 'h' times. On each of the 'h' projected versions of Q, K, and V, the attention function is applied parallelly, which outputs d_v dimension output. Each projection is called an Attention head. These outputs are concatenated and again projected to give out final values. Multi-head Attention elevates the model's ability to focus on different parts of the input sequence by taking in information from different representations jointly.

$$MultiHead(Q, K, V) = Concat\left(head_1, head_2 ....head_h\right) \tag{4}$$

where,

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{5}$$

5.2 LSTM Model

The Long-Short Term Memory (LSTM) model is a type of RNN model. It is frequently used in different NLP tasks because of its capability of learning long-range dependencies, and because of that, it is also used in predicting time-series data. It has three different cell gates that handle the flow of incoming information. The three gates are: Forget gate, Input gate, and Output gate. These gates decide whether to save the new information or to forget the previously saved information. LSTMs have feedback connections that differentiate them from more traditional feedforward neural networks. This paper uses a Bi-directional LSTM model for Bitcoin price prediction. The model contains two Bi-directional LSTM layers and a final linear layer.

6 Results

Most of the previous methods employed for predicting Bitcoin or any other cryptocurrency prices used a smaller data window. Most of them used an average of 4 years of historical data for prediction, while the proposed model used approximately 9 years of data. Using large time-series data is essential for the proper evaluation of a model. It helps us determine how well a model can capture long-range dependencies. The prediction task becomes more difficult with Bitcoin because of the frequent fluctuation in its price. Different types of models and methods have been proposed for predicting Bitcoin prices but the proposed performs better than most of them. The proposed model gets RMSE error of **181.1013**. Table 1 shows comparison between various models and Figure 2 shows the result of test data predictions.

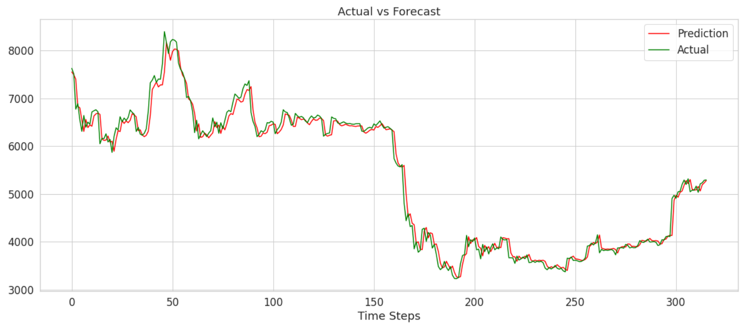

Figure 2: Test data result

Table 1: Comparison of proposed method with other methods

Year	Reference	Method	Data Window	Data parameters	Test Data	RMSE
2021	[22]	Bayesian regression model along with expert correction	2017-2018	External factors and trading data	NA	856.4
2020	[15]	hybrid CNN-LSTM model	2016-2018	External factors and trading data	20 datapoints(4%)	258.31
2022	[2]	multimodal Adaboost LSTM ensemble model	2016-2020	External factors, twitter sentiments and trading data	274 datapoints(15%)	243.47
2021	[3]	hybrid (Multi-scale Residual Convolutional) MRC-LSTM model	2015-2020	External factors and trading data	364 datapoints(20%)	261.44
2021	[4]	GRU model	2018-2021	historical data	251 datapoints(20%)	174.129
2021	[18]	Multiple-Input Cryptocurrency Deep Learning Model(MICDL)	2017-2020	historical data	152 datapoints(11%)	257.728
2022	**proposed model**	**Transformer-LSTM ensemble Model**	**2010-2019**	**historical data**	**316 datapoints(10%)**	**181.1013**

7 Discussion and future work

This paper proposes a Transformer-LSTM ensemble model which can forecast univariate time-series data. Firstly, the two models, transformer and LSTM are given a sequence of historical Bitcoin close price data as input, and they predicts the close price of the next day in the sequence. The final prediction is the average of the output of the two models. The predicted results gets a RMSE score of **181.1013**, which is one of the lowest error rate. While most of the previous studies used a small dataset, the proposed model uses a large dataset of almost 9 years. This shows that the proposed model is able to capture long-range dependencies and is also able to fit large training data very well. Transformer models are still not explored profoundly for predicting time-series data, and this paper is an attempt to do so.

For future work, the model can be modified for multi-variate time-series data. Just as stock prices depend on various external factors, the price of cryptocurrencies like Bitcoin also depends on different features like number of web searches, gold price, tweets, etc. By introducing such features to the transformer model, the model's accuracy can be increased, but one has to also remember the possibility of overfitting on the data.

Acknowledgments. I want to express great appreciation and gratitude to Professor Seema Mahajan *(HOD Computer Engineering, Indus University)* for her valuable and constructive suggestions during the planning and development of this research work. Her insight and expertise greatly assisted the research, and her willingness to give time so generously is strongly appreciated.

8 Statements and Declarations

Funding : Not applicable.

Conflict of interest : The author declares no conflict of interest.

