A Short Survey on Bitcoin Price Prediction

Stefano Bistarelli[1], Francesco Santini[1] and Luca Maria Tutino[1,2]

[1]Dipartimento di Matematica e Informatica, Università degli Studi di Perugia, Via Vanvitelli 1, 06123 Perugia (PG), Italy
[2]Dipartimento di Matematica e Informatica "Ulisse Dini", Università degli Studi di Firenze, Viale Giovanni Battista Morgagni, 67/a, 50134 Firenze (FI), Italy

Abstract

Bitcoin and other cryptocurrencies have emerged as decentralized digital payment systems driven solely by blockchain technology and supply and demand dynamics. Predicting their prices has become a significant field of research with important implications in automated trading. This study conducts a Systematic Literature Review (SLR) to identify the most common features and models used in recent works for Bitcoin price prediction. Employing an ad hoc method to automate the survey procedure, a summary table is constructed to extract pertinent information. The objective is to discern the performance of various features and models across different data ranges and resolutions, laying the groundwork for developing an effective automated trading system. By synthesizing findings from the literature, this study aims to provide insights into the state-of-the-art methodologies for cryptocurrency price prediction, facilitating informed decision-making for market agents.

Keywords

Bitcoin, Cryptocurrency, Machine Learning, Price prediction,

1. Introduction

Interest in cryptocurrencies, especially in Bitcoin, has returned to its peak due to the recent bull market, which has shown even the most skeptical that the price of Bitcoin exhibits a notable regularity linked to one of the primary (inflationary) characteristics of the algorithm: the periodic (quadrennial) decrease of the Mining Reward. Despite the difficulty in capturing this periodicity with existing models, it is far from certain that it does not exist.

Bitcoin [1] is a cryptocurrency created through open-source software operating on peer-to-peer networks, serving as an irreversible private payment platform. It lacks physical form, is not backed by any public entity, and thus does not require intervention from government agencies or other intermediaries like banks to conduct transactions. These transactions occur within the blockchain system, which acts as an open ledger, efficiently recording transactions between parties and leaving an indelible mark that cannot be erased. This makes blockchain a decentralized validation protocol that is difficult to manipulate, with a low risk of fraud.

Being a recently created financial product, Bitcoin still faces the challenge of high volatility (see Figure 1). According to [2], over seven years, from April 2015 to April 2022, the standard deviation of Bitcoin's daily return rate was 3.85%, 2.68 times that of gold's return rate during the

DLT 2024: 6th Distributed Ledger Technologies Workshop, May 14-15 2024, Turin, Italy
stefano.bistarelli@unipg.it (S. Bistarelli); francesco.santini@unipg.it (F. Santini); lucamaria.tutino@unifi.it (L. M. Tutino)
https://bista.sites.dmi.unipg.it/ (S. Bistarelli); https://francescosantini.sites.dmi.unipg.it/ (F. Santini)
0000-0001-7411-9678 (S. Bistarelli); 0000-0002-3935-4696 (F. Santini); 0009-0002-5135-9681 (L. M. Tutino)

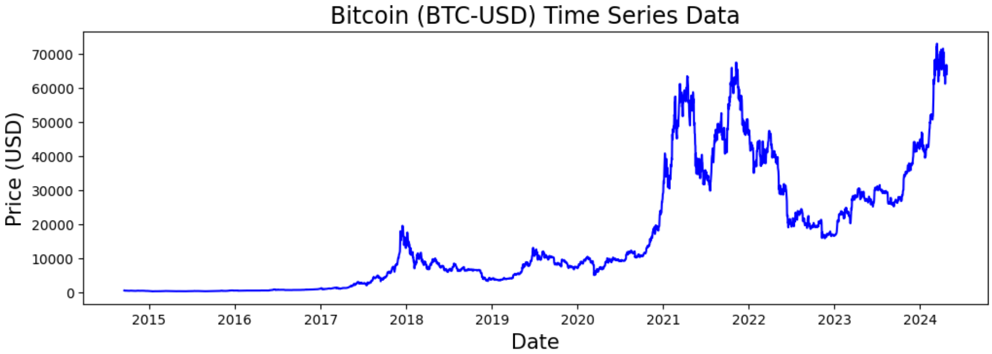


Figure 1: Time evolution of Bitcoin price.

same period, and 3.36 times that of the S&P500 index. Because of these significant price fluctuations, Bitcoin's function as a store of value and transactional currency has been questioned, and the findings of several researchers indicate that Bitcoin might represent a new asset class [3].

While benefiting from BTC's security and decentralization, understanding its trend to minimize floating risk poses a complex problem. Apart from this critical issue, owning cryptocurrencies can also bring high returns to appropriately managed portfolios.

This paper introduces a meticulous *Systematic Literature Review* (*SLR*, [4]) focusing on Bitcoin price prediction. Given the relatively nascent nature of Bitcoin and its unique characteristics, traditional predictive approaches derived from other asset classes (typically from the stock market) offer only partial applicability. However, the flexibility of machine learning approaches and the availability of extensive multidimensional and multimodal data present a promising avenue for forecasting Bitcoin prices. By undertaking this review, we seek to contribute to the decision-support literature by identifying pertinent methods, patterns, and promising avenues for further investigation. Additionally, the review aims to develop robust reporting guidelines to enhance transparency within the field.

Cryptocurrency price prediction is a complex task influenced by various factors such as volatility, trading volume, mining difficulty, market sentiment, regulatory developments, and technological advancements. Integrating these diverse factors poses a significant challenge for researchers and practitioners alike. However, advancements in statistical analysis, machine learning algorithms, and deep learning techniques offer promising solutions to address these challenges.

The SLR performed in this paper aims to contribute to understanding Bitcoin price prediction by trying to answer the following research questions:

RQ1: What features are used in the prediction, and from which sources are they taken?

RQ2: What models are used to make the predictions?

RQ3: What models perform best for different data ranges and resolutions?

