International Journal of Combinatorial Optimization Problems and Informatics, 17(2), 2026, 341-356. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2026.v17i2.1166

www.editada.org

Comparative Analysis of Bitcoin Price Prediction Models: A Systematic literature review

Luis Manuel Lagunez Rodríguez, José Luis López Martínez, Francisco Moo Mena

[1] Universidad Autónoma de Yucatán, Mérida, Yucatán, México
a20216349@alumnos.uady.mx, jose.lopez@correo.uady.mx, mmena@correo.uady.mx

Abstract. This paper presents a systematic literature review of machine learning models used for predicting Bitcoin prices. We identified and analyzed 26 different machine learning models notable for their effectiveness in predictive tasks. Out of these, 6 models are discussed in detail, focusing on their advantages, disadvantages, and potential hybrid approaches. Additionally, we collected 21 evaluation metrics, identifying 8 as the most relevant for cryptocurrency price prediction. In terms of datasets, we found a reliance on public sources such as Kaggle and Yahoo Finance; however, challenges related to the inconsistency in data availability remain. Lastly, we noted a lack of standardized procedures for comparing models, highlighting the need for the development of systematic methodologies to standardize evaluations in this field.

Keywords: Bitcoin, cryptocurrencies, price prediction, machine learning, evaluation metrics, datasets.

Article Info
Received July 1, 2025
Accepted Dec 11, 2026

1 Introduction

Bitcoin is understood as a form of electronic cash that allows for online payments and transfers with the advantages of being instant, anonymous, and without the need to use a third party for this service (Das Gupta, Kolla, Yadav, Arora, & Pandey, 2023). The value of Bitcoin is extremely volatile as it follows the principle of not being governed by any financial entity and being part of the decentralized market (DE-FI). Various researchers have undertaken the task of applying machine learning algorithms to study the behavior of Bitcoin's price over the years (Zhi & Chan, 2024), using information gathered from websites known as exchanges, which serve as intermediaries for the buying, selling, and trading of digital assets.

The main difficulty in predicting the behavior of Bitcoin's price lies in the vast number of variables that influence the buying and selling transactions of this cryptocurrency. Some authors have meticulously selected a series of variables, aiming to ensure that the information obtained is accurate and suitable for effectively training machine learning models, as seen in the research conducted by (Mahfooz & Phillips, 2024) and in (Singh, Pise, & Yoon, 2024).

In the present writing, a Systematic Literature Review (SLR) was conducted with the purpose of exploring the different studies on machine learning models and artificial intelligence techniques regarding the prediction of Bitcoin's price. This systematic review follows the PRISMA 2020 guidelines to ensure methodological rigour and transparency. The subsequent sections include a methodology section that details the use of the SLR, followed by a results and discussion section that addresses the research questions posed in the methodology. Finally, a conclusion is presented, analyzing the obtained results, along with a section on future work that identifies areas of opportunity for future research in the field of computer science.

This paper is structured as follows: Section 2 provides background information and outlines the method used to conduct this Systematic Literature Review. Section 3 defines the search strategy. In Section 4, we describe the procedures conducted in the study. Section 5 discusses the results. Finally, Section 6 offers the main conclusions and suggests directions for future research.

2 Research Method

A systematic literature review (SLR) is a comprehensive analysis of both the quantitative and qualitative aspects of previously published primary studies, defined as a synthesis of the available evidence. This type of review enables researchers to remain current on various topics of interest and to compare existing evidence according to the guidelines proposed by Kitchenham and Charters (Kitchenham et al., 2009) for software engineering. The objective of a systematic literature review is to identify, evaluate, and combine evidence from primary studies using a rigorous methodology.

The steps involved in conducting a proper systematic literature review in software engineering are as follows:

1. Search definition: This step involves establishing the research question, defining the scope of the review, setting the inclusion and exclusion criteria, and finally, creating the search string.

2. Search execution: This step focuses on selecting relevant primary studies and establishing the criteria for analysis.

3. Discussion of results: In this phase, characterization schemes (categories) are created, and the results obtained from the analysis are examined.

Following the methodology outlined by Carrizo and Moller (2018), the method used in this review is described below.

3 Search definition

To analyze the price variation of Bitcoin and its behavior in recent years, the following research questions have been formulated. These questions were developed through an iterative and evolutionary process over several rounds of peer review.

3.1 Research questions

Consequently, the research questions listed in Table 1 have been formulated to guide this study.

Table 1. Research questions

Research question	Motivation
RQ1. How many publications on Bitcoin price prediction models have been published in the last 5 years?	To assess the current academic interest in this field, identify research trends, and determine the ongoing relevance of cryptocurrency forecasting.
RQ2. What variables are considered in Bitcoin price prediction models?	It is crucial to understand the key factors that influence the highly volatile cryptocurrency market. Identifying these variables can improve the accuracy of predictive models.
RQ3. What models are used to predict the Bitcoin price?	Identifying the models used to predict Bitcoin prices is crucial for determining the most effective techniques in financial forecasting.
RQ4. What evaluation metrics are used in Bitcoin price prediction models?	Comparing the accuracy and effectiveness of Bitcoin price predictive models requires an understanding of key metrics.
RQ5. What datasets are used in Bitcoin price prediction models?	To know the datasets used for Bitcoin price predictive models, we must identify their characteristics for future proposals
RQ6. What procedures are used to compare Bitcoin price prediction models?	One objective of this research is to know how to compare Bitcoin price prediction models. This understanding could help researchers and investors establish a formal procedure for evaluating these models.

