



# Insights into Bitcoin and energy nexus. A Bitcoin price prediction in bull and bear markets using a complex meta model and SQL analytical functions

Adela Bâra<sup>1</sup> · Simona-Vasilica Oprea<sup>1</sup> · Mirela Panait<sup>2,3</sup>

Accepted: 18 April 2024  
© The Author(s) 2024

## Abstract

Cryptocurrencies are in the center of attention of investors, public authorities and researchers, but the interest has shifted from purely financial aspects regarding the way of trading, lack of regulation and supervision of transactions, volatility, correlation with other assets to aspects related to sustainability taking in account the high energy consumption generated by the mining process and the impact on environmental pollution. Bitcoin was chosen for the research considering the dominance that this financial asset has on the cryptocurrency market and its position as alpha currency. The article focuses on the relationship between Bitcoin transactions and energy consumption, for period 1st January 2019—31st of May 2022, this interval having significant price movements. The authors made a prediction of the Bitcoin price using a complex meta-model and SQL analytical functions. The analysis is based on 15 fundamental variables in order to forecast the price: Bitcoin data (prices and volume), electricity price and traded quantity on day-ahead market (DAM), gas price and traded quantity on DAM, inflation in EU, EU-ETS emissions certificates and oil prices. The study reveals the importance of the relationship Bitcoin—energy—carbon emissions, elements that capture the impact of the mining process on the environment from the perspective of energy consumption. Investors on the Bitcoin market must be aware not only of the importance of financial aspects on the price of cryptocurrencies (inflation, demand, offer), but also of other elements related to the evolution of energy prices (electricity, oil, gas, renewable energy) and the evolution of emissions certificates prices. Considering the promotion of the principles of sustainable development on the capital market, portfolio investors have become increasingly attentive to the social and environmental performance of financial assets. This study aims to make financial market players aware of the non-financial implications of their transactions. In addition, the energy transition and the reconfiguration of the energy mix are elements of impact on the cryptocurrency market through the technical levers involved in the mining process.

**Keywords** Bitcoin price prediction · Meta-model · Analytical functions · Bull/bear markets · Energy data

✉ Simona-Vasilica Oprea  
simona.oprea@csie.ase.ro

Adela Bâra  
bara.adela@ie.ase.ro

Mirela Panait  
mirela.matei@upg-ploiesti.ro

<sup>1</sup> Bucharest University of Economic Studies, Department of Economic Informatics and Cybernetics, no. 6 Piața Romană, Bucharest, Romania

<sup>2</sup> Petroleum-Gas University of Ploiești, Department of Cybernetics, Economic Informatics, Finance and Accountancy, no. 39 București Blvd, Ploiești, Romania

<sup>3</sup> Institute of National Economy, Romanian Academy, Calea Victoriei, Romania

## Variables

$C_{BTC}^h$	Close price of BTC – target for prediction
$H_{BTC}^{h-24}$	High price of BTC
$I_{EU}^{h-24}$	Inflation in EU
$L_{BTC}^{h-24}$	Low price of BTC
$N_{BTC}^{h-24}$	Number of trades
$O_{BTC}^{h-24}$	Open price of BTC
$P_{EUETS}^{h-24}$	EU-ETS emissions certificates prices
$P_{El}^{h-24}$	Electricity price on DAM
$P_{Gas}^{h-24}$	Gas price on DAM
$P_{Oil}^{h-24}$	Oil price
$QAV_{BTC}^{h-24}$	Quote Asset Volume
$Q_{El}^{h-24}$	Electricity traded quantity on DAM
$Q_{Gas}^{h-24}$	Gas traded quantity on DAM
$TBV_{BTC}^{h-24}$	Taker base volume
$TQV_{BTC}^{h-24}$	Taker quote volume

$V_{BTC}^{h-24}$	Total volume
$h$	Hour
$k$	Iterator of the time series
$n$	Number of rows
$p$	Stacking predictions
$q$	Window size
$T$	Training interval
$t$	Last price of the time series set to $h - 24$
$wd$	Weekday
$y$	Year
$F1 \div F15$	Analytical features
$m$	Month
$h-24$	Denotes the shifted values by 24 h from the target

### Acronyms

$R^2$	Coefficient of determination
BA	Base Asset
BTC	Bitcoin
DAM	Day-Ahead Market
ETS	Emissions Trading System
HGB	Histogram-based Gradient Boosting
LGB	Light Gradient Boosting
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
QA	Quote Asset
RF	Random Forest
RMSE	Root-Mean Squared Error
VR	Voting Regressor
XGB	EXtreme Gradient Boosting

## 1 Introduction

The financial market is in a complex process of metamorphosis after the appearance of cryptocurrencies. These financial assets have aroused the interest of portfolio investors both from the perspective of the speculative opportunities they offer but also due to the possibilities of use for hedging strategies and diversification of portfolios of classic financial instruments Salisu and Obiora, [38]). The scale of cryptocurrency transactions, the decentralized trading mechanism, the dramatic fluctuations recorded over time but also the lack of regulation have determined the increase in attention for this market from portfolio investors, public authorities, researchers and other categories of stakeholders Bourriet et al. [10]; Fang et al. [19]; García-Corral et al. [21]; Rao et al. [37].

Investors' interest in this financial asset has gradually made cryptocurrencies, and in particular Bitcoin, acquire the classic functions of money, namely currency of exchange, standard of value, currency of payment and Bitcoin as a

medium of exchange, a unit of account and a store of value. Researchers' interest in cryptocurrencies has increased over time, being addressed specific aspects such as the relationship between financial assets—cryptocurrencies, market efficiency of the crypto market, financing risks, market efficiency Tiwari et al. [49]; Şcheau et al. [44]; Wu et al. [59]; Aivaz et al. [5], investors herding behavior, cryptocurrency as safe haven, blockchain technologies, need to regulate transactions, money laundry and financing terrorism Alexandros, [3]; Dzidzikashvili and Kheladze, [16]; Şcheau and Achim, [43]; Wang et al. [57].

The emergence of cryptocurrencies was generated by two major trends recorded in the international economy, namely the digital revolution and the scandals and different crises that shook the international financial system and that weakened the confidence of investors and financial consumers in banking institutions in particular (central banks and commercial and investment banks). Cryptocurrency is a digital currency that is not issued by a central bank and is not based on users' trust in a certain financial entity, because it uses a decentralized ledger system (blockchain technology) and involves a mining process based on solving some hash problems by the computer. This specific Bitcoin mining process is similar to the process of gold extraction Bjerg [8]; Zimmer, [56]; Bibi, [9]; Sapra, N., and Shaikh [39] and investors' interest in this financial asset is reminiscent of the nineteenth century gold rush. The parallel between gold and cryptocurrencies is not limited only to the impact on the financial market, but also to the externalities generated on the labor market and the environment Maurushat and Halpin, [34]; Dennin, [13]. People's greed and the desire to obtain spectacular gains in very short periods of time have generated the intensification of the cryptocurrency production process and the ignoring of the negative externalities generated by the mining process (high energy consumption, carbon emissions, electronic waste and consistent frauds). In order to reduce the negative impact on the environment and society, however, concerted actions on the part of the authorities, private companies and portfolio investors are necessary. Crypto-culture encouraged by technical progress that allows novice investors to make transactions with the help of a mobile phone generates fraud among the population with a low level of financial education, and the social repercussions are dramatic Savona, [42]; Dulisse et al. [17]; Biju and Thomas, [7].

The mining is to excavation of cryptocurrencies “through the mathematical operations of a mining machine in the network (i.e., “hash rate”)” Yuan et al. [53]. The mining process requires considerable computational effort, specifically electricity and CPU time and the miners receive a fee as cryptocurrency. Validation of transactions is decentralized, and blockchain technology also offers the advantage of anonymity of the cryptocurrency owners and their transactions.

Launched in 2009, after the international financial crisis, Bitcoin was received with interest considering the failure of public financial policies and the activity of banks that led to the international financial crisis of 2008–2009. Its evolution started to become more and more interesting as certain psychological thresholds were reached, namely the threshold of 1000 USD in 2017, so that by the end of this year it would reach a value of 14,000. Another event of the black swan type, namely the COVID-19 crisis has generated an increasing interest of investors in cryptocurrencies Akbulaev and Abdulhasanov, [2]; Bibi, [9] and digital payments. Despite the reluctance and warnings issued by some central banks such as the Bank of England that drew attention to the risks involved in the use of cryptocurrencies to the detriment of official currencies, some countries, such as El Salvador and South Africa, have adopted Bitcoin as a legal tender and payment instrument. Among all cryptocurrencies, Bitcoin is by far the most used financial asset of this type, maintains the lion's share of market (Symitsi and Chalvatzis, [46]; Mohsin et al. [33]), which is why this currency was chosen to carry out this study.

