Bitcoin Price Prediction using Machine Learning

Utla Tarun   
*Department of Computer Science and Engineering  
Bapatla Engineering College (Autonomous)  
(Affiliated to Acharya Nagarjuna University)*Bapatla, India  
[tarunutla1234@gmail.com](mailto:tarunutla1234@gmail.com)

T Eswar Prasad   
*Department of Computer Science and Engineering  
Bapatla Engineering College (Autonomous)  
(Affiliated to Acharya Nagarjuna University)*Bapatla, India  
[eswarprasad195@gmail.com](mailto:eswarprasad195@gmail.com)Venkata Mounika  
*Department of Computer Science and Engineering  
Bapatla Engineering College (Autonomous)  
(Affiliated to Acharya Nagarjuna University)*Bapatla, India  
[venkatamounika5@gmail.com](mailto:venkatamounika5@gmail.com)

R L Pujitha   
*Department of Computer Science and Engineering  
Bapatla Engineering College (Autonomous)  
(Affiliated to Acharya Nagarjuna University)*Bapatla, India  
[pujarapuri@gmail.com](mailto:pujarapuri@gmail.com)

*Abstract*—**This research investigates the influences on Bitcoin's price volatility using machine learning techniques. We leverage technical indicators such as Exponential Moving Average (EMA), Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD) alongside data from Yahoo Finance to develop a sophisticated model for predicting daily price shifts. Through advanced analytics, this study provides investors with valuable insights to navigate the challenges of the cryptocurrency landscape and enhance their understanding of Bitcoin's market behavior. Additionally, we explore the temporal aspects of Bitcoin price movements, analyzing the impact of short-term versus long-term trends on prediction accuracy. This temporal analysis offers actionable insights into optimal entry and exit points, empowering investors to capitalize on market opportunities effectively. Furthermore, we extend our analysis to explore anomaly detection and risk management in Bitcoin trading.**

Keywords—Bitcoin, Price volatility, Machine learning, Technical indicators, EMA, RSI, MACD, Yahoo Finance

# Introduction

Cryptocurrency markets, led by Bitcoin, have disrupted traditional finance, offering novel investment opportunities while challenging established paradigms. Bitcoin, as the vanguard of this revolution, epitomizes decentralization, transparency, and financial autonomy. However, amidst its soaring ascent, Bitcoin's persistent price volatility poses both opportunities and risks for investors.

At the core of this initiative lies the acknowledgment of Bitcoin's distinctive characteristics, including its finite supply, decentralized nature, and the profound influence of market sentiment and external events. Understanding these fundamentals is imperative for constructing a robust predictive model capable of capturing the complexities of Bitcoin's price dynamics.

Ultimately, our project aspires to equip investors with actionable insights and decision-making tools, enabling them to navigate the intricacies of the cryptocurrency landscape with confidence and precision. By harnessing the transformative potential of machine learning and data analytics, we endeavor to unlock new frontiers in Bitcoin price prediction, ushering in a more informed and efficient market ecosystem.

# Literature survey

The cryptocurrency market is transforming the world of money and finance [1] and has seen significant growth in the last years [1], [2]. In particular, the number of cryptocurrencies reached more than 7000 in 2021 [3] and the crypto market capitalization hit $3 trillion the same year [3]. The banking and financial industry has taken notice of Blockchain benefits. The underlying technology behind every cryptocurrency is Blockchain technology. Blockchain is a distributed/decentralized database that is organized as a list of blocks, where the committed blocks are immutable. It has many attractive properties including transparency and security [2].

The crypto market has many good characteristics including high market data availability and no closed trading periods. However, it suffers from its high price volatility and relatively smaller capitalization. In crypto financial trading, data can be available to all traders. However, the analysis and the selection of this data make the difference between executing good trades and bad trades. Therefore, one of the main challenges in financial trading is to develop methods/approaches to extract meaningful knowledge and insights from the data. Furthermore, due to the high price volatility of cryptocurrency prices, forecasting the price becomes more challenging.

Up to now, few studies have attempted to create profitable trading strategies in the cryptocurrency market. In 2018, Saad et al. [4] provided a machine-learning model to predict Bitcoin prices. In particular, they made use of a regression model and involved many factors that impact the price of Bitcoin. However, they did not provide Buy and Sell signals, which are the most important in building a trading strategy/approach. Furthermore, they did not consider any kind of technical indicators. The use of technical indicators as features to feed machine learning models for financial trading has been successfully employed by many researchers [5], [6]. McNally et al. [7] proposed a machine learning model that makes a recurrent neural network, called the Long Short Term Memory Model (LSTM). LSTM achieves an accuracy of 52%, for classification. However, this is not acceptable for building a trading strategy. Our approach achieves 96%, which is quite sufficient.