3.3 Scope of the review

In this study, a collection of literature was conducted through digital libraries, which involved an automated keyword search. The following databases were selected for literature collection: ACM DL, Elsevier, IEEE Xplore, and Springer (see Table 2), as these databases are closely related to the research area and contain articles relevant to Bitcoin price prediction. To create the search string, the AND operator was used to join two groups of keywords, and the OR operator to alternate between keywords defined in the same group. The search string was limited to a maximum of eight logical operators due to technical restrictions of the Elsevier platform. The search sequence uses a combination of keywords highlighting the following specific points:

- The word Bitcoin about price.
- Tags that highlight the presence of prediction models.
- Metrics that allow quantifying the model comparison.
- It specifies that the results obtained are from the computing field.

Table 2. Data Sources

Database	Website
ACM DL	https://dl.acm.org/
Elsevier	https://www.sciencedirect.com
IEEE Xplore	https://ieeexplore.ieee.org/
Springer	https://link.springer.com

3.4 Inclusion and Exclusion Criteria

This section outlines the criteria used for selecting primary studies. The specific inclusion and exclusion criteria are clearly presented in Table 3 and Table 4, respectively.

Table 3. Inclusion criteria

ID	Description
IC1	Studies published in English.
IC2	Studies published between 2020 and 2024
IC3	Studies that are research articles published in peer-reviewed journals.
IC4	Studies that discuss Bitcoin price prediction

Table 4. Exclusion criteria

ID	Description
EC1	Articles whose content is impossible to obtain because they are private.
EC2	We must be diligent in our research and avoid studies that present data already published elsewhere to prevent data duplication.
EC3	Studies that focus on a population that does not meet the inclusion criteria
EC4	Studies that reference Bitcoin as a form of comparison in price prediction and do not work directly with Bitcoin, for example, articles that discuss stock market investments and only mention Bitcoin as a digital asset.
EC5	Studies that have too short a follow-up to obtain relevant results, as this may not provide a comprehensive understanding of the topic.
EC6	Studies that do not meet a minimum methodological standard

To ensure that the literature reviewed offers adequate evidence to address the research questions posed earlier, we implemented a systematic approach to identifying primary studies. This process followed the PRISMA 2020 flowchart model (PRISMA 2020: Preferred Reporting Items for Systematic Reviews and Meta-Analyses, 2020), which outlines the flow of information through the different phases of a systematic review (see Fig1).

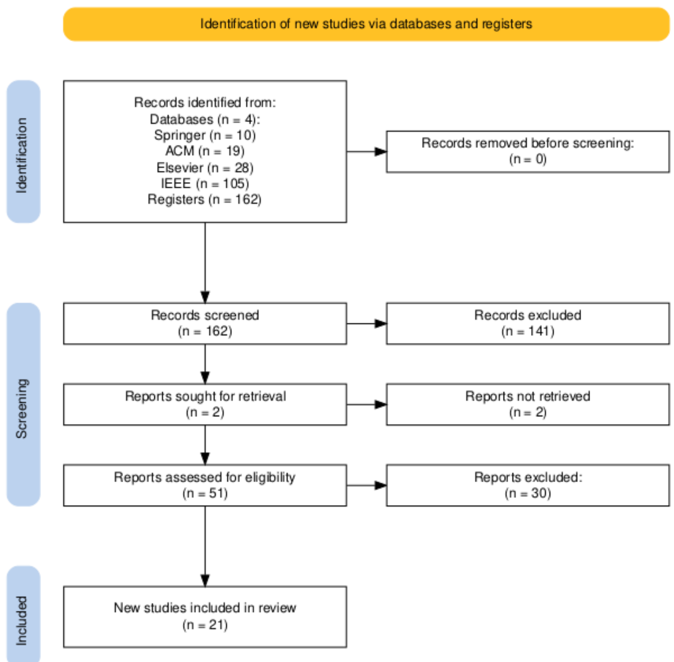

Fig. 1. Primary studies selected after applying the review protocol, flow diagram adapted from (Prisma, Prisma 2020 flow diagram, 2020; Haddaway et al., 2022).

Following PRISMA 2020 guidelines, the selection procedure comprised three distinct steps. During the identification phase, a predetermined search string was used to retrieve records from four specialized databases: ACM, Elsevier, IEEE, and Springer, ensuring comprehensive coverage. In the screening stage, titles and abstracts were assessed to remove duplicates and exclude studies unrelated to the computing sector. Finally, full-text articles were evaluated against specific inclusion and exclusion criteria (IC1-IC4 and EC1-EC6) during the eligibility phase.

This process enhances the reproducibility of the review by providing a transparent audit trail documenting the rationale for each study's inclusion or exclusion. Such a systematic approach minimizes selection bias and ensures that the synthesized evidence on Bitcoin price prediction remains transparent and methodologically robust.

3.5 Search String

The search string used to retrieve information from the databases mentioned earlier was as follows:

("Bitcoin price prediction") AND ("machine learning" OR "prediction models") AND ("variables" OR "datasets" OR "metrics" OR "comparison") AND ("software engineer" OR "computer science").

In this study, we associate the keywords "Bitcoin price prediction," "machine learning," and "prediction models" with the answer to Research Question 1 (RQ1). The term "variables" is related to Research Question 2 (RQ2). For Research Questions 3 (RQ3), 4 (RQ4), and 5 (RQ5), the relevant keywords are "machine learning" and "prediction models," along with "datasets," "metrics," and "comparison." Finally, we include the terms "software engineer" and "computer science" because this study focuses on the field of computing rather than the social or economic fields.

