

Prediction of Bitcoin Price Based on Transformer, LightGBM and Random Forest

Yihan Tu^{1,a,*}

¹*College of Economics and Management, Nanjing University of Aeronautics and Astronautics,
Nanjing, China*

a. guiqiu@ldy.edu.rs

**corresponding author*

Abstract: Contemporarily, cryptocurrency have been established and widely recognized as an alternative method of exchange currency. As computer technology advances, the trading of cryptocurrency is increasingly viewed as a popular and lucrative form of investment. On the other hand, the volatile nature of cryptocurrency results in sudden and unpredictable price fluctuations. Consequently, the necessity for developing a precise and dependable predictive model for portfolio management and optimization is acknowledged. At present, RNN and other deep learning models are commonly utilized in predicting stock prices, ensemble learning methods including Random Forest and Xgboost have been widely implemented and researched in the realm of investments, engagement in trading activities in various fields including gold and other markets. However, research on cryptocurrency investment is still insufficient and the practical application of the mentioned models has not yielded satisfactory returns. As machine learning and artificial intelligence continue to advance, the enhancement of machine learning models' mechanisms could significantly boost the accuracy in predicting cryptocurrency prices. Hence, the utilization of Transformer architecture alongside various established models was opted for in the prediction of Bitcoin's price. The attention mechanism embedded in the Transformer model has proven to be effective in anticipating the value of assets undergoing significant price shifts. Meanwhile, the difference in the accuracy of short-term and long-term price prediction of each model may also help and guide cryptocurrency investors in model selection and development during investment and market analysis.

Keywords: Cryptocurrency, machine learning, forecasting, transformer.

1. Introduction

In the mid-20th century, machine learning was first introduced. Alan Turing, a key figure in the field, posed critical inquiries regarding the capacity of machines to imitate facets of human cognition in publication 'Computing Machinery and Intelligence [1]. Between 2010 and 2020, there were significant advancements achieved in the realm of artificial intelligence. A milestone during this period was the emergence of deep learning as a predominant force. It highlighted the capabilities of deep learning and reinforcement learning in addressing complex issues, sparking a wave of similar techniques applied across various fields. Transformer has garnered significant attention in the fields of NLP and time series analysis. Advancements in NLP have been greatly influenced by the

emergence of Transformer models, with the introduction of groundbreaking models such as BERT standing out as a major milestone [2].

This groundbreaking technology introduced by Transformers has paved the way for a new approach to managing sequential data, greatly impacting the field of NLP and extending its influence beyond. In contrast to earlier approaches that tackled sequences incrementally, Transformers tackle sequences holistically, resulting in notable improvements in both efficiency and efficacy. As hardware performance has advanced significantly in the 21st century, the field of machine learning has experienced a surge in research interest. Initially, various industries have incorporated machine learning, including financial markets, crude oil market, gold market and the future market [3]. Moreover, the application of artificial intelligence algorithms and powerful computing to forecast cryptocurrency prices has emerged as a trending field for machine learning applications. The rise of cryptocurrencies as a modern electronic substitute for traditional currency exchange methods has significantly impacted emerging economies and global economy [4]. Having integrated into a majority of financial transactions, cryptocurrency trading is commonly viewed as a leading and auspicious form of lucrative investments. Nevertheless, this expanding financial sector experiences notable instability and considerable price fluctuation. Given the unpredictable factors and drastic price fluctuations involved in cryptocurrency forecasting, it is widely considered as one of the most formidable challenges in predicting time series data.