In addition to the purely financial aspects that cryptocurrency transactions entail, namely the risks generated by the lack of regulation and supervision, the safe haven status, speculative operations, researchers have also begun to analyze the relationship between Bitcoin and sustainable development and the implications that the mining process has on the environment from the perspective of energy consumption and electronic waste De Vries, [12]; Erdogan et al. [18]. The topics addressed are related to aspects such as Environment, Social and Governance (ESG) aspects, and energy transition Sundaram, [45]; Toumi et al. [50]. There are opinions regarding the positive impact of cryptocurrency transactions on accelerating the improvement of the energy mix by increasing the share of renewable energy Bastian-Pinto et al. [6]; Kumar, [28]; Yüksel et al. [54]; Miśkiewicz et al. [32]; Zheng et al. [55]. The problem of energy consumption generated by cryptocurrency transactions is also of interest from the perspective of central banks that are looking for technical solutions regarding future payment systems. The banking authorities are committed to promoting the principles of digital development, which is why reducing the energy consumption generated by payment systems is a challenge and future digital currencies can use certain technical elements specific to cryptocurrencies Derbali, et al. [14]; Agur et al. [1].

Our research reflects a shifting focus in cryptocurrency research from primarily financial to environmental sustainability concerns. Recent technological advancements like the multi-attention fusion residual convolutional neural network (MAR-CNN) and the deep variational autoencoder (DVAE) are relevant to this type of research. MAR-CNN is an advanced form of CNN that integrates multiple attention

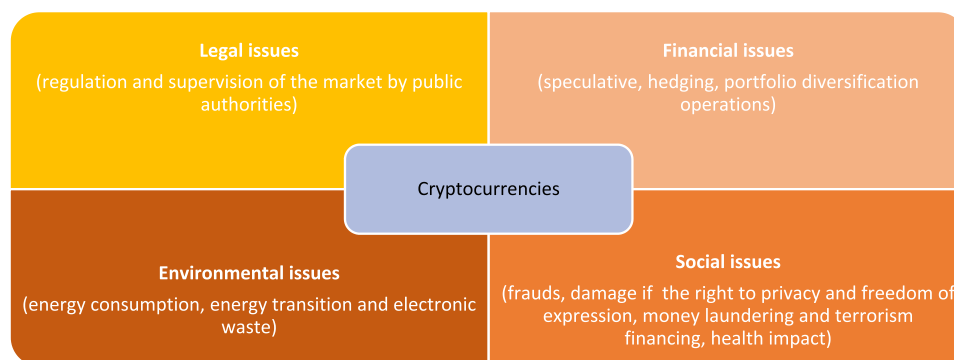
mechanisms with residual learning. The attention mechanisms help the network focus on specific features within the data that are most relevant for the task at hand, potentially improving accuracy and efficiency in pattern recognition. In residual learning, the network also learns from the modification of inputs through layers, which helps in combating the problem of vanishing gradients in deep neural networks. As for its applications in cryptocurrency research, MAR-CNN could be used to more accurately model the complex relationships and patterns between various inputs affecting Bitcoin mining energy consumption, such as electricity price, hardware efficiency and transaction volume. Further, it may enhance price prediction models by identifying and focusing on the most impactful variables from historical data, potentially leading to more accurate and robust predictive outcomes.

A DVAE is a type of generative model that learns to represent high-dimensional data in a lower-dimensional latent space. It is used to generate new data instances similar to those on which the model was trained. Variational autoencoders are particularly known for their ability to handle complex distributions and to generate new instances while controlling for specific factors in the data. Its applications in cryptocurrency research are in simulation and forecasting, and anomaly detection. By learning the distribution of variables affecting Bitcoin's energy consumption, a DVAE may simulate various future scenarios under different conditions, aiding in forecasting and strategic planning. DVAEs could be deployed to detect unusual patterns in energy usage that might indicate inefficiencies or changes in mining technology or strategy.

Given the complex interplay between Bitcoin transactions, energy consumption and carbon emissions, leveraging MAR-CNN and DVAE may significantly enhance the accuracy and depth of analysis. For instance, MAR-CNN could refine the analysis of how specific market variables (like gas prices or inflation) directly impact Bitcoin's energy consumption by focusing attention on the most relevant features during periods of significant price movement, while DVAE could generate potential future scenarios based on past data, helping to understand how changes in energy prices or regulation might impact Bitcoin mining.

This article focuses on the relationship between Bitcoin transactions and energy consumption. The authors propose a prediction method of the Bitcoin price using a complex Meta-Model (MM) and SQL analytical functions. Building a MM that stacks several ML models regressors as its sub-models brings the following advantages: i) diversity of models—combining different types of models can capture a wide range of patterns in the data and leads to better performance than any single model; ii) error reduction—the MM can learn to correct the errors of the individual sub-models, leading to improved accuracy; iii) robustness—the MM is

Fig. 1 Main issues



more robust to overfitting, especially if the sub-models are diverse. Predicting Bitcoin prices is a complex task due to the high volatility and unpredictability of the cryptocurrency market. By using a stacked model, the current approach leverages the strengths of each individual model. For instance, RF captures non-linear relationships, XGB handles outliers well and LGB works efficiently with large datasets.

One contribution of this paper consists of building a dataset that consists of Bitcoin data, energy data and other macroeconomics indicators (the realization of this data set is to be appreciated considering that the data have been collected from several sources, pre-processed and stored into a relational database, which will later allow their use within specific statistical models). Taking in account the plethora of scientific methods like MAR-CNN Yan et al. [51] and the DVAE Yan et al. [52] as potential price predictive or detection model, which can be used as an interdisciplinary application, in this paper we propose a prediction method that embeds the first three categories of the ensemble techniques: bagging (RF), boosting (HGB, LGB, XGB) and voting (VR). The fourth category—stacking is included to use the predictions of the five models as input to another model (that is known as meta-model) that learns from the results of the base models. Another contribution of this paper consists of building features using SQL analytic functions.

This article is structured on specific sections that ensure the presentation of international concerns regarding cryptocurrencies from a financial perspective, but also from a sustainable development view, the authors presenting the results of the main studies from the international scientific literature that allow identifying the research gap.

## 2 Literature review

The scale of cryptocurrency transactions and the interest of portfolio investors for both speculative and hedging operations were accompanied by the publication of scientific studies that focused on certain aspects related (Fig. 1) to the need to regulate transactions, the status of safe haven

for these financial assets or the impact of the Covid-19 crisis on this segment of the financial market Al-Shboul et al. [4]; Wen et al. [58]; Wang et al. [57]. In addition to the strictly financial and legal aspects of these cryptocurrencies, certain studies are focused on the impact that transactions with cryptocurrencies have on sustainable development (by energy consumption, energy transition and electronic waste perspectives), considering the mining process specific to these financial assets. Based on these considerations, certain studies focused the analysis of related risks of conduction mechanism between energy and Bitcoin markets Küfeoğlu and Özkuran, [27]; Li et al. [29]; Gurrib et al. [23]; Yuan et al. [53]. The lack of regulation and supervision accompanied by people's greed have generated numerous frauds that have mainly affected incipient investors, lacking a solid financial culture Maurushat and Halpin, [34]; Savona, [42]; Thakur, [48].

Cryptocurrency transactions have become a challenge for the energy transition process considering the high energy consumption generated by the mining process Goodkind et al. [22]; Nández Alonso et al. [35]. In addition, in order to face the increasingly intense competition on this market, players need more and more powerful computers, and giving up old computers fuels the waste process and generates new problems related to e-waste management. Having the goal of achieving carbon neutralization assumed by countries through various international treaties, some economies such as China have declared cryptocurrency transactions illegal Yuan et al. [53] because carbon footprints are considerable, the mining process being based on fossil fuels Foteinis, [20], Sarkodie and Owusu, [41], Long et al. [30]; Pagone et al. [36].

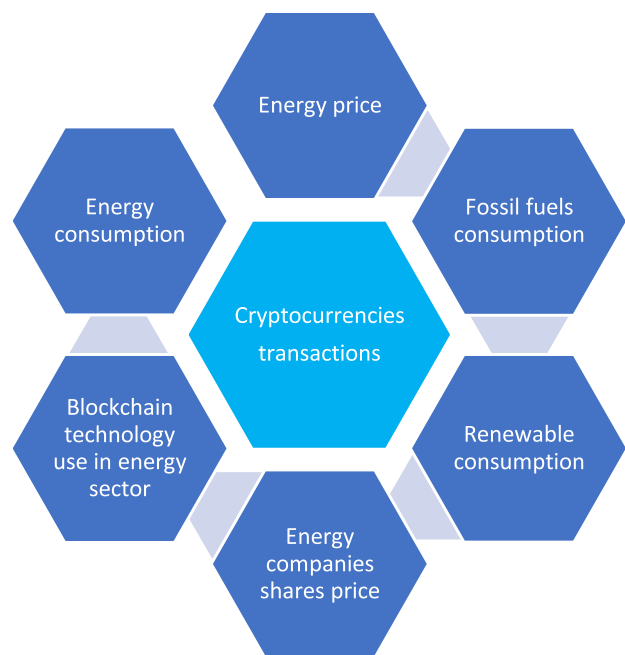
Using the quantile connectedness method, Yuan et al. [53] focused on their study on spillover effects among Bitcoin price, hash rate, electricity demand and energy consumption. The researchers draw the conclusions that there is strong relationship between energy consumption and the Bitcoin market, and the risk infection mechanism makes its presence felt. Certain studies have focused on evaluating the energy consumption generated by the mining process

specific to cryptocurrencies, with alarming results. Thus, the study carried out by Kohli et al. [26] energy consumption generated by Bitcoin is similar with Sweden and by Ethereum is nearly identical as Romania.

