BİTLİS EREN ÜNİVERSİTESİ
FEN BİLİMLERİ DERGİSİ

ISSN: 2147-3129 / e-ISSN: 2147-3188

Article Type	: Research Article	**Year**	: 2025
Received	: December 31, 2024	**Volume**	: 14
Revised	: February 14, 2025	**Issue**	: 1
Accepted	: March 5, 2025	**Pages**	: 561-582
DOI	: 10.17798/bitlisfen.1610560		

PREDICTING BITCOIN MINING ENERGY CONSUMPTION USING MACHINE LEARNING: A CASE FOR K-NEAREST NEIGHBORS REGRESSION

Nazmiye ELİGÜZEL [1], **Sena AYDOĞAN [2] ***

[1] *Gaziantep İslam, Science and Technology University, Department of Industrial Engineering, Gaziantep, Türkiye*

[2] *Gazi University, Department of Industrial Engineering, Ankara, Türkiye*

**** Corresponding Author:*** *senaaydogan@gazi.edu.tr*

ABSTRACT

The energy consumption of Bitcoin mining has emerged as a critical topic in cryptocurrency research, influenced by the significant environmental and economic impacts of blockchain activities. This study examines the energy consumption of Bitcoin mining with a dataset that includes essential blockchain variables such as overall hash rate, network difficulty, daily confirmed transactions, mempool size, average block size, and daily Bitcoin output. A new energy consumption indicator is proposed to contribute to the research domain. The proposed indicator better accurately reflects the dynamics of blockchain energy utilization. Various machine learning models, such as Random Forest, Gradient Boosting, Support Vector Regression, and Multi-layer Perceptron, are evaluated, with particular emphasis on k-Nearest Neighbors Regression (k-NNR). The k-NNR model surpassed all other models, with a $R2$ value of 0.80427 and a Mean Squared Error (MSE) of 0.00441, indicating its high prediction accuracy. Analysis of feature importance indicated that daily Bitcoin production and block size are significant determinants of energy use. The findings underscore the efficacy of k-NNR in energy modeling, offering insights into Bitcoin's energy dynamics and establishing a foundation for more energy-efficient blockchain systems.

Keywords: Bitcoin, Energy consumption, Hash rate, k-Nearest neighbor regression, Machine learning.

1 INTRODUCTION

Blockchain is a disruptive form of distributed ledgers, developed by Satoshi Nakamoto [1] in 2008. A blockchain is a distributed, decentralized, and securely maintained peer-to-peer network that facilitates the management of information and trust among users without requiring a trusted third party [2]. This could diminish the influence of central authorities within financial ecosystems, consequently lessening the necessity for intermediaries like banks or government agencies to maintain trust. Blockchain technology, as a distributed database, facilitates the sharing of digital events among all members in the blockchain. The information included in each block is encrypted with a "hash" value. Every block in a blockchain encompasses the hash value of the preceding block [2], [3]. This approach can diminish fraud attempts by enhancing transparency, thereby significantly reducing the likelihood of data manipulation. The cryptocurrency market has experienced substantial growth in recent years [4]. This growth could require the implementation of new regulations within financial markets. Since 2009, various cryptocurrencies have developed, beginning with Bitcoin, the initial prominent implementation of Satoshi Nakamoto's blockchain technology announced in 2008 [5]. Bitcoin, as articulated by Satoshi Nakamoto, is a decentralized peer-to-peer electronic cash system that employs a consensus protocol to avoid double spending across different nodes [5].

Bitcoin mining is a decentralized computing procedure in which transactions are authenticated and included into the public ledger, referred to as the blockchain. The procedure for generating bitcoins is termed mining, and the individuals involved are referred to as miners. All transactions are executed and recorded on a decentralized ledger: the blockchain [6]. This may result in heightened individual engagement within the financial system, motivating users to actively participate in the blockchain, thereby assisting in the mitigation of fraud and data manipulation. The energy consumption of cryptocurrencies has emerged as a prominent topic of discourse in recent years [6]. An increased electricity cost per bitcoin mining indicates increased consumption of energy [7]. Substantial discrepancies exist in the calculations of bitcoin's energy usage because to numerous uncertainties in the process, including the type of hardware utilized in mining and its operational duration. Estimating future energy consumption for Bitcoin mining is challenging, as Bitcoin values directly influence mining activities and, consequently, energy usage [6]. A significant number of recent studies have been released, all of which anonymously emphasize the escalating energy issue associated with bitcoin mining [4], [6], [8]–[10]. Several significant studies in the literature examine bitcoin mining and energy consumption, which we addressed in our research.

Maiti [7] examined the non-linear correlation between Bitcoin prices and energy consumption from 2010 to 2021 with a Threshold Regression model. It delineates six regimes of price fluctuations influenced by energy consumption thresholds, revealing substantial effects just in high consumption regimes, hence highlighting the disparate impact of energy on Bitcoin price dynamics. Sapra and Shaikh [4] examined the influence of market indices and Ethereum prices on Bitcoin energy use, employing Autoregressive Distributed Lag modeling with data from 2018 to 2023. The study indicates that rising Ethereum prices and cryptocurrency indices lead to increased Bitcoin energy consumption and emissions, implying the necessity for green investments in cryptocurrencies that employ alternatives to proof-of-work (PoW) techniques. Kevser [11] analyzed the correlation of geopolitical dangers, global economic policy uncertainty, and Bitcoin energy use. Analysis of data from 2011 to 2022, employing Hatemi-J causality tests, indicates that global concerns elevate Bitcoin demand and energy consumption, positioning Bitcoin as a safeguard in times of uncertainty while exacerbating its environmental consequences. Tissaoui et al. [12] utilized the Quantile Nonlinear Autoregressive Distributed Lags (QNARDL) model and Extreme Gradient Boosting (XGBoost) to evaluate the influence of Bitcoin prices on energy consumption. Short-term price increases result in rapid energy surges, however long-term impacts diminish energy consumption, with XGBoost surpassing conventional forecasting techniques. Sapra et al. [13] analyzed the causal relationships among Bitcoin's energy use, pricing, and market volatility. The study concludes that Bitcoin prices Granger-cause energy use, whereas the reverse is not true, positioning price as a net contributor and consumption as a recipient in market dynamics. Syzdykova [14] analyzed the energy requirements of Bitcoin mining, emphasizing its substantial contribution to world electricity usage, amounting to 204.5 Terawatt hours (TWh) per year by 2022. The research highlights the shortcomings of the PoW process and promotes the incorporation of renewable energy in mining operations. Kohli et al. [5] compared the energy usage of cryptocurrencies with centralized systems such as Visa, highlighting Bitcoin's environmental impact comparable to that of national energy usage. The research commends Ethereum 2.0's shift to proof-of-stake as a significant achievement in sustainability. Zaghdoudi et al. [15]employed machine learning methods, including XGBoost, to forecast energy usage affected by uncertainty indices. The results underscore economic policy uncertainty and geopolitical risks as major predictors, illustrating the effectiveness of sophisticated modeling tools. Bublyk et al. [16] projected Bitcoin's energy consumption would attain 142 TWh by 2026, incurring substantial environmental consequences. It advocates for the association of mining activities with renewable energy sources to reduce their environmental impact. Adewuyi et al. [17] conducted an investigation into Bitcoin's energy consumption employing structural break and non-linear analytics. It recognizes bubbles associated with market activities and advocates for sustainable investments and policies to mitigate the environmental impact of cryptocurrency mining.

Table 1 indicates summary of literature on Bitcoin energy consumption.

Table 1. Literature on Bitcoin Energy Consumption.

