Review of Statistical and Machine Learning Solutions to Forecast Cryptocurrency

# Introduction

The rapid growth and widespread adoption of cryptocurrencies have sparked considerable interest among academic researchers and industry professionals. As cryptocurrencies continue to evolve and establish themselves in the financial landscape, there is a pressing need to develop reliable and accurate forecasting methods that can predict their future behavior. Forecasting cryptocurrency prices and market trends presents numerous challenges due to the complex and volatile nature of these digital assets.

This survey paper aims to fulfill the need for a comprehensive review of the statistical and machine learning solutions proposed and employed in the field of cryptocurrency forecasting. By thoroughly analyzing and evaluating the existing literature, our objective is to provide valuable insights into the various methodologies, techniques, and models used in this domain. The primary goal of this survey is to offer a consolidated overview of the different approaches and their performance in predicting cryptocurrency prices, market trends, and other relevant variables. We will explore a wide range of statistical and machine learning techniques, delving into their strengths and weaknesses in the context of cryptocurrency forecasting.

Moreover, we will delve into the specific challenges associated with cryptocurrency forecasting, including data availability, data quality, market volatility, and regulatory factors. By addressing these challenges, we can identify the limitations of the current forecasting methodologies and identify potential areas for improvement. The insights gained from this survey will be invaluable to researchers, practitioners, and stakeholders who are interested in understanding the current state of cryptocurrency forecasting and exploring new avenues for enhancing prediction accuracy. By synthesizing the knowledge accumulated in this field, we aim to provide a solid foundation for future research and advancements in the development of effective forecasting solutions for cryptocurrencies.

# Background

The emergence and rapid growth of cryptocurrencies have captivated the attention of investors, traders, and researchers across the globe. Bitcoin, Ethereum, Litecoin, and numerous other digital currencies have disrupted traditional financial systems and introduced new possibilities for financial transactions and investments. The decentralized nature, cryptographic security, and potential for high returns associated with cryptocurrencies have sparked immense interest in this alternative asset class. However, the cryptocurrency market is characterized by unprecedented levels of volatility, making it a challenging and unpredictable landscape. Fluctuations in prices, market trends, and investor sentiments can occur within minutes or even seconds. As a result, individuals and organizations involved in cryptocurrency trading and investment require reliable forecasting methods to make informed decisions and manage risks effectively.

Traditional forecasting techniques used in conventional financial markets may not be directly applicable to cryptocurrencies due to their unique characteristics. Cryptocurrencies exhibit complex dynamics, non-linear patterns, and susceptibility to various external factors. Regulatory changes, news sentiment, market sentiment, and technological developments can significantly impact cryptocurrency prices and market behavior. To overcome these challenges, researchers and practitioners have turned to statistical and machine learning solutions to forecast cryptocurrencies. By leveraging historical price data, market indicators, and other relevant variables, these methodologies aim to capture and model the intricate patterns and dynamics of cryptocurrency markets. Statistical techniques such as time series analysis, regression models, and correlation analysis are utilized to identify underlying trends, patterns, and relationships within the data. In addition, machine learning algorithms are employed to extract complex patterns and make accurate predictions.

The application of statistical and machine learning solutions to cryptocurrency forecasting offers several advantages. These methods can handle large volumes of data, capture nonlinear relationships, and adapt to changing market conditions. By analyzing vast datasets encompassing historical prices, trading volumes, market sentiment, and other relevant factors, these approaches aim to uncover hidden patterns and exploit market inefficiencies for forecasting purposes.

# Survey Methodology

The area of cryptocurrency price prediction is still in its early stages and necessitates further research to delve deeper into this domain. When considering literature primarily focused on finance, encompassing finance, economics, business, management, and accounting, the number of publications in this specific area remains relatively limited. This paper presents a systematic literature review on cryptocurrency price prediction utilizing traditional statistical and ML techniques. A comprehensive search was conducted using topic-related keywords and short phrases such as 'cryptocurrency price prediction,' 'ML techniques (and) cryptocurrency price prediction,' 'cryptocurrency price prediction using ML techniques,' and 'DL (and) cryptocurrency price prediction.' Subsequently, the identified papers underwent screening and categorization into distinct groups, encompassing traditional statistical and ML techniques, and further subcategorized into various ML and DL techniques. The field of cryptocurrency price prediction still holds substantial potential for exploration and warrants attention due to its significant impact on the financial system.

