



Article

# The Impact of Academic Publications over the Last Decade on Historical Bitcoin Prices Using Generative Models

Adela Bâra<sup>1</sup> and Simona-Vasilica Oprea<sup>2,\*</sup>

<sup>1</sup> Academy of Romanian Scientists, Ilfov 3, 050044 Bucharest, Romania; bara.adela@ie.ase.ro

<sup>2</sup> Department of Economic Informatics and Cybernetics, Bucharest University of Economic Studies, No. 6 Piața Romană, 010374 Bucharest, Romania

\* Correspondence: simona.oprea@csie.ase.ro; Tel.: +40-752-29-44-22

**Abstract:** Since 2012, researchers have explored various factors influencing Bitcoin prices. Up until the end of July 2023, more than 9100 research papers on cryptocurrencies were published and indexed in the Web of Science Clarivate platform. The objective of this paper is to analyze the impact of publications on Bitcoin prices. This study aims to uncover significant themes within these research articles, focusing on cryptocurrencies in general and Bitcoin specifically. The research employs latent Dirichlet allocation to identify key topics from the unstructured abstracts. To determine the optimal number of topics, perplexity and topic coherence metrics are calculated. Additionally, the abstracts are processed using BERT-transformers and Word2Vec and their potential to predict Bitcoin prices is assessed. Based on the results, while the research helps in understanding cryptocurrencies, the potential of academic publications to influence Bitcoin prices is not significant, demonstrating a weak connection. In other words, the movements of Bitcoin prices are not influenced by the scientific writing in this specific field. The primary topics emerging from the analysis are the blockchain, market dynamics, transactions, pricing trends, network security, and the mining process. These findings suggest that future research should pay closer attention to issues like the energy demands and environmental impacts of mining, anti-money laundering measures, and behavioral aspects related to cryptocurrencies.



**Citation:** Bâra, A.; Oprea, S.-V. The Impact of Academic Publications over the Last Decade on Historical Bitcoin Prices Using Generative Models. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 538–560. <https://doi.org/10.3390/jtaer19010029>

Academic Editors: Albert Y.S. Lam and Yanhui Geng

Received: 24 January 2024

Revised: 24 February 2024

Accepted: 2 March 2024

Published: 6 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** cryptocurrency; research publication; topic modelling; latent Dirichlet allocation; sentiment analysis; Bitcoin prices

## 1. Introduction

The emergence of writing and money played a very important role in human history. They were tools that conveyed knowledge and intensified trade using a unified value that made it easier to exchange goods and services between people. The evolution of money probably began with clay inscriptions and continued with coined money, which are universal trading instruments built on people's trust. Blockchain and cryptocurrencies emerged after 2008 when Lehman Brothers went bankrupt, shocking financial markets and shaking confidence in banks. They represented, according to what Satoshi claimed, "the absolute autonomy of money" [1] and a response to the financial crisis. Digital technology's progress has shaped the performance and standards, imposing new payment methods and new money accepted by individuals, organizations and communities. Today, Bitcoin is the most popular virtual coin that has grabbed attention due to its spectacular price volatility. Out of the 22,900 crypto coins (with 8800 active), as of August 2023, Bitcoin's dominance had reached 49%. Most of its fluctuations were allegedly attributed to public opinion, influencers and social networks.

Bitcoin was born along with blockchain technology and while blockchain is perceived as beneficial to many fields, Bitcoin is highly controversial [2–4]. The high carbon footprint and energy consumption required by mining processes have led many researchers to call

it dirty [5–7]. It was estimated that Bitcoin mining is roughly responsible for 65.4 MtCO<sub>2</sub> yearly, which is comparable to Greece's emissions. Moreover, it was reported that the carbon intensity of Bitcoin mining increased by 17% in 2021 compared to 2020. In May 2021, specialized hardware devices (servers) worldwide consumed approximately 13 GW for mining processes. Inner Mongolia was one of the first Chinese provinces which banned crypto mining in March 2021 due to environmental concerns. Until June 2021, crypto mining bans were imposed in Sichuan and Xinjiang as well. In November 2021, Sweden also called for a ban on cryptocurrency mining as the usage of renewable energy to supply crypto mining would delay the energy transition towards a green environment [8]. Its price fluctuations have been remarkable, leading investors to rate it as a high-risk, less predictable investment tool. Security, privacy, money laundering and payments to the dark market are issues that have to be further investigated and tackled [9–12]. Aiming to find out the status quo of blockchain entrepreneurial innovations and research and how can these innovations contribute to the economic and sustainable development of society, these authors approached a combined method using web mining and extracting topics from unstructured data on the blockchain from Crunchbase, similar websites and research publications [13]. The analysis indicated that blockchain innovations have gone beyond the financial sector, addressing critical infrastructures such as the supply chain and energy delivery. More potential is envisioned in several promising application fields such as energy communities, renewable energy sources, digital twins, sustainable energy, voting, traceability or integrity verification. A blockchain solution was envisioned to solve the local energy trading aspects [14]. Furthermore, a conceptual architecture design of a blockchain solution for e-voting in elections at the university level was proposed in [15]. The power of a blockchain-based supply chain was further examined in [16].

Moreover, the way Bitcoin and other cryptocurrencies are perceived depends largely on the cultural perspective and ideological nuances at the regional or even at the country level. How cryptocurrencies are perceived in Japan and Sweden was analyzed in [17], examining attitudes and discourses on social media. These discourses focused on the role of blockchain technology in climate change, its environmental impact, the legitimate use of cryptocurrencies, politicians' tax evasion, payments to the dark markets or other ethical aspects. By using topic modelling in relation to the social media posts on cryptocurrencies, the authors studied the peoples' attitudes towards cryptocurrencies in these two countries, discerning advantages, opportunities and ethical challenges that were related to cryptocurrencies. The authors demonstrated that autonomy and money are well connected. They interpreted cryptocurrencies as a monetary innovation brought by society's development of enhanced freedom, trading and autonomy. Like in ancient Greece, cryptocurrencies can be a trading opportunity for over two billion unbanked poor people that are not part of the financial system. Trade would offer them an opportunity to overcome poverty and precarious living conditions. The results of the analysis showed that the perception of cryptocurrencies highly depends on the trust in public and private institutions. Although both Japan and Sweden have a high degree of trust in institutions, Sweden tends to value the principle of autonomy more compared to Japan, who seems to be more pragmatic, as shown in the online forums. In [18], critics argued that cryptocurrency mining wastes a lot of energy, while supporters claimed it is environmentally friendly. This study introduces mining domestic production (MDP) to evaluate Bitcoin mining's output and carbon emissions in China, comparing it to three traditional industries. The findings revealed that Bitcoin mining's energy efficiency is not always the worst. The paper offers a fresh viewpoint on assessing Bitcoin mining's profitability against carbon emissions, relative to other sectors. Additionally, it suggests that Bitcoin mining might help developing countries expand their electrical infrastructure and earn revenue. From the economic and sentiment analysis perspectives, several research papers are analyzed [19]. The differential influence of social media sentiment on cryptocurrency returns and price volatility during the COVID-19 pandemic were investigated in [20]. Whether cryptocurrencies provide a viable hedging mechanism for benchmark index investors was further studied in [21]. The

non-linear causal linkages of EPU and gold with major cryptocurrencies during bull and bear markets were examined [22], whereas herding behavior and price convergence clubs in cryptocurrencies during bull and bear markets were investigated in [23].

Although more than 9100 papers about Bitcoin and cryptocurrencies were written and indexed in Web of Science Clarivate platform, the relation between academic publications over the past 10 years and Bitcoin prices has yet to be examined. The scope of this paper is to reveal the main topics that concern researchers in relation to high-risk, less predictable investment instruments. By extracting the relevant topics, we aim to find gaps and potential new topics that have been less approached by the academic community. To identify topics, latent Dirichlet allocation (LDA) is employed. Furthermore, the document-topic density ( $\alpha$ ), word-topic density ( $\beta$ ) and the optimal number of topics ( $K$ ) are investigated. Equally interesting is the discovery of the sentiment of this community in relation to Bitcoin/cryptocurrencies and its effect on prices. Although sentiment analysis has been extensively performed on tweets, forums, other social media posts and reports [24], the sentiment of research publications has not been investigated yet. Additionally, it is intriguing to observe whether there are correlations between academic opinions and Bitcoin prices or connections between thousands of abstracts and prices.

This paper aims to answer the following three questions:

*(RQ1) What are the relevant topics investigated by researchers in relation to cryptocurrencies and potential gaps?*

*(RQ2) What is the academic community's sentiment in relation to Bitcoin/cryptocurrencies?*

*(RQ3) Is there a connection between academic publications and Bitcoin's price evolution?*

The contribution of this paper lies in its comprehensive approach to understanding the academic landscape surrounding Bitcoin and cryptocurrencies. The key contributions are as follows:

- (a) An extensive review of cryptocurrency research and identification of key themes: The study analyses over 9100 research papers on cryptocurrencies, providing a broad overview of the academic focus in this field since 2012. This extensive review helps in identifying the predominant themes and trends in cryptocurrency research. Employing LDA to analyze unstructured abstracts, the study identifies key topics within the vast body of cryptocurrency literature. This approach offers a structured way to understand the main areas of academic interest and research. By calculating perplexity and topic coherence metrics, the research determines the optimal number of topics, adding rigor to the topic modelling process and ensuring the relevance and clarity of the topics identified;
- (b) The use of advanced NLP techniques: The application of BERT-transformers and Word2Vec to process abstracts for predicting Bitcoin prices represents an innovative use of NLP in financial analysis;
- (c) An investigation of the academic influence on Bitcoin prices: The study explores a relatively unexamined area—the impact of academic publications on Bitcoin prices. The findings indicate a weak connection, providing insights into the influence of academic research on market dynamics;
- (d) The sentiment analysis of academic publications: Unlike prior studies that focused on the sentiment analysis of social media, this study extends the analysis to academic writings, offering a new perspective on how academic sentiment correlates with Bitcoin prices;
- (e) Highlighting research gaps and future directions: By extracting relevant topics and sentiments, the study identifies gaps and less-explored areas in cryptocurrency research, suggesting future directions, such as the impact of mining on energy and the environment, anti-money laundering measures, and behavioral aspects of cryptocurrency use. Overall, this paper makes a significant contribution by blending topic modelling, sentiment analysis, and financial prediction techniques to provide a multifaceted view of the academic discourse on cryptocurrencies and its potential impact on market behavior.