Study	Objective	Methodology	Variables	Time Period (Frequency)	Key Findings	Contribution
Maiti [7]	Investigate the non-linear relationship between Bitcoin prices and energy consumption.	Threshold Regression Model	Bitcoin price	November 2010 - October 2021 (Monthly)	Bitcoin prices are significantly influenced by energy consumption only in high-consumption regimes, highlighting uneven impacts.	Explores complex price-energy dynamics and identifies energy-related thresholds in price fluctuations.
Sapra & Shaikh [4]	Assess the impact of crypto indices and Ethereum prices on Bitcoin energy usage.	Autoregressive Distributed Lag (ARDL)	CBECI, Average Block size, Hash rate, No of transactions, Cix200, Etherium price	December 2018 - January 2023 (Monthly)	Rising Ethereum prices and crypto indices drive Bitcoin energy consumption and emissions. Suggests green investment in non-PoW mechanisms.	Advocates for environmentally sustainable crypto practices through alternative technologies.
Kevser [11]	Analyze the link between geopolitical risks, global policy uncertainty, and Bitcoin energy use.	Hatemi-J Causality Test	Global economic political uncertainty, geopolitical risk	May 2011 - February 2022 (Monthly)	Global uncertainties increase Bitcoin demand and energy consumption, positioning Bitcoin as a hedge but exacerbating environmental concerns.	Highlights Bitcoin's dual role as a financial hedge and an environmental burden in uncertain scenarios.
Tissaoui et al. [12]	Evaluate Bitcoin price effects on energy consumption.	QNARDL Model, XGBoost	Bitcoin price	1 July 2010 - 1 December 2022 (Daily)	Short-term price increases lead to energy surges, but long-term effects reduce consumption. XGBoost excels in prediction.	Demonstrates XGBoost's predictive accuracy in energy modeling.
Sapra et al. [13]	Explore causal relationships between Bitcoin energy use, prices, and market volatility.	Vector Auto Regression based Granger Causality, Diebold-Yilmaz Connectedness Analysis	CBECI, Crypto volatility index, bitcoin closing price	31 March 2019 - 30 March 2023 (Daily)	Bitcoin prices Granger-cause energy use; prices act as net contributors, while energy use is a recipient in market dynamics.	Provides insights into Bitcoin's market energy feedback loops.
Syzdykova [14]	Examine Bitcoin's energy consumption and mining inefficiencies.	Literature Review	-	-	Bitcoin's annual electricity usage reached 204.5 TWh by 2022. Advocates for renewable energy in mining.	Emphasizes PoW inefficiencies and recommends renewable energy adoption.
Kohli et al. [5]	Compare cryptocurrency energy use with centralized systems like Visa.	Comparative Analysis	-	-	Bitcoin's energy use is equivalent to that of some nations. Ethereum 2.0's shift to proof-of-stake is highlighted as a sustainability milestone.	Stresses the need for transitions to proof-of-stake and energy-efficient practices.
Zaghdoudi et al. [15]	Predict CBECI influenced by uncertainty indices using machine learning.	Machine Learning (XGBoost, Support Vector Regression, CatBoost)	CBECI, Economic politicy uncertainty index, geopolitical risk index, energy uncertainty index	1 July 2010 – 1 December 2022 (Quarterly)	Economic and geopolitical risks are significant predictors of Bitcoin energy use. XGBoost outperforms other models.	Demonstrates the utility of advanced machine learning tools for energy prediction.

Table 1 (Continued). Literature on Bitcoin Energy Consumption.

Study	Objective	Methodology	Variables	Time Period (Frequency)	Key Findings	Contribution
Bublyk et al. [16]	Forecast spending on digital transformation technologies and services worldwide	Regression Analysis	Total cryptocurrency market cap, bitcoin energy consumption, etherium energy consumption	2017 – 2022 (Quarterly)	Bitcoin energy consumption projected to reach 142 TWh by 2026; recommends linking mining to renewable energy sources.	Proposes renewable energy integration in mining operations.
Adewuyi et al. [17]	Investigate Bitcoin's energy use and carbon footprint, focusing on bubbles and structural breaks.	Structural Break and Non-Linear Analytics	Maximum and optimal electricity consumption, maximum and optimal average emissions, global economic policy index, no of transactions, the level of credit risk, VIX index, google trend, geopolitical risk, volatility, volume, bitcoin energy consumption, bitcoin carbon footprint	7 July 2010 – 4 December 2021 (Daily)	Identifies market-driven energy bubbles and advocates for sustainable investments and policies to reduce the environmental impact of mining.	Calls for regulatory and investment measures to mitigate crypto mining's ecological footprint.

Research gaps identified in literature and our contributions are as follows:

- Many studies utilize restricted factors such as price, market indices, or geopolitical threats, yet fail to comprehensively incorporate blockchain data (e.g., hash rate, network difficulty, mempool size). We employ various blockchain parameters, including hash rate, network difficulty, mempool size, block size, and daily confirmed transactions, to enhance our analysis of energy consumption.
- Current research frequently employs long-term aggregate data (monthly, yearly), but short-term, granular analyses (daily data) are infrequently conducted. We concentrate on detailed, short-term blockchain data (61 days), providing new insights into daily energy consumption trends.
- Limited studies emphasis on developing supplementary measures for energy consumption related to particular blockchain activities. We introduce an extra indicator for energy consumption, facilitating improved analysis of blockchain energy dynamics. The estimated energy consumption indicator, which is based on daily total hash rate and daily electricity usage, reflects the energy requirements of systems in a more dynamic

and realistic way. This approach not only refines energy estimations but also provides valuable insights for policymakers and researchers studying sustainable blockchain operations.

- Our study is distinguished by its comparative investigation of various machine learning techniques to model and forecast Bitcoin's energy consumption. We assess the efficacy of multiple algorithms, including prevalent methods such as Random Forest, Gradient Boosting, Support Vector Regression, and Multi-layer Perceptron. We concentrate on the k-Nearest Neighbors Regression (k-NNR), which remains inadequately examined in the current literature.

This paper is organized as follows: after the introduction, the Material and Method section elucidates the data collection and preprocessing procedures, the feature selection process, the development of an additional energy consumption metric, and detailed analysis of k-NNR. The Results and Discussion section delineates the findings from the model comparisons, underscores the preeminence of k-NNR, and elucidates the ramifications of the results within the framework of blockchain energy dynamics. Conclusion encapsulates the principal contributions of the study, delineates its limits, and proposes avenues for further research.

2 MATERIAL AND METHOD

This section outlines the processes of data collection and preprocessing, the feature selection stage, the calculation of an additional metric energy consumption, data normalization, k-NNR, and model evaluation. The stages applied in this section is summarized in Figure 1.

The dataset used in this study was obtained from publicly available blockchain data [18], including metrics such as total hash rate, network difficulty, daily confirmed transactions, mempool size, average block size, and daily Bitcoin output. The dataset was cleaned by removing any unnecessary spaces in column names and filtering out rows with missing data. The hash rate, an indicator of the processing power allocated to protecting a blockchain via proof-of-work consensus, is essential for thwarting various attacks [19]. In Bitcoin mining, difficulty quantifies the challenge miners face in locating a valid block, which is standardized across the whole network and is recalibrated every 2016 blocks [20]. Daily confirmed transactions indicates the daily volume of processed confirmed transactions. The mempool effectively illustrates the increase in transactions pending confirmation, serving as a leading signal of prospective cash flows that may influence bitcoin's trading volumes and market prices [21]. Block size denotes the quantity of transactions contained within the block [22]. Namely, the Bitcoin block size denotes the maximum data capacity of an individual block within the Bitcoin blockchain. Each block encompasses transaction data, and its size dictates the number of transactions that can be incorporated within a single block. Daily Bitcoin output denotes the total quantity of Bitcoin processed each day. Data for the specified entries was gathered from blockchain.com [18] over a period of 61 days, from October 4 to December 5. Bitcoin mining and energy consumption are influenced by variables such as market volatility, mining difficulty, and transaction volume. A 61-day dataset provides a sufficient time frame to analyze changes in these variables. In the literature, datasets spanning 30 to 90 days are frequently used for short- and medium-term forecasts [23], [24]. A 61-day period strikes a suitable balance for effective model training and testing in short-term predictions. These entries were recognized as essential for comprehending energy consumption within the blockchain network. Table 2 presents the sample data from our investigation, encompassing a duration of 14 days.