# Discussion of the Literature

The extensive literature on cryptocurrency forecasting can be broadly classified into two major categories namely statistical modelling and machine learning [1]. These two categories represent the primary approaches employed to develop predictive models and algorithms for forecasting cryptocurrency behavior.

## Statistical Modelling

According to Brooks (2019) [2], the conventional methods for predicting cryptocurrency prices typically involved the application of statistical and econometric models. These econometric approaches combine statistical and economic theories to estimate and forecast the values of various economic variables. When examining cryptocurrency price volatility and prediction using econometrics, researchers commonly employed statistical models on time-series data. In this section, we provide a review of the existing statistical and econometric techniques utilized for cryptocurrency price prediction.

In the past, classical approaches such as Holt-Winters exponential smoothing, as discussed by Chatfield and Yar (1988) [3], were employed for forecasting time-series data. This approach relies on linear assumptions and involves segregating input data into multiple trends, which are used to predict features that exhibit seasonal effects, such as sales. However, this approach is not suitable for accurately predicting cryptocurrency prices since cryptocurrencies do not exhibit seasonal effects.

In a study conducted by Sovbetov (2018) [4], the augmented Dickey-Fuller unit-root test and bound testing approach were utilized to investigate the factors influencing the prices of five cryptocurrencies (bitcoin, ethereum, dash, litecoin, and monero) from 2010 to 2018, using weekly data. The findings revealed that market beta, trading volume, and volatility exerted a significant influence on the prices of all five cryptocurrencies, both in the short and long run.

Roy, Nanjiba, and Chakrabarty (2018) [5]conducted a study using annual bitcoin data from 2013 to 2017, where they applied time-series models, including the autoregressive integrated moving-average (ARIMA) model, autoregressive model, and moving-average model, to forecast the bitcoin price. Their findings indicated that the ARIMA model outperformed the other models in predicting the bitcoin price.

Guo and Antulov-Fantulin (2018) [6] collected data on volatility and order book information related to bitcoin from September 2015 to April 2017. They proposed temporal mixture models to predict changes in the bitcoin price, and their models demonstrated better predictive performance compared to other models.

In a study by Abu Bakar, Rosbi, and Uzaki (2019) [7], a moving-average method was employed to forecast the bitcoin price. The experiment utilized data collected from October 1 to December 20, 2019. The results showed that the 2-day moving-average method yielded the most accurate predictions with the lowest mean absolute error (MAE) percentage across all observation periods.

Bhambhwani, Delikouras, and Korniotis (2019) [8] focused on extracting data from August 2015 to January 2019 to examine the fundamental drivers of cryptocurrency prices, including bitcoin, ethereum, monero, litecoin, and dash. They utilized the dynamic ordinary least-squares method and found that the prices of these currencies were dependent on factors such as computing power and network characteristics.

Bystrom and Krygier (2018) [9] conducted a study where they analyzed daily, weekly, and monthly data from 2011 to 2017 to explore correlations, regressions, vector autoregression (VAR), and impulse response functions techniques. They investigated the variables influencing changes in bitcoin prices and identified a relationship between these variables and the volatility of the trade-weighted USD currency index, as well as search pressures on bitcoin-related words on Google.

Kaya (2018) [10] employed correlation and regression analyses using weekly data from August 8, 2014, to May 4, 2018, obtained from the Bitstamp online cryptocurrency market. The study aimed to examine the impact of public interest, the volatility metric of the S&P 500 Index options, and political and regulatory news on cryptocurrency prices. The findings revealed that public interest emerged as the most influential factor driving cryptocurrency prices.

Kjærland, Khazal, Krogstad, Nordstrøm, and Oust (2018) [11] utilized econometric methods, including an autoregressive distributed lag model and the GARCH model, to investigate the determinants of bitcoin (BTC) price dynamics. The analysis involved daily spot rates for BTC/USD from January 1, 2013, to February 20, 2018. The results demonstrated that returns on the S&P 500 played a significant role in explaining the dynamics of BTC prices.

Phillips and Gorse (2018a) [12] employed the wavelet coherence approach to examine the co-movement between cryptocurrency prices and various factors, such as social media factors, Google search volume, and Wikipedia. The study utilized cryptocurrency price data from multiple exchanges spanning the period from 2010 to 2017. Additionally, social media factors derived from Reddit, Google search volume from Google Trends service, and data from Wikipedia to track the number of new users learning about a cryptocurrency were included. The findings revealed a positive relationship between cryptocurrency prices and their related factors.