4 Executions

In September 2024, a literature search was conducted using specialized databases (see Table 2), focusing on a five-year retrospective period. This search initially identified a total of one hundred sixty-two articles (see Fig. 1). After applying the established exclusion criteria, twenty-one primary studies were selected. Of these, four articles from ACM, three from Elsevier, thirteen studies from IEEE, and one from Springer.

The selection procedure is comprised of three distinct phases:

Phase 1: Primary studies are evaluated and filtered based on Inclusion Criteria 1 (IC1) and Inclusion Criteria 2 (IC2). Studies that do not satisfy Exclusion Criteria 1 (EC1) and Exclusion Criteria 3 (EC3) will be excluded from consideration.

Phase 2: In this step, primary studies are further assessed and filtered by Inclusion Criteria 3 (IC3) and Inclusion Criteria 4 (IC4).

Phase 3: Finally, primary studies that do not comply with Exclusion Criteria 2 (EC2), Exclusion Criteria 4 (EC4), and Exclusion Criteria 5 (EC5) will be eliminated from the selection process. This methodical approach ensures a rigorous selection of studies adhering to specified inclusion and exclusion criteria. The results from the previous phases are shown in Table 5, illustrating how the various inclusion and exclusion criteria were systematically applied to each database considered in the study.

Table 5. Phases of selection procedure:

Database	First results	Phase 1	Phase 2	Phase 3
ACM DL	19	4	4	4
Elsevier	28	23	10	3
IEEE Xplore	105	18	17	13
Springer	10	6	2	1
Total	**162**	**51**	**33**	**21**

After conducting searches in the database and applying the chosen inclusion and exclusion criteria, it became evident that the number of results found in the "First Results" section was relatively low, particularly when compared to research outputs in more established scientific fields. This limited number of initial results can be attributed to the emerging nature of Bitcoin-related research, especially in price prediction, which remains a developing and specialized topic within the broader domains of finance and artificial intelligence.

The limited amount of research available highlights the scientific community's ongoing efforts to better understand the volatility, market behavior, and underlying mechanisms of cryptocurrencies, especially Bitcoin. Despite its global popularity and growing adoption, Bitcoin remains in the early stages of academic exploration as a possible alternative to fiat currency.

To facilitate a systematic comparison, Table 6 provides a high-level synthesis of the 21 primary studies, highlighting the most representative model, dataset, and metric for each work. It is important to note that this table serves as a general roadmap; therefore, the models and metrics listed here represent the primary focus of each study. Detailed frequency distributions are analyzed more precisely in the subsequent sections corresponding to each Research Question (Section 5).

Table 6. Systematic summary of selected studies: Models, Datasets, and Metrics

Author (Year)	Model(s)	Dataset	Metrics	Main Contribution
Aljojo (2021)	NARX	Private	MSE, R^2	Analyzes the influence of timestamps on prediction.
Aljadani (2022)	DLCP2F (DL)	Binance/CMC	MSE, MAE	Proposed a multi-stage deep learning framework.
Choi (2022)	ResNet	Yahoo Finance	MSE, MAE	Focuses on explainability using gradient-based methods.
Das Gupta (2023)	LSTM, RNN	Kaggle	RMSE, MAE	Evaluation of sequential vs. time-series models.
Iqbal (2024)	CNN-LSTM	Binance	RMSE, MAE	Simultaneous classification-regression framework.
Jay (2020)	Stochastic NN	Kaggle	MSE, MAE	Integration of stochastic components in NNs.
Jones (2022)	LSTM	CMC	RMSE, MAE	Impact of training size on model accuracy.
Kalyani (2023)	ML/Hybrid	Blockchain	Accuracy	Prediction of return rates for financial products.
Liu (2024)	LLM/Sent.	News/X	F1, Acc.	Impact of sentiment analysis via Large Language Models.
Mahfooz (2024)	Exogenous	Yahoo Finance	RMSE, MAE	Forecasting using external/exogenous variables.
Muminov (2024)	DQN	Binance	POCID, Acc.	Reinforcement learning for market direction.
Nayak (2023)	RVFL+AEFA	Yahoo Finance	MAPE, RMSE	Optimization via electric field algorithms.
Oh (2022)	Dense Samp.	Yahoo Finance	MSE, MAE	Improved forecasting through dense time-series sampling.
Parekh (2022)	DL-GuesS	X/Twitter	Acc, F1	Sentiment-based prediction using deep learning.
Patel (2022)	Hybrid	Survey-based	Various	Comprehensive review of architectural advancements.
Sabry (2020)	ML Compar.	CMC	Acc, MSE	Challenges and opportunities in AI for Crypto.

Singh (2024)	Optimized MLP	CMC	MAE, RMSE	Feature subset optimization for higher accuracy.
Syed (2023)	Prognostic DL	Deephaven	RMSE, MAE	Split-second forecasting using high-frequency data.
Vilca Zuniga (2023)	Blockchain ML	Blockchain	Profitability	Maximizing portfolio returns during downtrends.
Zhang (2024)	TimesNet	Multi-source	MSE, MAE	Integration of social sentiment and market data.
Zhi (2024)	Transformer	Social Media	RMSE, MAPE	Clustering social data for enhanced prediction.

Note: CMC = CoinMarketCap. Accuracy is denoted as Acc. All studies utilize OHCLV data as the primary input feature.