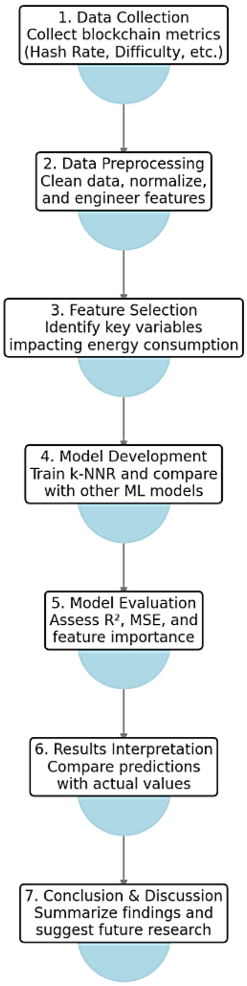

Figure 1. Methodological stages of the study.

Table 2. Sample Data from Bitcoin Mining Dataset.

Date	Total hash rate (TH/s)	Network difficulty	Daily confirmed transactions	Mempool size	Average block size (MB)	Daily Bitcoin output
5 November 2024	702351988.692	101646843652784	501641	111112.224	1.740	1268937.223
6 November 2024	773092476.762	101646843652784	553047	109742.375	1.670	902154.559
7 November 2024	788251152.777	101646843652784	497146	117335.380	1.700	953404.938
8 November 2024	717510664.707	101646843652784	577146	133364.323	1.630	551065.648
9 November 2024	702351988.692	101646843652784	558278	137806.078	1.550	808961.400
10 November 2024	747828016.737	101646843652784	462458	141248.557	1.640	1204667.966
11 November 2024	687193312.677	101646843652784	503677	143844.458	1.630	1383872.631
12 November 2024	737722232.727	101646843652784	570979	169088.875	1.610	1222908.306
13 November 2024	697299096.687	101646843652784	503566	186026.771	1.560	1058121.422
14 November 2024	641717284.633	101646843652784	562223	186044.266	1.570	928123.081
15 November 2024	757933800.747	101646843652784	602655	168264.260	1.610	524803.198
16 November 2024	793304044.782	101646843652784	598496	157410.719	1.670	444772.304
17 November 2024	757933800.747	101646843652784	673308	179352.193	1.710	888920.856
18 November 2024	789177471.480	101766294632436	810805	190610.734	1.600	943414.474

The supplied data delineates daily indicators pertaining to the Bitcoin network from November 5 to November 18, 2024. In previous studies (see Table 1), the major indicators - the total hash rate, network difficulty, daily confirmed transactions, mempool size, and daily Bitcoin output – have been used extensively to evaluate the energy consumption of the Bitcoin network. This dataset offers an overview of the Bitcoin network's performance and activity, emphasizing the relationship among hash rate, difficulty, transaction volume, and mining rewards over time. In this study, we add another indicator, namely energy consumption, which measures the energy needs of systems by looking at the daily total hash rate and daily power usage. This gives a more accurate and updated picture of energy use.

2.1 Energy Consumption Analysis of Bitcoin Mining

The energy usage of Bitcoin mining is closely associated with the processing power necessary for its PoW method. The energy requirement can be assessed by integrating the network's hash rate with the energy efficiency of mining equipment [6]. There are considerable differences in the calculations of bitcoin's energy consumption because to numerous uncertainties in the process, including the type of device utilized in mining and the duration of its operation [6].

We utilized data from the Cambridge Bitcoin Electricity Consumption Index (CBECI) [25]. CBECI offers current estimates of Bitcoin's daily energy requirements and an annualized electricity consumption prediction. Due to the decentralized structure of the network, the exact power demand cannot be ascertained; therefore, numerous assumptions were made, including hypothetical lower-bound and upper-bound estimates. These two borders represent an informed estimate, providing a more precise representation of the real power demand. The lower-bound estimate represents the potential minimum overall power requirement predicated on the optimal scenario that all miners consistently utilize the most energy-efficient equipment. The upper-bound estimate represents the theoretical maximum total power requirement predicated on the worst-case scenario wherein all miners consistently utilize the least energy-efficient hardware, provided that operating the equipment stays economically viable concerning electricity expenses. The estimate is predicated on the more plausible premise that miners utilize a blend of profitable hardware [25]. The data obtained from the CBECI on December 27, 2024, is as follows: Theoretical Lower Bound: an annual consumption of 92.610 TWh, Estimated Consumption: with an annual consumption of 180.970 TWh, Theoretical Upper Bound: with an annual consumption of 417.520 TWh. TWh refers terawatt-hour. The formula for daily electricity consumption is presented in Equation 1.

$$Daily\ electricity\ consumption\ (TWh) = \frac{Annualised\ consumption\ (TWh)}{365} \tag{1}$$

This study assessed daily electricity usage by utilizing estimated electrical consumption data in accordance with Equation 1.

$$Daily\ electricity\ consumption\ (TWh) = \frac{180.970}{365} = 0.496$$

The rationale for utilizing estimated energy consumption data is its enhanced realism. Daily energy consumption is determined by multiplying the daily total hash rate by the daily electricity usage, as specified in Equation 2.

$$Energy\ consumption = Total\ hash\ rate\ (TWh) * daily\ electricity\ consumption (TWh) \tag{2}$$

The determined energy consumption factor will serve as the dependent variable in the proposed study.

The purpose of this study's proposed extra indicator is to improve the precision of estimates regarding the energy usage of Bitcoin mining. This metric incorporates essential blockchain metrics that impact energy consumption, such as hash rate, block size, and transaction volume. This indication offers a more dynamic approach to energy estimation by reflecting fluctuations in network conditions, as opposed to traditional methods that focus exclusively on mining difficulty or electricity costs.

2.2 k- Nearest Neighbors Regression (k-NNR)

k-NNR is one of the most ancient and straightforward regression techniques [26]. k-NNR is a form of supervised learning technique. Supervised learning deduces a function (learner) from a dataset, which comprises a collection of training instances referred to as samples. Each sample consists of a pair including an input vector (instance) and the corresponding output value. Upon completing the training set, the learner aims to accurately ascertain the output for novel situations. [27]. k-NNR generates estimates by analyzing the results of the k nearest neighbors to the specified position. Consequently, to facilitate predictions using k-NNR, a metric for assessing the distance between the query point and instances from the sample cases is required [28]. The method calculates the distance between a query point and all data points in the training set to make a prediction. The Euclidean distance is one of the most prevalent methods for measuring this distance. Therefore, Euclidean distance is utilized in this study. The algorithm determines the k nearest neighbors to the query point utilizing the selected distance measure. In regression, the predictions are the mean of the outcomes of the k nearest neighbors [28]. Equation 3 is the formula for generating predictions.

$$y = \frac{1}{k}\sum_{i=1}^{k} y_i \tag{3}$$

where y is the predicted value, k is the number of neighbors, y_i is the energy consumption value at index i.