While the cryptocurrency market is able to benefit from abundant market data and continuous trading, it also encounters challenges like significant price fluctuations and low market valuation. In order for cryptocurrency financial transactions to succeed, thorough analysis and data selection are essential, making the development of machine learning models crucial for extracting meaningful features. In the realm of stock price prediction, the effectiveness of decision trees has been well-established [5]. LSTM, with its strong performance in recent years and extensive research, has emerged as a reliable method for stock price forecasting. Despite the limited exploration into generating profitable trading strategies in the cryptocurrency market, these models hold promise for further development. The attention mechanism has proven to be effective in dealing common issues with time series sequencing, such as making predictions on cryptocurrency prices. This implies that Transformer's accuracy in forecasting the value of the digital currency will surpass that of previous models. Transformer architecture greatly benefits from the presence of the attention mechanism, leading to a significant boost in its efficiency for making accurate predictions on time series data [6]. The mechanism of attention enables the model to shift its focus across various segments of the input sequence while generating an output, effectively capturing the relationships between distant words or events. In the time series prediction, Transformer leverages the attention mechanism to assign weights to prior occurrences in anticipation of forthcoming events. This becomes especially valuable in situations where recent occurrences may not hold the most significance for making a forecast. Transformer shows potential in time series prediction by utilizing their ability to understand sequential patterns. Various fields benefit from the utilization of predictive modeling with Transformers, including applications such as weather forecasting and anomaly detection. It possesses the ability to recognize recurring patterns and evolving trends, enabling accurate projections of forthcoming data points [7].

Lately, there has been a scarcity of comparative studies among various popular models in the realm of cryptocurrency price prediction, including an exploration of the Transformer model's effectiveness. Hence, this research is to assess the predictive accuracy of different architectural models in forecasting virtual currency prices by utilizing established machine learning models like Transformer, LightGBM, Random Forest, and mentioning OLS regression for comparison. Exploring the underlying factors contributing to the outstanding predictive performance of a specific model is the focus of this investigation.

2. Data and Method

The data selected for bitcoin price ranges from 2020-1-1 to 2024-8-5, sourced from yahoo finance records. With 1689 records in the dataset, the training set covers the period from 2020-1-1 to 2023-8-7, while the test set spans from 2023-8-8 to 2024-8-5. The features are mainly 'Close', 'Open', 'High', 'Low', and 'Volume', 'Fluctuation'. Generate variables 'wr','obv','obv_ma','diff12 ','dea9','macd9'. Technical indicators such as obv (energy tide indicator), obv_ma (moving average of obv), diff12, dea9 and macd9 play an important role in predicting bitcoin price. These indicators help traders identify market trends, momentum, and potential buy and sell points through different calculations and logic.

Four models including Transformer, LightGBM, Random Forest, and Ols were selected. Transformer model, an innovative neural network model originating from Google's self-attention mechanism, specializes in handling sequential data. In the field of NLP, the Transformer model is widely used due to its excellent parallel performance and short training time. The Transformer model achieves global modelling and inter-element correlation by assigning weights to each element in the input sequence and calculating weighted sums through the Self-Attention Mechanism. Each layer of the structure includes encoder and decoder components, incorporating various attention mechanisms and feed-forward neural network modules. The process of encoding the input sequence into a high-dimensional feature vector is handled by the Encoder, while the responsibility of decoding the vector into the target sequence falls on the Decoder. Moreover, the model integrates methods like residual connection and layer normalization to enhance overall performance. The main parameters are as following:

- `d_model`: the model dimension, i.e. the dimension of the input/output vectors of the encoder and decoder.
- `n_heads`: the number of heads in the multi-head attention mechanism for parallel processing of different parts of the input sequence.
- `d_ff`: the dimension of the feed-forward neural network layer, which is usually larger than `d_model` to increase the nonlinear capability of the model.
- `n_layers`: the number of layers of encoder and decoder, the more layers, the more complex the model, but may also lead to overfitting.