5 Discussion of Results

RQ1. *How many publications on Bitcoin price prediction models have been published in the last 5 years?*

Based on a review of primary studies in the literature, there has been a noticeable and sustained increase in the number of publications on Bitcoin price prediction models. Figure 2 shows a significant increase in publications in 2022, where the number of studies reached its highest point. This surge can be attributed to a combination of factors, including a growing interest in decentralized finance (DeFi), a recovery in research activity following the pandemic, and improved access to data and computing resources. However, the slight decline observed in 2023 and 2024 may indicate either a stabilization of the research trend or a shift in focus toward other emerging technologies. This fluctuation suggests that while the topic of Bitcoin price prediction is becoming increasingly relevant, it is still in development and does not yet exhibit the publication consistency found in more established research fields. Therefore, this trend reflects the growing academic and professional interest in applying mathematics and computer science to this field. The information presented in Figure 3 shows the countries that have conducted these studies, which can provide insight into the trends surrounding Bitcoin price prediction. A global distribution of the countries contributing to this research reveals significant participation from Asia, particularly from India (seven studies), South Korea (seven studies), and Saudi Arabia (four studies), which are leading the way in contributions.

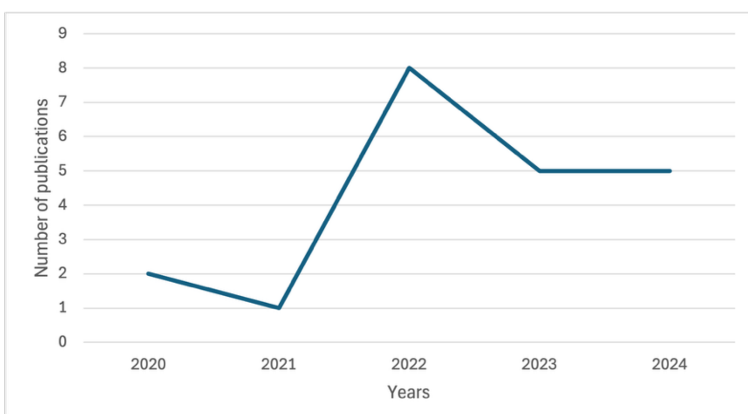

Fig. 2. Publications on Bitcoin Price Prediction Models Over the Past Five Years.

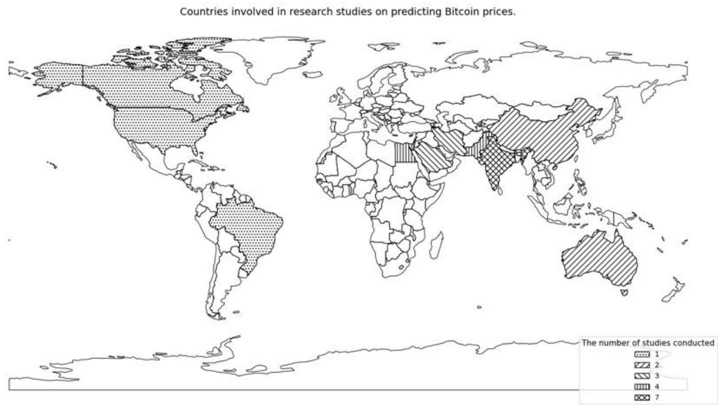

Fig. 3. Countries participating in studies on bitcoin price prediction

RQ2. *What variables are considered in Bitcoin price prediction models?*

Bitcoin, a digital currency that underpins decentralized finance, falls within the realm of economics and, by extension, the social sciences. Compiling the various factors that influence Bitcoin's price is a complex endeavor. Research has identified a total of thirty-nine variables employed across multiple prediction models. For this study, we have selected the six most representative variables from the reviewed studies, as reported by Zhang & Sang (2024) and Sabry, Labda, Erbad, & Malluhi (2020). The selected variables are described in Table 6. Figure 4 shows the previously defined concepts. In addition to these core market variables, several reviewed studies have also incorporated external factors to enhance model performance and account for non-technical influences on Bitcoin's price. These include sentiment indicators (e.g., social media trends or news sentiment), macroeconomic signals (such as interest rates or inflation data), blockchain-specific metrics (like hash rate or transaction volume), and even Google Trends data. Although these features are not as widely utilized as the variables listed in Table 5, they reflect the interdisciplinary nature of Bitcoin price modeling and the importance of considering both technical and behavioral signals in predictive analytics.

Table 7. The selected variables

Variable	Description
Open	The price at which the asset began trading during a given period, which helps gauge initial market sentiment.
High	The highest price reached by the asset during the specified period.
Close	The price at which the asset finished trading at the end of the designated period.
Low	The lowest price recorded for the asset within the same time frame.
Volume	The total number of units traded during the period, indicating market activity.
Date	The time reference (e.g., day, hour, minute) corresponding to the recorded price and volume data.

Note: The first five variables constitute the **OHCLV** (Open, High, Close, Low, Volume) dataset, which is the standard benchmark in financial forecasting. All 21 primary studies analyzed in this SLR consistently use these variables as the core inputs to their machine learning models.

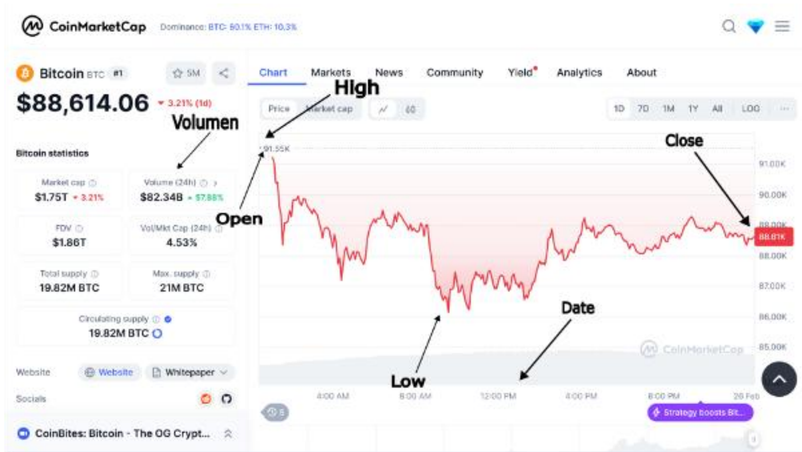

Fig. 4. Daily bitcoin time frame. Image taken from (Coinmarketcap, s.f.).