The novelty in this paper lies in several key areas: (1) The unexplored relationship between academic publications and Bitcoin prices: While there has been extensive research on various factors influencing Bitcoin prices, this study is novel in its examination of the direct relationship between academic publications and Bitcoin price movements. It explores whether scientific writing in this domain has any significant impact on the market; (2) The use of advanced analytical NLP techniques: The employment of LDA to analyze unstructured abstracts from over 9100 research papers is a novel approach. This method, combined with perplexity and topic coherence metrics, is used to determine the optimal number of topics; (3) BERT-transformers and Word2Vec for price prediction: The study innovatively uses BERT-transformers and Word2Vec to vectorize (process) abstracts and assess their potential in predicting Bitcoin prices. This application of advanced NLP technologies in the financial analysis of cryptocurrencies represents a novel approach; (4) The sentiment analysis of the academic community: While sentiment analysis has been widely conducted on social media platforms, this study breaks new ground by focusing on the sentiment within academic research publications and its potential effect on Bitcoin prices; (5) The identification of research gaps and emerging topics: By extracting and analyzing topics from a vast number of research papers, the study identifies gaps and new topics that have been less explored by the academic community, offering directions for future research; (6) A comprehensive scope of analysis: The paper's aim to answer three research questions about relevant topics, academic sentiment, and the connection between research and Bitcoin prices covers a comprehensive scope that has not been collectively explored before in the context of cryptocurrency research. Thus, the novelty of this study is its comprehensive and multidimensional analysis of the relationship between academic research and the Bitcoin market.

The current research is structured in five sections: in the Section 1 a brief introduction regarding our purpose, motivation and research questions is presented. Furthermore, we expose three research questions that will be answered in the Section 5. In Section 2, the most relevant previous publications are summarized indicating the main findings. The research methodology is described in Section 3, and it is focused on the LDA technique implementation to extract the relevant topics from the abstracts. Results are presented in Section 4 and conclusions in Section 5.

## 2. Literature Review

One of the input datasets included in this paper consisted of the list of 9105 publications on Bitcoin and cryptocurrency extracted from the Web of Science Clarivate platform (<https://clarivate.com/>, accessed 30 July 2023). Therefore, we exhaustively searched the most recent publications related to the topic modelling, sentiment analysis, research publications, academic community opinion and prediction related to the cryptocurrencies in general and Bitcoin in particular. Extracting news from media, social posts and even research content and the implementation of natural language processing (NLP) techniques have been extensively investigated over the past decade. Moreover, topic modelling was one of the NLP applications studied in several papers.

### 2.1. Topic Modelling

The relation between the international news and Bitcoin prices was analyzed in [25]. The news released from 2018 and 2020 were investigated using an LDA-based topic modelling approach. The text of 4218 cryptocurrency-related news articles from 60 countries were searched for significant topics, identifying 18 relevant topics related to security, financial issues, the market and economy. The results indicated that the news led to an impact of their content on the Bitcoin price variations' increasing volatility. After the publication, there was a negative impact on the next 24 h of prices.

Blockchain technology has been evolved together with cryptocurrencies and has found applicability in various fields [26]. It is one of the promoters of cryptocurrencies. Therefore, numerous research papers focused on investigating this technology. More

than 900 blockchain-related papers published by the IEEE, Springer and ACM publishing houses along with other databases were selected and their texts formed a corpus that was analyzed using LDA. The scope was to identify topics that require more investigation. The authors found 15 research directions or gaps that can be analyzed in relation to blockchain technology. Warehouse receipts are key in supply chain finance [27]. Traditional pledging methods are inefficient and insecure, leading to issues like duplicate pledges. To combat this, the authors introduce a Blockchain-based Digital Asset Platform (BDAP) with enhanced security and multi-party certification. Testing shows the BDAP's efficiency with an average response time of 1.441 s under a high user load. However, its adoption is slow, with limited bank participation, suggesting a gradual change in the industry's traditional mindset. Ref. [28] analyzed blockchain technology through six core layers: its application, contract, actuator, consensus, network, and data. It reviews the literature to understand the blockchain's global applications, finding China's focus on practical applications in industries and smart cities. In contrast, international research focuses on the blockchain in finance, integrating crypto assets with traditional industries like payment systems. The findings highlight the interplay between smart cities and cryptocurrencies, suggesting a mutually reinforcing relationship. Other applications of LDA and sentiment analysis were also found [29,30], revealing a broad spectrum of implementation.

Data from firms' reporting files are examined in [31], focusing on the blockchain technologies, cryptocurrencies and related applications. Both disclosures about blockchain and cryptocurrencies were analyzed using text-processing techniques. LDA was applied to extract relevant topics from the two types of disclosures. Five topics were underlined by the LDA: solutions related to blockchain implementation, factors that may bring risks, business models, and market-related services: payment and transactions. Furthermore, the two types of disclosures were analyzed from the relevance point of view and the authors found out that the blockchain solution and risk issues brought a positive value relevance, whereas disclosures regarding Bitcoin transactions and other cryptocurrencies brought a negative value relevance. Therefore, the firms positively valued blockchain applications, whereas cryptocurrencies and their related issues became less appreciated.

The literature related to blockchain technology published between 2013 and 2018 was investigated in [32], aiming to understand the status of blockchain research and plan the next research steps. An LDA was applied, revealing that most of the research was in the computer science and business research area. The main topics of the current research are related to technology architecture, financial applicability, privacy and security issues and smart applications. Moreover, the research envisioned three next stages: technology, business application and integration with IoT, AI and other technologies. Additionally, the authors discovered two strong connections between the blockchain and cryptocurrencies and between the blockchain and various applications. The software developers' posts from two blockchain-related Stack Exchange sites were investigated with an LDA to identify relevant topics [33] in an attempt to find the challenges that software developers encounter. The results showed that the posts regarding the blockchain increased in occurrence, while those regarding mining cryptocurrencies decreased. The authors considered that more documentation on blockchain development would support developers' communities that may lack supporting materials and guidelines. Moreover, there is an increasing interest in blockchain technology, spurred by Bitcoin's success and its expanding range of applications [34]. More studies emphasized the need for high-performance data interaction within blockchain systems. An overview of blockchain's development is provided, examining research efforts to enhance blockchain performance from on-chain, off-chain, and cross-chain interaction technology perspectives. Special attention is given to cross-chain interaction technologies, with the results showing that cross-chain technologies significantly enhanced blockchain performance.

The sentiment analysis, Bitcoin price and topic modelling are interconnected. Sentiment analysis provides insight into the emotional tone of discussions surrounding Bitcoin. Positive sentiment can attract more buyers and investors, potentially leading to price in-

creases, while negative sentiment can lead to price drops. Topic modelling helps identify the key subjects being discussed in relation to Bitcoin. By understanding the topics driving sentiment, investors can anticipate potential market-moving events and reactions. Both sentiment analysis and topic modelling contribute to the broader understanding of market dynamics, which can be leveraged by traders, investors, and analysts to make informed decisions about Bitcoin. It is important to note that while these methods can provide valuable insights, the cryptocurrency market is complex and influenced by a wide range of factors. Therefore, a comprehensive approach to analysis, including technical, fundamental, and sentiment-based insights, is often used to make informed decisions in this volatile market.

## 2.2. Sentiment Analysis

Twitter sentiment was analyzed to find whether the price of Bitcoin would increase or decrease and its magnitude [35]. The authors examined emotions conveyed by the tweets and their volume. The challenge was to set the time window in which the emotion is transmitted and impacts the price and, thus, becomes a real predictor. The study managed to predict the Bitcoin price direction and magnitude with good accuracy. Moreover, the NFTs were monitored, and the prediction was extended to the NFTs, namely CryptoPunks, which is one of the most well-known collections on the market. Other authors investigated Elon Musk's activity on Twitter in January 2021 and its impact on Bitcoin prices [36]. The analysis involved sentiment analysis on tweets after January 2021 and data extracted from Binance to understand the relationship between tweets and price evolution. The results showed that the volume of tweets did significantly increase. The volume of tweets was strongly and positively correlated with Bitcoin's price evolution. However, the effect of the tweets' sentiment on price proved to be weak; therefore, sentiment was not included as a credible predictor. Moreover, the study aimed to understand the influence of social media influencers that may affect Bitcoin's evolution and have a significant impact on its price. A similar study analyzed the effects of the COVID-19 pandemic and public sentiment on the cryptocurrencies' prices. The authors implemented the exponential generalized autoregressive conditional heteroskedasticity model, a quantile regression and a sentiment analysis showing that cryptocurrencies were less predictable during the COVID-19 pandemic [37].

A light gradient boosting machine (LGBM) classifier was employed in [38]. The results showed that 78.06% and 94.03% of hourly and daily bullish market movements can be attributed to public tweets, whereas 83.08% and 94.60% of hourly and daily bearish market movements could be justified by public tweets. The authors examined the interval from September 2017 to September 2022, analyzing more than 28 million tweets containing Bitcoin as a keyword. Machine learning (ML) and NLP techniques were applied to vectorize text, calculate semantic distances, extract keywords and encode variables. Furthermore, ML and NLP were employed to extract insights from tweets and expert ratings obtaining the best results of assessing the sentiment with a support vector machine (SVM) classifier. The authors used 68,281 tweets from 57 initial coin offerings (ICO) from 4 industries (the cryptocurrency, platform, business services, and entertainment) [39].

The Bitcoin price direction was predicted using a linear discriminant analysis-based classifier and sentiment analysis [40]. Price information and crypto-related news headlines were collected to forecast the day-ahead price direction. The SVM outperformed the other forecasting algorithms. However, all of them show better accuracy when predicting an increase in the day-ahead price compared to forecasting a decrease in the price. The inclusion of news sentiment increased the accuracy by 0.585. News and information from social groups in financial markets were investigated in [41]. The authors analyzed the relationships between moods and price time series to predict prices. The sentiment moods were calculated using the probability distribution of the news and were vectorized using a BERT-based transformer language model. Recurrent neural networks were employed to forecast the prices. Additionally, 15,000 tweets on cryptocurrency were analyzed from the emotional point of view evaluating the emotion score for anger, anticipation, disgust, fear,

joy, sadness, surprise, trust and the two basic sentiments: negative and positive. A total of 53,077 sentiments were extracted from the 15,000 tweets. The results showed that the data sample consisted in positive sentiments that might lead to an increase in investment in the decentralized finance market. This study indicated how ML algorithms measured the emotions from tweets to understand the implications of cryptocurrencies' prices [42].

A few months of Twitter data before and after the COVID-19 pandemic were analyzed in [43], applying the latent semantic analysis and decomposition of singular values. Its scope was to analyze the public's opinion on Bitcoin and cryptocurrencies ex-ante and ex-post the COVID-19 pandemic and understand the relevant topics that led to negative emotions related to cryptocurrencies and trading. The aim was to underline and share the findings (topics). Furthermore, an analysis of the change in discussions on social media related to the Bitcoin price was presented in [44]. The authors proposed a Word2Vec vectorized topic model that identifies which topics on social networks had an impact on Bitcoin prices from 2017 to 2018 which were characterized as a higher volatile interval. The authors tested the words' change in frequency within four intervals and compared four word2vec models (combining a continuous bag of words, cbow, skip-gram, sg, Hierarchical Softmax, HS, and Negative Sampling, NEG) to evaluate their consistency and performance. Eight topics were identified when prices shifted to a decreasing slope: wallet, fork, transfer, posts, block size, confirmation, exchanges, and password; whereas three to five topics were identified when prices shifted to an increasing slope: startup, East Asia, ICO, competition, and lighting network.