In Section 2, we report some of the related work in the literature that surveys Bitcoin forecasting; Section 3 describes the survey methodology step by step, from the choice of the database to the extraction of the desired information in the final dataset of paper; in Section 4, we give the results analyzing our records using pies, occurrences histograms, and heat maps. Finally, Section 5 presents the final findings and ideas about possible future work.

2. Related work

In literature, some recent survey works that have an objective similar to ours, i.e. Bitcoin price prediction, are: Jaquart et al. 2020 [5]; Khedr et al. 2021 [6]; Patel et al. 2022 [7]; Mezquita et al. 2022 [8]; Fang et al. 2022 [9]; Zhang et al. 2024 [10].

In [5], the authors conducted a literature search following the guidelines proposed by Webster and Watson [11]. They initiated the process by querying various interdisciplinary research databases. This search aimed to match their defined search terms in titles, abstracts, or keywords, leading to the identification of an initial set of 101 publications by April 2019. Each publication's title and abstract were then analyzed, resulting in the exclusion of 76 papers that did not explicitly align with the scope of the literature review. Subsequent forward and backward searches yielded eight additional relevant articles, bringing the total to 33 papers for in-depth review. The authors then derived key concepts for paper categorization through an initial screening of 10 recent peer-reviewed conference proceedings and journal papers, reviewed these papers, and developed a set of initial ideas for classification, which were refined throughout the screening process. The final synthesis identified four distinct concepts: Method, Features, Prediction Interval, and Prediction. The authors reflect on the significance of Bitcoin pricing research via machine learning, highlighting the lack of transparency and comparability across the reviewed literature. They propose recommendations for future researchers to address these issues, including structured disclosure of model configurations, publication of models and data to open research repositories, and benchmarking against other published models.

The contribution of [6] lies in its comprehensive analysis of articles about cryptocurrency price prediction, spanning the period from 2010 to 2020. It sheds light on the challenges inherent in traditional approaches and proposes a shift towards Machine Learning (ML) and Deep Learning (DL) paradigms for more effective solutions. This study offers valuable insights into future research directions and potential solutions to these challenges, filling a critical gap in existing literature and providing a foundation for further exploration.

The work in [8] aims to provide a comprehensive overview of the current state of the cryptocurrency market and the patterns discernible from its activity, explicitly focusing on price prediction. Through a review of 13 papers, various approaches to predicting cryptocurrency prices are explored, including analysis of social network and news activity, machine learning algorithms, and market behavior in response to regulatory actions. The works examined employ diverse methods, such as machine learning and statistical analysis, to forecast market closing prices and analyze factors like sentiment analysis of tweets and news, investor reactions to regulatory announcements, and the underlying technology of each cryptocurrency. It's concluded that some methods display promise in predicting market prices accurately and that external factors such as regulations and social sentiment, coupled with pump-and-dump schemes, complicate prediction tasks.

In [9], the authors give an overview and analysis of the research work on cryptocurrency trading. This survey presents a vocabulary of the definitions and current state of the art. The paper comprehensively surveys 146 cryptocurrency trading papers and analyses the research distribution that characterizes the cryptocurrency trading literature. Research distribution among properties and categories/technologies is analyzed in this survey. The study also summarises the datasets used for experiments and examines the research trends and opportunities in cryptocurrency trading. Future research directions and opportunities are discussed.

The work in [10] provides a comprehensive review of deep learning methods applied in cryptocurrency research, spanning price prediction, portfolio management, bubble analysis, abnormal trading detection, trading regulations, and initial coin offerings. The contribution of this study lies in four key aspects: conducting a literature review on deep learning models across multiple financial applications, offering an overview of cryptocurrency history and significant currencies, comprehensively reviewing deep learning models in cryptocurrency research across diverse tasks, discussing findings from reviewed studies, and outlining future research directions. The review highlights the popularity of deep learning models such as convolutional neural networks, recurrent neural networks, deep belief networks, and deep reinforcement learning in financial applications, particularly for cryptocurrency price prediction. Researchers have developed LSTM-based derivative models like the SAM-LSTM model, which outperform traditional prediction models. Deep learning models such as DRL, LSTM, RF, RNN, GRU, and MLP have also been employed to analyze Bitcoin asset portfolios, detect market bubbles, identify abnormal trading behaviors, and supervise market transactions. Some studies have also combined multiple deep-learning models to enhance prediction accuracy. Overall, this study contributes to advancing multidisciplinary research on cryptocurrency and deep learning, providing insights into the efficacy of deep learning models across various cryptocurrency-related tasks and paving the way for future research in this domain.

In [7], it is stated that in the realm of cryptocurrency price prediction, researchers deploy a combination of time-series models like ARIMA and sophisticated deep learning algorithms such as LSTM and GRU, crucial for capturing the intricate patterns and trends in cryptocurrency price movements, which are characterized by high volatility and non-linearity. Moreover, comprehensive statistical analyses of cryptocurrency price histories are imperative to understand these digital assets' underlying dynamics and risks. The researchers also pay close attention to market activities, including the actions of companies, startups, organizations, and influential individuals within the cryptocurrency ecosystem, as these factors can significantly influence price fluctuations. However, despite the advancements in predictive modeling, there remains a notable gap in existing research concerning the interdependencies among different cryptocurrencies. This oversight is addressed by the authors of this study, who propose a novel fusion model that integrates LSTM and GRU architectures. By considering hierarchical dependencies among cryptocurrencies, their model demonstrates superior predictive accuracy compared to conventional approaches. Furthermore, the authors emphasize the importance of ongoing research efforts to enhance prediction accuracy through federated and distributed learning methodologies.

3. Methodology

The first phase of the work was dedicated to developing the review protocol. This part is particularly critical, as effectively and precisely defining all the steps allows for optimizing and expediting the entire process. An initial scoping exercise was conducted to glean a preliminary understanding of the existing research landscape and to ascertain suitable search strategies, such as identifying appropriate databases, defining time periods, selecting relevant search terms, and considering language restrictions. We then developed our protocol, structuring the review in steps as follows:

- **database selection**: selection of the database to use in the analysis;
- **string development**: development of the string for the database search, using logic operators and suitable restrictions;
- **inclusion criteria**: selection of the subset of papers to use in the last phase of the survey;
- **extracted information**: extraction and analysis of the gathered information.