Using a VAR-AGARCH model, Symitsi and Chalvatzis, [46] studied the spillover effects of Bitcoin on clean energy, fossil fuel energy companies and technology companies. The results of this research suggested that there are long-run volatility effects of Bitcoin on energy companies and significant return spillovers from energy to this cryptocurrency. Maiti [31], using a discrete threshold regression mode, focused the research on Bitcoin prices and Bitcoin energy consumption relationship in the interval November 2010 and October 2021. This scholar demonstrated that the impact of energy consumption on cryptocurrency prices is not uniform for the analyzed period.

Corbet et al. [11] focused their study on Bitcoin-energy markets relationship taking in account the growing interest of portfolio investors for cryptocurrencies and difficulties noticed in the mining of these assets. The conclusion of the study is that Bitcoin transactions have an impact on large electricity and utilities markets. So, the demand for energy and emissions of carbon dioxide are rising, these results being registered also by Dogan et al. [15]. Based on bootstrap Granger causal relationship tests, Su et al. [47] investigate the influences between Bitcoin and oil markets. The main conclusions of the study are that investors speculate on the relationship between these two markets in order to build diversified portfolios that ensure risk reduction. Similar results were obtained by Huynh et al. [25], using transferring entropy method for US and European crude oil indices and cryptocurrencies. In addition, the blockchain technology used by Bitcoin, as an essential component of the Fourth Industrial Revolution, can be solution for new technological strategies to promote the cut of international transaction fees.

Using a time-varying Granger causality test for period Sept 17, 2014, to October 12, 2021, Dogan et al. [15] explored the relationship between Bitcoin, energy and carbon emissions in order to analyze the impact of this cryptocurrency on energy transition and the environment. The causal relationship identified between Bitcoin, energy and carbon emissions is essential for tailoring public policies not only on the financial market but also on other components related to the twin transition. This study highlights the need to decarbonize cryptocurrencies using clean energy in mining process, but also to accelerate the digitalization of the energy sector, which can be based on blockchain technology. Sarker et al. [40] used nonlinear ARDL method and the Granger causality test for period 2013–2021 to analyze the asymmetric effects of climate policy uncertainty and energy prices on Bitcoin prices. Researchers draw attention to the impact that price fluctuations on the energy market have on

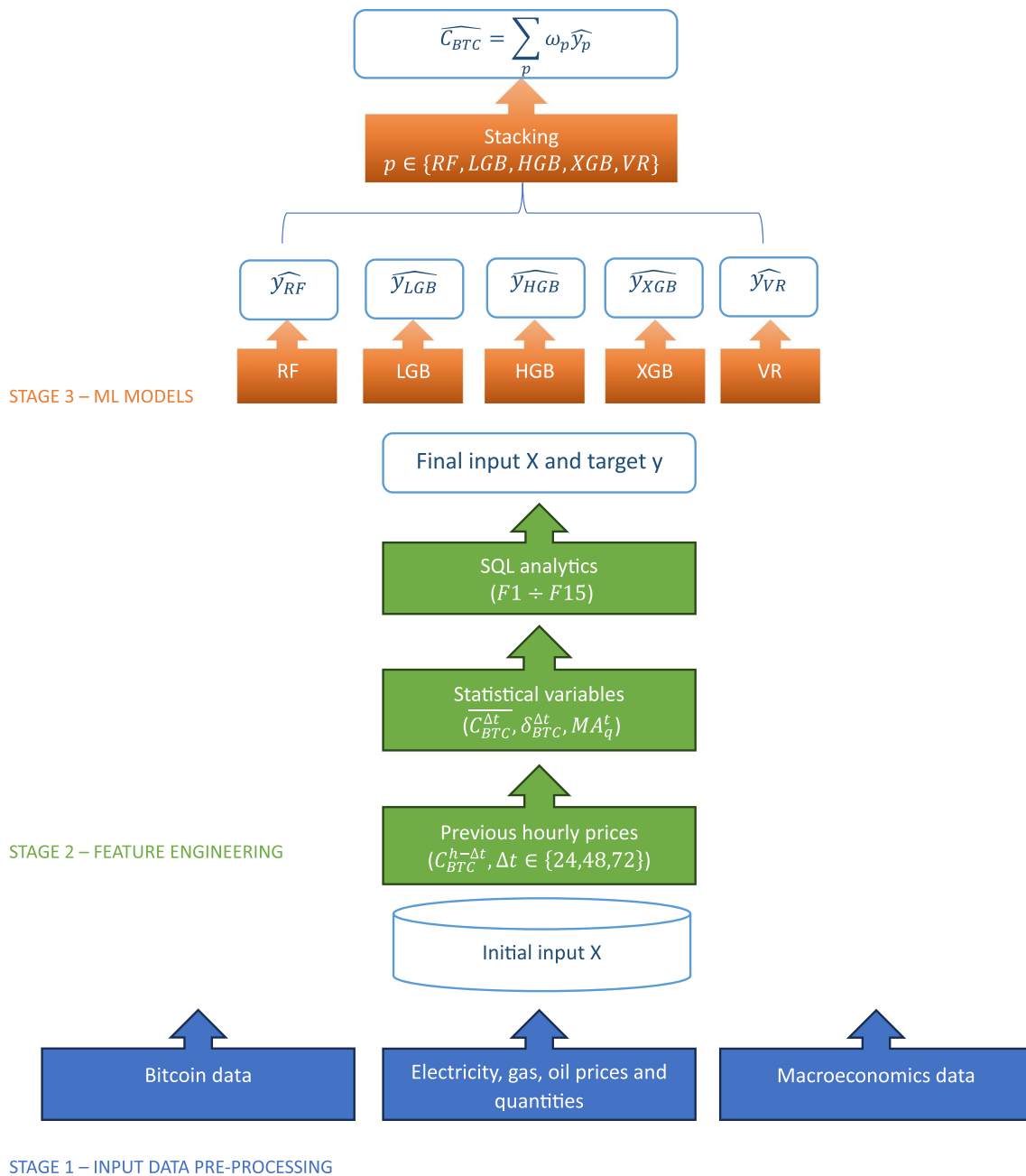


**Fig. 2** Main implications of cryptocurrencies transactions on energy sector

Bitcoin, that is an energy intensive asset. Investors in the cryptocurrency market should be much more attentive to developments on the energy market to better anticipate the trend of these financial assets price.

So, taking in account the results of these scientific studies, the analysis of the Bitcoin-energy relationship is of particular interest not only from the perspective of the connections that exist between these strategic assets, but also from the point of view of the impact that cryptocurrency transactions have on the energy market (on the price of certain assets and on the price of shares issued by energy companies) and energy transition process (Fig. 2). The high consumption generated by the mining process brings challenges that can determine the attraction of new sources of energy, especially renewables, considering that it is desired to reduce the negative externalities on the environment. In addition, the price prediction for Bitcoin is important to reduce the herd behavior of investors. Considering the energy-Bitcoin relationship, investors can better base their trading decisions on the energy market to obtain certain effects on Bitcoin and to optimize the composition of portfolios that can contain both classic assets (such as oil) and financial assets such as cryptocurrencies Su et al. [47]. Besides the negative externalities on the environment, specialists also analyze the social impact of cryptocurrency mining, considering that the pollution generated has direct repercussions on the health of the population. In this line, Goodkind et al. [22] had launched the concept of crypto damages—as the impact cryptocurrencies has on human health and climate. In conclusion, the blockchain technology used





**Fig. 3** Stages of the proposed methodology

to create bitcoin is considered "primary fuel for the global network of money transmission" Hashemi Joo et al. [24]. The paradox is that "while blockchain can serve us, Bitcoin threatens our survival" Sapra and Shaikh [39].

### 3 Methodology

The proposed methodology consists of three stages as shown in Fig. 3:

- Stage 1 – the datasets consisting in resources prices and traded quantities, macroeconomics and Bitcoin data are collected from several sources, pre-processed and stored into a relational database.
- Stage 2 – an extensive feature engineering stage is performed to extend the initial input with more features and enhance the model with new analytical and statistical variables that lead to the increase of the forecast accuracy.

- Stage 3 – a set of five ML models are trained and tested, and their predictions are stacked using a meta-model. This model adjusts these predictions with a set of weights to provide the final prediction of the Bitcoin price.