First, all features are normalized to a range of 0 to 1 using Min-Max scaling in order to guarantee comparability across variables of different scales. This action enhanced the results' interpretability and lessened bias in the regression models. Network difficulty, daily confirmed transactions, mempool size, average block size, and daily Bitcoin output are utilized as input variables to predict the energy consumption. Using equations 1 and 2, the dependent variable, energy consumption, is calculated by multiplying the daily electricity usage by the overall hash rate. To assess model performance, the dataset is split into subsets for testing (20%) and training (80%). k-NNR is utilized as the modeling technique because of its interpretability and efficacy in analyzing linear relationships. Through trial and error, we determine that the number of neighbors, k, is 8. Figure 2 demonstrates the impact of varying the number of neighbors (k) on the test R^2 score k-NNR across different training set sizes (85, 80, and 75).

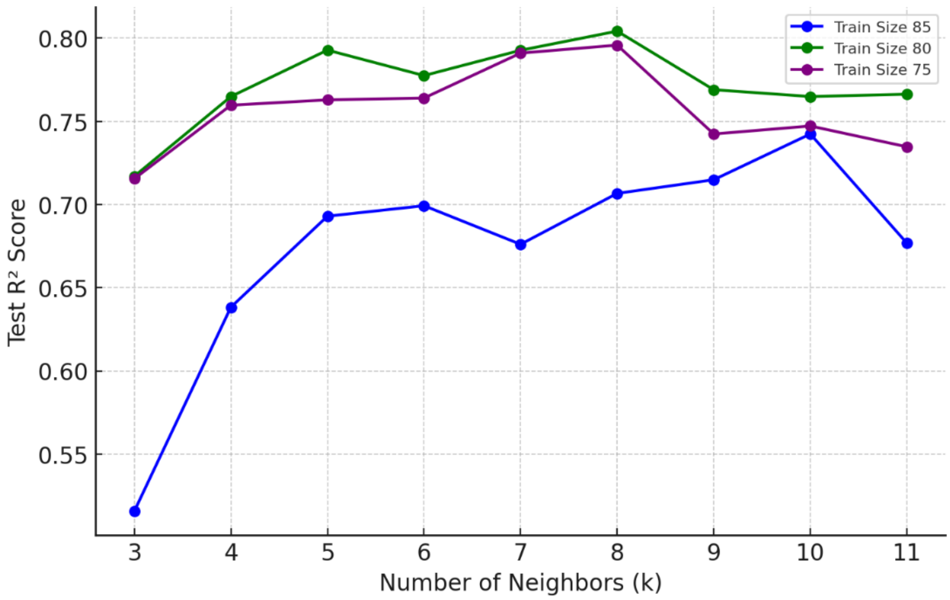

Figure 2. Test R2 score vs. number of neighbors(k) for k-NNR.

This analysis highlights that the optimal number of neighbors for k-NNR lies around k = 7 to 9, depending on the training set size. Based on this trend, we selected k=8 as the optimal value, as it provides a strong balance between generalization and performance across different training sizes.

To ensure the reproducibility of our study, we use a fixed random seed (42) for train-test splits and model training. Features are normalized using Min-Max scaling, and energy cost per transaction is calculated. Python programming software is utilized for analysis. Experiments are performed on a system with an 11th Gen Intel Core i7-11370H processor (3.30GHz) and 16GB RAM.

3 RESULTS AND DISCUSSION

This section delineates the outcomes of the k-NNR model employed to forecast energy consumption in Bitcoin mining. The k-NNR model produced a R^2 (coefficient of determination) value of 0.804, signifying that the model accounts for 80% of the variance in energy consumption. This elevated R^2 indicates the efficacy of the k-NNR model in elucidating the correlation between the chosen features network difficulty, daily confirmed transactions, mempool size, average block size, and daily Bitcoin output and the goal variable, energy consumption. The Mean Squared Error (MSE), which quantifies the average squared deviation between anticipated and actual values, is determined as 0.004. The little error underscores the model's accuracy in predictions and its capacity to generalize effectively to novel inputs. The conjunction of a high R^2 and low MSE substantiates the dependability and precision of the k-NNR model in evaluating energy consumption within blockchain networks. These results confirm the significance of preprocessing operations, including normalization and feature selection, which assured the model efficiently utilized the input data.

The results are corroborated by feature importance analysis and comparisons of anticipated and actual performance. These visuals and data elucidate the model's efficacy and the significance of each feature. Figure 3 depicts the simulated feature importance for the k-NNR model, highlighting the relative contributions of input variables to energy consumption prediction.

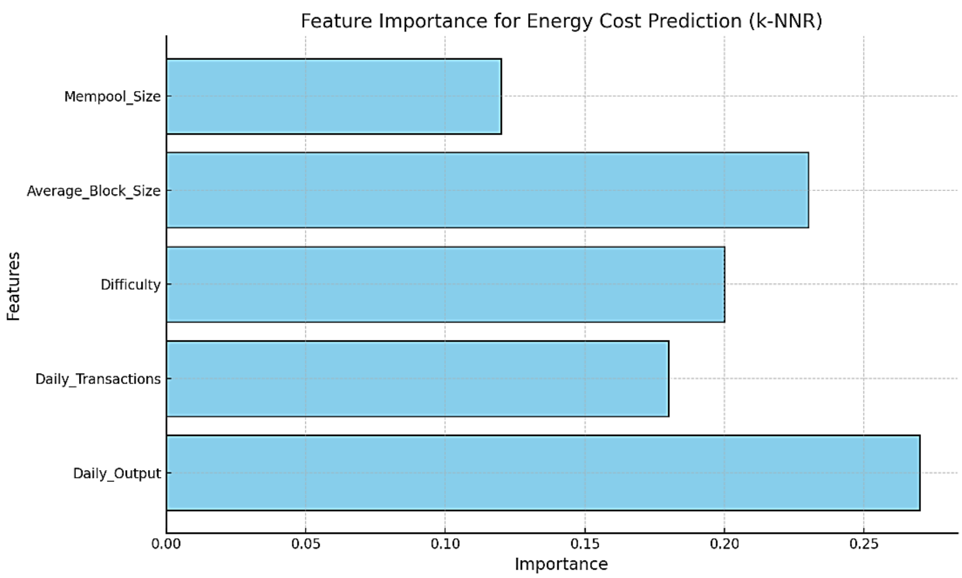

Figure 3. Feature Importance for Energy Consumption Prediction.

Daily Bitcoin output proves to be the most significant variable, underscoring its essential function in influencing energy usage. Average block size and Network difficulty demonstrates substantial contributions, highlighting their influence on mining operations and energy consumption. Daily confirmed transactions and mempool size exhibites moderate significance, indicating their relevance to network activity while demonstrating a lesser direct impact on energy use. Figure 4 displays the scatter plot comparing estimated energy consumption with actual values for the test dataset. The red dashed line denotes the optimal fit where forecasts align precisely with the actual values.