3.1. Database selection

After the preliminary search, as the first step, we select the database to use in the review. Bibliographic databases (BDBs) play a vital role in systematic reviews, offering access to standardized descriptions of documents and advanced query systems and also additional services such as citation metrics, bibliometric data, and links to repositories. The most popular databases for Computer Science research are Google Scholar[1], Elseviers Scopus[2], and Clarivate Web of Science (WoS[3]), the last two mainly used to ensure a curated selection of articles. The earliest and the most current content coverage comparisons have attested that Scopus offers more comprehensive coverage than WoS. In general, it was shown that there was a significant overlap between the information indexed by Scopus and WoS, with Scopus, even if the overlap between was determined to vary significantly across disciplines like pointed out in [12], indexing more unique sources than WoS [13].

Elsevier's Scopus contains peer-reviewed scientific output in books, journal articles, conference proceedings, and more than 90 million abstracts, bibliometric data, and bibliographic citations. Academic subjects covered include health sciences (medicine, veterinary, professions health); life sciences (biology, agriculture, biochemistry, pharmacy) = 17%; physical sciences (chemistry, physics, mathematics, engineering, information technology, social sciences) = 27%; and social sciences and humanities (art, literature, psychology, economics, and social sciences) = 31%. Leveraging its advanced search features, Scopus allowed us to narrow our search parameters effectively and efficiently. Additionally, the platform's extensive database afforded us greater confidence in the reliability and relevance of the articles retrieved.

[1]Scholar: https://scholar.google.com/
[2]Scopus: https://www.scopus.com/
[3]WoS: https://www.webofscience.com/

3.2. String development

Following the database selection, we worked on creating the search string on Scopus. This component must be carefully curated to balance finding a broader pool of publications and methodically choosing a smaller set of papers that satisfy inclusion criteria.

Our search string consisted of the three main key terms: *"Bitcoin"*, *"Price"*, and *"Prediction"*. Based on our research objectives, these terms were complemented with alternatives using the OR logic operator and asterisks to include all the forms of the words prediction and forecasting. We also requested our papers to be written in English, to be *articles* or *conference papers*, to be published before 2024, and to belong to the four subject areas of *Computer Science*, *Engineering*, *Mathematics* and *Economics, Econometrics, and Finance*. Additionally, we decided to limit our search to the titles of the paper to limit the number of results.

The specific query used was:

```
TITLE ( ( cryptocurrency OR bitcoin ) AND price AND ( predict* OR forecast* )
    ) AND PUBYEAR < 2024 AND ( LIMIT-TO ( SUBJAREA , "COMP" ) OR LIMIT-TO (
    SUBJAREA , "ENGI" ) OR LIMIT-TO ( SUBJAREA , "DECI" ) OR LIMIT-TO (
    SUBJAREA , "MATH" ) OR LIMIT-TO ( SUBJAREA , "ECON" ) ) AND ( LIMIT-TO (
    LANGUAGE , "English" ) ) AND ( LIMIT-TO ( DOCTYPE , "cp" ) OR LIMIT-TO (
    DOCTYPE , "ar" ) )
```

The search produced 372 articles since 2017, an initial dataset that will be further refined in the next phase of our analysis.

Starting from this dataset of articles, we generated a csv table using an integrated tool in Scopus, which will be the basis for subsequent analyses. The fields of the table are:

- **Authors**: the authors of the paper;
- **Title**: the title of the paper;
- **Year**: year of publication;
- **Cited by**: number of citation;
- **Source Title**: name of the publishing journal;
- **DOI**: Digital Object Identifier of the paper;
- **Link**: URL of the article;
- **Abstract**: brief summary of the paper;
- **References**: references of the article.

The table was then imported into a Colab Notebook to automate the various survey phases as much as possible and extract useful information in histograms and pie charts.

Figure 2 depicts the number of articles published annually within the timeframe considered by our survey; the research trend is generally upward, albeit with irregular growth, partly due to the extreme volatility of Bitcoin and other cryptocurrency prices. The first paper is by Jang et al., and it dates back to 2017 [14]).

The distribution of categories among articles and conference papers is almost equally portioned, 49% of the articles being Conference papers, while the rest articles.

Figure 3 shows the distribution of citation numbers through the dataset of papers, from which we notice that the majority of documents, with few exceptions barely visible in the histogram, have a low number of citations.

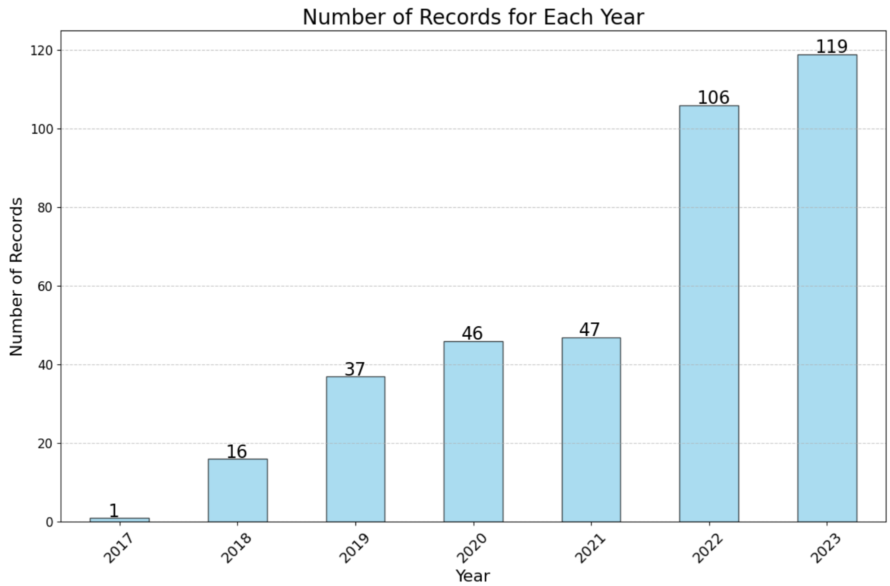


Figure 2: Number of papers in the initial database for each year in the time interval our survey considers.