RQ3. *What models are used to predict the Bitcoin price?*

The review of articles identified 26 machine learning models (see Fig. 5) used in Bitcoin price prediction (see Syed et al., 2023). (Jones & Demirel, 2022) mentions the differences that occur when experiments are performed with different training sizes. (Zhi & Chan, 2024) shows how a machine learning model is trained, as well as the metrics used to compare its performance with other experiments. The models analyzed exhibit at least one of the following characteristics:

- Comparison of one model with another to evaluate its performance.
- Detailed study of a model, including an analysis of its advantages and disadvantages.
- Combination of two models to create a hybrid approach, the results of which were compared with those of other models.

The frequency analysis presented in Figure 5 indicates a distinct trend toward the adoption of deep learning architectures, especially those designed for sequential data processing. Although the numerical distribution highlights the prevalence of models such as LSTM, a more comprehensive technical comparison is necessary to elucidate the rationale for selecting these models in the 21 studies examined.

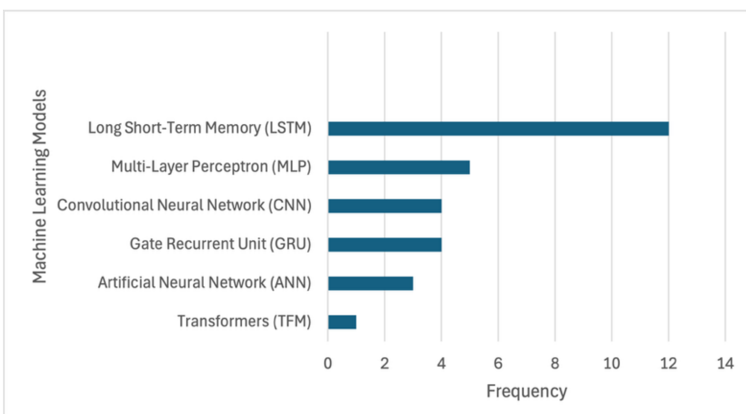

Fig. 5. Frequency table of machine learning models used for Bitcoin price prediction.

Consequently, Table 8 offers a comparative synthesis of these models, outlining their fundamental characteristics, advantages, and limitations within the context of Bitcoin price forecasting. This qualitative mapping facilitates a clearer understanding of the factors influencing researchers' transitions between traditional neural networks and advanced hybrid or attention-based architectures.

Table 8. Comparative analysis of models used in Bitcoin price prediction: Features, benefits, and constraints

Model	Description	Advantage	Limitations
Long Short-Term Memory	Recurrent neural network designed to process temporal data sequences.	Excellent capacity to retain long-term dependencies in time-series.	High computational cost and risk of overfitting on small datasets.
Multi-Layer Perceptron	Basic architecture of dense layers that maps inputs to non-linear outputs.	Versatile and easy to implement for identifying general market patterns.	Difficulty capturing the sequential and temporal nature of prices.
Convolunational Neural Network	Networks that use filters to extract local features from data.	Highly efficient at detecting local patterns and trends in price charts.	Not intrinsically designed for long-term sequential ordering.
Gate Recurrent Unit	Simplified version of LSTM with fewer control gates.	Faster training than LSTM with similar predictive performance.	Lower memory capacity for extremely long sequences compared to LSTM.
Artificial Neural Network	Combination of two or more architectures (e.g., CNN-LSTM).	Leverages the strengths of each model to improve accuracy and robustness.	Increased complexity in hyperparameter tuning and architecture design.
Transformers	Models based on attention mechanisms that process data in parallel. & Superior ability to capture complex relationships in global data.	Superior ability to capture complex relationships in global data.	Requires large data volumes to outperform traditional recurrent networks.

RQ4. *What evaluation metrics are used in Bitcoin price prediction models?*

During the literature synthesis, it was noted that 21 evaluation metrics are employed to assess model performance. Of these, four metrics are commonly used for evaluating prediction models. Additionally, four metrics specifically applied to Bitcoin price prediction were identified in studies by Aljojo, Alshutayri, Aldhahri, Almandeel, and Zainol (2021), Patel (2022), and Nayak, Das, Dehuri, and Cho (2023). Figures 6 and 7 illustrate the most frequently used evaluation metrics across the reviewed studies. Metrics such as the Coefficient of Determination (R^2) and F1 Score are widely used in machine learning tasks, as they provide insight into a model's predictive power and classification performance, respectively. However, their use in financial time-series forecasting is less common, as they do not always reflect economic or investment significance.

In contrast, Figure 7 highlights the dominance of regression-based error metrics, such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), which are particularly relevant for predicting continuous values. These metrics are widely adopted in Bitcoin price prediction due to their ability to quantify the deviation between expected and actual prices in intuitive and interpretable ways.