LightGBM is an optimized Gradient Boosting Decision Tree (GBDT) implementation developed by Microsoft, known for its faster training speed, low memory usage and high accuracy. It uses a histogram algorithm and a Leaf-wise growth strategy to effectively solve the training problem with large data volumes, and supports parallelized learning and direct processing of category features. LightGBM discretizes continuous floating-point features into k discrete values through a histogram algorithm, and constructs a histogram of width k . The histogram is then divided into k discrete values. During the training process, LightGBM examines the dataset sequentially, accumulating statistics for each discrete value present in the histogram in order to identify the most favorable dividing point. At the same time, it uses a Leaf-wise growth strategy to find the leaf with the largest splitting gain from all the current leaves for splitting, thus improving the training efficiency. Main parameters are as following:

- `boosting_type`: type of boosting algorithm, default is 'gbdt'.
- `num_leaves`: number of leaf nodes, used to control the complexity of the tree.
- `learning_rate`: learning rate, which controls the step size of each iteration.
- `feature_fraction`: the proportion of features randomly selected in each iteration to prevent overfitting.

- **bagging_fraction**: the proportion of randomly selected samples in each iteration, again used to prevent overfitting.

Random Forest, as a composite model that incorporates numerous decision trees, enhances the accuracy and consistency of the overall model by amalgamating the predictive outcomes yielded by the various decision trees. Random Forest constructs multiple decision trees by randomly sampling and randomly selecting features. When constructing each tree, it first randomly selects a subset from the original dataset as the training set and randomly selects a portion of features from all the features to train the model. Finally, the random forest integrates the predictions of all decision trees by majority voting or averaging. Main parameters are as follows:

- **n_estimators**: the number of decision trees, i.e., the number of trees in the forest.
- **max_depth**: the maximum depth of the tree, used to control the complexity of the tree.
- **max_features**: the maximum number of features to be considered when constructing the tree.
- **bootstrap**: whether or not sampling with put-back is used to select the training set.
- **n_jobs**: number of jobs running in parallel, used to accelerate model training.

OLS, short for Ordinary Least Squares, by minimizing the sum of the squares of the errors, it determines the most optimal function to correspond with the given dataset. It finds the line of best fit by solving a set of equations such that the sum of the squares of the differences between the actual observations and the model predictions is minimized. This usually involves solving a system of linear equations or minimizing a quadratic function. Main parameters are as follows:

- **Beta coefficients**: model parameters, i.e., slopes of fitted lines (for univariate linear regression) or coefficient matrices (for multivariate linear regression).
- **Intercept term**: the intercept of the fitted line, indicating the predicted value of the dependent variable when the independent variable is zero.
- **Residual**: the difference between the actual observations and the model predictions, used to assess the effectiveness of the model fit.

R^2 and MAE play an important role in assessing the performance of a model, each providing a measure of the model's predictive accuracy from a different perspective. R^2 is used to measure how well the model fits the observed data, i.e. the proportion of the variance of the target variable that can be explained by the model. Its value ranges from 0 to 1. The closer R^2 is to 1, the better the model fits the data, and the closer the model's predictions are to the actual observations. R^2 measures the model's fit by comparing it to a mean model, providing a relative assessment. The MAE calculates the average of the absolute values of the differences between the predicted and true values and reflects the overall level of error in the model's predictions. The smaller the MAE, the more accurate the model's predictions. Compared to MSE (Mean Square Error), MAE has better robustness to outliers because it does not significantly increase the overall error level due to individual extreme error values.

3. Results and Discussion

3.1. Feature Engineering

Through the original data, generating the relevant technical analysis of the indicators required for subsequent training calculations. The correlation coefficients are shown in Fig. 1. By observing the correlation coefficients of each feature, it is found that the correlation between the indicators 'Close', 'Open', 'High', 'Low'. The correlation between these indicators is very high, and 'Close' is chosen as the main feature of analysis in the subsequent analysis and the rest features are eliminated to prevent the problem of multicollinearity.