Wiedmer (2018) [13] conducted an investigation into the determinants of cryptocurrency prices using a panel of 17 cross-sections. The study employed unit-root and cointegration tests and estimated the effects using vector error correction models, dynamic ordinary least squares, and fully modified ordinary least squares. Causality flows were tested through weak exogeneity and Granger causality tests. The findings indicated that Metcalfe's law, community factors, and search engine queries exerted a significant influence on cryptocurrency prices.

Troster, Tiwari, Shahbaz, and Macedo (2019) [14] utilized GARCH and generalized autoregressive score (GAS) models to predict bitcoin returns and risks. Out-of-sample performance was compared between the two models. The results demonstrated that the GAS model with a heavy-tailed distribution provided superior out-of-sample predictions, and its flexibility contributed to its robustness.

In the study by Bartolucci et al. (2020) [15], the "butterfly effect" was employed to examine cryptocurrency prices, whereby insights from the "issues" of open-source projects were utilized to enhance cryptocurrency price prediction. The analysis involved sentiment analysis, politeness, and emotions derived from GitHub statistics in the field of econometrics. Various schemes integrating financial data, sentiment analysis, and a range of internal and external factors were employed in the studies. Bitcoin emerged as the most commonly studied and popular cryptocurrency among researchers.

## Machine Learning

In recent years, the application of machine learning (ML) techniques in cryptocurrency price prediction has garnered significant research interest due to its ability to capture general trends and fluctuations. Figure 9 provides an illustration of the utilization of ML in cryptocurrency price prediction. Numerous ML techniques have been employed in this domain, including classification, regression, deep learning (DL), and reinforcement learning (RL) models. The following discussion highlights the ML techniques employed in cryptocurrency price prediction, with a dedicated section focusing on DL and RL models due to their distinct characteristics and widespread adoption. Some researchers have specifically focused on comparing different ML methods in classification and regression tasks.

Derbentsev, Datsenko, Stepanenko, and Bezkorovainyi (2019) [16] proposed a short-term forecasting model that utilized an ML approach to predict cryptocurrency prices of ripple, bitcoin, and ethereum. Rane and Dhage (2019) [17] conducted a comparative analysis of various ML algorithms to select the optimal technique for predicting bitcoin prices. They presented a survey of different ML techniques to determine the most suitable method for bitcoin price prediction, and efforts were made to improve the accuracy of less precise techniques. Chakraborty and Roy (2019) [18] introduced STORJ tokens for transactions within the STORJ network and utilized different time-series models, including Box-Jenkins models, neural networks (NN), and a switching regression model, for predicting future prices.

These studies highlight the diverse ML approaches employed in cryptocurrency price prediction, indicating the increasing focus on developing accurate and effective models in this field.

The use of advanced technologies provides several key advantages, including the flexibility to obtain estimated equations for each regime.

Greaves and Au (2015) [19] utilized blockchain data to predict bitcoin prices by employing support vector machines (SVM), artificial neural networks (ANN), linear regression, and logistic regression. Among these models, an NN classifier with two hidden layers achieved the highest price accuracy of 55%, followed by logistic regression and SVM. Additionally, the research explored the performance of several tree-based models and the K-nearest neighbors approach. However, the study revealed limited predictability when using only blockchain data for training and prediction. The researchers concluded that incorporating features directly extracted from bitcoin exchanges, such as financial flow features, would likely enhance the accuracy of bitcoin price prediction.

Li et al. (2019) [20] examined Twitter signals to predict price fluctuations in ZClassic. They collected tweets on an hourly basis for a duration of 3.5 weeks and classified each tweet as positive, negative, or neutral. These tweets were then used to create weighted or unweighted indices. The researchers trained an extreme gradient boosting (GB) regression tree model and compared its predictions with historic price data.

Mittal, Dhiman, Singh, and Prakash (2019) [21] employed various ML techniques, including linear regression, polynomial regression, recurrent neural networks (RNN), and long short-term memory (LSTM), to analyze the correlation between bitcoin price and Twitter and Google search patterns. Among the analyzed factors, tweet sentiment analysis yielded the worst results. However, when LSTM, RNN, and polynomial regression were applied to Google Trends data and tweet volumes, improved performance accuracy was observed.