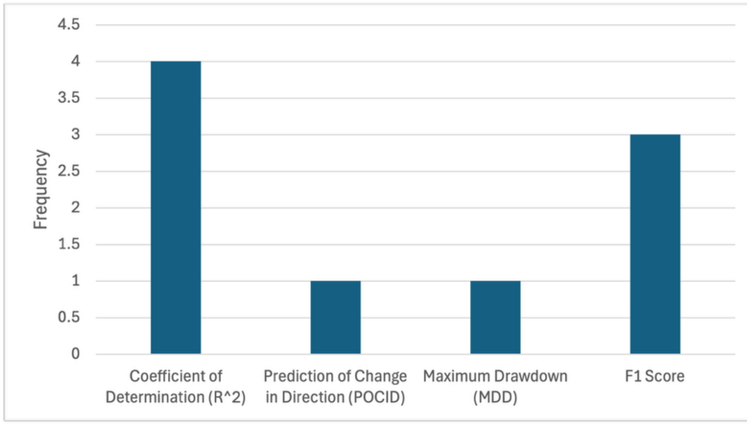

Fig. 6. Frequency table of evaluation metrics used specifically for bitcoin price prediction.

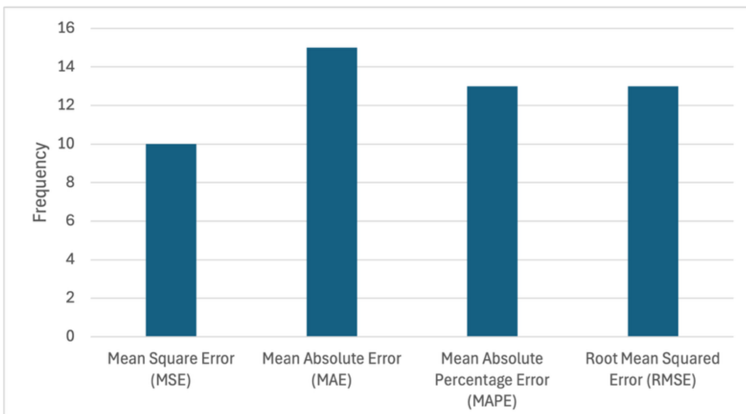

Fig. 7. Frequency table of evaluation metrics used in machine learning models for bitcoin price prediction.

A wide range of 21 evaluation metrics are used to validate Bitcoin price prediction models, according to a synthesis of the research. Regression-based error metrics clearly predominate, as shown by the frequency analysis in Figures 6 and 7, but it is crucial to classify these tools based on their particular use in the financial industry.

Table 9 provides a comparative overview that distinguishes between measures designed to assess the economic relevance and directional predictability of Bitcoin price fluctuations and general performance metrics that evaluate technical model accuracy. This categorization is significant because financial measurements (Section B) yield insights into market utility and risk management, while classic error metrics (Section A) quantify deviations from actual prices.

Table 9. Comparative analysis of evaluation metrics: General vs. Bitcoin-specific

Metric	Description	Advantages	Limitations
Section A: General Evaluation Metrics (Model Performance)			
Mean Square Error (MSE)	Measures average squared difference between values.	Penalizes larger errors; useful for high-volatility.	Highly sensitive to outliers in Bitcoin data.
Mean Absolute Error (MAE)	Calculates average magnitude of errors.	Direct measure of error in price units; easy to interpret.	Does not account for error direction.
Mean Absolute Percentage Error (MAPE)	Measures percentage of error relative to actual price.	Scale-independent; allows comparison across different price levels.	Can be problematic if actual values are near zero.
Root Mean Squared Error (RMSE)	Square root of the average of squared errors.	Same units as the target variable; emphasizes large errors.	Like MSE, it is very sensitive to extreme volatility spikes.
Section B: Specific Bitcoin/Financial Metrics (Market Utility)			
Coefficient of Determination (R^2)	Statistical fit or classification trend measures.	Standardized for comparing ML performance.	May not reflect economic significance in trading.
Prediction of Change in Direction (POCID)	Percentage of Change in Direction.	Evaluates ability to predict price movement direction.	Does not quantify the magnitude of the change.
Maximum Drawdown (MDD)	Maximum Drawdown.	Crucial for risk, measuring largest peak-to-trough decline.	Only focuses on losses, not on overall profitability.
F1 Score	Balance between precision and recall for trends.	Robust for evaluating classification of price directions.	Less intuitive for regression-based price forecasting.

RQ5. *What datasets are used in Bitcoin price prediction models?*

When analyzing the origin of the datasets used in various studies, it was noted that some authors did not specify where their information was sourced. This raises the possibility that they may have created their own datasets, which were not made publicly available alongside their published work. In Figure 8, Kaggle is referenced twice as a data source. The first instance represents information obtained from social network X, which was compiled for market sentiment analysis. The second instance refers to data collected from Binance, a cryptocurrency exchange that provides historical pricing information for various cryptocurrencies. Additionally, Yahoo Finance relies on CoinMarketCap as its primary source of information. However, it is evident that there were more queries on Yahoo Finance than on CoinMarketCap. This disparity may be attributed to the relative accessibility of information on one platform compared to the other (Muminov, Sattarov, & Na, 2024; Oh & Lee, 2022).

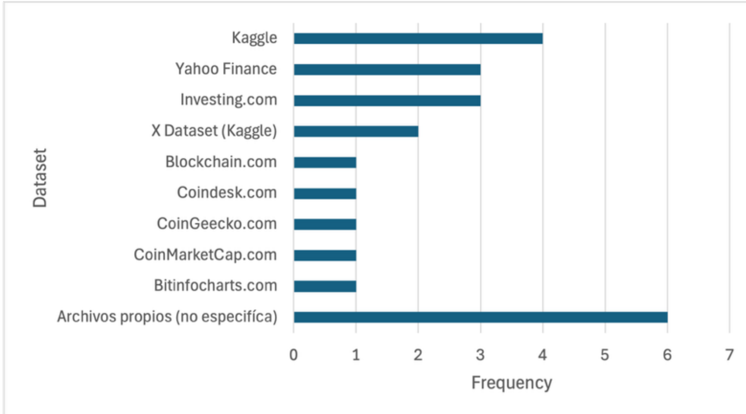

Fig. 8. Frequency table of datasets used in machine learning models for Bitcoin price prediction.