In this paper, we contribute to developing profitable trading strategies by proposing a new approach that integrates various features including technical indicators and historical data. The key contribution of the proposed approach is providing buy and sell signals with high accuracy along with the prediction of future prices. We evaluate the proposed approach through the analysis of Bitcoin cryptocurrency.

# Mathematical modelling

In this section, we provide a mathematical modelling of the proposed approach.

## Notations & Definitions

Table I below provides the definitions of parameters and abbreviation.

|  |  |
| --- | --- |
| **Notation** | **Description** |
|  | Total number of observations/samples |
|  | Number of samples in the training set |
|  | Number of samples in the test set |
|  | Total number of input features |
|  | Close price at time |
|  | Opening price at time |
|  | High price at time |
|  | Low price at time |
|  | Volume of the cryptocurrency that is being in trade at time |
|  | Span |
|  | Relative strength index at time ti within a time period α |
|  | Moving average convergence divergence at time |
|  | Exponential moving average at time ti within a period of time α |
|  | Price rate of change at time ti within a period of time α |
|  | Stochastic oscillator at time ti within a period of time α |
|  | Momentum at time within a period of time α |
|  | Bollinger bands at with a period of time α |
|  | Set of real numbers |
|  | Set of targets, |

## Mathematical model

In this section, we present a mathematical modeling of our approach.

Let ( denotes a single sample/observation. The set of samples is represented by:

Where

Since we consider both technical indicators and Blockchain historical data to predict the price, we need to combine/merge different data sets. Specifically, technical indicators and historical data are input to our model. Our feature vector in a given time t can be expressed as follow:

Let us generalize our model by stacking all features vectors in one matrix X. This matrix can be expressed as follows:

Where

The output matrix can be expressed as follows:

Where .

## Model Integration

Our model integrates both classification and regression techniques to predict cryptocurrency prices. The classification component predicts whether the price will increase or decrease (1 or 0), while the regression component predicts the exact price value.

The integration of classification and regression components can be represented mathematically as follows:

### Classification Component

Where represents the predicted class label for sample i and is the input feature vector.

### Regression Component

Where represents the predicted class label for sample i and is the input feature vector.

### Integration

Let's generate the combined equation considering the classification output to determine the investment decision (buy or sell), without necessarily combining it with the regression prediction.

Where

* represents the final prediction for sample
* is the regression prediction for sample
* is the classification prediction for sample
* If the classification model predicts a buy signal , the final prediction is the regression prediction , indicating the estimated future price.
* If the classification model predicts a buy signal , the final prediction is the regression prediction , indicating that no investment is advised.

## Features

In this section, we present different features. In particular, we used historical market data and technical analysis indicators.

### Historical Data

Regarding the historical data, we consider the date, close price, and volume.

#### Date

The date field acts as a **temporal reference point** within your Bitcoin price prediction model. It allows the model to capture trends and seasonality patterns in historical data, potentially improving prediction accuracy.

#### Close Price

Close price refers to the price at which a cryptocurrency closes at a given time period.

#### Volume

Volume is the number of units traded in the market during a given time period*.*

### Technical Analysis Indicators

Technical analysis is a trading discipline employed to evaluate investments and identify trading opportunities by analyzing statistical trends gathered from trading activity, such as price movement and volume [10]. In this work, we consider the exponential moving average, moving average convergence divergence, relative strength index, momentum, price rate of change, and stochastic oscillator.

#### Exponential Moving Average:

The exponential moving average (EMA) was first introduced by Roberts (1959) [11]. It is a type of moving average (MA) that places a greater weight and significance on the most recent data points. EMA is also referred to as the exponentially weighted moving average. represents the closing price at time represents the close price at time t2, and gradually represents the close price at time tm. Knowing that measures the step time (e.g., 1 minute, 15 minutes, 1 day). EMA can be expressed recursively as follows:

#### Moving Average Convergence Divergence

Moving Average Convergence Divergence (MACD) is a technical indicator created by Gerald Appel in 1970 [12]. MACD helps investors understand the movement of the price (i.e., the market will be in bullish or bearish movement) [12]. Usually, MACD is calculated by subtracting the 26-period EMA from the 12-period EMA. Formally, it is expressed as follows:

Where is the period EMA and

#### Relative Strength Index

The relative strength index (RSI) is a technical indicator used to chart the current and historical strength or weakness of a stock/market based on the closing price of a recent trading period. It was originally developed by J. Welles Wilder [13]. RSI is classified as a momentum oscillator, which is an indicator that varies over time within a band. Technically, RSI is typically used for 14 days and is measured on a scale from 0 to 100. RSI takes the values 70 and 30 with high and low levels of the market [13]. RSI within a band α (α usually equals 14) can be mathematically expressed as follows:

Where and represent average gain over α-days and average loss over α-days, respectively.