Cryptocurrency ecosystems and social media environments were investigated in [45]. A Hawkes' model was employed together with other NLP techniques to conceive an empirical analysis to figure out the relationship between cryptocurrencies' prices and social media from January 2019 onwards, concentrating on the rise of the Ethereum and Bitcoin. The relationship between the shifts in cryptocurrencies' prices, sentiment and topic conversations on social media was investigated using the Hawkes' model. The results showed that some topics and sentiment in social media forego certain price movements. Topics such as trading, governments and exchange currency (Ethereum cryptocurrency) negatively impacted Ethereum and Bitcoin prices. Discussions related to investments led to price increases, whereas discussions related to decentralized and technological applications led to price decreases. Furthermore, 14.156 posts from the Reddit blockchain platform were analyzed and topics such as currency inflation, smart contracts auditing, crypto comparison (Bitcoin, Ethereum, Hyperledger Fabric, Cardano, and the comparison between Stellar and Ripple proved to be more frequent than others) and blockchain trends (which wallet is the best for storing tokens, most arguments concerning Stellar and Ethereum trends) were revealed in [46]. The authors explored the users' topics extracted from conversations about blockchain trends, technologies, platforms, smart contract development and auditing using topic modelling and the BERT-transformer model.

The influence on Bitcoin prices from popular social networks makes prices extremely volatile and speculative. However, the influences of users are not equal and a hypertext-induced topic selection algorithm was used to split the dataset into two groups based on the users' influence [47]. Topic modelling was employed to extract topics from the discourses of the two groups. Significant differences in opinions were identified. Additionally, the results indicated that the opinions of the leaders were not in line with the majority. Moreover, in research, there is an inertia between the emergence of technology and publications [48]. However, substantial research has been undertaken regarding blockchain and Bitcoin. Their potential development and impact as well as their positive roles have to be enhanced. Solutions and cross-disciplinary exchanges are also required when it comes to energy consumption and the carbon footprint. The next steps in terms of research concepts, applications and future exploration were envisioned. Analyzing the sentiments of users and extracting the topics from social media via large data streaming are essential to making timely decisions. The model proposed in [49] provided a dynamic and scalable topic modelling over data streaming and a sentiment analysis at the topic level. Posts related to

Ethereum, Bitcoin and Facebook from Twitter were investigated, and the results offered a large data-scale topic model on social media streaming and a sentiment analysis at the topic level.

### 2.3. Crypto Price Prediction

To explore the extensive scope of analysis in predicting Bitcoin prices is one of our objectives. Researchers considered a variety of factors to forecast Bitcoin prices with maximum accuracy. The year 2021 witnessed significant volatility, with rapid transitions between bull and bear markets. This period saw both substantial losses and gains, prompting researchers to identify the key variables and indices that were most influential on Bitcoin price movements. Contrary to predictions that the fluctuations seen in 2021 would not recur, by the end of 2023, there was a sudden surge in Bitcoin prices, indicating the onset of a new bull market. However, Bitcoin is still considered a high-risk investment instrument whose price is difficult to predict. Using long short-term memory (LSTM), 47 input variables for 10 months were trained to predict the Bitcoin price movements resulting in an error of 3.52% [50]. Considering the transaction data, namely the buy and sell orders from Coinbase, which is one of the major digital currency exchange platforms, the investors' sentiment was measured. Then, using a bootstrapped quantile regression approach, the authors mentioned a relevant relation between sentiment and Bitcoin prices [51]. Additionally, day-ahead Bitcoin price predictions using random forest regression and LSTM, including variables which have impact on the Bitcoin price prediction, were depicted in [52]. A three-year dataset was considered (from 2015 to 2018), three stock market indexes (NASDAQ, S&P500, DJI) from the US, oil prices, CO<sub>2</sub> certificates and ETH prices were proven to impact the Bitcoin prices. However, from 2018 onwards, JP225, which is a Japanese stock market index, and ETH prices increased in importance. The accuracy of the prediction model depended on the interval and the number of lags. The model with one lag provided the best accuracy. The relation between Bitcoin and US stock markets was further investigated in [53], estimating the Bitcoin capacity to forecast the US composite and sectoral stock indices using daily data from November 2017 to December 2021. The results showed that the Bitcoin price is a significant predictor for the US stock market volatility. The findings also indicated an inverse relation between Bitcoin prices and the US stock prices emphasizing the importance of tracking Bitcoin prices for both practitioners and policy makers.

The effects of the interest rate on the Bitcoin price were investigated in [54] using a structural vector autoregressive (SVAR) model. The input dataset was extracted from January 2012 to October 2022 considering the following variables: VIX, which is the Chicago Board Options Exchange's (CBOE) Volatility Index—a popular measure of the stock market expectation of volatility that is based on S&P500 index options, the interest rate spread, positive real interest rate and negative real interest rate, gold price, oil price and DXY, which is the U.S. Dollar Index that tracks the strength of the dollar against six currencies. DXY goes up when the U.S. dollar gains strength compared to other currencies. Based on the variance decomposition, the negative real interest rate shocks had a stronger impact than the positive real interest rate on the Bitcoin price. A long-run Bitcoin price prediction using regressive methods and ML methods, namely SVM, was investigated in [55] showing the Bitcoin trend to stock-to-flow and providing insights into Bitcoin's scarcity/abundance by comparing its current stock with the flow of new stock entering the market over a certain period of time and potential price trends. Both these approaches are data-driven. Addressing the challenge of identifying one-time change addresses in Bitcoin address clustering, an emerging issue in social computing, ref. [56] introduced an innovative clustering method. Traditional research in this area, often limited to specific transaction types, struggles with low recognition rates and high false positives. Their method, utilizing multi-conditional recognition, was tested with on-chain Bitcoin transaction data. It outperforms existing heuristics, identifying at least 12.3% more one-time change addresses.

As Bitcoin is one of the means of financing terrorism, the authors of [57] explored the Bitcoin prediction from an terrorism-based event perspective. They used a vector

autoregressive (VAR) model to examine the financing capacity of Bitcoin and risk hedging. The results showed that terrorist attacks/incidents and brutality have a significant impact on Bitcoin price. These aspects indicated that the investors were concerned about the number of casualties and the frequency of the attacks. The effect of Bitcoin price on terrorism was also analyzed, but the influence was not significant. Econometrics were further employed to predict Bitcoin volatility using two-component CARR, GARCH and CGARCH models [58]. To estimate the effectiveness of the investors' sentiment in predicting the prices, the authors of [59] used the Bitcoin Misery Index (BMI) that results from trading cryptocurrencies and eliminating individuals judgements. The BMI indicates buy and sell opportunities for Bitcoin investors. When the BMI falls below 27, it indicates a strong buy signal. The higher the BMI is, the more likely the Bitcoin price may fall. Furthermore, they employed a bagged support vector regression to predict the Bitcoin prices. The study spanned from March 2018 to May 2022 and the results indicated that the index improved the accuracy of the prediction model. Additionally, a feature selection method enhanced the results of the prediction model providing the prices for the next 30 days. The relation between the Bitcoin price volatility and market concentration was described in [60], revealing the supply side of the Bitcoin ecosystem. It was demonstrated that the bounded market concentration in pooled mining capped the Bitcoin price variation obtaining a robust prediction model.

Moreover, the researchers investigated the connectedness between Bitcoin price and the energy consumption for mining processes that may have an impact on Bitcoin prices [61]. The input dataset was extracted from January 2014 to July 2021 and the authors used an artificial neural network (ANN) to predict the Bitcoin prices taking into account the energy consumption. The study showed that there is a strong correlation between the prices and the consumed energy. ANNs are also used in [62] to predict the prices of Bitcoin. The dataset spanned from January 2018 to September 2018 and was extracted from coindesk. A combination of price-related and lagged features was created to improve the prediction capabilities of ANNs. The results indicated that the Fitnet network with a trainlm Matlab function and 30 hidden neurons outperformed the other ANNs. In [63], the unstructured data from financial news and deep learning (LSTM) were employed to predict the Bitcoin prices. This method outperformed the combinations between other ML algorithms and datasets that did not include financial news. The Facebook prophet model was proposed in [64] to predict the prices of cryptocurrencies. The results were compared to LSTM and ARIMA. The FB prophet proved to outperform LSTM and ARIMA models as these two models have drawbacks that make them incompatible with a performant cryptocurrency price prediction. In [65], the Bitcoin price was used as a predictor for the currency exchange rates prediction. The proposed prediction model consisted of an autoregressive distributed lag (ADL) providing good results for daily horizons.

The emergence of Bitcoin futures contracts, the COVID-19 pandemic and their impact on the Bitcoin price were analyzed in [66]. The results indicated that Bitcoin's future contracts positively impacted its returns, but it had no impact on price volatility. According to [66], the COVID-19 pandemic did not influence Bitcoin returns nor volatility. However, strong movements took place between Bitcoin prices and COVID-19-related events (such as lockdowns, viruses' evolution, the number of dead, vaccination stages, etc.). As numerous research studies struggled to identify the variables that significantly contribute to Bitcoin or other crypto prices and sometimes the results remained inconsistent, the authors of [67] analyzed a wide variety of variables that were considered in previous research using an extreme bounds analysis (EBA), which is a large-scale sensitivity analysis capable of analyzing model uncertainty aspects. The results obtained using EBA indicated that market supply and demand, policy uncertainty and public interest are robust features capable of explaining the price fluctuations. Furthermore, global macroeconomic and financial events may explain Bitcoin price movements.

The text synthesizes research findings across three conceptual frameworks central to understanding Bitcoin and cryptocurrency: topic modeling, sentiment analysis and price

prediction, drawing on a variety of data sources including academic publications, social media, and news articles.

In topic modeling, the aim is to decipher the thematic structure within a vast corpus of text data related to Bitcoin and cryptocurrencies. Researchers employ methods like LDA to parse content from diverse sources, uncovering prevalent topics such as blockchain technology's applications, financial issues, market dynamics, and security concerns. This approach reveals the broad spectrum of themes discussed in the context of a blockchain and cryptocurrency, pointing out specific areas that require further research and the impact of news and social media on market perceptions.

Sentiment analysis seeks to measure the emotional tone and public sentiment towards Bitcoin and cryptocurrencies based on online discourse. By analyzing social media posts, news headlines and other text-based data, researchers assess how sentiment influences Bitcoin prices. The application of learning techniques in sentiment analysis shows a clear link between public sentiment and Bitcoin price fluctuations, demonstrating how positive or negative sentiments can significantly affect market movements. The analysis extends to examining the impact of influential figures and global events, such as the COVID-19 pandemic, on market sentiment and cryptocurrency prices.

Price prediction focuses on forecasting Bitcoin price movements through various analytical methods, including statistical models, machine learning algorithms and deep learning frameworks. These models incorporate a wide range of variables, from market indices and sentiment data to transaction volumes and global economic indicators, striving to predict price changes with accuracy. Despite the challenges posed by the volatile and speculative nature of Bitcoin, some models have achieved notable success in forecasting price movements. Research in this area also explores the role of external factors, such as global events and technological advancements, in influencing Bitcoin prices.