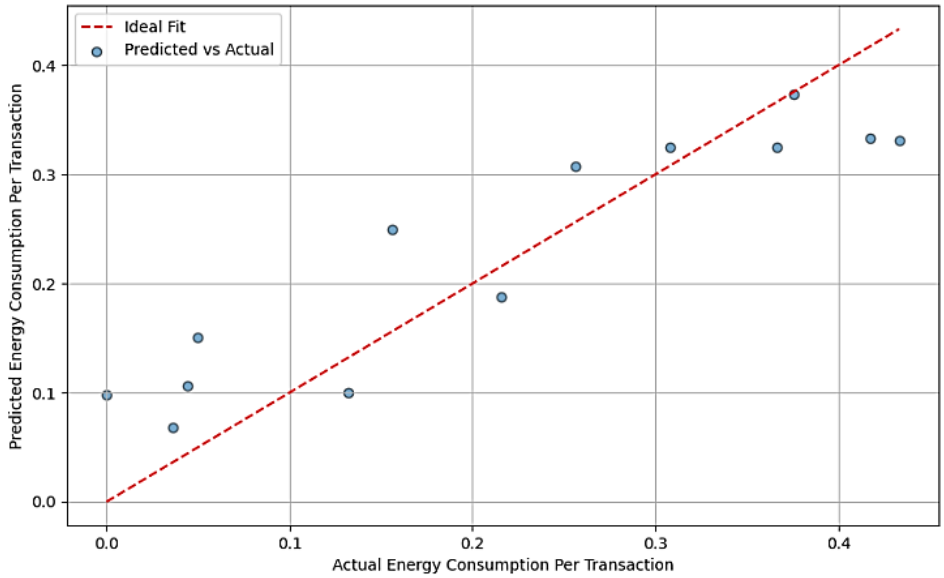

Figure 4. Predicted vs. Actual Energy Consumption (TWh) Per Transaction.

The majority of points are concentrated near the optimal fit line, signifying high prediction accuracy of the model. The R^2 value of 0.804 indicates the model's robust predictive performance, accounting for 80% of the variance in energy consumption. This performance confirms the efficacy of k-NNR regression in examining blockchain data.

3.1 Comparison of various machine learning methods

This study evaluated and contrasted various machine learning algorithms according to their efficacy in regression tasks. Figure 5-8 present a comparative analysis of multiple machine learning models based on their training R^2, test R^2, training MSE, and test MSE across different training sizes.

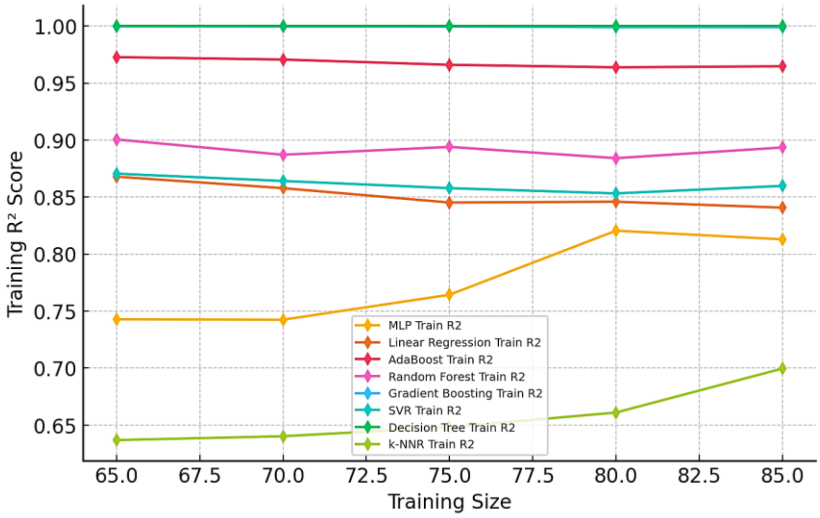

Figure 5. Training R^2 score vs. training size.

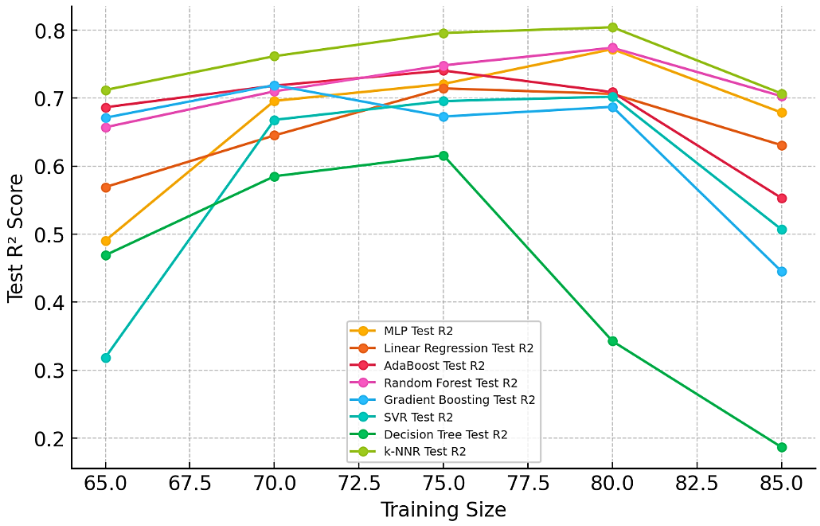

Figure 6. Test R^2 vs. training size.

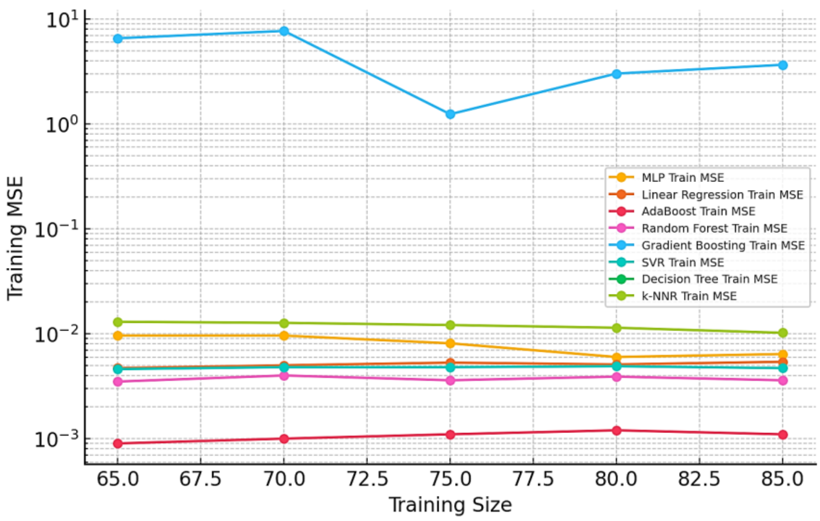

Figure 7. Training MSE vs. training size.

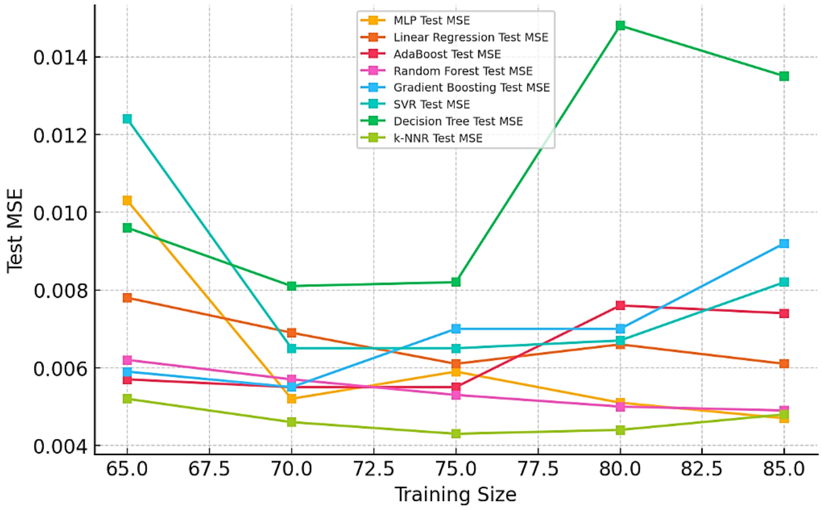

Figure 8. Test MSE vs. training size.