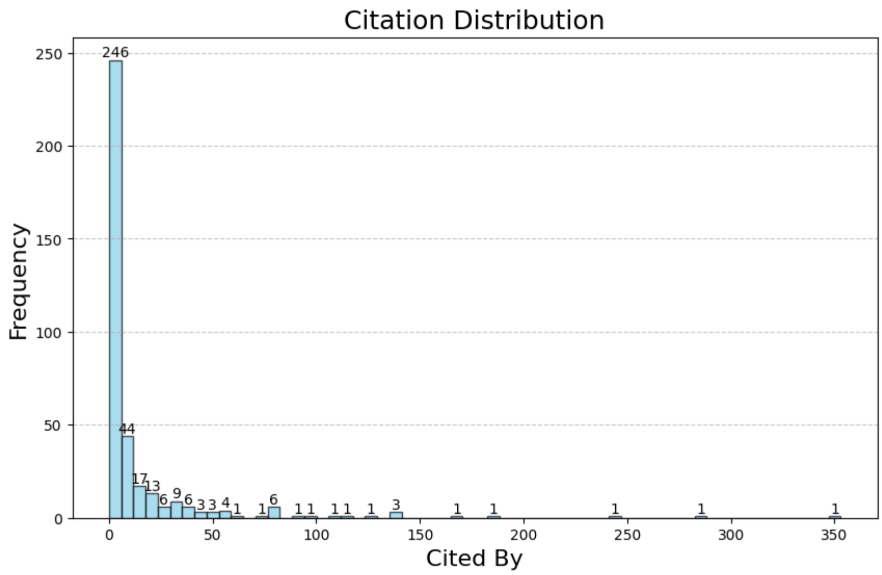


Figure 3: Distribution of the citations.

Figure 4 shows the distribution of categories among the different subjects, Computer Science being the top one, followed by Engineering and Mathematics.

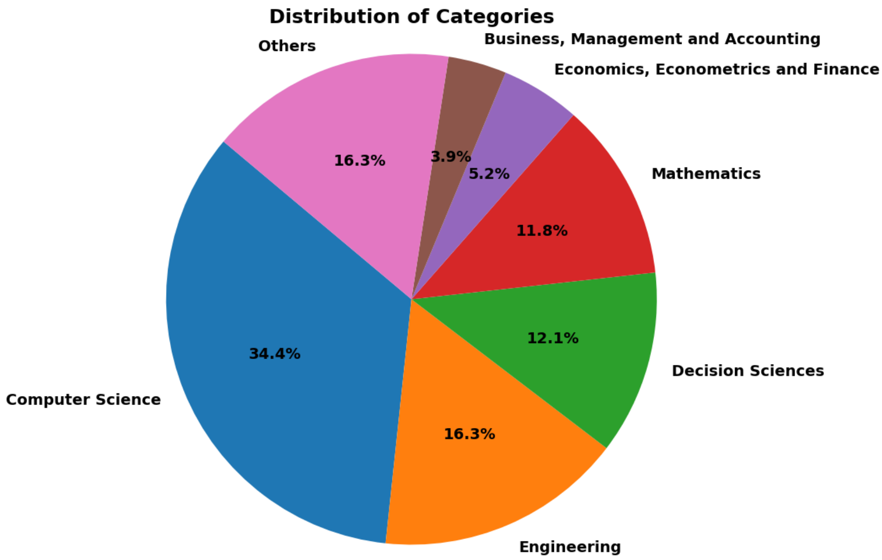


Figure 4: Distribution of subjects.

3.3. Inclusion criteria

Due to the large number of articles, we have decided to focus our more in-depth analyses on a subset of approximately 100 articles, aiming to consider the most relevant ones, especially those from the last three years, 2021-2023, the period most pertinent to our analysis.

We used the articles' citations, crafting a normalized index to consider mostly more recent and relevant articles, given by:

$$\mathrm{S} = \frac{\mathrm{C}}{(2024 - \mathrm{y})^2}$$

that is a score S proportional to the number of citations C (as counted 8 Feb 2024), and inversely proportional to the square of the difference between 2024 (now) and the year y of publication of the considered paper. In this way, we eliminated articles without citations from the second part of our analysis. We also excluded surveys that did not develop their model and papers that did not include BTC price as one of the study targets.

Figure 5 shows the number of selected and discarded articles year by year. We notice that, especially in 2022 and 2023, with our inclusion criteria, discarded articles are way more than the selected ones. In the future, we will consider more papers for the final dataset.

3.4. Extracted information

The analysis aims to compile and organize some of the data on Bitcoin price prediction in the literature. To do this, we created a second table by starting with the first and manually adding additional fields for each of the 102 articles chosen. The method takes time since it takes a lot of reading to obtain the necessary information from all this material.

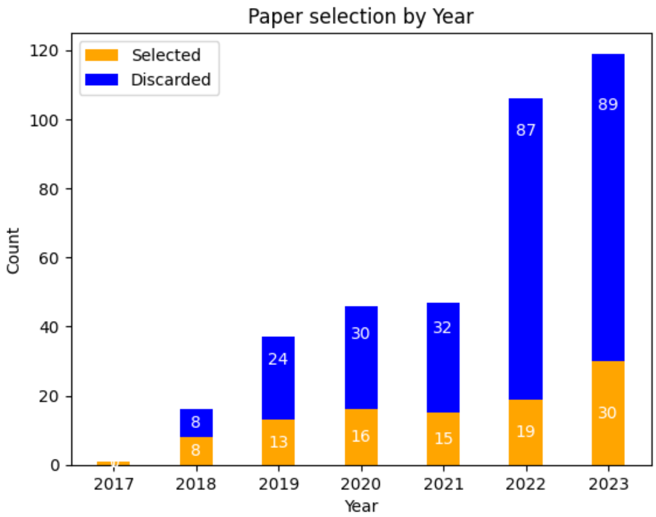


Figure 5: Number of selected and discarded papers to form the dataset for the final analyses.