While significant strides have been made in predictive analytics, the unpredictable nature of Bitcoin prices highlights the ongoing need for exploring new data sources, analytical techniques and theoretical frameworks. This continuous exploration aims to improve prediction accuracy and identify factors driving Bitcoin and other cryptocurrencies.

As noticed in the literature review, although the three subsections approach distinct themes, such as topic modelling, sentiment analysis and price prediction, they often mingle, and numerous researchers combine them to extract valuable insights for investors. As mentioned above, Bitcoin and cryptocurrencies are in the limelight of research and grab huge interest from both researchers and industry. However, one can notice that numerous studies were concerned about sentiment analysis, price prediction and its determinants and only a few studies were focused on LDA and top modelling approaches. To the best of our knowledge, no previous study has included unstructured data from research publication and its relationship with Bitcoin prices.

### 3. Methodology

The link between sentiment analysis, Bitcoin price, and topic modelling lies in their potential to provide insights into the dynamics of the cryptocurrency market and the factors that underly it. Sentiment analysis involves analyzing textual data, such as news articles, social media posts, and online discussions, to determine the overall sentiment or emotional tone of the content. In the context of the cryptocurrency market, sentiment analysis can help gauge market participants' attitudes, opinions, and emotions about Bitcoin and other cryptocurrencies. Positive sentiment might be associated with optimism and a belief in price appreciation, while negative sentiment could indicate concerns and potential price declines. By monitoring sentiment, traders and investors can gain insights into market sentiment shifts that might impact Bitcoin prices. The price of Bitcoin is determined by supply and demand dynamics within the market. It can be further influenced by a wide range of factors, including market sentiment, potential regulatory developments, macroeconomic trends, technological advancements, and more. Sentiment analysis can play a role in understanding how emotional factors can impact buying and selling decisions,

thereby affecting Bitcoin price movements. Positive sentiment might lead to increased demand and rising prices, while negative sentiment could lead to selling pressure and price drops.

In order to perform a price prediction, the input data in text format is processed using BERTs (Bidirectional Encoder Representations from Transformers) and Word2Vec that involve distinct processes due to their different architectures and purposes in NLP. A BERT is first pre-trained on a large corpus of text using tasks such as masked language modelling (MLM) and next sentence prediction (NSP). In MLM, random words in a sentence are masked for the model to predict, while in NSP, a BERT learns to predict if one sentence logically follows another. After pre-training, the BERT is fine-tuned for specific tasks like sentiment analysis or question answering. This involves training on a smaller, task-specific dataset, updating the entire model's weights but much quicker than pre-training. For inference, the input text is tokenized, and special tokens are added [68,69]. The BERT then generates a contextual representation of each token, useful for various downstream tasks. Word2Vec, on the other hand, can be trained using either the continuous bag of words (CBOW) model or the skip-gram model. CBOW predicts a word based on its context, while skip-gram does the opposite. Once trained, Word2Vec provides a vector for each word, capturing semantic meanings and relationships. These vectors can be used for text similarity, clustering, or as features in machine learning models [70,71]. Unlike the BERT, Word2Vec only provides word-level embeddings and does not consider a sentence-level context. The BERT and Word2Vec differ significantly. The BERT provides context-sensitive embeddings where the representation of a word changes based on its context, whereas Word2Vec offers static embeddings. The BERT is more complex and resource-intensive compared to the simpler and less computationally demanding Word2Vec. In terms of application, the BERT is versatile for a range of NLP tasks, especially those requiring an understanding of context or sentence structure, while Word2Vec is more limited to word-level analysis. The BERT and Word2Vec are both potent in the field of NLP and are suited for different tasks and challenges in language processing.

Topic modelling is a technique used to extract underlying topics or themes from a collection of documents. In the context of cryptocurrency, topic modelling can help identify the key subjects or trends that are being discussed in news articles, social media posts, and online forums. By analyzing these topics, traders and investors can gain a better understanding of the factors driving sentiment and influencing Bitcoin price movements. For example, topics related to regulatory news, crypto-crime events, environmental impacts, technological advancements, market adoption, ICO, or macroeconomic events can all impact sentiment and subsequently influence Bitcoin price.

Revealing relevant topics from large texts can be a complex task. On one hand, they show the mainstream, but on the other hand they also indicate the gaps in the research field. Latent Dirichlet allocation (LDA) helps to explain a large text dataset by identifying the unobserved groups or topics that are also known as latent. In the text, words form a large bag of words in which each word is allocated to a topic. LDA is applied to gain insights into the realm of unstructured data. Initially, the LDA was proposed in genetics by Pritchard et al. [72] and then in machine learning by David Blei, Andrew Ng and Michael Jordan in the *Journal of Machine Learning Research* in 2003 [73]. It was also refined and described by Jelodar et al. Research papers (from 2003 to 2016) related to topic modelling based on LDA were investigated to discover the research progress, trends, challenges and applications in various sciences [74].

LDA is a probabilistic technique that aims at identifying the main topics in large collections of text data. When LDA is employed, probabilities are calculated using the Bayesian method and the expectation maximization (EM) algorithm that is based on iterations in order to find the maximum likelihood. The EM performs (1) an expectation step (known as *E*) which generates a function *L* for the expectation of the log-likelihood and (2) a maximization step (known as *M*) which computes the parameters  $\omega$  that maximize the expected log-likelihood identified at the *E* step. These parameters are calculated to identify

the distribution of the latent variables. Having a set of observed variables  $X$ , a set of topics  $K$  and the model parameters  $\omega$ , the likelihood function  $L$ , the maximum likelihood estimate (MLE) is calculated by maximizing the marginal likelihood of  $X$ :

$$L(\omega; X) = p(X|\omega) = \int p(X|K, \omega)p(K|\omega)dK \tag{1}$$

where  $X$ —observed variables;

$K$ —the latent variables (topics);

$\omega$ —the model parameters;

$L$ —the likelihood function.

The EM algorithm iteratively alternates the two steps ( $E$  and  $M$ ). In the expectation step, the expected value  $Q(\omega|\omega^{(t)})$  is defined:

$$Q(\omega|\omega^{(t)}) = E_{K \sim p(\cdot|X, \omega^{(t)})} [\log p(X, K|\omega)] \tag{2}$$

In the  $M$  step, the parameters  $\omega$  that maximize the expected values are found:

$$\omega^{(t+1)} = \underset{\omega}{\operatorname{argmax}} Q(\omega|\omega^{(t)}) \tag{3}$$

$$\omega^{(t+1)} = \underset{\omega}{\operatorname{argmax}} E_{K \sim p(\cdot|X, \omega^{(t)})} [\log p(X, K|\omega)] \tag{4}$$

LDA identifies the latent topics in the text similar to the exploratory factor analysis (EFA) applied to numeric features. Moreover, a confirmatory factor analysis (CFA) takes latent variables and estimates whether the latent features via a measurement model explain the observed variables [75,76]. The LDA derives from the probabilistic latent semantic analysis (pLSA). While both methods are similar and require the user to specify the number of topics (similar to the K-means clustering method), LDA compared to pLSA offers a better disambiguation of words and a more accurate allocation of documents to latent topics.

The total probability of the LDA model is described in Equation (5):

$$P(W, Z, \omega, \varphi; \alpha, \beta) = \prod_{i=1}^K P(\varphi_i; \beta) \prod_{j=1}^A P(\omega_j; \alpha) \prod_{t=1}^N P(Z_{j,t}|\omega_j) P(W_{j,t}|\varphi Z_{j,t}) \tag{5}$$

$A$ —the number of abstracts;

$N$ —the number of words in a given abstract  $j$ ;

$K$ —the number of topics;

$\alpha$ —the parameter of the Dirichlet per-document topic distributions;

$\beta$ —the parameter of the Dirichlet per-topic word distribution;

$\omega$ —the  $\theta$   $i$  topic distribution for abstract  $j$ ;

$\varphi$ —the  $\varphi$   $k$  word distribution for topic  $i$ ;

$Z$ —the  $z$   $i$   $j$ —topic for the word  $t$  in abstract  $j$ ;

$W$ —a  $w$   $i$   $j$ —specific word.

The two types of analyses (the relevant topic extraction using LDA and the Bitcoin price prediction) are presented below following the data collecting, pre-processing, processing, analyzing and interpreting stages. The process of obtaining the relevant topics is performed in seven steps:

*I. Topic extraction using Latent Dirichlet Allocation*

Step 1: Input dataset WOS as dataframe1 (df1). Details regarding the data source are provided in Section 4;

Step 2: Pre-process the Abstract column from WOS (eliminating punctuation, lowercasing characters, removing stopwords, etc., using a pipeline);

Step 3: Perform an exploratory data analysis (EDA) on the pre-processed Abstract—WordCloud;

- Step 4: Create the dictionary and obtain the corpus that is a term document frequency;  
 Step 5: Set the number of topics, building the LDA model using the corpus and the number of topics;  
 Step 6: Calculate the coherence score and perplexity to find the optimal number of topics;  
 Step 7: Visualizing the topic using the inter-topic distance map and top 30 most salient terms.

#### II. Bitcoin price prediction

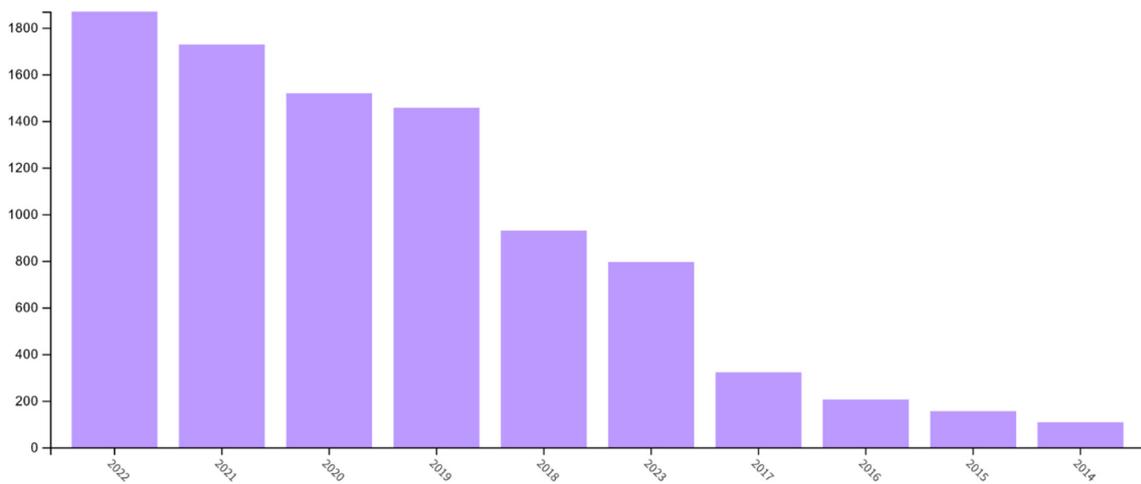
- Step 1: Input dataset WOS as dataframe1 (df1), input dataset Bitcoin prices as dataframe2 (df2). Details regarding the data source are provided in Section 4;  
 Step 2: Merge df1 and df2 using the Month and Year columns;  
 Step 3: Pre-process the Abstract column from WOS (eliminating punctuation, lowercasing characters, removing stopwords, etc., using a pipeline);  
 Step 4: Use TextBlob to analyze sentiment. Polarity is obtained from the pre-processed Abstract;  
 Step 5: Calculate and graphically visualize the percentages of positive, neutral and negative sentiments;  
 Step 6: The pre-processed Abstract column is vectorized using the BERT-transformer and Word2Vec (vector\_size = 100, window = 5, workers = 4, sg = 0), using the skip-gram algorithm;  
 Step 7: Apply several regressors (random forest, histogram gradient boosting, eXtreme Gradient Boosting, light gradient boosting, linear regression, and a voting regressor) to predict the Bitcoin price. Test the result for different time horizons. Calculate performance metrics (MAE, RMSE, MAPE) and graphically visualize the results.