The datasets used in the 21 primary studies demonstrate a strong reliance on open-access repositories and financial platforms. As shown in Figure 8, platforms such as **Kaggle** and **Yahoo Finance** are the most frequent sources, primarily because they provide historical OHCLV data that is already structured for machine learning tasks.

A critical finding of this review is the lack of standardization in data granularity and timeframes. Some studies utilize high-frequency data, such as minute or hourly intervals, while others rely on daily closing prices, which directly affects the comparability of model performance. Additionally, approximately 28% of the reviewed articles (6 studies) use private or non-specified datasets, creating significant challenges for reproducibility. This disparity indicates that the field would benefit from a standardized, publicly available benchmark dataset that includes not only price data but also social sentiment and macroeconomic signals to ensure fair model evaluation.

RQ6. *What procedures are used to compare Bitcoin price prediction models?*

In the reviewed studies, authors compared Bitcoin price prediction models using various metrics such as those presented above. Some authors have used comparative tables to highlight the different machine learning models, along with the results obtained for each metric (Iqbal et al., 2024; Kalyani, Parvathy, Abdeljaber, Murthy, & Acharya, 2023), accompanied by graphical material to facilitate the visualization of the information provided. However, the way the results were presented varied according to each author's preferences, leading to a lack of uniformity in the comparison (Vilca Zuniga et al., 2023; Parekh et al., 2022).

The reviewed literature did not identify any standard procedures or methodologies for comparing Bitcoin price prediction models. This gap presents an opportunity to establish and standardize results by employing systematic approaches based on software engineering (Liu et al., 2024).

6 Conclusions and future work

The models identified showcase a wide range of approaches, focusing on comparisons, as well as the analysis of advantages and disadvantages, and proposing hybrid methods. However, significant challenges arise from the lack of uniformity in comparison methodologies and the dependence on non-standardized datasets.

The evaluation metrics used in different studies vary greatly among authors, which complicates direct comparisons of the approaches. While websites like Yahoo Finance and Kaggle are commonly utilized, the creation of specific datasets and their limited availability hinder the reproducibility of these studies.

The analysis of the 21 primary studies reveals several critical gaps in methodological maturity. Table 10 synthesizes these findings by contrasting current key results with the challenges that impede standardization in the field. This summary provides a technical justification for the lack of reproducibility and the considerable variety of metrics previously discussed.

Table 10. Synthesis of findings and identified research gaps in Bitcoin price prediction

Focus Area	Key Findings	Identified Gaps
Model Evaluation	Use of 21 distinct metrics, predominantly regression-based (MSE, MAE).	Lack of uniformity: Absence of a standard benchmark hinders direct comparison between hybrid and traditional models.
Data Transparency	High reliance on Kaggle and Yahoo Finance, but 28% of studies use non-specified sources	Reproducibility issues: Limited availability of specific datasets pre vents the validation of experimental results.
Variable Selection	Universal adoption of OHCLV data as the core input for ML models.	Interdisciplinary gap: Limited integration of social sentiment and macroeconomic signals in standard models.
Comparison Methods	Ad-hoc comparison procedures based on individual author preferences.	Methodological void: Need for systematic software engineering approaches to formalize model evaluation.

Additionally, no standardized comparison methodology has been identified, indicating an opportunity to employ software engineering-based approaches that could facilitate a more structured and transparent evaluation of predictive models.

The findings presented in this study highlight several opportunities for improving and developing Bitcoin price prediction models. First, it is recommended to develop standardized methodologies for evaluating predictive models. This would enable more accurate and fair comparisons when applying artificial intelligence techniques to Bitcoin price forecasting. Another important aspect is the generation and dissemination of standardized and accessible datasets. This promotes the reproducibility of studies. Since the data can vary based on the collection process, it is essential to establish standardization to facilitate future research.

Appendix

The references of the studies included in the review can be consulted at the following link: https://github.com/LuisLagunez/Articulos-SLR

References

Aljadani, A. (2022). DLCP2F: A DL-based cryptocurrency price prediction framework. *Discover Computing, 1*(1), 1–15. https://doi.org/10.1007/s44163-022-00036-2

Aljojo, N., Alshutayri, A., Aldhahri, E., Almandeel, S., & Zainol, A. (2021). A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of timestamp influence on Bitcoin value. *IEEE Access, 9*, 148611–148624. https://doi.org/10.1109/ACCESS.2021.3124629

Carrizo, D., & Moller, C. (2018). Methodological structures of systematic literature review in software engineering: A systematic mapping study. *Ingeniare. Revista Chilena de Ingeniería, 26*(Suppl. 1), 45–54. https://doi.org/10.4067/S0718-33052018000500045

Choi, H., Jung, C., Kang, T., Kim, H. J., & Kwak, I. Y. (2022). Explainable time-series prediction using a residual network and gradient-based methods. *IEEE Access, 10*, 110537–110550. https://doi.org/10.1109/ACCESS.2022.3213926

CoinMarketCap. (s. f.). *CoinMarketCap*. Recuperado el 5 de diciembre de 2024 de https://coinmarketcap.com/