The added fields (with the legenda), shown in tables 1 and 2 in the Appendix A, are:

- **Crypto**: objective (dependent) variables of the prediction (i.e., price of BTC, ETH, etc.);
- **Features**: independent variables (features) used for the prediction;
- **Models**: models (algorithms) used for the prediction;
- **Evaluation metrics**: metrics used to evaluate the performance of the models for Bitcoin price prediction;
- **Data range**: time interval analyzed in the considered paper;
- **Data resolution**: resolution(s) of the model (es. 1d, 1 min, etc.);
- **Source**: online source where historical data are retrieved.

Manually inserted data must be standardized to add records for later suitable statistical analysis intelligently. Appending a legend listing all the traits and models examined is crucial. Many articles utilize ensemble models, which are combinations, and in some cases hybrid (in the sense that they combine statistical methods with machine learning), of various models to obtain predictions. However, this possibility is not considered in our analysis as we currently aim only to understand, through statistical analysis, the most commonly used models without considering the various hybrid combinations. In this case, the configurations would be numerous, and the analysis would fail to extrapolate almost any regularity.

4. Results

In the final stage of the review, we try to answer the research questions by extracting the desired information from the definitive set comprising approximately 100 papers described at the end of the previous section. We imported this second refined dataset into the Colab notebook, where all the survey steps are implemented and automated. This allowed us to import our dataset,

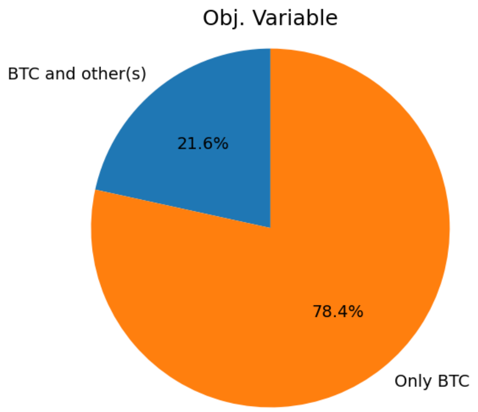
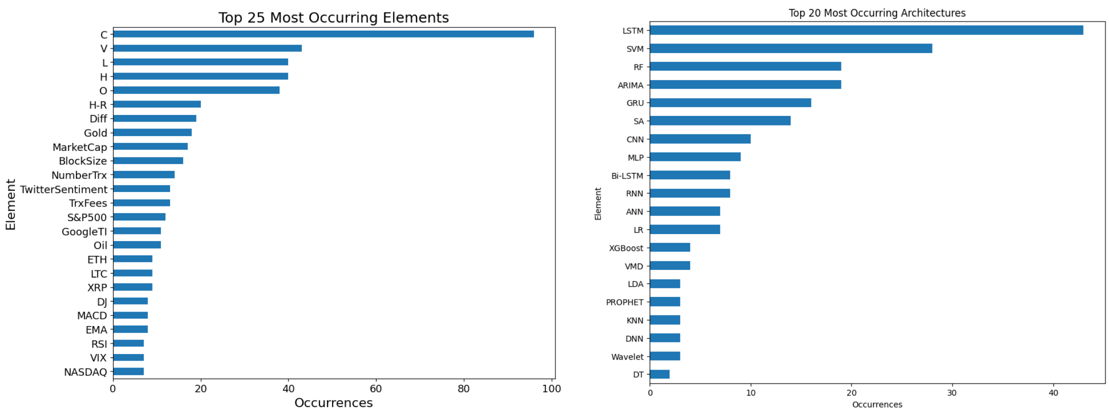


Figure 6: Distribution of the objective of the price prediction (dependent variable).

initially in CSV format, convert it into a pandas data frame, and interact with it more efficiently, given the numerous tools available in this Python environment. We then created various graphs.

Figure 6 shows the percentage of articles predicting only Bitcoin and those predicting Bitcoin and other cryptocurrencies. Keeping in mind that we discarded papers that did not predict the price of Bitcoin, it shows that the vast majority of the considered papers only predict the price of BTC. At the same time, the rest also predict other crypto prices, like ETH, LTC, and XRP, that is primarily cryptocurrencies with high capitalization.



(a) Occurrences of predictive features.

(b) Occurrenceses of predictive models.

Figure 7: Occurrences of predictive features and models architecture.

Figure 7a displays the list and occurrences of the 25 most frequently used features. We notice, as expected, that the most utilized features are the candle data (OHCL) and the volume of transactions V [15], followed by Market Capitalization and some on-chain features like Hash Rate and Difficulty [16]. Also, macroeconomic indexes like S&P500, the price of metals like gold or other materials [17], the price of other cryptocurrencies (ETH, LTC, etc., [18]), and