### 3.1 Stage 1—Input data pre-processing

The data sample interval spans from 1st of January 2019 to 31st of May 2022. In this interval, significant price movements took place. Bull and bear markets explain price trends and are associated with significant price movements. Bull markets are intervals in which the price goes upwards, while the opposite is true for bear markets. To forecast the Bitcoin close price ( $BTC\_USD - C_{BTC}^h$ ), as the relationship between energy, Bitcoin prices and their volatility became notorious, three main categories of features or variables are considered: 1) Bitcoin data (prices and traded volumes), 2) electricity, gas, oil prices and quantities and 3) macroeconomics (such as: inflation, EUETS price). They were collected from various sources like databases, files and online sources. For transparency and replicability, the input data was stored on the GitHub platform.<sup>1</sup> These categories consist of 15 fundamental variables that are known in the previous day of forecasting ( $h - 24$ ) and are used by the ML algorithms to provide the forecast of the Bitcoin close price for the next day ( $C_{BTC}^h$ ). Therefore, the initial dataset is composed of: Open-High-Low prices ( $O_{BTC}^{h-24}, L_{BTC}^{h-24}$ ), Total volume ( $V_{BTC}^{h-24}$ ), Quote Asset Volume ( $QAV_{BTC}^{h-24}$ ), Number of trades ( $N_{BTC}^{h-24}$ ), Taker base volume ( $TBV_{BTC}^{h-24}$ ), Taker quote volume ( $TQV_{BTC}^{h-24}$ ), Electricity price on DAM ( $P_{El}^{h-24}$ ), Gas price on DAM ( $P_{Gas}^{h-24}$ ), Inflation in EU ( $I_{EU}^{h-24}$ ), EU-ETS emissions certificates prices ( $P_{EUETS}^{h-24}$ ), Oil price ( $P_{Oil}^{h-24}$ ), Electricity traded quantity on DAM ( $Q_{El}^{h-24}$ ), Gas traded quantity on DAM ( $Q_{Gas}^{h-24}$ ). Quote volume is used in sell orders when a trader knows beforehand how much quote coin, he/she will sell. Base volume is used in buy orders when a trader is not aware of how much quote coin, he/she can buy considering the available funds. On a trading exchange, assets are traded in pairs. One asset (Base Asset—BA) is traded against the other (Quote Asset QA). The price indicates how much of the QA is required to buy 1 unit of the BA. A taker BA volume represents how much of the total BA volume is contributed by the taker orders. The total volume is equal to the taker buy plus taker sell.

For pre-processing, we perform the following steps: a) Data cleaning addresses any issues that could compromise the quality of the data analysis. It involves identifying and correcting errors like typographical mistakes, and inconsistencies of the data entries. This step also includes the treatment of missing values through methods like

interpolation, where gaps are filled based on surrounding data values. Furthermore, this stage covers the removal of duplicate records that may skew the results of the analysis; b) Data transformation includes changing the data types—for instance, converting timestamps from strings to datetime objects—or aggregating data points, such as summing up hourly transactions to obtain totals, to better suit the analysis needs; c) Data merging and integration combines the datasets from the following sources: Bitcoin trading prices,<sup>2</sup> Romanian electricity market volumes and prices,<sup>3</sup> gas volumes and prices,<sup>4</sup> inflation rates,<sup>5</sup> oil prices<sup>6</sup> and EU-ETS emissions certificates prices.<sup>7</sup> All variables were recorded at hourly resolution and merged into a single, comprehensive dataset based on timestamp, facilitating coherent analysis across all variables; d) Data partitioning divides the dataset into yearly subsets for conducting time-based analyses and for setting up training and testing sets in ML algorithms. Partitioning helps in evaluating the performance of predictive models under unbiased conditions by training them on one subset of the data and testing them on another; e) Data normalization and standardization involve scaling numerical data to ensure that it falls within a particular range, such as between 0 and 1 (normalization) or having a mean of 0 and a standard deviation of 1 (standardization). Such scaling is critical for ML models that are sensitive to the magnitude of data, ensuring that no variable unduly influences the model due to its scale; f) Data encoding converts categorical data into a numerical format since ML models handle numeric inputs. Techniques like one-hot encoding or label encoding are employed depending on the nature of the categorical data; g) Error checking and validation involves rigorously checking the dataset for errors and validating its accuracy to ensure that the data is fit for ML models. This includes applying specific data validation rules, such as constraints on data values and relationships, to prevent logical inconsistencies in the dataset.

By following these detailed preprocessing steps, the dataset is effectively prepared to be used in Step 2 and enhanced with more derived features. A data sample is showcased in Table A1 in Appendix A.

<sup>1</sup> <https://github.com/simonavoprea/Bitcoin-Energy-Nexus>

<sup>2</sup> <https://www.investing.com/crypto/bitcoin/btc-usd>

<sup>3</sup> [https://www.opcom.ro/pp/grafice\\_ip/raportPIPSiVolumTranzacti onat.php?lang=ro](https://www.opcom.ro/pp/grafice_ip/raportPIPSiVolumTranzacti onat.php?lang=ro)

<sup>4</sup> <https://www.brm.ro/piata-spot-gn/>

<sup>5</sup> <https://www.rateinflation.com/inflation-rate/euro-area-historical-inflation-rate/>

<sup>6</sup> <https://www.macrotrends.net/2480/brent-crude-oil-prices-10-year-daily-chart>

<sup>7</sup> <https://www.investing.com/commodities/carbon-emissions-historical-data>

### 3.2 Stage 2—Features engineering

The input of the ML algorithms  $X^h$  is initially composed of the 15 raw or fundamental features described and pre-processed in Step 1 that represent the recorded hourly

values of the previous day. The hour ( $h$ ), month ( $m$ ), weekday ( $wd$ ) and year ( $y$ ) are also extracted from the timestamp and added to the input variables to capture the time variations. For day  $d$  of prediction, the fundamental features are lagged by 24 h ( $h-24$ ).

$$X^h = [P_{El}^{h-24}, Q_{El}^{h-24}, P_{Gas}^{h-24}, Q_{Gas}^{h-24}, P_{Oil}^{h-24}, P_{EUETS}^{h-24}, J_{EU}^{h-24}, O_{BTC}^{h-24}, H_{BTC}^{h-24}, L_{BTC}^{h-24}, N_{BTC}^{h-24}, V_{BTC}^{h-24}, QAV_{BTC}^{h-24}, TBV_{BTC}^{h-24}, TQV_{BTC}^{h-24}, m, h, wd, y] \tag{1}$$

Feature engineering focuses on the Bitcoin price moving average, 24, 48, 72-shifted vectors, their mean value and SQL Analytical functions.

#### 3.2.1 Statistical variables

The input is filled with 11 statistical features. Previous hourly prices ( $C_{BTC}^{h-\Delta t}$ ) with a lag ( $\Delta t$ ) of 24 to 72 h are calculated by shifting the prices. The averages of the previous hourly prices for 3 consecutive days ( $C_{BTC}^{\Delta t}$ ) are calculated using Eq. (2):

$$\overline{C_{BTC}^{\Delta t}} = \frac{\sum_{\Delta t \in \{24, 48, 72\}} C_{BTC}^{h-\Delta t}}{3} \tag{2}$$

The hourly variations of the previous Bitcoin prices for the last 3 consecutive days ( $\delta_{BTC}^{\Delta t}$ ) are determined using Eq. (3):

$$\delta_{BTC}^{\Delta t} = 100 - \frac{C_{BTC}^{h-\Delta t}}{C_{BTC}^{h-\Delta t-24}} \times 100 \tag{3}$$

Moving average refers to a series of averages of fixed size subsets of the total set of observations. It is also known as rolling average with a specified window size ( $q$ ).

$$MA_q^t = \frac{1}{q} \sum_{k=t-q+1}^t C_{BTC}^k \tag{4}$$

where:

- $t$  represent the last price of the time series and it is set to  $h - 24$
- $C_{BTC}^k$  represent the values of time series,  $k = t - q + 1, t$
- $q \in \{3, 7, 10, 20, 50\}$  represents the window size.

#### 3.2.2 SQL analytical functions

SQL provides analytical functions like RANK, ROW\_NUMBER, LEAD, LAG, and window functions that enable

advanced analytical calculations. SQL analytical functions, also known as window functions, perform calculations across a set of rows related to the current row in a dataset. These functions enable calculations beyond simple aggregation and grouping. They are particularly useful for tasks like ranking, cumulative sums, moving averages, etc. Analytical functions are usually used in extracting and reporting on data stored in relational databases and data warehouses. They process data based on groups of records but differ from aggregate or group functions by returning multiple results for each individual group. The group of records to which the analysis is applied is called a window and is defined using an analysis clause. The window determines the range of records to be analyzed for each current record. The size of the window can be determined either physically, by specifying the number of records in the group, or logically, by conditions on the field values. Analytic operations are the last processed in an SQL query before the ORDER BY clause.

The structure of an analytic function is as follows: ANALYTIC FUNCTION () OVER (*analytical clause*). The analytical clause can have the following subclauses: PARTITION BY (*expression1, expression2, etc.*) ORDER BY *expression/position/alias* [ASC/DESC] [null first/last] *window clause*, where window clause can be: ROWS/RANGE [BETWEEN] {UNBOUNDED PRECEDING}/{CURRENT ROW}/{*expression* PRECEDING/FOLLOWING} [AND] {UNBOUNDED PRECEDING}/{CURRENT ROW}/{*expression* PRECEDING/FOLLOWING}.