Thus, in this paper, we process text data using a range of NLP techniques to draw out pertinent insights. To transform the text into numerical features, we employ advanced vectorization methods such as the BERT and Word2Vec. These methods enable us to generate numerical vectors, facilitating the training of various machine learning models, including random forest, histogram gradient boosting, eXtreme Gradient Boosting, light gradient boosting, linear regression, and a voting regressor. While there are alternative vectorization techniques available, such as TF-IDF (term frequency-inverse document frequency), one-hot encoding or n-grams, we found them to be less efficient and less suitable for handling large data volumes. Additionally, we experiment with different configurations of the LDA to achieve optimal outcomes.

## 4. Results

One of the datasets was extracted from the Web of Science Clarivate platform on 30 July 2023. Two words were included in the search: Bitcoin and cryptocurrency. The total number of research papers that met the searching criteria summed up to 9105 and the number of columns totalized 72 (such as: Publication Type, Journal Abbreviation, Publisher, Authors, Affiliations, Abstract, Keywords, Article Title, Cited Reference Count, Times Cited, WoS Core, Times Cited, All Databases, ISSN, eISSN, Publication Date, Volume, Issue, DOI, DOI Link, Number of Pages, WoS Categories, Web of science Index, Research Areas, Highly Cited Status, UT (Unique WOS ID), Year, etc.).

The abstract was analyzed and converted from text to features. A total of 357 articles were excluded from the list as their abstracts did not exist. Moreover, 3839 articles were without a publication date that was necessary to be converted into a valid month. Therefore, a random function between one and twelve was applied. A total of 2361 papers were in Computer Science Information Systems, 2001 in Computer Science Theory Methods, and 1568 in Business Finance. Researchers from the USA wrote 1800 papers, 1647 were from China and 818 were from England. A total of 5704 papers were in the articles category, 2911 were proceeding papers and 309 were in the early access category. The number of publications related to Bitcoin and other cryptocurrencies progressively increased over time (as in Figure 1).



**Figure 1.** The frequency of publications from 2014 to 2022 in descending order. Source: Web of Science.

A total of 2080 of papers were published by the IEEE publisher, 1740 by Elsevier and 1272 by Springer Nature. A total of 4058 publications were included in the Computer Science main domain, 2908 in Business Economics, while 1571 in Engineering.

The other dataset used in this research consists of the monthly prices of Bitcoin extracted from Investing.com (accessed on 23 June 2023). The Bitcoin prices started to increase in 2017 and then again in 2019, but the highest jump was recorded in 2021. The price curve in 2021 had two humps followed by a steep valley in 2022 down to almost 15,000 USD. Today, prices are half of what they used to be in 2021 when they went up to almost 70,000 USD (as in Figure 2).



**Figure 2.** Analyzing the price evolutions of Bitcoin. Source: authors.

A rough estimation of the topics included in the abstracts can be obtained rapidly using WordCloud. The size of the words is directly proportional to their frequencies. The largest font size in Figure 3 belongs to blockchain, followed by Bitcoin, cryptocurrencies, transaction, network, protocol, system, application, etc. Using nltk and Spacy Python libraries, pre-processing tasks were applied to the text. First, the punctuation was removed, letters were lowercased and regular stopwords plus several words specific to research papers were removed.



The baseline coherence score for the default LDA model is calculated, but furthermore, a sensitivity test is created to compute the optimal model hyperparameters: the number of topics ( $K$ ), document-topic density—Dirichlet hyperparameter alpha: ( $\alpha$ ), and the word-topic density—Dirichlet hyperparameter beta: ( $\beta$ ). Therefore, the optimal  $\alpha = 0.01$ ,  $\beta = 0.91$  with a corresponding  $K = 6$  and  $C_s = 0.385$ , obtaining an improvement of 27.06%.

Using the genism and pyLDAvis libraries, six well-delimited topics with small overlaps are distinguished in the following LDA diagrams (Figure 5). The first and the largest topic that consisted in 32.1% of the tokens is undoubtedly related to *blockchain*, data, distributed/decentralized technology, system security, smart contracts and consensus (Figure 5a).

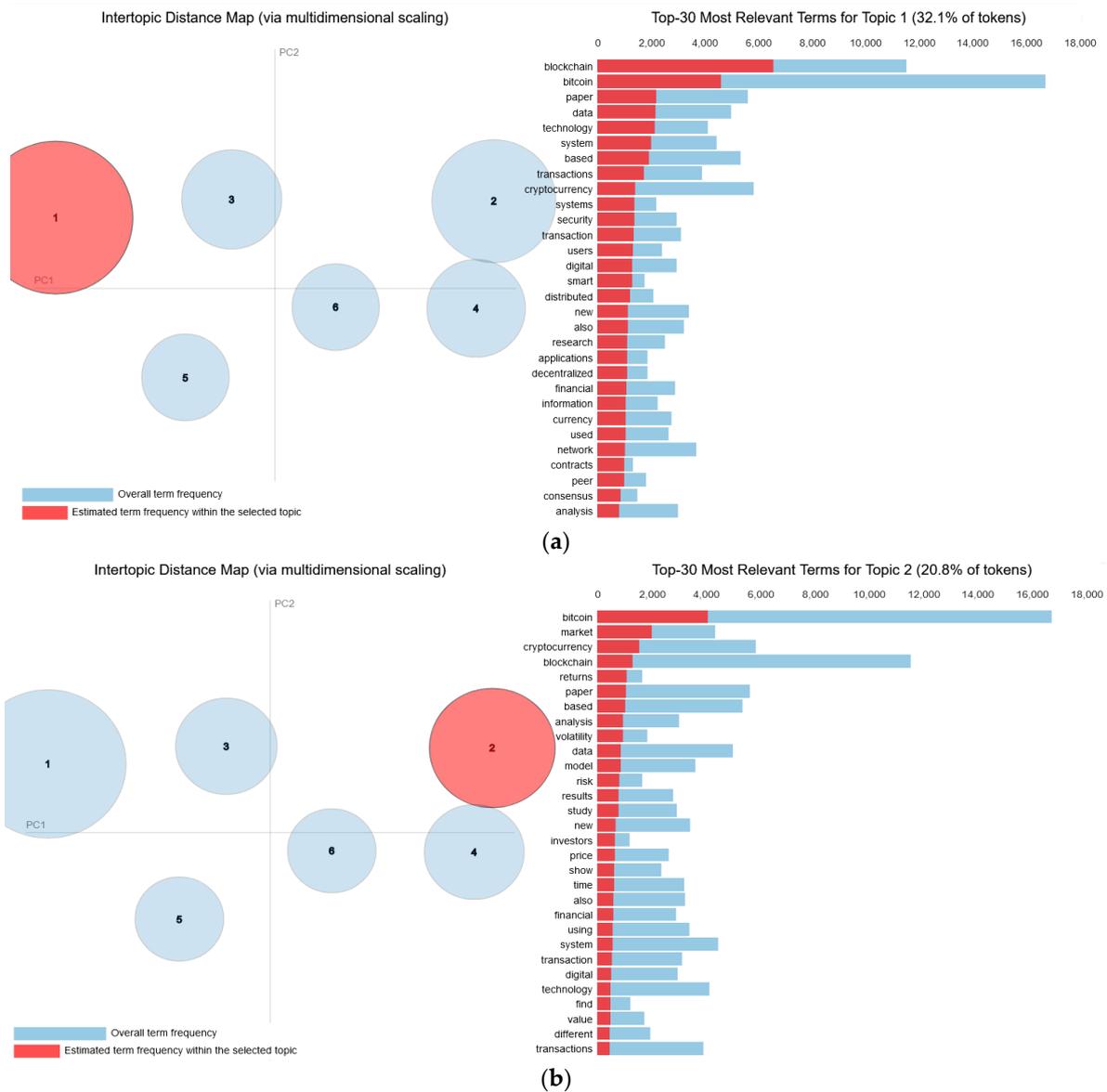


Figure 5. Cont.