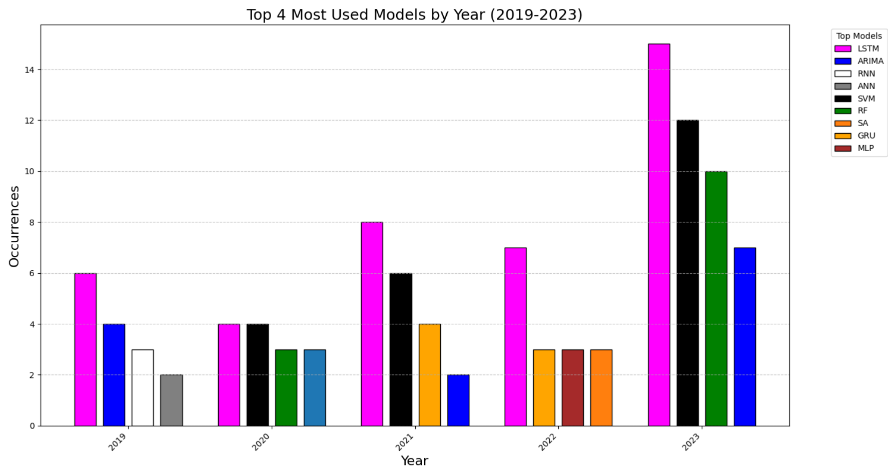
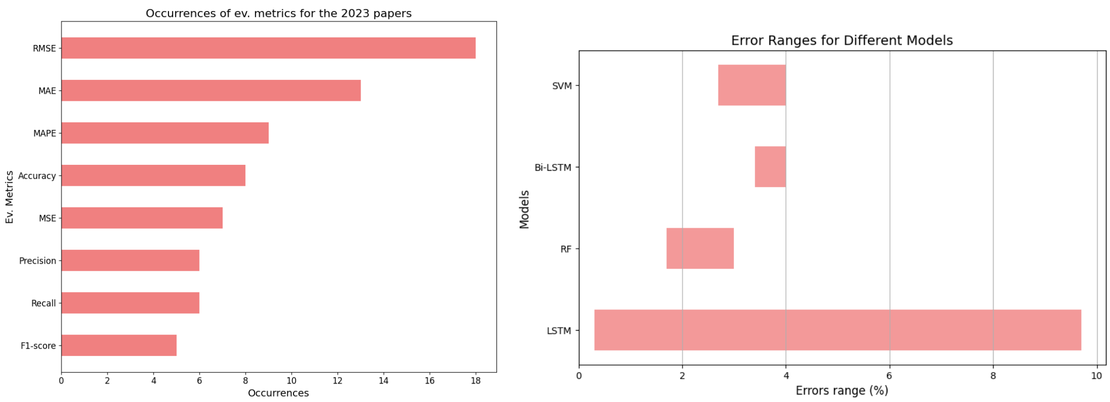


Figure 8: Evolution of the top 4 most used architectures throughout time.

many price-related technical indicators like MACD and RSI are used for the forecasting of Bitcoin price [19]. In the works that utilize Sentiment Analysis (SA), Google Trends Index and Twitter Sentiment Score are mostly considered [20]. Figure 7b shows the occurrences of the top prediction model architectures. Deep learning models based on LSTM and GRU are the most widely used [21, 22], followed by the machine learning techniques, SVM and RF [23], and by the ARIMA time series forecasting statistical method [24].

Figure 8 shows the evolution of the top 4 most used architectures through time. We notice that LSTMs are always at the first position and that machine learning algorithms, like RF and SVM, have re-gained prominent positions in 2023. The same applies to the ARIMA statistical forecasting method, often used as a benchmarking model.



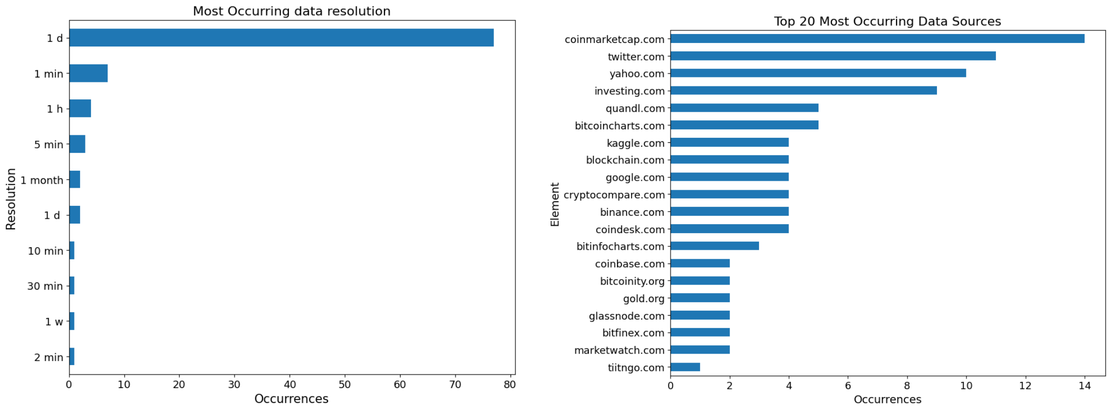
(a) Most Occurring evaluation metrics.

(b) Error ranges of the models.

Figure 9: Figure 9a shows the occurrences of the evaluation metrics used in 2023 papers, while figure 9b shows the error ranges of some regression models.

The metrics used to evaluate the performance of the price prediction depend on the paper's approach, that is, whether the forecasting is given in the form of regression (R) or classification (C). In the former case, metrics like MSE, RMSE, MAE, and MAPE are used, while in the latter, Accuracy, Precision, Recall, and F1-score are used. From figure 9a we can see that the RMSE is the most frequent used metric, suitable to regression models, followed by MSE and MAPE. For classification tasks, the most used metric is the Accuracy.

Figure 9b shows the error ranges for the models used in 2023 papers. It has to be pointed out that is very difficult to compare different models that in general give predictions for different dataset ranges and resolution, so we decided to give error ranges for the most current models to give an idea of their effectiveness.



(a) Most Occurring data resolution.

(b) Top 20 Most Occurring Data Sources

Figure 10: Most occurrent data resolutions and sources.

Figure 10a and Figure 10b show respectively the top occurrent data resolutions and sources. The most frequent forecasting time interval is one day, and the most popular site for retrieving financial data is coinmarketcap[4]. In contrast, for Sentiment Analysis, which uses textual datasets, the most common source site is X (i.e., Twitter[5]).

Figure 11 shows the dataset ranges' heat map. We notice that the datasets begin in 2009, the genesis year of blockchain technology and Bitcoin, and have a peak at the beginning of 2018, right after the first massive adoption phase, when the capitalization of the Crypto market reached for the first time the level of 1 T$.