#### Momentum

Momentum (MOM) measures the velocity of a stock price over a while, which means the speed at which the price is moving; typically we use the close price [14]. MOM helps investors identify the strength of a trend [14]. Formally, the momentum can be expressed as follows:

Where is number of days.

#### Price rate of change

The Price Rate Of Change (PROC) measures the most recent change in price. It can be expressed as follows:

where is the price rate of change at time and is the number of periods to look back.

#### Stochastic Osillator

A stochastic oscillator is a popular technical indicator for generating overbought and oversold signals. Usually the current value of the stochastic indicator is denoted by and it is computed as follows:

where represents the most recent closing price, represents the lowest price traded during the previous periods, and represents the highest price traded during the previous periods.

#### Bollinger Bands

Bollinger Bands are a popular technical analysis tool used by traders to assess price volatility and identify potential buy or sell signals in the market. Here’s a breakdown with a formula:

* *Middle band:* A simple moving average (SMA) of the closing price over a user-defined period (e.g., 15 days).
* *Upper band:* The middle band plus a certain number of standard deviations (typically 2) away.
* *Lower Band:* The middle band minus the same number of standard deviations.

Where is a user-defined coefficient, Standard deviation of the closing prices over the calculation period . is a Simple moving average of closing at time .

#### Signal

Let Y be a random variable that takes the value (Buy and Sell, respectively). To generate Buy or Sell signals, we employ a technical indicator called Moving Average . identifies the trend of the market. The rule that generates Buy and Sell signals at time consists of comparing two moving averages. Formally, the rule is expressed as follows:

Where

is the length of the short (long) . We denote the indicator with lengths s and l by . In this paper, we consider the because of its high accuracy.

# Machine learning Models

Here, we discuss the machine learning models employed in our approach. These models address both classification and regression tasks:

## Classification Models

### Logistic Regression:

A common choice for binary classification problems. we use a linear function, y = m\*X + c, followed by the sigmoid function σ to get an output y, denoted by y’, such that 0 < y’<1 and c ∈ R.

### Random Forest:

An ensemble method using decision trees, often effective for various classification tasks. Random Forest (RF) classifier is a combination of decision trees [16]. RF uses averaging over all the tree predictors to improve the accuracy and control over-fitting [17]. Furthermore, one of the main qualities of RF is that it helps us to measure the importance of each feature.

### XGBoost:

(eXtreme Gradient Boosting) is a powerful tree-based ensemble method for classification tasks. It sequentially builds decision trees, each focusing on correcting the errors of its predecessor. This approach leverages regularization techniques and gradient boosting to prevent overfitting and efficiently handle complex data structures.

where represents the prediction of the model after adding the tree. represents the prediction of the model after adding tree. is a learning rate hyperparameter to prevent overfitting. represents the prediction made by the tree based on the input features

## Regression Models

### Random Forest Regressor

A Random Forest Regressor (RFR) has emerged as a prominent ensemble learning technique for financial forecasting tasks due to its effectiveness in handling complex and high-dimensional data. This research explores the application of RFR for Bitcoin price prediction, a domain characterized by inherent volatility and non-linear relationships.

### XGBoost Regressor

A Gradient Boost Approach XGBoost addresses these challenges by leveraging ensemble learning and gradient boosting. It constructs a series of weak learners (typically decision trees) sequentially, where each tree focuses on correcting the errors of its predecessors. This sequential learning approach progressively improves the model's prediction accuracy.

# Result & Evaluation

We incorporate these two different models in our projects which helps the investors to predict the price and also know when to buy and when to sell.

For the data, we stream real-time historical market data from Yahoo Finance website. The data is from 01 January, 2015 to 01 April, 2024 and with a time step of one day (i.e., γ = 1 day) .we consider 1 day because it gives us good accuracy. We split the data into 70% for training and 30% for testing.

## Evaluation Parameters

To evaluate and measure the robustness and the goodness of the proposed approach, we present different evaluation parameters:

### Accuracy

It is the fraction of predictions our model got right Formally; accuracy can be expressed as follows:

where is the number of true positives and is the number of true negatives.

### Precision

It is expressed as follows:

where F P is the number of false positives.