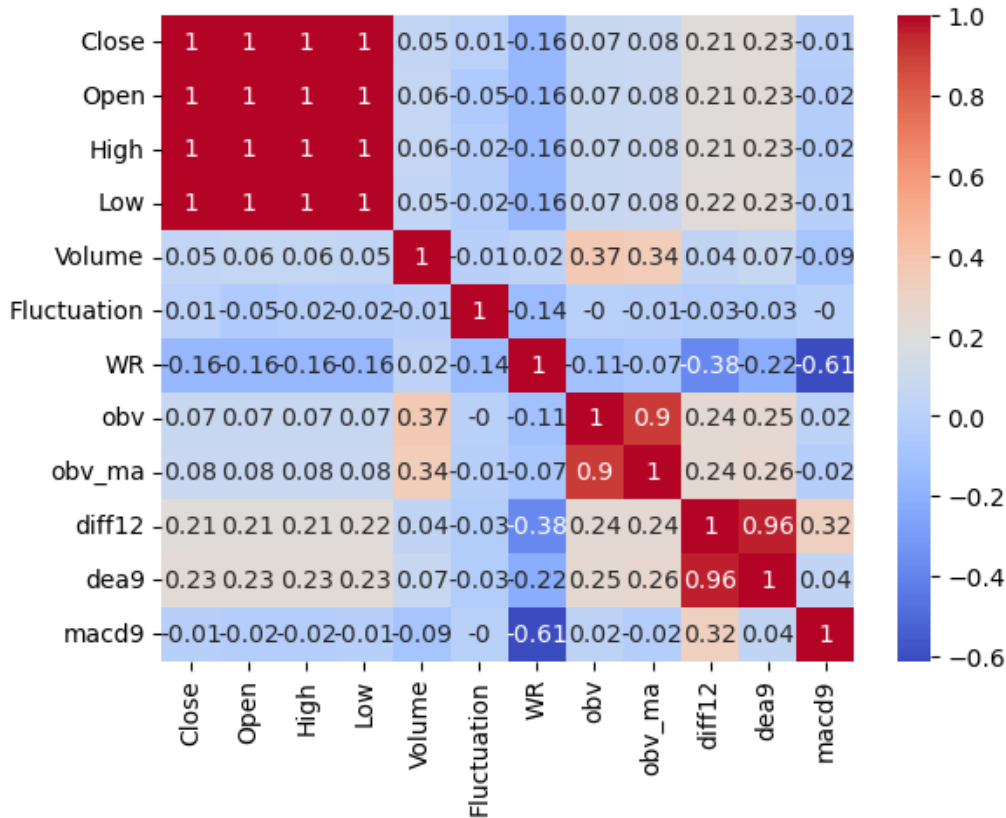


Figure 1: Correlation coefficients (Photo/Picture credit: Original).

3.2. Model Performances

LightGBM, Random forests, and Ols models have a period of better fit in the early part of the prediction results and a large difference in the late part of the predictions from the true values towards the true values. Transformer's results have a mediocre data fit in the early part of the results and a better fit between the predicted and true values in the middle and late part of the predictions. The results are summarized in Fig. 2. Khedr suggests that cryptocurrencies do not have seasonal effects and therefore are difficult to predict using statistical methods [8]. Tanwar suggests that although the use of the Transformer model results in longer computation time, it has a higher prediction accuracy [9]. Referring to Hafid used Moving Average Convergence Divergence (MACD) as one of the features for training, the difference is that compared to the features used to train the XGBoost Regressor model [10], this paper uses the variables to train the Transformer model, and the accuracy of the results is higher. will be higher. Technical indicators and a neural network model are utilized by Khaniki in order to capture temporal dynamics and extract essential features from raw cryptocurrency data [11]. Nascimento utilizes a transformer encoder-decoder model based on multi-head attention for forecasting the values of digital assets (Dogecoin) and demonstrates favorable predictive performance [12]. So, it's reasonable for this paper to apply transformer to Bitcoin for both of them are cryptocurrencies. King demonstrates the importance of introducing variables through numerical simulation of algorithmic trading, and prediction by machine learning models [13].