If the window clause is omitted, then by default it applies RANGE BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW. Using the ROWS BETWEEN 1 PRECEDING AND 1 FOLLOWING clause determines the analysis of the previous and immediately following value of the current record. A list sample of analytical functions is showcased in Table 1.

Thus, the first three analytical functions ( $F1 \div F3$ ) – the mean BTC prices from the same year ( $y$ ) and month ( $m$ ) (columns in ORDER BY) with the current value from the same hour (column in PARTITION BY) considering the preceding and following  $n$  values are depicted in Eq. (5).



**Table 1** Examples of analytical functions

Analytical Feature	Depiction
RANK(), DENSE_RANK(), NTILE()	These functions rank rows based on specified criteria. RANK assigns the same rank to rows with equal values and leaves gaps, DENSE_RANK assigns the same rank to equal values without gaps, and NTILE divides the result set into a specified number of roughly equal parts
LEAD(), LAG()	These functions access data from subsequent (LEAD) or preceding (LAG) rows in a dataset based on a specified ordering. They are used to calculate differences or trends in data
SUM(), AVG()	While SUM() and AVG() can be used as aggregate functions, they can also be used as window functions to calculate running totals and averages over a set of rows
PERCENTILE_CONT(), PERCENTILE_DISC()	These functions calculate the specified percentile: PERCENTILE_CONT calculates the continuous percentile, while PERCENTILE_DISC calculates the discrete percentile
CUME_DIST()	This function calculates the cumulative distribution returning the fraction of rows that are less than or equal to the current value

$$F1 \div F3 : AVG(C_{BTC}^h)OVER (PARTITION BY h ORDER BY y, m ROWS BETWEEN n PRECEDING AND n FOLLOWING) \tag{5}$$

where  $n \in \{1, 3, 5\}$ .

The fourth analytical function ( $F4$ ) – the mean BTC prices of previous values relative to the current value from the same year, month, at the same hour—is described in Eq. (6).

$$F4 : AVG(C_{BTC}^h)OVER (PARTITION BY h ORDER BY y, m ROWS BETWEEN unbounded preceding AND current row) \tag{6}$$

The mean BTC prices of next values relative to the current value from the same year, month, at the same hour represent

the fifth feature showcased in Eq. (7).

$$F5 : AVG(C_{BTC}^h)OVER (PARTITION BY h ORDER BY y, m ROWS BETWEEN current row AND unbounded following) \tag{7}$$

The sixth feature consists of the mean of BTC price from the same month between which there is a  $\pm 5$  difference relative to the current value. It is illustrated in Eq. (8).

$$F6 : AVG(C_{BTC}^h)OVER (PARTITION BY y, m, h ORDER BY m RANGE BETWEEN 5 preceding AND 5 following) \tag{8}$$

The sixth feature consists of the mean of BTC price from the same month between which there is a  $\pm 10$  difference relative to the current values in Eq. (9).

$$F7 : AVG(C_{BTC}^h)OVER (PARTITION BY y, m, h ORDER BY m RANGE BETWEEN 10 preceding AND 10 following) \tag{9}$$

The eighth and ninth features represent the minimum Bitcoin value and the maximum Bitcoin value from the same hour, with less or equal values relative to the current value as in Eq. (10) and Eq. (11).

---


$$F8 : \text{MIN}(C_{BTC}^h) \text{ OVER (PARTITION BY } h \text{ ORDER BY } y, m, h \text{ ROWS BETWEEN unbounded preceding AND current row)} \quad (10)$$

$$F9 : \text{MIN}(C_{BTC}^h) \text{ OVER (PARTITION BY } h \text{ ORDER BY } y, m, h \text{ ROWS BETWEEN unbounded preceding AND current row)} \quad (11)$$


---

The minimum Bitcoin value and the maximum Bitcoin value from the same year, month, hour, with equal or greater values relative to the current value are shown in Eq. (12) and Eq. (13) and represent features F10 and F11.

$$F10 : \text{MIN}(C_{BTC}^h) \text{ OVER (PARTITION BY } h \text{ ORDER BY } y, m, h \text{ ROWS BETWEEN unbounded preceding AND current row)} \quad (12)$$

$$F11 : \text{MIN}(C_{BTC}^h) \text{ OVER (PARTITION BY } h \text{ ORDER BY } y, m, h \text{ ROWS BETWEEN current row AND unbounded following)} \quad (13)$$


---

The standard deviation of the Bitcoin price from the same year and month (column in ORDER BY) with the current value from the same hour (column in PARTITION BY) represents the twelfth feature and it is given in Eq. (14).

$$F12 : \text{STDDEV}(C_{BTC}^h) \text{ OVER (PARTITION BY } h \text{ ORDER BY } y, m \text{ ROWS BETWEEN } 1 \text{ preceding AND } 1 \text{ following)} \quad (14)$$

F13 calculates the Bitcoin price percentile and returns a value that represents the percent of values that are less or equal than the current value as in Eq. (15). F14 calculates the percent rank of the current value within a specific hour as in Eq. (16).

$$F13 : \text{CUME\_DIST}() \text{ OVER (PARTITION BY } h \text{ ORDER BY } C_{BTC}^h) \quad (15)$$

$$F14 : \text{PERCENT\_RANK}() \text{ OVER (PARTITION BY } h \text{ ORDER BY } C_{BTC}^h) \quad (16)$$


---

F15 computes the ratio of a value to the sum of a set of values as in Eq. (17).

$$F15 : \text{RATIO\_TO\_REPORT}(C_{BTC}^h) \text{ OVER (PARTITION BY } y, m, h) \quad (17)$$

Finally, the input of the ML algorithms is completed with the above-calculated features:

$$X^h = [X^h, C_{BTC}^{h-24}, C_{BTC}^{h-48}, C_{BTC}^{h-72}, \overline{C_{BTC}^{\Delta t}}, \delta_{BTC}^{\Delta t}, MA_q^t, F1 \div F15] \quad (18)$$

The closing hourly BTC prices represent the target feature ( $y^h = C_{BTC}^h$ ) of the ML algorithms.

### 3.3 Stage 3—Forecasting ML models

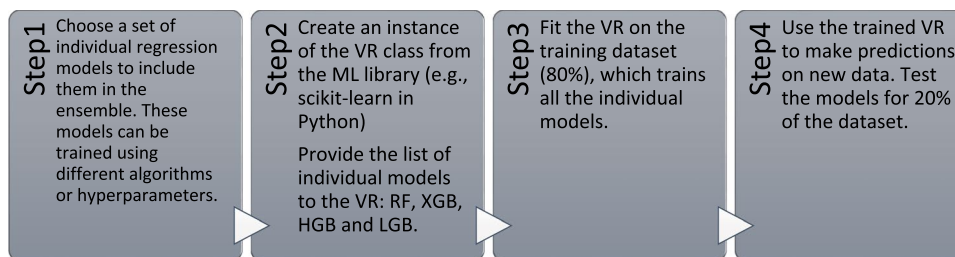
A stochastic forecasting method that combines four ensemble ML algorithms: Random Forest (RF), eXtreme Gradient Boosting (XGB), Histogram-based Gradient Boosting (HGB) and Light Gradient Boosting (LGB) Machine using

a Voting Regressor (VR) model is proposed in this paper. The proposed model combines the predictions of multiple individual models to make a final prediction, resulting in improved performance and generalization. Therefore, VR combines the predictions of multiple models to make a more accurate prediction than any single model. The idea behind using VR is that different models might capture different aspects of the underlying data patterns, and combining their predictions can lead to a more accurate overall prediction.

The main four steps of implementing a VR are as follows (as in Fig. 4):

- *Step 1*—Choose a set of individual regression models to include them in the ensemble. These models can be trained using different algorithms or hyperparameters. The effectiveness of a Voting Regressor depends on the diversity and quality of the individual models included in the ensemble. Experiments with different combinations of models and assessment of the ensemble's performance on a validation dataset are performed.

Fig. 4 Steps in setting VR



- *Step 2* – We create an instance of the VR class from the ML library (e.g., scikit-learn in Python) and provide the list of individual models to the VR. When choosing the models, the most accurate models were selected.
- *Steps 3 and 4*—Fit the VR on the training dataset (80%), which trains all the individual models. Use the trained VR to make predictions on new data. Test the models for 20% of the dataset.