### Recall

It is expressed as follows:

where F N is the number of false positives

## Result & Analysis

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **LR** | **RF** | **XGB** |
| Score | 80.6 | 88.2 | 87.1 |

Table V.1 K-Fold Cross Validation

Table V.1 provides a K-fold cross-validation comparison among the different proposed machine learning models. This comparison is based on the score (accuracy) presented in table . This score is calculated as the average of the accuracy of 13 folds.

Table shows that the three models provide three different score. We choose XGB (an ensemble model) to forecast crypto market since it has the ability to deal with very larger sizes of data, a large number of features, and an expected non-linear relationship between the predicted variable and the features [19].

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Accuracy | Precision | Recall | F1-score | Support |
| 0 |  | 0.97 | 0.95 | 0.96 | 422 |
| 1 |  | 0.97 | 0.98 | 0.97 | 592 |

Table V.2 Classification Report

Table V.2 shows the classification report of the proposed approach. It shows the accuracy, precision, and the recall.

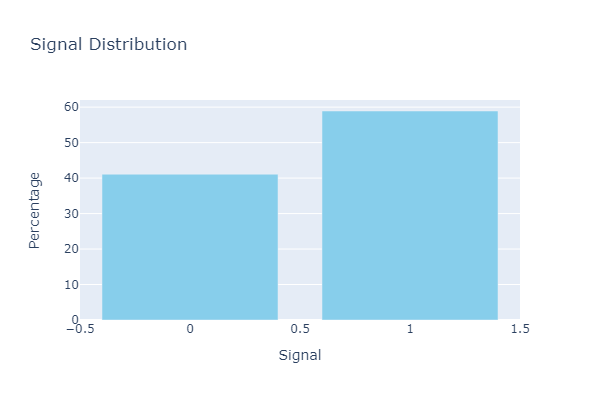


Figure . Signal

Particularly, Figure V.1 shows that the predicted variable’s class 1 is slightly more bigger than 58% of the time, meaning there are more buy signals than sell signals. The predicted variable is relatively unbalanced

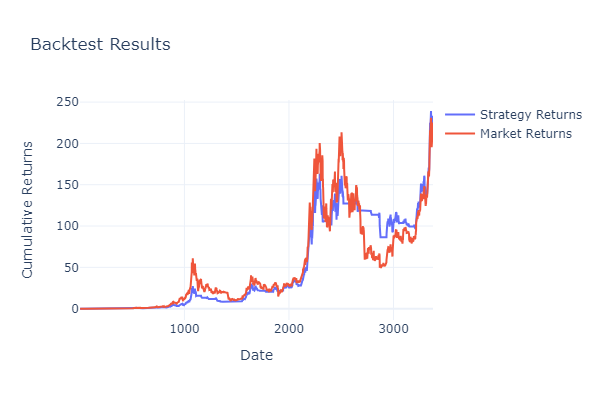


Figure . Back Testing result

Figure V.2 shows that the predicted returns are very close to the actual returns. This means that the proposed approach performs well for predicting Bitcoin.

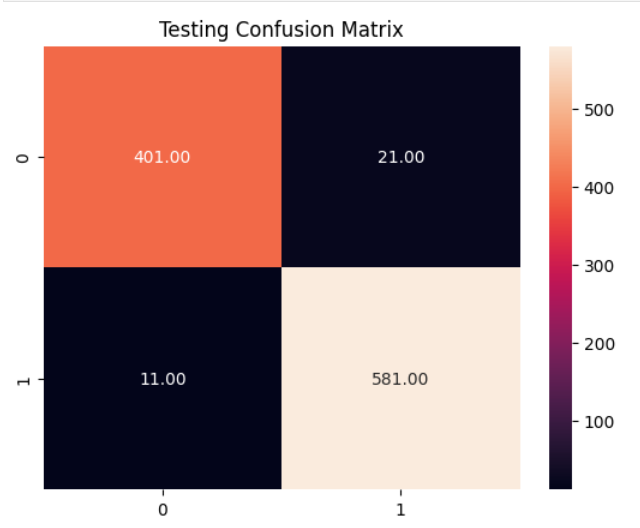


Figure . Bitcoin Confusion Matrix

Figure V.3 shows the confusion matrices corresponding to Bitcoin.

Our research revolves around gathering and analyzing data from the cryptocurrency market, with a keen focus on identifying the most significant factors that drive price fluctuations. Through our investigation, we've found that technical indicators wield considerable influence, surpassing even traditional metrics like closing price and volume (as depicted in Figure 4). While our methodology has yielded promising results in terms of accuracy, precision, and recall, there's still room for refinement. Our ultimate aim is to bridge the gap between our predicted returns and the actual returns observed in the market, striving for ever-closer alignment with our strategy's performance.