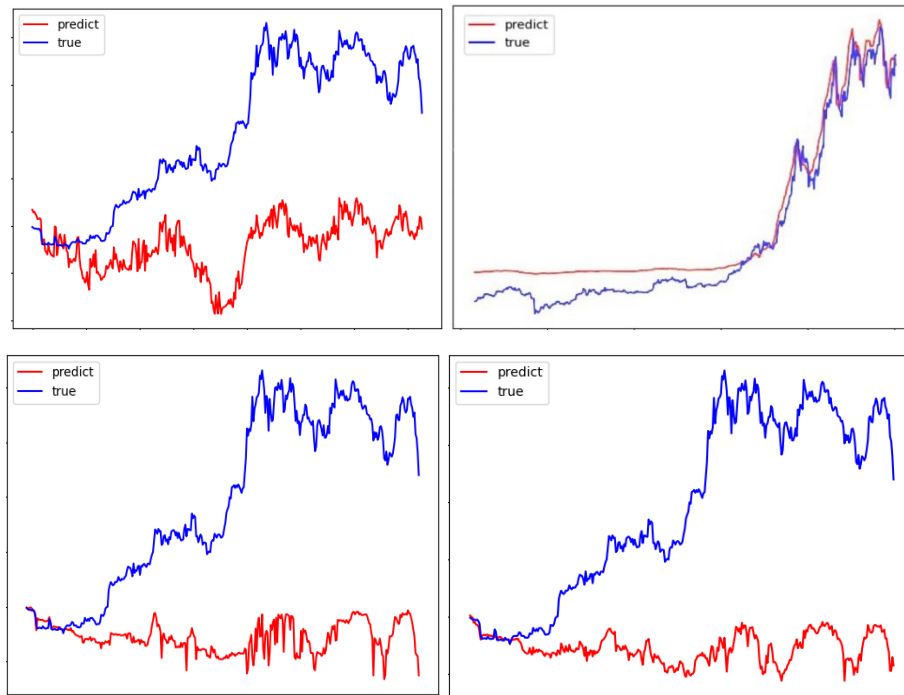


Figure 2: Performances of the Models: OLS, Transformer, LightGBM, Random Forest (Photo/Picture credit: Original).

3.3. Comparison and Implications

In machine learning and statistical modelling, differences in the performance of different models on data are often related to the characteristics of the model itself, the characteristics of the data, and the parameter settings during model training. In terms of model properties, LightGBM and Random Forest, both are tree-based integration models, particularly adept at handling complex relationships, interactivity, and large sets of data. However, their ability to capture long-term dependencies in time series data might not be as strong compared to the Transformer model. Moreover, when the data exhibits significant time trends or cyclical variations, these models may encounter challenges in fully capturing such patterns. Conversely, the ordinary least squares (OLS) model assumes a linear correlation between the independent and dependent variables. In cases where the data exhibits non-linear patterns or intricate relationships, OLS might struggle to provide accurate predictions. Utilizing the self-attention mechanism, Transformer model effectively captures the intricate temporal dependencies present in sequential data. This ability enables effective handling of sequential data, particularly in cases involving intricate temporal patterns and extended historical trends.

In terms of data characteristics, in cases where the data exhibits a smooth trend initially but later undergoes significant fluctuations or changes in direction, traditional tree-based models and OLS regression may struggle to adapt, as they primarily focus on capturing local patterns and linear relationships. In contrast to traditional tree-based models and OLS, the Transformer model, with its unique self-attention mechanism, demonstrates enhanced flexibility to address the dynamic shifts within the data sequence, particularly in later segments. The capacity of a model to generalize is reflected in its performance on new, unseen data. If the model overfitted the training data during training its performance on the test data may be influenced. The Transformer model performs well in the later stages due to its strong representation learning ability, which may because the fact that it succeeds in capturing the key features in the data and thus performs better in the later stages.

3.4. Limitations and Prospects

For model optimization, for Transformer, one can try to adjust the model parameters (e.g., learning rate, tree depth, attention mechanism, etc.), as well as adopt more sophisticated feature engineering methods to improve the generalization ability and stability of the model. For Random Forests, the number and depth of decision trees can be adjusted, as well as introducing feature importance assessment to optimize feature selection. For OLS, the introduction of non-linear terms or the use of other more sophisticated regression models can be considered instead. In terms of parameter selection, this paper only adopts the relevant factors used for quantitative analysis of technical aspects, and does not choose the relevant factors of fundamentals. In the subsequent research, one can consider combining these two factors to conduct a comprehensive analysis. After the model is deployed, continuous monitoring of the model's predictive performance should be carried out, with necessary adjustments and optimizations being made according to the actual situation. Regular assessments and examinations are conducted on the model to ensure its capability to adjust to variations in data and fulfill business requirements.