Availability of data and material : The data will be made available on request.

References

[1] A. Azari. Bitcoin price prediction: An arima approach, 2019.

[2] Z. Boukhers, A. Bouabdallah, M. Lohr, and J. Jürjens. Ensemble and multimodal approach for forecasting cryptocurrency price, 2022.

[3] Q. Guo, S. Lei, Q. Ye, and Z. Fang. Mrc-lstm: A hybrid approach of multi-scale residual cnn and lstm to predict bitcoin price. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2021.

[4] M. J. Hamayel and A. Y. Owda. A novel cryptocurrency price prediction model using gru, lstm and bi-lstm machine learning algorithms. *AI*, 2(4):477–496, 2021.

[5] . Haviluddin, R. Alfred, J. H. Obit, M. H. A. Hijazi, and A. A. A. Ibrahim. A performance comparison of statistical and machine learning techniques in learning time series data. *Advanced Science Letters*, 21(10):3037–3041, oct 2015.

[6] M. A. Hossain, R. Karim, R. Thulasiram, N. D. B. Bruce, and Y. Wang. Hybrid deep learning model for stock price prediction. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1837–1844, 2018.

[7] H. Jang and J. Lee. An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. *IEEE Access*, 6:5427–5437, 2018.

[8] V. Karalevicius, N. Degrande, and J. D. Weerdt. Using sentiment analysis to predict interday bitcoin price movements. *The Journal of Risk Finance*, 19(1):56–75, jan 2018.

[9] S. M. Kazemi, R. Goel, S. Eghbali, J. Ramanan, J. Sahota, S. Thakur, S. Wu, C. Smyth, P. Poupart, and M. A. Brubaker. Time2vec: Learning a vector representation of time. *CoRR*, abs/1907.05321, 2019.

[10] K. Khare, O. Darekar, P. Gupta, and V. Z. Attar. Short term stock price prediction using deep learning. In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, pages 482–486, 2017.

[11] Y. B. Kim, J. G. Kim, W. Kim, J. H. Im, T. H. Kim, S. J. Kang, and C. H. Kim. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PLOS ONE*, 11(8):e0161197, aug 2016.

[12] L. Kristoufek. BitCoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific Reports*, 3(1), dec 2013.

[13] D. Kumar and S. Rath. *Predicting the Trends of Price for Ethereum Using Deep Learning Techniques*, pages 103–114. 01 2020.

[14] S. Lahmiri and S. Bekiros. Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solitons Fractals*, 118:35–40, 2019.

[15] Y. Li and W. Dai. Bitcoin price forecasting method based on CNN-LSTM hybrid neural network model. *The Journal of Engineering*, 2020(13):344–347, jul 2020.

[16] B. Lim, S. Arık, N. Loeff, and T. Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021.

[17] M. Liu, S. Ren, S. Ma, J. Jiao, Y. Chen, Z. Wang, and W. Song. Gated transformer networks for multivariate time series classification, 2021.

[18] I. E. Livieris, N. Kiriakidou, S. Stavroyiannis, and P. Pintelas. An advanced cnn-lstm model for cryptocurrency forecasting. *Electronics*, 10(3), 2021.

[19] M. Matta, M. I. Lunesu, and M. Marchesi. Bitcoin spread prediction using social and web search media. 06 2015.

[20] S. McNally, J. Roche, and S. Caton. Predicting the price of bitcoin using machine learning. In *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, pages 339–343, 2018.

[21] M. Mudassir, S. Bennbaia, D. Unal, and M. Hammoudeh. Time-series forecasting of bitcoin prices using high-dimensional features: a machine learning approach. *Neural Computing and Applications*, jul 2020.

[22] B. M. Pavlyshenko. Bitcoin price predictive modeling using expert correction. In *2019 XIth International Scientific and Practical Conference on Electronics and Information Technologies (ELIT)*, pages 163–167, 2019.

[23] E. Pintelas, I. E. Livieris, S. Stavroyiannis, T. Kotsilieris, and P. Pintelas. Investigating the problem of cryptocurrency price prediction: A deep

learning approach. In I. Maglogiannis, L. Iliadis, and E. Pimenidis, editors, *Artificial Intelligence Applications and Innovations*, pages 99–110, Cham, 2020. Springer International Publishing.

[24] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman. Stock price prediction using lstm, rnn and cnn-sliding window model. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1643–1647, 2017.

[25] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun. Transformers in time series: A survey, 2022.

[26] N. Wu, B. Green, X. Ben, and S. O'Banion. Deep transformer models for time series forecasting: The influenza prevalence case.

[27] G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff. A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery amp; Data Mining*, KDD '21, page 2114–2124, New York, NY, USA, 2021. Association for Computing Machinery.

[28] H. Zhao., M. Crane., and M. Bezbradica. Attention! transformer with sentiment on cryptocurrencies price prediction. In *Proceedings of the 7th International Conference on Complexity, Future Information Systems and Risk - COMPLEXIS,*, pages 98–104. INSTICC, SciTePress, 2022.