Das Gupta, S., Kolla, T., Yadav, R., Arora, M., & Pandey, M. (2023). Forecasting cryptocurrency prices using sequential and time series models. En *Proceedings of the 4th International Conference on Information Management & Machine Intelligence (ICIMMI '22)* (Article 91, pp. 1–6). ACM. https://doi.org/10.1145/3590837.3590928

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams. *Campbell Systematic Reviews, 18*(2), e1230. https://doi.org/10.1002/cl2.1230

Iqbal, M., Iqbal, A., Alshammari, A., Ali, I., Maghrabi, L. A., & Usman, N. (2024). Sell or HODL cryptos: Cryptocurrency short-to-long term projection using simultaneous classification-regression deep learning framework. *IEEE Access, 12*, 118169–118184. https://doi.org/10.1109/ACCESS.2024.3448234

Jay, P., Kalariya, V., Parmar, P., Tanwar, S., Kumar, N., & Alazab, M. (2020). Stochastic neural networks for cryptocurrency price prediction. *IEEE Access, 8*, 70294–70308. https://doi.org/10.1109/ACCESS.2020.2990659

Jones, J., & Demirel, D. (2022). Long short-term memory for Bitcoin price prediction. En *Proceedings of the 6th International Conference on Information System and Data Mining (ICISDM '22)* (pp. 25–30). ACM. https://doi.org/10.1145/3546157.3546162

Kalyani, K., Parvathy, V. S., Abdeljaber, H. A., Murthy, T. S., & Acharya, S. (2023). Effective return rate prediction of blockchain financial products using machine learning. *Computers, Materials & Continua, 74*(1), 2303–2316. https://doi.org/10.32604/cmc.2023.033162

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering. *Information and Software Technology, 51*(1), 7–15. https://doi.org/10.1016/j.infsof.2008.09.009

Liu, C., Arulappan, A., Naha, R., Mahanti, A., Kamruzzaman, J., & Ra, I.-H. (2024). Large language models and sentiment analysis in financial markets: A review, datasets, and case study. *IEEE Access, 12*, 134041–134061. https://doi.org/10.1109/ACCESS.2024.3445413

Mahfooz, A., & Phillips, J. L. (2024). Conditional forecasting of Bitcoin prices using exogenous variables. *IEEE Access, 12*, 44510–44526. https://doi.org/10.1109/ACCESS.2024.3381516

Muminov, A., Sattarov, O., & Na, D. (2024). Enhanced Bitcoin price direction forecasting with DQN. *IEEE Access, 12*, 29093–29112. https://doi.org/10.1109/ACCESS.2024.3367719

Nayak, S. C., Das, S., Dehuri, S., & Cho, S.-B. (2023). An elitist artificial electric field algorithm based random vector functional link network for cryptocurrency prices forecasting. *IEEE Access, 11*, 57693–57716. https://doi.org/10.1109/ACCESS.2023.3283571

Oh, I.-S., & Lee, J.-S. (2022). Dense sampling of time series for forecasting. *IEEE Access, 10*, 75571–75580. https://doi.org/10.1109/ACCESS.2022.3191668

Parekh, R., et al. (2022). DL-GuesS: Deep learning and sentiment analysis-based cryptocurrency price prediction. *IEEE Access, 10*, 35398–35409. https://doi.org/10.1109/ACCESS.2022.3163305

Patel, N., et al. (2022). Fusion in cryptocurrency price prediction: A decade survey on recent advancements, architecture, and potential future directions. *IEEE Access, 10*, 34511–34538. https://doi.org/10.1109/ACCESS.2022.3163023

PRISMA. (2020). *PRISMA 2020: Preferred reporting items for systematic reviews and meta-analyses*. https://www.prisma-statement.org/prisma-2020

Sabry, F., Labda, W., Erbad, A., & Malluhi, Q. (2020). Cryptocurrencies and artificial intelligence: Challenges and opportunities. *IEEE Access, 8*, 175840–175858. https://doi.org/10.1109/ACCESS.2020.3025211

Singh, S., Pise, A., & Yoon, B. (2024). Prediction of Bitcoin stock price using feature subset optimization. *Heliyon, 10*(7), e28415. https://doi.org/10.1016/j.heliyon.2024.e28415

Syed, S., Iqbal, A., Mehmood, W., Syed, Z., Khan, M., & Pau, G. (2023). Split-second cryptocurrency forecast using prognostic deep learning algorithms: Data curation by Deephaven. *IEEE Access, 11*, 128644–128654. https://doi.org/10.1109/ACCESS.2023.3331652

Vilca Zuniga, E. W., Ranieri, C. M., Zhao, L., Ueyama, J., Zhu, Y., & Ji, D. (2023). Maximizing portfolio profitability during a cryptocurrency downtrend: A Bitcoin blockchain transaction-based approach. *Procedia Computer Science, 222*, 539–548. https://doi.org/10.1016/j.procs.2023.08.192

Zhang, Z., & Sang, Q. (2024). Enhancing Bitcoin log-return prediction through TimesNet and sentiment analysis integration. En *Proceedings of the 4th International Conference on Machine Learning and Computer Application (ICMLCA 2023)* (pp. 82–91). ACM. https://doi.org/10.1145/3650215.3650232

Zhi, Y., & Chan, T.-H. H. (2024). Clustering social media data for Bitcoin price prediction with transformer model. En *Proceedings of the 6th International Conference on Machine Learning and Natural Language Processing (MLNLP ’23)* (pp. 135–140). ACM. https://doi.org/10.1145/3639479.3639508