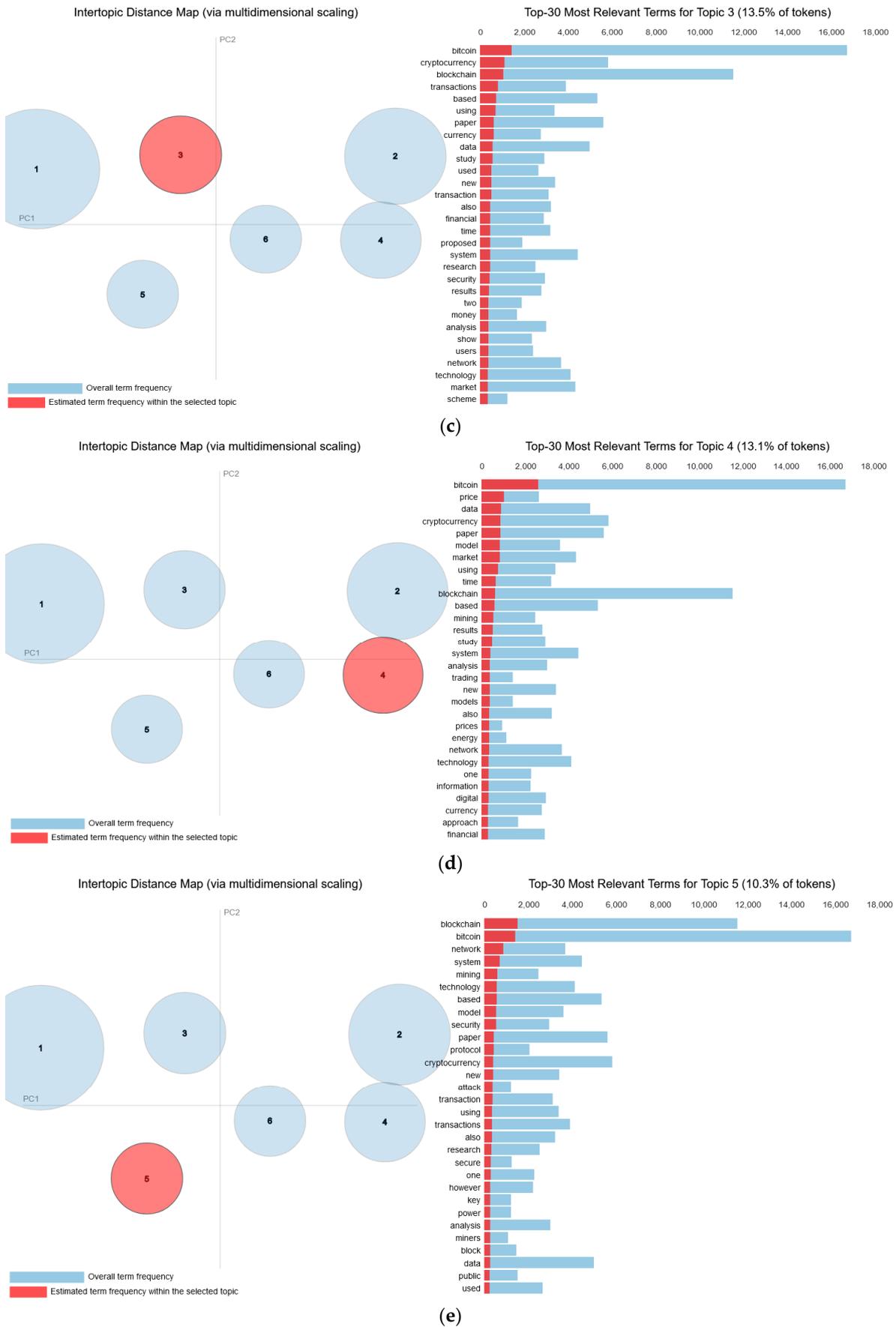
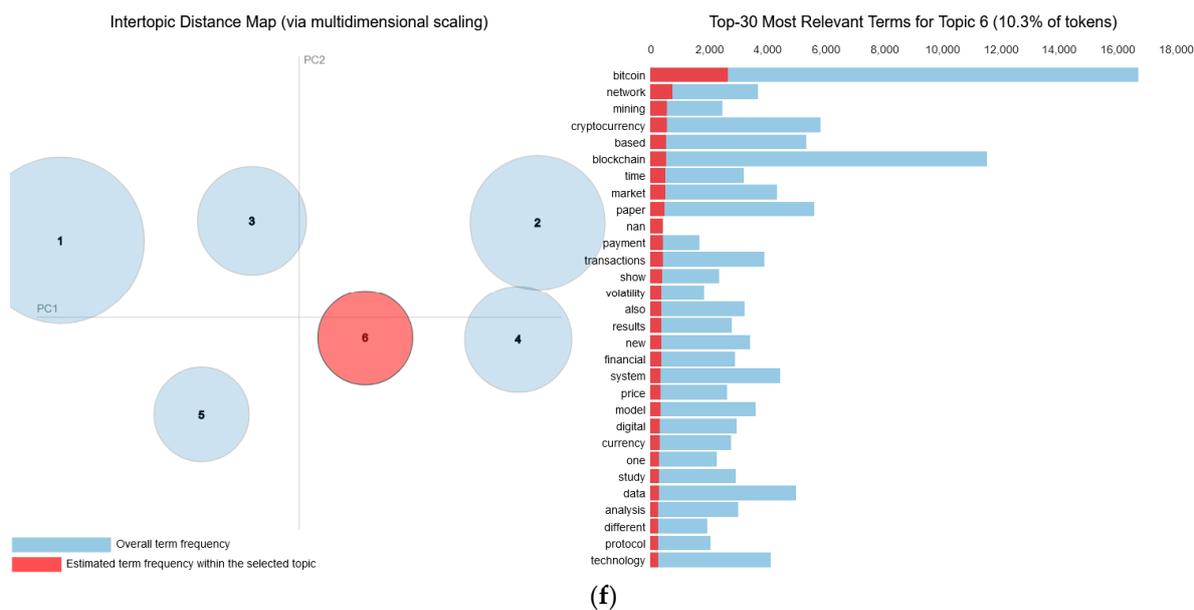


Figure 5. Cont.



**Figure 5.** LDA diagrams.

The second largest topic that consisted in 20.8% of the tokens focuses on the cryptocurrency *market*, returns, volatility, risk investors, financial aspects, price and value (Figure 5b). The third topic category that consisted in 13.5% of the tokens is related to cryptocurrency *transactions*, data, financial and system security (Figure 5c). The fourth topic that consisted in 13.1% of the tokens focuses on the Bitcoin *price*, pricing models, market, mining and energy (Figure 5d). The fifth topic that consisted in 10.3% is related to blockchain *network* security, mining technology, system security, attack, protocol, and miners (Figure 5e). The sixth topic is equal to the fifth topic in terms of tokens and is focused on the *mining* processing, market, payment, transactions, volatility, and financial issues (Figure 5f).

Topic modelling visualization is a type of statistical model for discovering the abstract “topics” that occur in a collection of documents. It is often used in text mining. In the following paragraphs, we break down what these visualizations generally represent: (1) *Inter-topic distance map* (via multidimensional scaling): The map shows different topics, represented as numbered circles. The size of each circle usually corresponds to the prevalence of the topic in the dataset. The distance between any two circles typically represents the similarity between the topics; closer circles suggest more similarity. The positioning along the axes indicates how the topics are spread; (2) *Top 30 most relevant terms for topics*: Each bar chart corresponds to a topic and lists the top 30 most relevant terms that characterize that topic. The length of the bars represents the term frequency within the topic. The red bars indicate terms that are more specific to the topic, while the blue bars indicate terms that are also relevant but not unique to the topic. The percentage of tokens indicates the weight of the topic in the overall dataset.

To compare, analyze, and interpret topic modelling visualizations, we look for patterns in the prevalence of topics, the uniqueness of terms to specific topics, and how topics might be related based on their proximity in the inter-topic distance map. The following aspects are analyzed to interpret subplots in Figure 5: (a) Comparing topic prevalence: The size of the circles in the distance map is checked to see which topics are more prevalent. Larger circles indicate a higher overall term frequency within the dataset; (b) Analyzing term relevance: To understand each topic, we review the terms listed in the bar charts. The terms further to the right are more relevant to the topic; (c) Identifying unique terms: The red bars represent terms that are not only frequent but also more specific to the topic, potentially giving a better sense of what distinguishes one topic from another; (d) Interpreting topic relationships: We examine which topics are close to each other on the map. Topics that are closer to each other might share some common terms or themes; (e) Topic contribution:

The percentage of tokens next to each bar chart title indicates how much of the text data was assigned to each topic, giving a sense of the topic's importance or dominance in the dataset. By following these steps, we construct a detailed understanding of the topics and how they relate to each other.

LDA is a form of unsupervised learning which is particularly well-suited for analyzing and categorizing large volumes of unlabeled text. It assumes that each document in a corpus can be described by a distribution of topics, and each topic can be described by a distribution of words. Each subplot in Figure 5 represents a visualization that come from LDA topic modelling. To analyze each subplot, we consider: (1) The inter-topic distance map: Each circle (bubble) represents a different topic discovered in the dataset. In the maps or subplots (a–f), the topics are numbered one through six. The size of a bubble reflects the prevalence of the topic. For instance, Topic 1 and Topic 2 are the largest bubble, suggesting they are prominent topics in the dataset. The distance between bubbles shows how different the topics are from each other; topics that are closer may share more common terms (Topic 2—market, and Topic 4—Bitcoin price), while those that are further apart are more distinct (Topic 2 and Topic 5—blockchain network security); (2) Top 30 most relevant terms: For each topic, there is a corresponding bar chart showing the 30 most relevant terms based on their frequency and distinctiveness. The x-axis indicates the frequency of terms, while the color coding (red to blue) represents how exclusive the terms are to the topic (with red being more exclusive). From the subplots in Figure 5, we further make several observations: (1) Topic size and spread: Larger topics (such as Topic 1 and Topic 2) have a wider spread of terms, indicating a broader discussion within the dataset. Smaller topics might represent more niche areas of discussion; (2) Term specificity: Red bars represent terms that are not only frequent but highly specific to the topic, giving us insight into what makes each topic unique; (3) Dominant terms: Common terms across different topics may represent overarching themes in the dataset, while unique terms give specific character to a topic; (4) Topic proximity: Topics that are close to each other on the distance map may have overlapping content or be related in some thematic way (only Topic 2 and Topic 4 indicate a small overlap, the rest of the topics being clearly distinct). As seen in Figure 5, in terms of the LDA results, the dataset contains a variety of discussions related to cryptocurrency, with some topics focusing on technical aspects, others on applications and on economic or trading perspectives.

## 5. Conclusions

More than 9100 research papers about Bitcoin and cryptocurrencies were published and indexed in Web of Science Clarivate platform until the end of July 2023. In this paper, we investigated the effect that publications may have on Bitcoin prices or the relationship between the academic community's opinion and Bitcoin prices. Furthermore, this paper aimed to uncover the main topics related to cryptocurrencies in general and Bitcoin in particular. LDA was applied to identify the relevant topics from the unstructured data of the research abstracts. The model hyperparameters were tuned in order to find the optimal number of topics.

Returning to the research questions exposed in the introduction, after merging and analyzing the two datasets, we are able to provide data-driven answers to (RQ1) What are the main topics investigated by the researchers in relation to cryptocurrencies and potential gaps? Blockchain, market, transactions, price, network security, and mining process are the six most predominant topics related to cryptocurrencies. Perplexity and topic coherence were calculated to optimally set the right number of topics.

To tune the number of topics and the parameters of the LDA method, the process proved to be time consuming (taking about 12 h) to process and test different scenarios. By tuning the LDA model, we obtained the optimal number of topics that provides the best coherence of topics and perplexity scores.

(RQ2) What is the academic community sentiment in relation to Bitcoin/cryptocurrencies? The abstracts are vectorized with BERT-transformers and word2vec and the potential of

abstracts to predict Bitcoin prices is analyzed. The sentiment analysis (SA) is performed and the relationship between SA and Bitcoin prices is showcased. Using the SA and investigating the polarity, we found that the majority of the abstracts conveyed a positive sentiment (86%), whereas 12% were negative and less than 2%. However, the findings indicate that academic publications have a minimal impact on Bitcoin prices, proving a weak link. This suggests that fluctuations in Bitcoin prices are not driven by scholarly research in this specific area. (RQ3) Is there a connection between academic publications and Bitcoin's price evolutions? Based on the Pearson correlation, the relation between the SA extracted from the abstracts and the Bitcoin prices is very weak or practically no evident connection could be detected. Further, the potential of the academic writing to influence the movements of Bitcoin prices is limited.

The six topics identified in this study suggest the mainstream or the relevant research on Bitcoin and cryptocurrencies. However, they also suggest gaps in the literature. For instance, the mining process that influences energy consumption and environmental aspects should be more investigated in the future. As Bitcoin accounts for roughly half of the energy demand of all cryptocurrencies, more energy-efficient consensus mechanisms are needed. Moreover, anti-money laundering measures, misbehaviors and crypto-crime related events should be further studied.