The Decision Tree Regression model attains a R^2 value approaching 1, indicating a significant likelihood of overfitting. This indicates that the model retains training data instead of generalizing patterns. Gradient Boosting Regression shows a large gap between training and test MSE. Gradient Boosting appears to exhibit somewhat elevated test errors across all training sizes, indicating a potential failure to adequately capture the complexity of the data, resulting in under fitting. It appears that Support Vector Regression is unable to adequately capture data variance. Regardless of training amount, k-NNR consistently maintains one of the lowest test MSE values. According to the stability of its test MSE, k-NNR is not severely over fitted or under fitted. k-NNR has a more moderate training R^2 value than Decision Tree, which has an R^2 near 1, indicating overfitting. This indicates that k-NNR is not memorizing the training data, implying it is probably generalizing effectively to novel data. k-NNR constantly ranks among the foremost models regarding test R^2 performance.

The outcomes are presented in Table 3 below, whereby the algorithms are evaluated based on two principal metrics: R^2 and MSE. Elevated R^2 values represent enhanced predictive accuracy, and diminished MSE values indicate reduced prediction mistakes.

Table 3. Comparison of various machine learning algorithms.

Methods	Training Size	Test Size	Training R^2	Test R^2	Training MSE	Test MSE
Multi-layer Perceptron Regression	85	15	0.8129	0.6787	0.0064	0.0047
	80	20	0.8205	0.7721	0.0060	0.0051
	75	25	0.7643	0.7208	0.0081	0.0059
	70	30	0.7424	0.6959	0.0096	0.0052
	65	35	0.7427	0.4904	0.0096	0.0103
Linear Regression	85	15	0.8407	0.6307	0.0054	0.0061
	80	20	0.8459	0.7062	0.0051	0.0066
	75	25	0.8452	0.7143	0.0053	0.0061
	70	30	0.8578	0.6451	0.0050	0.0069
	65	35	0.8679	0.5691	0.0047	0.0078
AdaBoost Regression	85	15	0.9647	0.5529	0.0011	0.0074
	80	20	0.9638	0.7087	0.0012	0.0076
	75	25	0.9660	0.7405	0.0011	0.0055
	70	30	0.9706	0.7183	0.0010	0.0055
	65	35	0.9727	0.6865	0.0009	0.0057
Random Forest Regression	85	15	0.8935	0.7028	0.0036	0.0049
	80	20	0.8841	0.7740	0.0039	0.0050
	75	25	0.8940	0.7482	0.0036	0.0053
	70	30	0.8871	0.7099	0.0040	0.0057
	65	35	0.9005	0.6571	0.0035	0.0062
Gradient Boosting Regression	85	15	0.9989	0.4450	3.6616	0.0092
	80	20	0.9991	0.6871	3.0207	0.0070
	75	25	0.9996	0.6729	1.2383	0.0070
	70	30	0.9997	0.7191	7.7044	0.0055
	65	35	0.9998	0.6711	6.5582	0.0059
Support Vector Regression	85	15	0.8598	0.5067	0.0047	0.0082
	80	20	0.8532	0.7021	0.0049	0.0067
	75	25	0.8578	0.6956	0.0048	0.0065
	70	30	0.8641	0.6680	0.0048	0.0065
	65	35	0.8705	0.3185	0.0046	0.0124
Decision Tree Regression	85	15	1.0	0.1865	0	0.0135
	80	20	1.0	0.3424	0	0.0148
	75	25	1.0	0.6158	0	0.0082
	70	30	1.0	0.5850	0	0.0081
	65	35	1.0	0.4689	0	0.0096
k-NNR	85	15	0.6996	0.7067	0.0102	0.0048
	80	20	0.6610	**0.8042**	0.0114	0.0044
	75	25	0.6481	0.7958	0.0121	0.0043
	70	30	0.6402	0.7617	0.0127	0.0046
	65	35	0.6368	0.7119	0.013	0.0052

Among the assessed approaches, the k-NNR proved to be the most efficient model, attaining the greatest R^2 value of 0.804 and the lowest MSE of 0.004. The results reveal that the k-NNR exhibited enhanced predicted accuracy and reduced error relative to other models, establishing it as the ideal selection for this investigation.

The performance of ensemble models, including Random Forest Regression and AdaBoost Regression, is commendable; yet, their findings are somewhat inferior to those of the k-NNR. Linear Regression, Support Vector Regression, and Multi-layer Perceptron Regression exhibit satisfactory performance but do not attain equivalent accuracy levels. The Gradient Boosting Regression and Decision Tree Regression exhibit inferior performance, with the former demonstrating a comparatively lower R^2 and a larger MSE, while the latter is the least effective of all models. The restricted application of the k-NNR technique in forecasting Bitcoin mining energy usage can be ascribed to various factors. The majority of research in the domain supports tree-based methodologies (Random Forest, Gradient Boosting) and Support Vector Regression because of their capacity to manage intricate nonlinear interactions. This study's findings indicate that k-NNR surpasses previous models, exhibiting a high R^2 of 0.804 and a low MSE of 0.004, thereby establishing it as a formidable alternative. Its efficacy is rooted in its capacity to identify local patterns in energy use, adeptly represent short-term variances, and sustain a robust equilibrium between training and testing performance, hence avoiding the overfitting observed in models such as Decision Tree Regression. Due to its simplicity, interpretability, and robust generalization capabilities, k-NNR offers a largely overlooked but highly effective method for blockchain energy modeling. In conclusion, the comparison analysis identifies the k-NNR as the most effective method for the specified regression problem.

This study's findings highlight numerous innovative additions to Bitcoin energy consumption modeling. This study presents a novel energy consumption indicator that offers a more accurate and dynamic depiction of blockchain energy usage, in contrast to prior research that predominantly depends on economic indicators or long-term aggregated data. The suggested indicator integrates critical blockchain variables—namely hash rate, network difficulty, mempool size, block size, and daily confirmed transactions—to provide a more thorough comprehension of energy consumption trends in Bitcoin mining. This innovative method improves the precision of energy assessments and offers a significant metric for researchers and policymakers seeking to boost energy efficiency in blockchain systems.

Our study presents key innovations and differences relative to current research on k-nearest neighbors (KNN) in bitcoin prediction applications. The following is existing research on KNN in blockchain:

Table 4. Existing research on KNN in bitcoin applications.