These studies demonstrate the application of ML techniques in leveraging blockchain data, Twitter signals, and Google search patterns to enhance cryptocurrency price prediction accuracy.

Atsalakis, Atsalaki, Pasiouras, and Zopounidis (2019) [22] employed a hybrid neuro-fuzzy controller named PATSOS to predict the daily price change trend of bitcoin. This approach demonstrated superior performance compared to two other computational intelligence models, one based on a simpler neuro-fuzzy approach and the other based on artificial neural networks (ANNs). The researchers also noted that the PATSOS system exhibited robust performance when applied to other cryptocurrencies as well.

Attanasio, Garza, Cagliero, and Baralis (2019) [23] conducted a comparison of various classification algorithms, including support vector machines (SVM), naive Bayes, and random forest (RF), to forecast the next-day price trends of a specific cryptocurrency. The results revealed that forecasting models based on a series of forecasts outperformed a single classification model due to the volatility and heterogeneity of financial instruments in cryptocurrencies.

These studies highlight the effectiveness of the PATSOS hybrid neuro-fuzzy controller in predicting bitcoin price trends and the advantage of employing multiple classification models for forecasting cryptocurrency price trends.

# Trends and Challenges

The field of cryptocurrency price prediction is witnessing several emerging trends in statistical and machine learning solutions. One prominent trend is the increased utilization of deep learning techniques, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), to capture complex patterns and temporal dependencies in cryptocurrency price data. Ensembling and hybrid models, which combine multiple forecasting models, have also gained popularity to improve prediction accuracy and robustness. Furthermore, researchers are exploring advanced feature engineering and selection techniques to leverage diverse data sources, including market data, social media sentiment, and news sentiment, to enhance the predictive power of models. These trends indicate a growing focus on incorporating cutting-edge methodologies and leveraging diverse data sources for more accurate cryptocurrency price forecasting.

Despite the advancements in statistical and machine learning solutions, several challenges remain in the field of cryptocurrency price prediction. Data quality and availability are significant concerns, given the volatility of cryptocurrency markets and the need for reliable and high-frequency data. Limited historical data, data inconsistencies, and the presence of outliers pose challenges in building robust prediction models. Additionally, model generalization and transferability are key challenges as cryptocurrencies exhibit unique characteristics, making it difficult to generalize models trained on one cryptocurrency to others. Bridging the gap between theoretical advancements and practical applications also requires interdisciplinary research and collaboration among experts from finance, economics, computer science, and data science. Addressing these challenges will be crucial for developing reliable and robust forecasting models in this rapidly evolving field of cryptocurrency price prediction.

# Conclusion

This survey paper has provided a comprehensive overview of the statistical and machine learning solutions employed for cryptocurrency price prediction. The review highlights the significant progress made in this field, with a wide range of techniques and approaches being explored. From traditional statistical models to advanced machine learning algorithms, researchers have shown a keen interest in developing accurate and reliable models to forecast cryptocurrency prices. Throughout the review, we have observed key trends, such as the adoption of deep learning techniques, ensembling and hybrid models, and the incorporation of diverse data sources. These trends reflect the continuous efforts to improve the predictive power of models and capture the complex dynamics of cryptocurrency markets. Additionally, the survey has shed light on the challenges faced in this domain, including data quality and availability issues, limited historical data, model generalization, and interdisciplinary collaboration.As the field of cryptocurrency continues to evolve, it is crucial to address these challenges and push the boundaries of research. Future work should focus on refining existing models, exploring innovative methodologies, and enhancing data quality and availability. Collaborations between researchers from different disciplines will be essential to develop robust and transferable prediction models.