4. Conclusion

To sum up, the discussion in this study involves the comparison of the predictive capabilities among four popular machine learning models, Transformer, LightGBM, Random Forests, and Ordinary Least Squares (OLS), when applied to forecast fluctuations in Bitcoin price. After conducting in-depth data analysis and evaluating various models, it is proven that the Transformer model excels in navigating through the intricacies of this volatile market environment, particularly in accurate long-term price predictions. Its forecast precision and consistency surpass the capabilities of traditional models. The results presented in this study not only showcase the significant promise of the Transformer model in predicting cryptocurrency price movements, but also broaden the scope of its application in analyzing financial markets. It provides an opportunity for investors to utilize advanced tools for a more informed investment approach, enabling them to predict cryptocurrency price movements more accurately and efficiently. Through the integration of forecasts generated by the Transformer model with in-depth market analysis, essential research, and diverse information sources, a more holistic evaluation of market risks and opportunities can be achieved, leading to optimized asset allocation and appreciation.

References

- [1] Alan, M.T. (1950) *Computing machinery and intelligence*. *Mind*, 59(236), 433–460
- [2] Jacob, D., Chang, M., Kenton, L. and Kristina, T. (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [3] Chen, J. (2023) *Analysis of Bitcoin Price Prediction Using Machine Learning*. *Journal of Risk Financial Management*, 16, 51.
- [4] Nasir, M.A., Huynh, T.L.D., Nguyen, S.P., Duong, D. (2019) *Forecasting cryptocurrency returns and volume using search engines*. *Financial Innovation*, 5, 2.
- [5] Abu Bakar, N. and Rosbi, S. (2017) *Autoregressive integrated moving average (ARIMA) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction*. *International Journal of Advanced Engineering Research and Science*, 4(11), 130–137.
- [6] Vaswani, A. (2017) *Attention is all you need*. *Advances in neural information processing systems*, 30.
- [7] Haryono, A.T., Sarno, R. and Sungkono, K.R. (2023) *Transformer-Gated Recurrent Unit Method for Predicting Stock Price Based on News Sentiments and Technical Indicators*. *IEEE Access*, 11, 77132–77146.
- [8] Khedr, A.M. Arif, I., El-Bamany, M., Alhashmi, S.M., and Sreedharan, M. (2021) *Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey*. *Intelligent Systems in Accounting, Finance and Management*, 28(1), 3-34.

- [9] Tanwar, A. and Kumar, V. (2022) *Prediction of cryptocurrency prices using transformers and long short term neural networks*. 2022 *International Conference on Intelligent Controller and Computing for Smart Power (ICICCCSP)* pp. 1-4.
- [10] Hafid, A., Ebrahim, M., Alfatemi, A., Rahouti, M., and Oliveira, D. (2024) *Cryptocurrency Price Forecasting Using XGBoost Regressor and Technical Indicators*. *arXiv preprint arXiv:2407.11786*.
- [11] Khaniki, M.A.L., and Manthouri, M. (2024) *Enhancing Price Prediction in Cryptocurrency Using Transformer Neural Network and Technical Indicators*. *arXiv preprint arXiv:2403.03606*.
- [12] Nascimento, E.G.S., de Melo, T.A., and Moreira, D.M. (2023) *A transformer-based deep neural network with wavelet transform for forecasting wind speed and wind energy*. *Energy*, 278, 127678.
- [13] King, J.C., Dale, R., and Amigó, J.M. (2024) *Blockchain metrics and indicators in cryptocurrency trading*. *Chaos, Solitons and Fractals*, 178, 114305.