The practical implications of this study, which analyzed over 9100 research papers on Bitcoin and cryptocurrencies, are multifaceted:

- (a) Guidance for future research. The identification of six predominant topics (blockchain, market dynamics, transactions, pricing, network security, and the mining process) highlights areas that have been the main focus of academic research. This can guide future studies to either build upon these areas or explore less-researched topics;
- (b) Highlighting research gaps. The study points out gaps in the literature, especially regarding the environmental impact of the mining process and the need for more energy-efficient consensus mechanisms. This suggests a direction for new research endeavors, particularly in addressing the sustainability challenges of cryptocurrencies;
- (c) Influence of academic research on the market. The finding that academic publications have minimal impact on Bitcoin prices provides valuable insights for investors, market analysts, and academics. It suggests that while academic research contributes to the understanding of cryptocurrencies, it does not significantly influence market movements;
- (d) Sentiment analysis insights. The study reveals that the majority of academic sentiments toward Bitcoin/cryptocurrencies are positive. However, the weak correlation between this sentiment and Bitcoin prices can inform researchers and practitioners about the limited role of academic opinion in influencing market dynamics. Thus, while the study indicates that academic research does not significantly impact Bitcoin prices, it contributes significantly to the broader understanding of cryptocurrencies, highlighting key areas for future research and policy considerations.

One of the limitations of the current research is that it is confined to academic publications indexed in the Web of Science Clarivate platform. This may exclude relevant research published in other databases or platforms, potentially leading to a limited understanding of the entire body of cryptocurrency research. Another limitation is that the current research concentrates more on technical aspects like identifying topics, connections, and sentiment analysis, potentially underrepresenting economic interpretations and theories related to cryptocurrencies.

In future work, we aim to conduct our cryptocurrency study in conjunction with multiple social media and research platform data analyses, which may offer more timely insights, and intend to incorporate a deeper economic interpretation in sync with the dynamic pace of Bitcoin.

**Author Contributions:** Conceptualization, A.B. and S.-V.O.; methodology, A.B.; validation, A.B. and S.-V.O.; formal analysis, A.B. and S.-V.O.; investigation, A.B. and S.-V.O.; resources, A.B.; data

curation, A.B.; writing—original draft preparation, A.B. and S.-V.O.; writing—review and editing, A.B. and S.-V.O.; supervision, A.B.; project administration, S.-V.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper was supported by Academy of Romanian Scientists, Ilfov 3, 050044 Bucharest, Romania, project title: “Soluții informatice pentru analiza impactului rețelelor de social media asupra instrumentelor investiționale cu grad ridicat de risc: cryptomonedă și bursă”.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data will be made available upon reasonable request.

**Acknowledgments:** This paper was supported by Academy of Romanian Scientists, Ilfov 3, 050044 Bucharest, Romania, project title: “Soluții informatice pentru analiza impactului rețelelor de social media asupra instrumentelor investiționale cu grad ridicat de risc: cryptomonedă și bursă”.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Seow, K.T. Supervisory Control of Blockchain Networks. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *50*, 159–171. [[CrossRef](#)]
- Parino, F.; Beiró, M.G.; Gauvin, L. Analysis of the Bitcoin Blockchain: Socio-Economic Factors behind the Adoption. *EPJ Data Sci.* **2018**, *7*, 38. [[CrossRef](#)]
- Liu, Y.; Zhang, L.; Zhao, Y. Deciphering Bitcoin Blockchain Data by Cohort Analysis. *Sci. Data* **2022**, *9*, 136. [[CrossRef](#)] [[PubMed](#)]
- Prybila, C.; Schulte, S.; Hochreiner, C.; Weber, I. Runtime Verification for Business Processes Utilizing the Bitcoin Blockchain. *Futur. Gener. Comput. Syst.* **2017**, *107*, 816–831. [[CrossRef](#)]
- Dias, R.; Alexandre, P.; Teixeira, N.; Chambino, M. Clean Energy Stocks: Resilient Safe Havens in the Volatility of Dirty Cryptocurrencies. *Energies* **2023**, *16*, 5232. [[CrossRef](#)]
- Sapra, N.; Shaikh, I. Impact of Bitcoin Mining and Crypto Market Determinants on Bitcoin-Based Energy Consumption. *Manag. Financ.* **2023**, *49*, 1828–1846. [[CrossRef](#)]
- Sharif, A.; Brahim, M.; Dogan, E.; Tzeremes, P. Analysis of the Spillover Effects between Green Economy, Clean and Dirty Cryptocurrencies. *Energy Econ.* **2023**, *120*, 106594. [[CrossRef](#)]
- de Vries, A.; Gellersdörfer, U.; Klaaßen, L.; Stoll, C. Revisiting Bitcoin’s Carbon Footprint. *Joule* **2022**, *6*, 498–502. [[CrossRef](#)]
- Zaghloul, E.; Li, T.; Mutka, M.W.; Ren, J. Bitcoin and Blockchain: Security and Privacy. *IEEE Internet Things J.* **2020**, *7*, 10288–10313. [[CrossRef](#)]
- BA, M. The Effect of Propagation Delay on the Dynamic Evolution of the Bitcoin Blockchain. *Digit. Commun. Netw.* **2019**, *6*, 157–166. [[CrossRef](#)]
- Chawki, M. The Dark Web and the Future of Illicit Drug Markets. *J. Transp. Secur.* **2022**, *15*, 173–191. [[CrossRef](#)]
- van Wegberg, R.; Oerlemans, J.-J.; van Deventer, O. Bitcoin Money Laundering: Mixed results? An Explorative Study on Money Laundering of Cybercrime Proceeds Using Bitcoin. *J. Financ. Crime* **2018**, *25*, 419–435. [[CrossRef](#)]
- Böhmecke-Schwafert, M.; Moreno, E.G. Exploring Blockchain-Based Innovations for Economic and Sustainable Development in the Global South: A Mixed-Method Approach Based on Web Mining and Topic Modeling. *Technol. Forecast. Soc. Chang.* **2023**, *191*, 122446. [[CrossRef](#)]
- Oprea, S.-V.; Bâra, A. Devising a Trading Mechanism with a Joint Price Adjustment for Local Electricity Markets Using Blockchain. Insights for Policy Makers. *Energy Policy* **2021**, *152*, 112237. [[CrossRef](#)]
- Oprea, S.-V.; Bâra, A.; Andreescu, A.-I.; Cristescu, M.P. Conceptual Architecture of a Blockchain Solution for E-Voting in Elections at the University Level. *IEEE Access* **2023**, *11*, 18461–18474. [[CrossRef](#)]
- Azzi, R.; Chamoun, R.K.; Sokhn, M. The Power of a Blockchain-Based Supply Chain. *Comput. Ind. Eng.* **2019**, *135*, 582–592. [[CrossRef](#)]
- Grassman, R.; Bracamonte, V.; Davis, M.; Sato, M. Attitudes to Cryptocurrencies: A Comparative Study Between Sweden and Japan. *Rev. Socionetwork Strat.* **2021**, *15*, 169–194. [[CrossRef](#)]
- Liu, F.; Wang, L.; Kong, D.; Shi, C.; Feng, Z.; Zhou, J.; Liu, J.; Li, Z. Is There More to Bitcoin Mining than Carbon Emissions? *Heliyon* **2023**, *9*, e15099. [[CrossRef](#)] [[PubMed](#)]
- Kyriazis, N.A.; Papadamou, S.; Tzeremes, P. Are Benchmark Stock Indices, Precious Metals or Cryptocurrencies Efficient Hedges against Crises? *Econ. Model.* **2023**, *128*, 106502. [[CrossRef](#)]
- Kyriazis, N.; Papadamou, S.; Tzeremes, P.; Corbet, S. The Differential Influence of Social Media Sentiment on Cryptocurrency Returns and Volatility during COVID-19. *Q. Rev. Econ. Financ.* **2022**, *89*, 307–317. [[CrossRef](#)]
- Kyriazis, N.; Papadamou, S.; Tzeremes, P.; Corbet, S. Can Cryptocurrencies Provide a Viable Hedging Mechanism for Benchmark Index Investors? *Res. Int. Bus. Financ.* **2023**, *64*, 101832. [[CrossRef](#)]
- Papadamou, S.; Kyriazis, N.A.; Tzeremes, P.G. Non-Linear Causal Linkages of EPU and Gold with Major Cryptocurrencies during Bull and Bear Markets. *North Am. J. Econ. Financ.* **2020**, *56*, 101343. [[CrossRef](#)]