Predicting cryptocurrency prices poses a formidable challenge due to the myriad unconventional factors at play, including the impact of social media trends and the intricate dynamics of investor psychology. To enhance our forecasting capabilities, future endeavors will explore avenues for incorporating a broader array of features into our models. These may include factors such as cash flow dynamics, mining rates, and transaction volumes. Additionally, our research trajectory will involve the exploration of diverse machine learning algorithms to further refine our predictive capabilities. By expanding the scope of our analysis and refining our modeling techniques, we aim to develop a more robust framework for forecasting cryptocurrency prices, thereby advancing our understanding of this dynamic and rapidly evolving market landscape.

# Conclusion

In this paper, we utilized two techniques to address the challenges of cryptocurrency market analysis. Firstly, we employed predictive modeling to forecast future price movements. This involved leveraging historical market data alongside a range of technical indicators. Secondly, we developed a signal generation mechanism aimed at translating these forecasts into actionable buy or sell signals. These signals serve as guidance for investment decisions, helping determine whether it's opportune to invest or abstain. By integrating these two approaches, we aimed to provide a comprehensive framework that enhances decision-making processes and empowers investors to navigate the intricate dynamics of cryptocurrency markets effectively.

# References

|  |  |
| --- | --- |
| [1] | G. Hileman and M. Rauchs, “Global cryptocurrency benchmarking study,” Cambridge Centre for Alternative Finance, vol. 33, no. 1, pp. 33–113, 2017. |
| [2] | P. Treleaven, R. G. Brown, and D. Yang, “Blockchain technology in finance,” Computer, vol. 50, no. 9, pp. 14–17, 2017. |
| [3] | CoinGecko, “Cryptocurrency prices, charts, and crypto market cap,” Available at https://www.coingecko.com/, Accessed: 02-Jul-2022. |
| [4] | M. Saad, J. Choi, D. Nyang, J. Kim, and A. Mohaisen, “Toward characterizing blockchain-based cryptocurrencies for highly accurate predictions,” IEEE Systems Journal, vol. 14, no. 1, pp. 321–332, 2019. |
| [5] | Y. Shynkevich, T. M. McGinnity, S. A. Coleman, A. Belatreche, and Y. Li, “Forecasting price movements using technical indicators: Investigating the impact of varying input window length,” Neurocomputing, vol. 264, pp. 71–88, 2017. |
| [6] | P. Oncharoen and P. Vateekul, “Deep learning for stock market prediction using event embedding and technical indicators,” in 2018 5th international conference on advanced informatics: concept theory and applications (ICAICTA). IEEE, 2018, pp. 19–24. |
| [7] | 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). IEEE, 2018, pp. 339–343. |
| [8] | P. Jay, V. Kalariya, P. Parmar, S. Tanwar, N. Kumar, and M. Alazab, “Stochastic neural networks for cryptocurrency price prediction,” Ieee access, vol. 8, pp. 82 804–82 818, 2020. |
| [9] | H. Singh and P. Agarwal, “Empirical analysis of bitcoin market volatility using supervised learning approach,” in 2018 Eleventh International Conference on Contemporary Computing (IC3). IEEE, 2018, pp. 1–5. |
| [10] | S. B. Achelis, “Technical analysis from a to z,” 2001. |
| [11] | ] J. M. Lucas and M. S. Saccucci, “Exponentially weighted moving average control schemes: properties and enhancements,” Technometrics, vol. 32, no. 1, pp. 1–12, 1990. |
| [12] | G. Appel, Technical analysis: power tools for active investors. FT Press, 2005. |
| [13] | J. W. Wilder, New concepts in technical trading systems. Trend Research, 1978. |
| [14] | L. K. Chan, N. Jegadeesh, and J. Lakonishok, “Momentum strategies,” The Journal of Finance, vol. 51, no. 5, pp. 1681–1713, 1996. |
| [15] | ] C. Cortes and V. Vapnik, “Support-vector networks,” Machine learning, vol. 20, no. 3, pp. 273–297, 1995. |
| [16] | T. K. Ho, “Random decision forests,” in Proceedings of 3rd international conference on document analysis and recognition, vol. 1. IEEE, 1995, pp. 278–282. |
| [17] | A. Geron, ´ Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. ” O’Reilly Media, Inc.”, 2019 |
| [18] | Binance, “Binance api,” Available at https://www.binance.com/en/binance-api, Accessed: 16-Jul-2022. |
| [19] | G. Biau and E. Scornet, “A random forest guided tour,” Test, vol. 25, no. 2, pp. 197–227, 2016. |