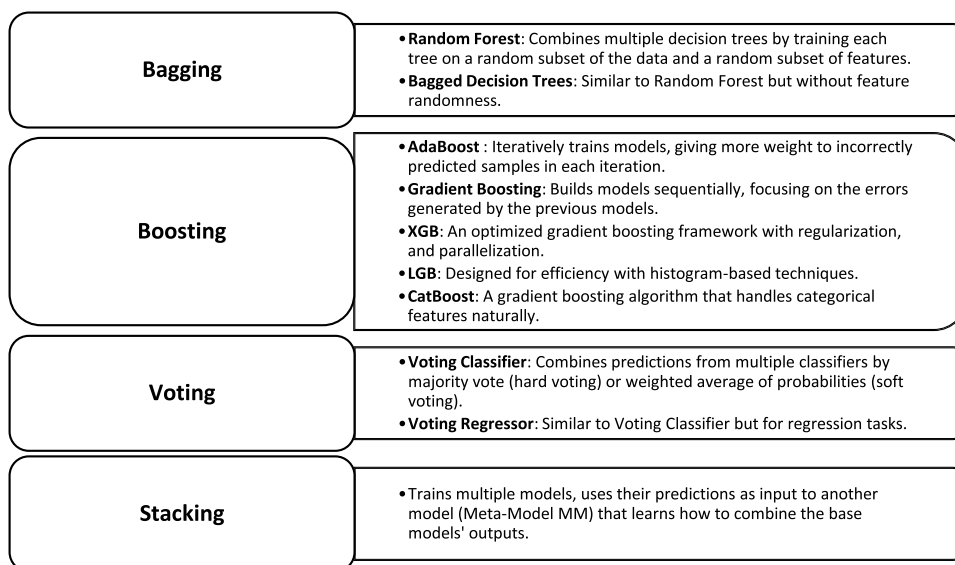
Ensemble ML algorithms combine the predictions of multiple individual models to create a stronger, more accurate model. They reduce overfitting and increase the robustness of ML models. There are four main categories of ensemble techniques briefly described in Fig. 5.

Light Gradient Boosting Machine (LGB) like other gradient boosting frameworks, builds an ensemble of weak learners (usually decision trees) in a sequential manner and incorporates several optimizations to make it faster and more scalable. Each new tree corrects the errors made by the previous ones, leading to improved

predictive performance. It is known for its speed, memory efficiency, and high performance on a wide range of datasets. It uses a histogram-based approach to bucket the data, which reduces memory usage and speeds up training. This approach discretizes the feature values into bins, reducing the number of unique values to consider during the split finding process. LGB utilizes multiple CPU cores for parallel training and supports GPU acceleration to speed up the training process.

Histogram-based Gradient Boosting (HGB) is a technique used in gradient boosting algorithms to improve their efficiency and speed. As mentioned, this technique is a core component of LGB. Traditional gradient boosting algorithms build trees using a depth-wise approach. They iteratively split data points at each level of the tree based on a specific feature’s values. However, this approach can be computationally expensive and memory-intensive, especially with large datasets. HGB as in LGB bins feature values in discrete bins or histograms during the data preprocessing step. Then, during the tree-building process, instead of searching for optimal splits for individual feature values, the algorithm works with the pre-binned histograms.

Fig. 5 Four main categories of ensemble methods



eXtreme Gradient Boosting (XGB) proved high predictive capabilities and feature-rich implementation. Like LGB, XGB is based on the gradient boosting as it iteratively builds an ensemble of weak learners, and it introduces a regularized learning objective that combines the loss function with penalties that prevent overfitting and improves the generalization of the model. Furthermore, XGB, similarly to LGB, utilizes multiple CPU cores and even distributed computing environments, making it highly scalable and efficient for large datasets and accelerates using GPUs, which significantly speeds up the training process. XGB allows customization of the criteria for finding optimal split points during tree construction.

Random Forest (RF) is an extension of decision tree algorithms that aims to improve their predictive performance and reduce overfitting. It consists of a collection of individual decision trees. Each tree is trained on a random subset of the training data and independently makes predictions. During the construction of each tree, a random subset of features is considered for splitting at each node, reducing overfitting. For each tree, a random subset of the training data is sampled with replacement. This process is known as bagging. The combination of random feature selection and bagging contributes to the diversity of the trees in the ensemble. For classification, the predictions of each individual tree are combined through a majority vote, while for regression tasks, the predictions are usually averaged. It is less sensitive to outliers and noise in the data compared to single decision trees. The training of individual trees can be parallelized, making Random Forests computationally efficient.

Thus, the proposed model embeds the first three categories of the ensemble techniques: bagging (RF), boosting (HGB, LGB, XGM) and voting (VR). The fourth category—stacking is included in order to use the predictions of the five models as input to another model (that is known as Meta-Model MM) that learns from the results of the base models.

The first training dataset  $T$  spans from 1st of January until 30th of April 2021, whereas the first testing dataset spans from 1st of May until 7th of May 2021. The ML algorithms (RF, LGB, HGB, XGB and VR) are trained on the  $T$  interval and provide for the next 7 days an individual forecast denoted by  $\widehat{y}_p^{h+\Delta t}$ , where  $\Delta t = 1, 168$  and  $p \in \{RF, LGB, HGB, XGB, VR\}$ . Each individual model provides an estimation. These estimations are further weighted with a set of weights ( $\omega_i$ ) and used as input for the stacking ML model. This model is a Meta-Model (MM) that mediates the predictions of the sub-models using the weights to improve the accuracy of the final prediction. Instead of training the MM on the input set  $X^h$ , it

is trained on the predictions ( $\widehat{y}_p^h$ ) of its sub-models using the same target ( $y^h = C_{BTC}^h$ ) for the same training period as the sub-models. Thus, it determines the weights to adjust each sub-model and learns how to optimally combine these predictions. The goal is to leverage the strengths of each individual model and mitigate their weaknesses, thereby enhancing prediction accuracy. After training, the predictions ( $\widehat{y}_p^{h+\Delta t}$ ) of the initial five ML models are weighted with the corresponding coefficient to forecast the hourly Bitcoin price ( $C_{BTC}^{h+\Delta t}$ ) for the next 7 days:

$$\widehat{C}_{BTC}^{h+\Delta t} = \sum_p \omega_p \times \widehat{y}_p^{h+\Delta t}, \forall p \in \{RF, LGB, HGB, XGB, VR\}, \Delta t = 1, 168 \quad (19)$$

The stacking ML model can be a simpler ML model such as Linear Regression (LR) or Decision Tree (DT) regressor or another ensemble model, such as Gradient Boosting Regressor (GBR). Therefore, the individual estimations are used as input features in order to obtain a final prediction using a stacking model, such as LR or DT. In the stacking approach, each sub-model's prediction is given a specific weight, which signifies the relative importance or trustworthiness of that model's output in the context of the combined prediction. These weights are critical as they determine how much influence each model's prediction has on the final output. To determine the weights, the method of least squares errors is a common objective function in regression problems. The least squares error function quantifies the difference between the observed values and the predicted values. Therefore, the weights are obtained by minimizing the Sum of the Squared Errors (SSE) between the actual target or the Bitcoin prices during the training interval and the stacked predictions  $p$ . By minimizing this sum, the fitting process seeks to find the model parameters (weights in this case) that result in the best possible predictions, in terms of being closest to the actual data (Bitcoin close price).

$$\omega_p = \operatorname{argmin}_{\omega_p} \sum_{h \in T} \left( C_{BTC}^h - \sum_p \omega_p \times \widehat{y}_p^h \right)^2 \quad (20)$$

Using SSE has several advantages: i) it is straightforward to implement and computationally efficient to optimize, which is particularly valuable when dealing with multiple predictors as in stacking; ii) SSE function is convex in terms of the regression coefficients (weights), ensuring that the optimization process finds a global minimum, thus the optimal set of weights; iii) squaring the errors penalizes larger discrepancies more severely, focusing the model on avoiding large mistakes in prediction. By optimizing these

weights to minimize SSE, the stacking model effectively learns the best way to combine the sub-models' outputs to replicate the actual Bitcoin prices as closely as possible. This results in a robust model that improves upon the performance of its constituent models, leading to more reliable and accurate forecasts.

Ensemble methods, like RF, XGB, LGB, HGB and VR, are advanced ML techniques known for their high accuracy and robustness. They have different parameters and structures. RF consists of a collection of decision trees. Each tree is built from a random subset of the training data. The key parameters are: number of trees, max depth of trees, min samples split, min samples leaf. Each tree makes a prediction, and the final output is typically the average of these predictions. XGB is a gradient boosting framework that uses a collection of decision trees. Its key parameters are: number of boosting rounds, max depth of trees, learning rate, subsample ratio of the training instance. It uses a more regularized model formalization to control over-fitting, which gives it better performance. LGB structure is similar to XGB, but uses a histogram-based algorithm for faster training and reduced memory usage. Its main parameters are: number of leaves, max depth, learning rate, feature fraction. HGB is a gradient boosting approach that uses histograms to represent the continuous feature values in each split. Its key parameters are: max iter (number of boosting stages), max depth, learning rate. HGB is designed to be faster than traditional gradient boosting by using histograms to speed up the calculation of potential splits. VR is not a tree-based model, but an ensemble that combines

predictions from multiple different regression models. Its key parameter depends on the individual regressors used within the ensemble. The final prediction can be either the average or a weighted average of all the regressors' predictions. In ensemble methods, the focus is more on how the individual models (like trees in RF or boosting rounds in XGB) are combined and optimized for the best predictive performance.