Authors	Application	Methods Used	Approach	Key Findings
Chevallier et al. [29]	Bitcoin Price Forecasting	KNN, ANN, SVM, Random Forest, AdaBoost, Ridge Regression	Segmentation of Bitcoin with Alternative Assets	Random Forest and AdaBoost Performed Best
Gu et al. [30]	Investment Model	KNN, ANN, Grey Prediction, LSTM	Comparison of Various Prediction Models	LSTM was Found Superior
Cortez et al. [31]	Crypto vs. Fiat Market Liquidity	KNN, ARMA, GARCH	Bid-Ask Spread Prediction	KNN Performed Better for Short-Term Predictions
Da Silva et al. [32]	Bitcoin Price Forecasting	KNN, SVR, ANN, GLM, Cubist	VMD-STACK Framework for Multi-Step Forecasting	Ensemble Learning Improved Prediction
Mayo & Elgazzar [33]	Cryptocurrency Price Prediction	KNN, ANN, SVM, Naïve Bayes, Random Forest	Analyzing Supply-Side Factors for Prediction	ANN and Random Forest Performed Best
Freeda et al. [34]	Bitcoin Price Forecasting	KNN, Random Forest, Gaussian Naïve Bayes, SVM, RNN	Comparison with Deep Learning Models	RNN Outperformed KNN
Ahmed et al. [35]	Bitcoin Price Prediction	KNN, XGBoost, Gradient Boosting, Random Forest, Linear Regression, SVM	Performance Evaluation Across Models	Gradient Boosting Achieved Highest Accuracy
Benjamin et al. [36]	Crypto Investment Strategy	KNN, Random Forest, Linear Regression	Financial Market Prediction	Random Forest Was More Effective
Jenifel et al. [37]	Bitcoin Price Forecasting	KNN, Linear Regression, Ridge Regression, Decision Tree, Random Forest, SVM, Neural Networks	Performance Analysis of Various ML Models	KNN Was Tested but Not Best Performing
Kawli et al. [38]	Cryptocurrency Price Prediction	KNN, LSTM, Bayesian Regression, SVM, Random Forest	Multi-Asset Price Forecasting	Random Forest and LSTM Achieved Best Results
Akyildirim et al. [39]	Bitcoin Futures Price Prediction	KNN, Logistic Regression, Naïve Bayes, Random Forest, SVM, Extreme Gradient Boosting	High-frequency Intraday Data Analysis	SVM Outperformed KNN in Prediction Accuracy
Li et al. [40]	Stock and Bitcoin Price Forecasting	Mask-LSTM, Mask-BiLSTM, Mask-GRU, KNN	Feature Fusion for Time-Series Prediction	Hybrid Model Outperformed Individual Models

The proposed study differs significantly from existing research by focusing on Bitcoin mining energy consumption rather than price prediction, fraud detection, or transaction classification. Unlike previous studies that rely on market indicators or macroeconomic factors, we incorporate direct blockchain metrics such as hash rate, network difficulty, mempool size, and daily Bitcoin output to develop a more precise energy consumption model. Additionally, while k-NNR has been underutilized in blockchain energy research, our study demonstrates its superior performance over Random Forest, Gradient Boosting, and Support Vector Regression, achieving the highest predictive accuracy ($R^2 = 0.804$, MSE = 0.004).

4 CONCLUSION AND SUGGESTIONS

This research employed the k-NNR model to forecast energy usage in Bitcoin mining, using essential parameters such network difficulty, daily confirmed transactions, mempool size, average block size, and daily Bitcoin output. The findings indicated that the k-NNR model attained robust predictive efficacy, evidenced by a R^2 value of 0.80 and a MSE of 0.004. These results underscore the model's reliability and precision in elucidating the intricate links between input features and energy consumption. The results emphasize the importance of daily Bitcoin production and average block size as key determinants of energy usage, highlighting the relevance of these parameters in mining activities. The study offers significant insights into Bitcoin's energy dynamics, although certain limits should be recognized. The investigation was confined to a 61-day dataset, which, although providing precise short-term insights, may not reflect longer-term trends or anomalies in blockchain activity. Although the 61 days dataset provides sufficient insights for short-term trend analysis, longer datasets may be required to capture seasonal and long-term fluctuations in Bitcoin mining energy consumption. The analysis presupposes that mining equipment and energy efficiency are constant, perhaps failing to capture real-world fluctuations. Further study can address these shortcomings by augmenting the dataset with longitudinal data, integrating fluctuations in mining hardware efficiency, and evaluating supplementary machine learning techniques. This study's findings establish a robust foundation for formulating energy-efficient solutions in bitcoin mining and directing sustainable blockchain operations.

Future research could enhance forecast accuracy by merging real-time blockchain indicators with external economic and environmental variables. Subsequent research could investigate the effects of alternate consensus processes, such as proof-of-stake, on energy efficiency. Enhancing the comparison study with deep learning models could provide more profound insights into energy use trends.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

Artificial Intelligence (AI) Contribution Statement

This manuscript was entirely written, edited, analyzed, and prepared without the assistance of any artificial intelligence (AI) tools. All content, including text, data analysis, and figures, was solely generated by the authors.

Contributions of the Authors

Nazmiye Eligüzel: Performed the computations, developed the model, wrote the manuscript.

Sena Aydoğan: Collected the data, reviewed the literature, revised the manuscript.

REFERENCES

[1] S. Nakamato, "Bitcoin: A peer-to-peer electronic cash system," *Decentralized Bus. Rev.*, 2008.

[2] H. Alshahrani *et al.*, "Sustainability in Blockchain: A Systematic Literature Review on Scalability and Power Consumption Issues," *Energies*, vol. 16, no. 3, 2023, doi: 10.3390/en16031510.

[3] N. Eligüzel, "An analysis of the integration of sustainability concepts into blockchain technology," *Int. J. Appl. Methods Electron. Comput.*, vol. 11, no. 3, pp. 158–164, 2023, doi: 10.58190/ijamec.2023.43.

[4] N. Sapra and I. Shaikh, "Impact of Bitcoin mining and crypto market determinants on Bitcoin-based energy consumption," *Manag. Financ.*, vol. 49, no. 11, pp. 1828–1846, 2023, doi: 10.1108/MF-03-2023-0179.

[5] V. Kohli, S. Chakravarty, V. Chamola, K. S. Sangwan, and S. Zeadally, "An analysis of energy consumption and carbon footprints of cryptocurrencies and possible solutions," *Digit. Commun. Networks*, vol. 9, no. 1, pp. 79–89, 2023, doi: 10.1016/j.dcan.2022.06.017.

[6] S. Küfeoğlu and M. Özkuran, "Energy Consumption oF Bitcoin Mining," in *Cambridge Working Papers in Economics: 1948*, .

[7] M. Maiti, "Dynamics of bitcoin prices and energy consumption," *Chaos, Solitons Fractals X*, vol. 9, p. 100086, 2022, doi: 10.1016/j.csfx.2022.100086.

[8] A. de Vries, "Bitcoin's energy consumption is underestimated: A market dynamics approach," *Energy Res. Soc. Sci.*, vol. 70, no. August, p. 101721, 2020, doi: 10.1016/j.erss.2020.101721.

[9] D. Das and A. Dutta, "Bitcoin's energy consumption: Is it the Achilles heel to miner's revenue?," *Econ. Lett.*, vol. 186, p. 108530, 2020, doi: 10.1016/j.econlet.2019.108530.

[10] J. Li, N. Li, J. Peng, H. Cui, and Z. Wu, "Energy consumption of cryptocurrency mining: A study of electricity consumption in mining cryptocurrencies," *Energy*, vol. 168, pp. 160–168, 2019, doi: 10.1016/j.energy.2018.11.046.

[11] M. Kevser, "Bitcoin as an Alternative Financial Asset: Relations Between Geopolitical Risk, Global Economic Political Uncertainty, and Energy Consumption," *PAMUKKALE J. EURASIAN Socioecon. Stud.*, vol. 9, no. 2, pp. 117–131, 2022.

[12] K. Tissaoui, T. Zaghdoudi, S. Boubaker, B. Hkiri, and M. Talbi, "Testing the Nonlinear Long- and Short-Run Distributional Asymmetries Effects of Bitcoin Prices on Bitcoin Energy Consumption: New Insights

through the QNARDL Model and XGBoost Machine-Learning Tool," *Energies*, vol. 17, no. 2810, 2024.

[13] N. Sapra, I. Shaikh, D. Roubaud, M. Asadi, and O. Grebinevych, "Uncovering Bitcoin's electricity consumption relationships with volatility and price: Environmental Repercussions," *J. Environ. Manage.*, vol. 356, no. January, p. 120528, 2024, doi: 10.1016/j.jenvman.2024.120528.

[14] A. Syzdykova, "Bitcoin Mining and Energy Consumption of Bitcoin," *J. Econ. Soc. Res.*, vol. 10, no. 19, pp. 1–15, 2023.