5. Conclusion

The primary contribution of our paper lies in the recent analysis of articles about Bitcoin price prediction, encompassing both traditional statistical and machine learning techniques. Extracting insights from a dataset of articles published between 2017 and 2023, we observe the nascent stage of the cryptocurrency domain, evident from most papers emerging in the past three years. This observation suggests fertile ground for future research endeavors in this

[4]Coinmarketcap: https://coinmarketcap.com
[5]Twitter: https://twitter.com/

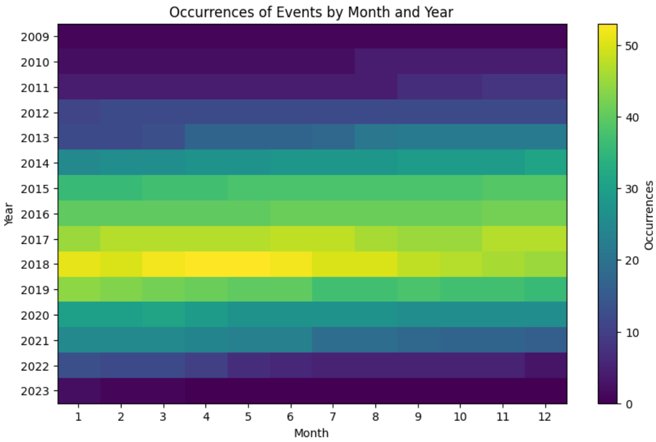


Figure 11: Heat-map of the datasets ranges.

domain. By consolidating relevant papers and analyzing their contributions, we aim to lay a foundational framework to support forthcoming research in Bitcoin price forecasting.

Despite the inherent challenges posed by the volatile nature of cryptocurrencies and their susceptibility to various internal and external factors, researchers have endeavored to develop robust prediction models. Time-series-based models, such as ARIMA and other regression-based approaches, have traditionally been employed for price prediction due to their reliance on historical price data. With the advent of deep learning algorithms, researchers have integrated these advanced techniques into their methodologies, utilizing variants such as RF, SVM, LSTM, GRU, and hybrid or ensemble models. These efforts signify a continuous evolution in pursuing more accurate and reliable cryptocurrency price forecasts.

In the future, we plan to leverage the information gathered from the present survey to define and assemble a multimodal (financial and textual) dataset and design and implement our portfolio of efficient predictors. The final goal is to build an advanced Automatic Trading System to optimize the management of Crypto assets and support short—and long-term investments.

Acknowledgments

The authors are members of the Gruppo Nazionale Calcolo Scientifico-Istituto Nazionale di Alta Matematica (GNCS-INdAM). This work has been partially supported by: GNCS-INdAM, CUP_E53C22001930001 and CUP_E53C23001670001;European Union - Next Generation EU PNRR MUR PRIN - Project J53D23007220006 EPICA: “Empowering Public Interest Communication with Argumentation”; University of Perugia - Fondo Ricerca di Ateneo (2020, 2021, 2022) - Projects BLOCKCHAIN4FOODCHAIN, FICO, AIDMIX, “Civil Safety and Security for Society”; European Union - Next Generation EU NRRP-MUR - Project J97G22000170005 VITALITY: “Innovation, digitalisation and sustainability for the diffused economy in Central Italy”; Piano di Sviluppo e Coesione del Ministero della Salute 2014-2020 - Project I83C22001350001 LIFE: “the itaLian system Wide Frailty nEtwork” (Linea di azione 2.1 “Creazione di una rete nazionale per

le malattie ad alto impatto" - Traiettoria 2 "E-Health, diagnostica avanzata, medical devices e mini invasività").

References

[1] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, Decentralized business review (2008).

[2] J. Chen, Analysis of bitcoin price prediction using machine learning, J. Risk Financial Management (2023).

[3] F. Glaser, K. Zimmermann, M. Haferkorn, M. Weber, M. Siering, Bitcoin - asset or currency? revealing users' hidden intentions, ECIS 2014 Proceedings - 22nd European Conference on Information Systems (2014).

[4] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman, Systematic literature reviews in software engineering – a systematic literature review, Information and Software Technology 51 (2009) 7–15. URL: https://www.sciencedirect.com/science/article/pii/S0950584908001390. doi:https://doi.org/10.1016/j.infsof.2008.09.009, special Section - Most Cited Articles in 2002 and Regular Research Papers.

[5] P. Jaquart, D. Dann, C. Martin, Machine learning for bitcoin pricing - A structured literature review, in: N. Gronau, M. Heine, H. Krasnova, K. Poustcchi (Eds.), Entwicklungen, Chancen und Herausforderungen der Digitalisierung: Proceedings der 15. Internationalen Tagung Wirtschaftsinformatik, WI 2020, Potsdam, Germany, March 9-11, 2020. Zentrale Tracks, GITO Verlag, 2020, pp. 174–188.

[6] A. M. Khedr, I. Arif, M. El-Bannany, S. M. Alhashmi, M. Sreedharan, Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey, Intelligent Systems in Accounting, Finance and Management 28 (2021) 3–34.

[7] N. P. Patel, R. Parekh, N. Thakkar, R. Gupta, S. Tanwar, G. Sharma, I. E. Davidson, R. Sharma, Fusion in cryptocurrency price prediction: A decade survey on recent advancements, architecture, and potential future directions, IEEE Access 10 (2022) 34511–34538.

[8] Y. Mezquita, A. B. Gil-González, J. Prieto, J. M. Corchado, Cryptocurrencies and price prediction: A survey, Lecture Notes in Networks and Systems 320 LNNS (2022) 339 – 346. Cited by: 9.

[9] F. Fang, C. Ventre, M. Basios, L. Kanthan, D. Martinez-Rego, F. Wu, L. Li, Cryptocurrency trading: a comprehensive survey, Financial Innovation 8 (2022). Cited by: 121; All Open Access, Gold Open Access, Green Open Access.

[10] J. Zhang, K. Cai, J. Wen, A survey of deep learning applications in cryptocurrency, iScience 27 (2024) 108509.