### 3.4 Evaluation of the accuracy of the models

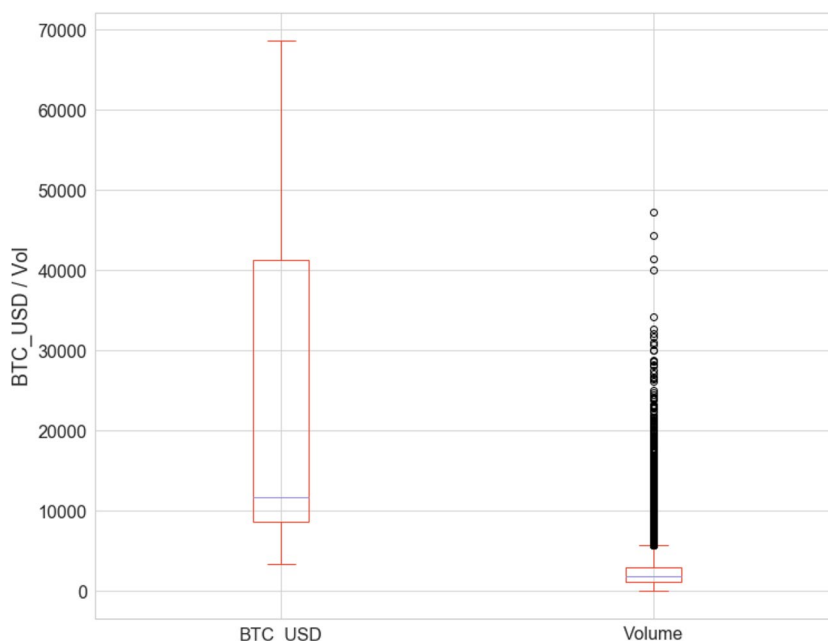
For assessing the accuracy of the stacking model, the following well-known metrics are calculated for the training and testing intervals: Root-Mean Squared Error (RMSE), coefficient of determination ( $R^2$ ), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE):

$$RMSE = \frac{1}{T} \sum_{h=1}^T \left( C_{BTC}^h - \widehat{C}_{BTC}^h \right)^2 \tag{21}$$

$$R^2 = 1 - \frac{\sum_{h=1}^T \left( C_{BTC}^h - \widehat{C}_{BTC}^h \right)^2}{\sum_{h=1}^T \left( C_{BTC}^h - \overline{C}_{BTC} \right)^2} \tag{22}$$

$$MAPE = \frac{1}{T} \sum_{h=1}^T \frac{\left| C_{BTC}^h - \widehat{C}_{BTC}^h \right|}{\left| C_{BTC}^h \right|} \times 100\% \tag{23}$$

**Fig. 6** Variations of Bitcoin price and volume during January 2019-May 2022





$$MAE = \frac{1}{T} \sum_{h=1}^T \left| C_{BTC}^h - \widehat{C}_{BTC}^h \right| \quad (24)$$

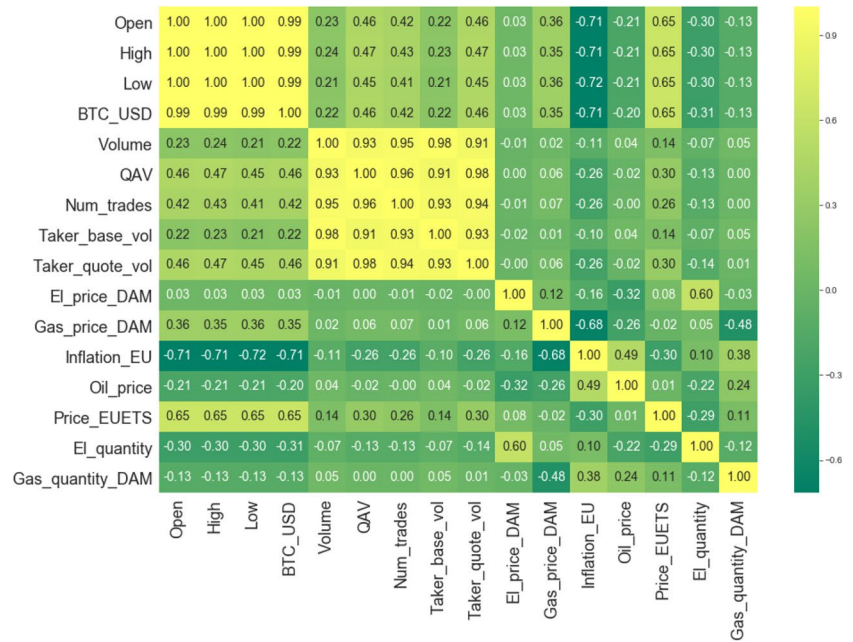
The same metrics are used to evaluate the accuracy of the ML models by replacing the  $\widehat{C}_{BTC}^h$  with the corresponding predictions  $\widehat{y}_p^h$ , where  $p \in \{RF, LGB, HGB, XGB, VR\}$ .

## 4 Results

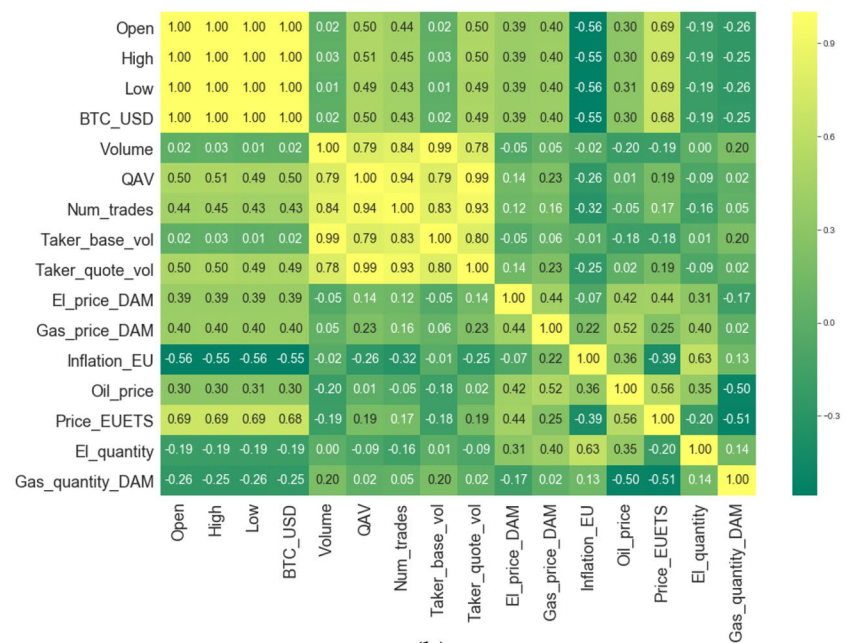
### 4.1 Case study description

The input data was collected from various websites. For Bitcoin data, energy data and macroeconomics (oil prices and inflation) several websites were visited, and the values were either downloaded or scraped, as mentioned in subsection 3.1. For instance, the electricity data was scraped from OPCOM website. The Bitcoin price and traded volume variations are depicted as boxplots in Fig. 6. Usually,

**Fig. 7** Heatmaps (a) 2019, (b) 2020, (c) 2021, (d) 2022

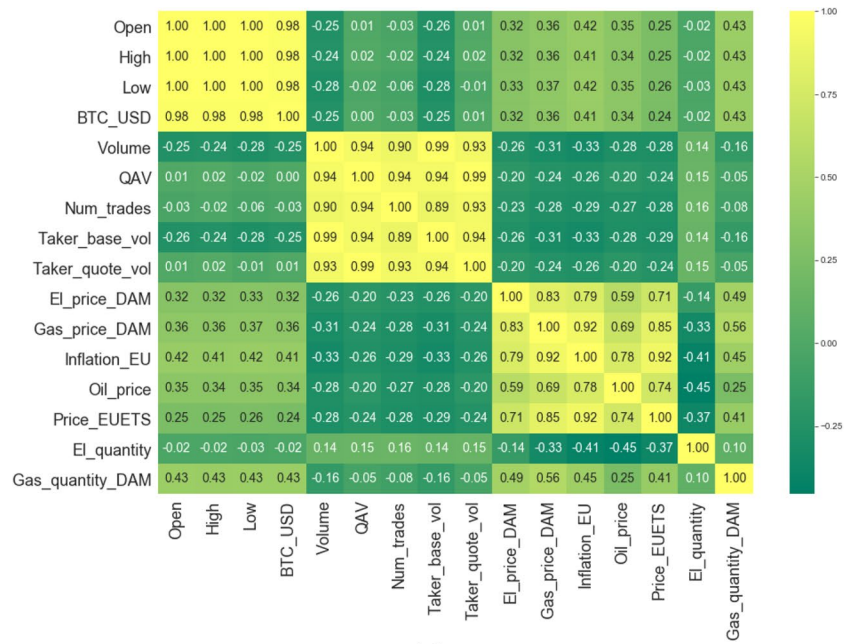


(a)