23. Papadamou, S.; Kyriazis, N.A.; Tzeremes, P.; Corbet, S. Herding Behaviour and Price Convergence Clubs in Cryptocurrencies during Bull and Bear Markets. *J. Behav. Exp. Financ.* **2021**, *30*, 100469. [[CrossRef](#)]
24. Shahzad, M.K.; Bukhari, L.; Khan, T.M.; Islam, S.M.R.; Hossain, M.; Kwak, K.-S. BPTE: Bitcoin Price Prediction and Trend Examination Using Twitter Sentiment Analysis. In Proceedings of the 2021 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 20–22 October 2021; pp. 119–122.
25. Coulter, K.A. The Impact of News Media on Bitcoin Prices: Modelling Data Driven Discourses in the Crypto-Economy with Natural Language Processing. *R. Soc. Open Sci.* **2022**, *9*, 220276. [[CrossRef](#)]
26. Sharma, C.; Sharma, S.; Sakshi. Latent DIRICHLET Allocation (LDA) Based Information Modelling on BLOCKCHAIN Technology: A Review of Trends and Research Patterns Used in Integration. *Multimed. Tools Appl.* **2022**, *81*, 36805–36831. [[CrossRef](#)] [[PubMed](#)]
27. Liu, F.; Feng, Z.; Qi, J. A Blockchain-Based Digital Asset Platform with Multi-Party Certification. *Appl. Sci.* **2022**, *12*, 5342. [[CrossRef](#)]
28. Liu, F.; Fan, H.Y.; Qi, J.Y. Blockchain Technology, Cryptocurrency: Entropy-Based Perspective. *Entropy* **2022**, *24*, 557. [[CrossRef](#)]
29. Cotfas, L.-A.; Crăciun, L.; Delcea, C.; Florescu, M.S.; Kovacs, E.-R.; Molănescu, A.G.; Orzan, M. Unveiling Vaccine Hesitancy on Twitter: Analyzing Trends and Reasons during the Emergence of COVID-19 Delta and Omicron Variants. *Vaccines* **2023**, *11*, 1381. [[CrossRef](#)]
30. Kovacs, E.-R.; Cotfas, L.-A.; Delcea, C. January 6th on Twitter: Measuring Social Media Attitudes towards the Capitol Riot through Unhealthy Online Conversation and Sentiment Analysis. *J. Inf. Telecommun.* **2023**, 1–22. [[CrossRef](#)]
31. Yen, J.-C.; Wang, T. Stock Price Relevance of Voluntary Disclosures about Blockchain Technology and Cryptocurrencies. *Int. J. Account. Inf. Syst.* **2021**, *40*, 100499. [[CrossRef](#)]
32. Zhou, L.; Zhang, L.; Zhao, Y.; Zheng, R.; Song, K. A Scientometric Review of Blockchain Research. *Inf. Syst. E-Bus. Manag.* **2020**, *19*, 757–787. [[CrossRef](#)]
33. Alahi, I.; Islam, M.; Iqbal, A.; Bosu, A. Identifying the Challenges of the Blockchain Community from Stackexchange Topics and Trends. In Proceedings of the 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), Milwaukee, WI, USA, 15–19 July 2019.
34. Liu, F.; He, S.; Li, Z.; Xiang, P.; Qi, J.; Li, Z. An Overview of Blockchain Efficient Interaction Technologies. *Front. Blockchain* **2023**, *6*, 996070. [[CrossRef](#)]
35. Lade, M.; Welekar, R.; Dadiyala, C. Bitcoin Price Prediction and NFT Generator Based on Sentiment Analysis. *Int. J. Next-Gener. Comput.* **2023**, *14*. [[CrossRef](#)]
36. Zaman, S.; Yaqub, U.; Saleem, T. Analysis of Bitcoin's Price Spike in Context of Elon Musk's Twitter Activity. *Glob. Knowl. Mem. Commun.* **2022**, *72*, 341–355. [[CrossRef](#)]
37. Bashir, H.A.; Kumar, D. Investor Attention, Twitter Uncertainty and Cryptocurrency Market amid the COVID-19 Pandemic. *Manag. Financ.* **2022**, *49*, 620–642. [[CrossRef](#)]
38. Wang, F.; Gacesa, M. Semi-Strong Efficient Market of Bitcoin and Twitter: An Analysis of Semantic Vector Spaces of Extracted Keywords and Light Gradient Boosting Machine Models. *Int. Rev. Financ. Anal.* **2023**, *88*, 102692. [[CrossRef](#)]
39. Chursook, A.; Dawod, A.Y.; Chanaim, S.; Naktasukanjn, N.; Chakpitak, N. Twitter Sentiment Analysis and Expert Ratings of Initial Coin Offering Fundraising: Evidence from Australia and Singapore Markets. *TEM J.* **2022**, *11*, 44–55. [[CrossRef](#)]
40. Gurrib, I.; Kamalov, F. Predicting Bitcoin Price Movements Using Sentiment Analysis: A Machine Learning Approach. *Stud. Econ. Financ.* **2021**, *39*, 347–364. [[CrossRef](#)]
41. Farimani, S.A.; Jahan, M.V.; Fard, A.M.; Tabbakh, S.R.K. Investigating the Informativeness of Technical Indicators and News Sentiment in Financial Market Price Prediction. *Knowl.-Based Syst.* **2022**, *247*, 108742. [[CrossRef](#)]
42. Hassan, M.K.; Hudaefi, F.A.; Caraka, R.E. Mining Netizen's Opinion on Cryptocurrency: Sentiment Analysis of Twitter Data. *Stud. Econ. Financ.* **2021**, *39*, 365–385. [[CrossRef](#)]
43. Dwivedi, D.N.; Vemareddy, A. Sentiment Analytics for Crypto Pre and Post Covid: Topic Modeling. In *International Conference on Distributed Computing and Intelligent Technology*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2023.
44. Burnie, A.; Yilmaz, E. An Analysis of the Change in Discussions on Social Media with Bitcoin Price. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Paris, France, 21–25 July 2019.
45. Ortu, M.; Vacca, S.; Destefanis, G.; Conversano, C. Cryptocurrency Ecosystems and Social Media Environments: An Empirical Analysis through Hawkes' Models and Natural Language Processing. *Mach. Learn. Appl.* **2021**, *7*, 100229. [[CrossRef](#)]
46. Ibba, G.; Vaccargiu, M. Analysis of Users' Most Discussed Topics and Trends on Blockchain Technologies and Smart Contracts. In Proceedings of the 2023 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER), Taipa, Macao, 21–24 March 2023.
47. Mulahuwaish, A.; Loucks, M.; Qolomany, B.; Al-Fuqaha, A. Topic Modeling Based on Two-Step Flow Theory: Application to Tweets about Bitcoin. *IT Prof.* **2023**, *25*, 52–63. [[CrossRef](#)]
48. Treiblmaier, H.; Swan, M.; de Filippi, P.; Lacity, M.; Hardjono, T.; Kim, H. What's Next in Blockchain Research? An Identification of Key Topics Using a Multidisciplinary Perspective. *ACM SIGMIS Database Database Adv. Inf. Syst.* **2021**, *52*, 27–52. [[CrossRef](#)]
49. Pathak, A.R.; Pandey, M.; Rautaray, S. Topic-Level Sentiment Analysis of Social Media Data Using Deep Learning. *Appl. Soft Comput.* **2021**, *108*, 107440. [[CrossRef](#)]

50. Schulte, M.; Eggert, M. Predicting Hourly Bitcoin Prices Based on Long Short-Term Memory Neural Networks. In *Innovation through Information Systems: Volume II: A Collection of Latest Research on Technology Issues*; Lecture Notes in Information Systems and Organisation; Springer: Cham, Switzerland, 2021.
51. Koutmos, D. Investor Sentiment and Bitcoin Prices. *Rev. Quant. Financ. Account.* **2022**, *60*, 1–29. [[CrossRef](#)]
52. Chen, J. Analysis of Bitcoin Price Prediction Using Machine Learning. *J. Risk Financ. Manag.* **2023**, *16*, 51. [[CrossRef](#)]
53. Bouri, E.; Salisu, A.A.; Gupta, R. The Predictive Power of Bitcoin Prices for the Realized Volatility of US Stock Sector Returns. *Financ. Innov.* **2023**, *9*, 62. [[CrossRef](#)] [[PubMed](#)]
54. Köse, N.; Ünal, E. The Asymmetric Effects of the Interest Rate on the Bitcoin Price. *Financ. Czech J. Econ. Financ.* **2023**, *73*, 189–217. [[CrossRef](#)]
55. Caliciotti, A.; Corazza, M.; Fasano, G. From Regression Models to Machine Learning Approaches for Long Term Bitcoin Price Forecast. *Ann. Oper. Res.* **2023**, 1–23. [[CrossRef](#)]
56. Liu, F.; Li, Z.; Jia, K.; Xiang, P.; Zhou, A.; Qi, J.; Li, Z. Bitcoin Address Clustering Based on Change Address Improvement. *IEEE Trans. Comput. Soc. Syst.* **2023**; early access. [[CrossRef](#)]
57. Song, Y.; Chen, B.; Wang, X.-Y. Cryptocurrency Technology Revolution: Are Bitcoin Prices and Terrorist Attacks Related? *Financ. Innov.* **2023**, *9*, 29. [[CrossRef](#)] [[PubMed](#)]
58. Xinyu, W.; Shenghao, N.; Haibin, X. Forecasting Bitcoin Volatility Using Two-Component CARR Model. *Econ. Comput. Econ. Cybern. Stud. Res.* **2020**, *54*, 77–94. [[CrossRef](#)]
59. Hajek, P.; Hikkerova, L.; Sahut, J.-M. How Well Do Investor Sentiment and Ensemble Learning Predict Bitcoin Prices? *Res. Int. Bus. Financ.* **2023**, *64*, 101836. [[CrossRef](#)]
60. Jia, D.; Li, Y. Bounded Pool Mining and the Bounded Bitcoin Price. *Financ. Res. Lett.* **2023**, *52*, 103529. [[CrossRef](#)]
61. Bejan, C.A.; Bucerzan, D.; Crăciun, M.D. Bitcoin Price Evolution versus Energy Consumption; Trend Analysis. *Appl. Econ.* **2022**, *55*, 1497–1511. [[CrossRef](#)]
62. Aghashahi, M.; Bamdad, S. Analysis of Different Artificial Neural Networks for Bitcoin Price Prediction. *Int. J. Manag. Sci. Eng. Manag.* **2022**, *18*, 126–133. [[CrossRef](#)]
63. Jakubik, J.; Nazemi, A.; Geyer-Schulz, A.; Fabozzi, F.J. Incorporating Financial News for Forecasting Bitcoin Prices Based on Long Short-Term Memory Networks. *Quant. Financ.* **2022**, *23*, 335–349. [[CrossRef](#)]
64. Rathore, R.K.; Mishra, D.; Mehra, P.S.; Pal, O.; Hashim, A.S.; Shapi’I, A.; Ciano, T.; Shutaywi, M. Real-World Model for Bitcoin Price Prediction. *Inf. Process. Manag.* **2022**, *59*, 102968. [[CrossRef](#)]
65. Feng, W.; Zhang, Z. Currency Exchange Rate Predictability: The New Power of Bitcoin Prices. *J. Int. Money Financ.* **2023**, *132*, 102811. [[CrossRef](#)]
66. Cevik, E.I.; Gunay, S.; Dibooglu, S.; Yıldırım, D.Ç. The Impact of Expected and Unexpected Events on Bitcoin Price Development: Introduction of Futures Market and COVID-19. *Financ. Res. Lett.* **2023**, *54*, 103768. [[CrossRef](#)]
67. Ahmed, W.M.A. Robust Drivers of Bitcoin Price Movements: An Extreme Bounds Analysis. *North Am. J. Econ. Financ.* **2022**, *62*, 101728. [[CrossRef](#)]
68. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the NAACL HLT 2019—2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019.
69. Bello, A.; Ng, S.-C.; Leung, M.-F. A BERT Framework to Sentiment Analysis of Tweets. *Sensors* **2023**, *23*, 506. [[CrossRef](#)]
70. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of the 1st International Conference on Learning Representations, ICLR 2013—Workshop Track Proceedings, Scottsdale, AZ, USA, 2–4 May 2013.
71. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Distributed Representations of Words and Phrases and Their Compositionality. *Neural Inf. Process. Syst.* **2006**, *26*.
72. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **2000**, *155*, 945–959. [[CrossRef](#)] [[PubMed](#)]
73. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022. [[CrossRef](#)]
74. Jelodar, H.; Wang, Y.; Yuan, C.; Feng, X.; Jiang, X.; Li, Y.; Zhao, L. Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey. *Multimed. Tools Appl.* **2019**, *78*, 15169–15211. [[CrossRef](#)]
75. Bâra, A.; Oprea, S.-V. What Makes Electricity Consumers Change Their Behavior? Influence of Attitude and Perceived Impact of DR Programs on Awareness. *Kybernetes*, **2023**; ahead-of-print. [[CrossRef](#)]
76. Oprea, S.-V.; Bâra, A. A Measurement Model for Electricity Consumers’ Awareness with Covariance Structure Analyses. A Solid Pillar for Boosting Demand Response Programs. *Sustain. Energy Technol. Assess.* **2022**, *53*, 102738. [[CrossRef](#)]
77. Chang, J.; Boyd-Graber, J.; Gerrish, S.; Wang, C.; Blei, D.M. Reading Tea Leaves: How Humans Interpret Topic Models. In Proceedings of the Advances in Neural Information Processing Systems 22, Vancouver, BC, Canada, 7–10 December 2009.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.