[15] T. Zaghdoudi, K. Tissaoui, M. Maâloul, Y. Bahou, and N. Kammoun, "Can Economic, Geopolitical and Energy Uncertainty Indices Predict Bitcoin Energy Consumption? New Evidence from a Machine Learning Approach," *Energies*, vol. 12, no. 3245, 2024.

[16] Y. Bublyk, O. Borzenko, and A. Hlazova, "Cryptocurrency energy consumption: Analysis, global trends and interaction," *Environ. Econ.*, vol. 14, no. 2, pp. 49–59, 2023, doi: 10.21511/ee.14(2).2023.04.

[17] A. O. Adewuyi, B. A. Wahab, A. K. Tiwari, and H. X. Do, "Do bitcoin electricity consumption and carbon footprint exhibit random walk and bubbles? Analysis with policy implications," *J. Environ. Manage.*, vol. 367, no. August, p. 121958, 2024, doi: 10.1016/j.jenvman.2024.121958.

[18] "Blockchain.com," 2024. https://www.blockchain.com/explorer/charts (accessed Dec. 14, 2024).

[19] D. Kim, D. Ryu, and R. I. Webb, "Does a higher hashrate strengthen Bitcoin network security?," *Financ. Innov.*, vol. 10, no. 1, 2024, doi: 10.1186/s40854-023-00599-8.

[20] X. Zhang, R. Qin, Y. Yuan, and F. Y. Wang, "An Analysis of Blockchain-based Bitcoin Mining Difficulty: Techniques and Principles," *Proc. 2018 Chinese Autom. Congr. CAC 2018*, pp. 1184–1189, 2018, doi: 10.1109/CAC.2018.8623140.

[21] A. Mikhaylov, H. Dinçer, S. Yüksel, G. Pinter, and Z. A. Shaikh, "Bitcoin mempool growth and trading volumes: Integrated approach based on QROF Multi-SWARA and aggregation operators," *J. Innov. Knowl.*, vol. 8, no. 3, 2023, doi: 10.1016/j.jik.2023.100378.

[22] B. Aygün and H. Arslan, "Block size optimization for PoW consensus algorithm based blockchain applications by using whale optimization algorithm," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 30, pp. 406–419, 2022, doi: 10.3906/elk-2105-217.

[23] M. Karimuzzaman, S. Afroz, M. M. Hossain, and A. Rahman, "Modelling COVID-19 cases and deaths with climate variables using statistical and data science methods," *Soft Comput.*, vol. 28, no. 21, pp. 12561–12574, 2024, doi: 10.1007/s00500-024-10352-7.

[24] A. Vij, K. Saxena, and A. Rana, "Prediction in Stock Price Using of Python and Machine Learning," *2021 9th Int. Conf. Reliab. Infocom Technol. Optim. (Trends Futur. Dir. ICRITO 2021*, pp. 1–4, 2021, doi: 10.1109/ICRITO51393.2021.9596513.

[25] "Cambridge Bitcoin Electricity Consumption Index," *University of Cambridge*, 2024. https://ccaf.io/cbnsi/cbeci (accessed Dec. 27, 2024).

[26] B. Nguyen, C. Morell, and B. De Baets, "Large-scale distance metric learning for k-nearest neighbors regression," *Neurocomputing*, vol. 214, pp. 805–814, 2016, doi: 10.1016/j.neucom.2016.07.005.

[27] Y. Song, J. Liang, J. Lu, and X. Zhao, "An efficient instance selection algorithm for k nearest neighbor regression," *Neurocomputing*, vol. 251, pp. 26–34, 2017, doi: 10.1016/j.neucom.2017.04.018.

[28] J. Tanuwijaya and S. Hansun, "LQ45 stock index prediction using k-nearest neighbors regression," *Int. J. Recent Technol. Eng.*, vol. 8, no. 3, pp. 2388–2391, 2019, doi: 10.35940/ijrte.C4663.098319.

[29] J. Chevallier, D. Guégan, and S. Goutte, "Is It Possible to Forecast the Price of Bitcoin?," *Forecasting*, vol. 3, no. 2, pp. 377–420, 2021, doi: 10.3390/forecast3020024.

[30] A. Gu *et al.*, "Empirical Research for Investment Model Based on VMD-LSTM," *Math. Probl. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/4185974.

[31] K. Cortez, M. D. P. Rodríguez-García, and S. Mongrut, "Exchange market liquidity prediction with the k-nearest neighbor approach: Crypto vs. fiat currencies," *Mathematics*, vol. 9, no. 1, pp. 1–15, 2021, doi: 10.3390/math9010056.

[32] R. G. Da Silva, M. H. Dal Molin Ribeiro, N. Fraccanabbia, V. C. Mariani, and L. Dos Santos Coelho, "Multi-step ahead Bitcoin Price Forecasting Based on VMD and Ensemble Learning Methods," *Proc. Int. Jt. Conf. Neural Networks*, 2020, doi: 10.1109/IJCNN48605.2020.9207152.

[33] D. Mayo and H. Elgazzar, "Predicting Cryptocurrency Price Change Direction from Supply-Side Factors via Machine Learning Methods," *2022 IEEE World AI IoT Congr. AIIoT 2022*, pp. 330–336, 2022, doi: 10.1109/AIIoT54504.2022.9817249.

[34] S. E. Freeda, T. C. E. Selvan, and I. G. Hemanandhini, "Prediction of Bitcoin Price using Deep Learning Model," *Proc. 5th Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2021*, no. Iceca, pp. 1702–1706, 2021, doi: 10.1109/ICECA52323.2021.9676048.

[35] F. U. Ahmed, M. Ahmed, F. H. Mahi, S. H. Abdullah, and S. A. Suha, "A Comparative Performance Evaluation of Bitcoin Price Prediction Using Machine Learning Techniques," *2023 Int. Conf. Inf. Commun. Technol. Sustain. Dev. ICICT4SD 2023 - Proc.*, pp. 194–198, 2023, doi: 10.1109/ICICT4SD59951.2023.10303490.

[36] J. J. Benjamin, R. Surendran, and T. Sampath, "A professional strategy for Bitcoin and Ethereum using Machine Learning for Investors," *4th Int. Conf. Inven. Res. Comput. Appl. ICIRCA 2022 - Proc.*, no. Icirca, pp. 798–804, 2022, doi: 10.1109/ICIRCA54612.2022.9985532.

[37] M. Geetha Jenifel, R. Anita Jasmine, and D. Umanandhini, "Bitcoin Price Predictive Dynamics Using Machine Learning Models," *2024 15th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2024*, pp. 1–6, 2024, doi: 10.1109/ICCCNT61001.2024.10725542.

[38] D. Kawli, A. Chaudhari, P. Ingale, G. Telange, and A. Banik, "Crypto-Visionary Price Forecasting System," *Proc. 5th Int. Conf. Smart Electron. Commun. ICOSEC 2024*, no. Icosec, pp. 1645–1650, 2024, doi: 10.1109/ICOSEC61587.2024.10722487.

[39] E. Akyildirim, O. Cepni, S. Corbet, and G. S. Uddin, "Forecasting mid-price movement of Bitcoin futures using machine learning," *Ann. Oper. Res.*, vol. 330, no. 1–2, pp. 553–584, 2023, doi: 10.1007/s10479-021-04205-x.

[40] Y. Li *et al.*, "MLBGK: A Novel Feature Fusion Model for Forecasting Stocks Prices: MLBGK: A Novel Feature Fusion Model…: Y. Li et al.," *Comput. Econ.*, no. 0123456789, 2024, doi: 10.1007/s10614-024-10796-x.