# References

|  |  |
| --- | --- |
| [1] | A. Khedr, I. Arif, M. El‐Bannany, S. Alhashmi and M. Sreedharan, "Cryptocurrency price prediction using traditional statistical and machine‐learning techniques: A survey. Intelligent Systems in Accounting,," *Finance and Management,* 2021. |
| [2] | C. Brooks, "Introductory econometrics for finance," *Cambridge, UK: Cambridge University Press,* 2019. |
| [3] | C. Chatfield and M. Yar, "Holt-winters forecasting: some practical," *Journal of the Royal Statistical Society: Series D (The Statistician),* vol. 37, no. 2, pp. 129-140, 1988. |
| [4] | Y. Sovbetov, "Factors influencing cryptocurrency prices: Evidence," *Journal of Economics and Financial Analysis,* vol. 2, pp. 1-27, 2018. |
| [5] | S. Roy, S. Nanjiba and A. Chakrabarty, "Bitcoin price forecasting using time series analysis," *21st International Conference of Computer and Information Technology (ICCIT),* pp. 1-5, 2018. |
| [6] | T. Guo and N. Antulov-Fantulin, "Predicting short-term Bitcoin price fluctuations from buy and sell orders," *A hybrid model for bitcoin prices prediction using hidden Markovmodels and optimized LSTMnetworks.,* pp. 721-728, 2018. |
| [7] | N. Abu Bakar, S. Rosbi and K. Uzaki, "Forecasting cryptocurrency price movement using moving average method: A case study of bitcoin cash," *International Journal of Advanced Research,* vol. 7, no. 12, pp. 609-604, 2019. |
| [8] | S. Bhambhwani, S. Delikouras and G. M. Korniotis, "Do fundamentals drive cryptocurrency prices?," https://news.miami.edu/miamiherbert/\_news-assets/files/2019/12/crypto.pdf, 2019. |
| [9] | H. Bystrom and D. Krygier, "What drives bitcoin volatility?," https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3223368, 2018. |
| [10] | Y. Kaya, "Analysis of cryptocurrency market and drivers of the bitcoin price: Understanding the price drivers of bitcoin under speculative environment," *KTH Royal Institute of Technology,* 2018. |
| [11] | F. Kjærland, A. Khazal, E. A. Krogstad, F. Nordstrøm, r. G. and A. Oust, "An analysis of bitcoin's price dynamics," *Journal of Risk and Financial Management,* vol. 11, no. 4, p. 63, 2018. |
| [12] | R. C. Phillips and D. Gorse, "Cryptocurrency price drivers: Wavelet coherence analysis revisited," *PloS ONE,* vol. 13, no. 4, 2018a. |
| [13] | J. Wiedmer, "The price of cryptocurrencies: an empirical analysis," *(MSc thesis), Univerität Bern, Bern, Switzerland,* 2018. |
| [14] | V. Troster, A. K. Tiwari, M. Shahbaz and D. N. Macedo, "Bitcoin returns and risk: A general GARCH and GAS analysis," *Finance Research Letters,* vol. 30, p. 187–193, 2019. |
| [15] | S. Bartolucci, G. Destefanis, M. Ortu, N. Uras, M. Marchesi and R. Tonelli, "The butterfly ‘‘affect’’: Impact of development practices on cryptocurrency prices," *EPJ Data Science,* vol. 9, no. 1, p. 21, 2020. |
| [16] | V. Derbentsev, N. Datsenko, O. Stepanenko and V. Bezkorovainyi, "Forecasting cryptocurrency prices time series using machine learning approach," *SHS Web of Conferences,* vol. 02001, p. 65, 2019. |
| [17] | P. V. Rane and S. N. Dhage, "Systematic erudition of bitcoin price prediction using machine learning techniques.," *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS),* pp. 594-598, 2019. |
| [18] | D. Chakraborty and A. Roy, "Time series methodology in storj token prediction," *2019 International Conference on Data Mining Workshops (ICDMW),* pp. 224-231, 2019. |
| [19] | A. Greaves and B. Au, "Using the bitcoin transaction graph to predict the price of bitcoin," 2015. [Online]. Available: https://pdfs.semanticscholar.org/a0ce/864663c100582805ffa88918910da89add47.pdf. |
| [20] | T. R. Li, A. Chamrajnagar, X. Fong, N. Rizik and F. Fu, "Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model," *Frontiers in Physics,* vol. 7, p. 98, 2019. |
| [21] | A. Mittal, V. Dhiman, A. Singh and C. Prakash, "Short-term bitcoin price fluctuation prediction using social media and web search data," *2019 Twelfth International Conference on Contemporary Computing (IC3),* pp. 1-6, 2019. |
| [22] | G. S. Atsalakis, I. G. Atsalaki, F. Pasiouras and C. Zopounidis, "Bitcoin price forecasting with neuro-fuzzy techniques," *European Journal of Operational Research,* vol. 276, no. 2, pp. 770-780, 2019. |
| [23] | G. Attanasio, P. Garza, L. Cagliero and E. Baralis, "Quantitative cryptocurrency trading: exploring the use of machine learning techniques," *5th Workshop on Data Science for Macro-modelling with Financial and Economic Datasets (DSMM'19),* pp. 1-6, 2019. |