[11] J. Webster, R. T. Watson, Analyzing the past to prepare for the future: Writing a literature review, MIS quarterly (2002) xiii–xxiii.

[12] M. Visser, N. J. Van Eck, L. Waltman, Large-scale comparison of bibliographic data sources: Scopus, web of science, dimensions, crossref, and microsoft academic, Quantitative science studies 2 (2021) 20–41.

[13] P. Wouters, M. Thelwall, K. Kousha, L. Waltman, S. De Rijcke, A. Rushforth, T. Franssen,

P. Wouters, et al., The metric tide, Literature review. Supplementary report I to the independent review of the role of metrics in research assessment and management (2015).
[14] H. Jang, J. Lee, An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information, IEEE Access 6 (2018) 5427–5437.
[15] M. J. Hamayel, A. Y. Owda, A novel cryptocurrency price prediction model using gru, lstm and bi-lstm machine learning algorithms, AI (2021).
[16] S. Ranjan, P. Kayal, M. Saraf, Bitcoin price prediction: A machine learning sample dimension approach, Computational Economics (2023).
[17] J. W. Goodell, S. Ben Jabeur, F. Saâdaoui, M. A. Nasir, Explainable artificial intelligence modeling to forecast bitcoin prices, International Review of Financial Analysis 88 (2023) 102702.
[18] Y. Zhu, J. Ma, F. Gu, J. Wang, Z. Li, Y. Zhang, J. Xu, Y. Li, Y. Wang, X. Yang, Price prediction of bitcoin based on adaptive feature selection and model optimization, Mathematics (2023).
[19] L. Cocco, R. Tonelli, M. Marchesi, Predictions of bitcoin prices through machine learning based frameworks, PeerJ Computer Science (2021).
[20] Z. Zhou, Z. Song, H. Xiao, T. Ren, Multi-source data driven cryptocurrency price movement prediction and portfolio optimization, Expert Systems with Applications 219 (2023) 119600.
[21] V. Kumar T, S. Santhi, K. G. Shanthi, G. M, Cryptocurrency price prediction using lstm and recurrent neural networks, in: 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2023, pp. 1–5.
[22] A. Dutta, S. Kumar, M. Basu, A gated recurrent unit approach to bitcoin price prediction, Juornal of Risk and Financial Management (2020).
[23] N. Maleki, A. Nikoubin, M. Rabbani, Y. Zeinali, Bitcoin price prediction based on other cryptocurrencies using machine learning and time series analysis, Scientia Iranica (2020).
[24] M. Poongodi, V. Vijayakumar, N. Chilamkurti, Bitcoin price prediction using arima model, International Journal of Internet Technology and Secured Transactions 10 (2020) 396–406.
[25] A. A. Oyedele, A. O. Ajayi, L. O. Oyedele, S. A. Bello, K. O. Jimoh, Performance evaluation of deep learning and boosted trees for cryptocurrency closing price prediction, Expert Systems with Applications 213 (2023) 119233.
[26] X. Sun, M. Liu, Z. Sima, A novel cryptocurrency price trend forecasting model based on lightgbm, Finance Research Letters 32 (2020) 101084.
[27] P. L. Seabe, C. R. B. Moutsinga, E. Pindza, Forecasting cryptocurrency prices using lstm, gru, and bi-directional lstm: A deep learning approach, Fractal and Fractional (2023).
[28] Z. Chen, C. Li, W. Sun, Bitcoin price prediction using machine learning: An approach to sample dimension engineering, Journal of Computational and Applied Mathematics 365 (2020) 112395.
[29] S. McNally, J. Roche, S. Caton, Predicting the price of bitcoin using machine learning, in: 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), 2018, pp. 339–343.

A. Tables and legenda

Paper	Prediction Obj.	Features	Models	Ev. Metrics
[25]	BTC, ETH, BNB, LTC, XLM, DOGE	O, H, L, C, V, SMA, EMA, WP	ADAB, GBM, XGB, MLP, GRU, CNN	MAPE = 1-3%
[26]	BTC, 41 other cryptos	DJ, S&P500, HangSengI, USDFI, ShanghaiStockI, ShenzhenCompI, RMB, FTSEChinaA50, Oil, O, H, L, C, V	LightGBM, SVM, RF	Accuracy = 48-62%
[27]	BTC, ETH, LTC	O, H, L, C, AdjC	LSTM, GRU, Bi-LSTM	RMSE = 1031-1280; MAPE = 0,036-0,057
[28]	BTC	BlockSize, H-R, Diff, NumberTrx, ConfTrx, MempoolTrxCount, MempoolSize, MarketCap, EstTrxValue, TrxFees, GoogleTI, Gold	LSTM, SVM, RF, XGB, LR	Accuracy = 48-66%, Precision =49-72%, Recall = 35-74%, F1-score = 48-78%
[29]	BTC	O, H, L, C, Diff, H-R	LSTM, RNN, ARIMA	Accuracy=50-52%, RMSE=5-53%

Table 1
Part 1: First five columns

Dataset range	Dataset resolution	Source
01/01/2018 - 02/03/2022	1 d	yahoo.com, bitfinex.com, investing.com
01/01/2018 - 30/06/2018	1 d	investing.com
01/01/2018 - 01/01/2023	1 d	yahoo.com
02/02/2017 - 01/02/2019	1 d, 5 min	coinmarketcap.com
19/08/2013 - 19/07/2016	1 d	coindesk.com, blockchain.info

Table 2
Part 2: Last two columns.

O: open price of the considered interval; **H**: max price; **L**: min price; **C**: closing price; **V**: volume of transactions; **H-R**: Hash rate; **Diff**: mining difficulty; **NumberTrx**: number of transactions; **ConfTrx**: confirmed transactions; **MempoolTrxCount**: Mempool transaction count; **TrxFees**: transaction fees; **GoogleTI**: Google Trends Index; **DJ**: Dow Jones Index; **MACD**: Moving Average Convergence Divergence; **RSI**: Relative Strength Index.