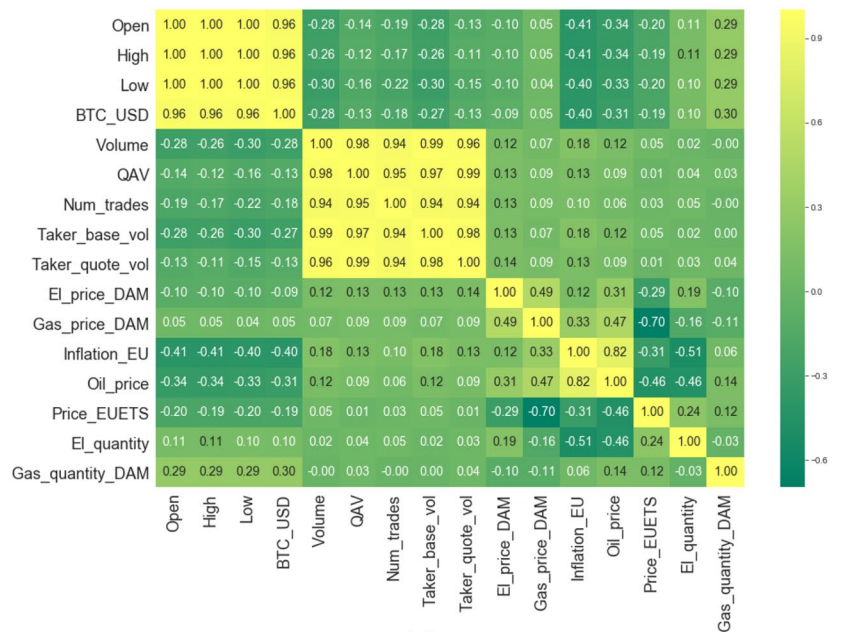


(b)

Fig. 7 (continued)



(c)



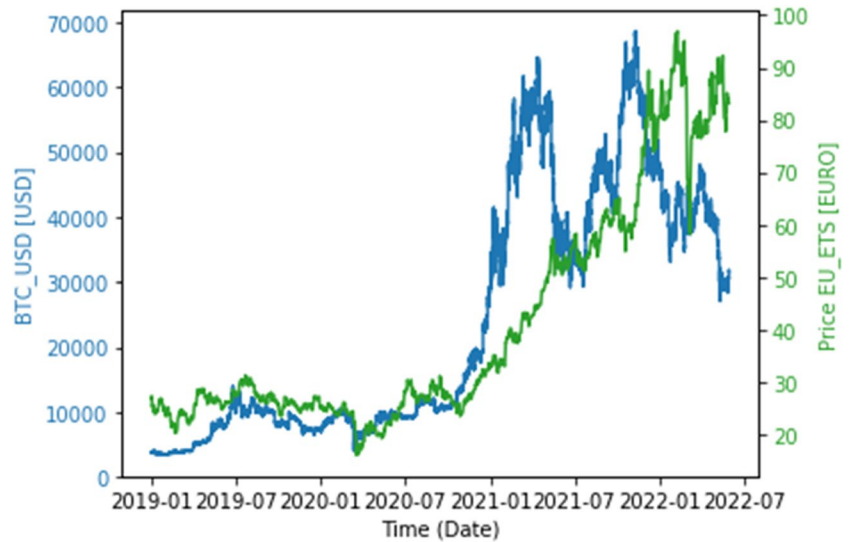
(d)

according to the interquartile range, the price varied between 1,000 and 40,000 USD. The median price was just over 1,000 USD, but it reached almost 70,000 USD for short intervals. The variation of the traded volume was much less than that of price, but outliers were more frequent.

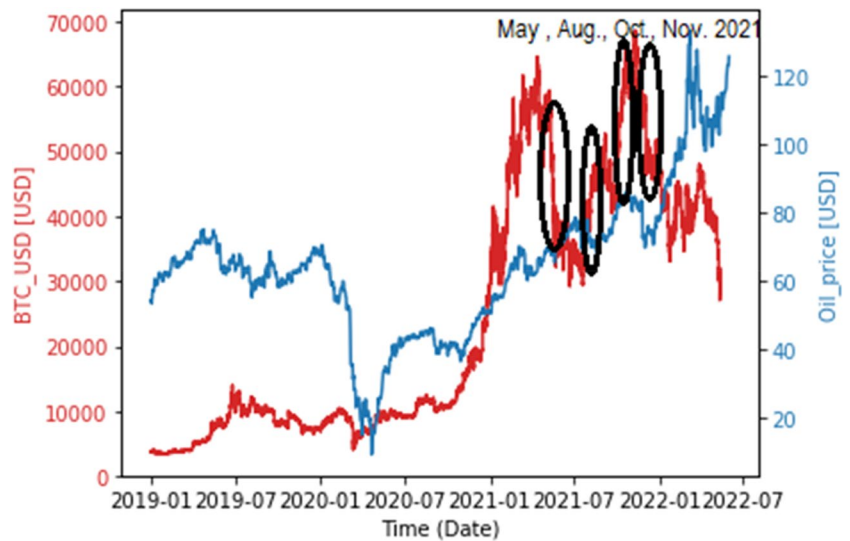
The correlations between Bitcoin price and other prices and macroeconomics indicators are analyzed in Fig. 7. Even from 2019, when the Bitcoin price was

much lower, a strong correlation between the price and the Price\_EUETS was recorded (0.65). Furthermore, a strong inverse correlation was noticed between Inflation\_EU and the price. There was no correlation between the electricity price on DAM and the price (0.03), and the correlation between gas price on DAM and the price was 0.35. The other correlations in 2019 were of medium or weak intensity (as in Fig. 7a). In 2020, the correlations between the Bitcoin price and electricity (0.39) and gas

**Fig. 8** Evolution of Bitcoin price and Price\_EUETS



**Fig. 9** Evolution of Bitcoin price and oil price



prices (0.40) on DAM intensified (as in Fig. 7b). The strong correlation between the price and Price\_EUETS maintained (0.68), while the inverse correlation with the Inflation decreased from -0.71 in 2019 to -0.55 in 2020. The weak correlation with the oil price changed the polarity from -0.20 in 2019 to 0.30 in 2020.

In relation to energy prices, in 2021, the Bitcoin price showcased ample swings (Fig. 7c). Most of the correlations were of medium intensity (Gas\_quantity\_DAM switched from -0.25 in 2020 to 0.43 in 2021, Inflation\_EU switched from -0.55 in 2020 to 0.41 in 2021). The stronger correlation with Price\_EUETS from 2019 and 2020 weakened from 0.68 in 2020 to 0.24 in 2021. In the first five months from 2022, the absent correlation between the price and the

electricity price on DAM returned (-0.09). A medium correlation with Inflation\_EU (-0.41) was noticeable, while the rest of the correlations were weak (as in Fig. 7d).

The evolution of Bitcoin price and the Price\_EUETS is showcased in Fig. 8. They both started to increase from the end of 2020. The highest values recorded for the Bitcoin price (the blue curve) in 2021 were followed in 2022 by the highest values recorded for the Price\_EUETS (the green curve). The evolution of Bitcoin price and oil price is showcased in Fig. 9. While the oil price was much higher than the Bitcoin price in 2019, they met in the first semester of 2020 when the oil price dropped at the beginning of COVID-19 pandemic, and both started to increase. The maximum values recorded by the Bitcoin prices in the

**Table 2** Implementation of the analytical functions

Analytical Feature	Depiction
<b>F1</b>	<p>Mean BTC price from the same year and month (column in ORDER BY) with the current value from the same hour (column in PARTITION BY) considering the preceding and following values</p> <pre>SELECT date_mm, year, month, hour, BTC_USD, AVG(BTC_USD) OVER (PARTITION BY hour ORDER BY year, month ROWS BETWEEN 1 PRECEDING AND 1 FOLLOWING) AS F1 FROM bit ORDER BY date_mm, year, month, hour;</pre>
<b>F2</b>	<p>Mean BTC price from the same year and month (column in ORDER BY) with the current value from the same hour (column in PARTITION BY) considering the preceding and following 3 values</p> <pre>SELECT date_mm, year, month, hour, BTC_USD, AVG(BTC_USD) OVER (PARTITION BY hour ORDER BY year, month ROWS BETWEEN 3 PRECEDING AND 3 FOLLOWING) AS F2 FROM bit ORDER BY date_mm, year, month, hour;</pre>
<b>F3</b>	<p>Mean BTC price from the same year and month (column in ORDER BY) with the current value from the same hour (column in PARTITION BY) considering the preceding and following 5 values</p> <pre>SELECT date_mm, year, month, hour, BTC_USD, AVG(BTC_USD) OVER (PARTITION BY hour ORDER BY year, month ROWS BETWEEN 5 PRECEDING AND 5 FOLLOWING) AS F3 FROM bit ORDER BY date_mm, year, month, hour;</pre>
<b>F4</b>	<p>Mean BTC price of previous values relative to the current value from the same year, month, at the same hour</p> <pre>SELECT date_mm, year, month, hour, BTC_USD, AVG(BTC_USD) OVER (PARTITION BY hour ORDER BY year, month ROWS BETWEEN unbounded preceding AND current row) AS F4 FROM bit ORDER BY date_mm, year, month, hour;</pre>
<b>F5</b>	<p>Mean BTC price of next values relative to the current value from the same year, month, at the same hour</p> <pre>SELECT date_mm, year, month, hour, BTC_USD, AVG(BTC_USD) OVER (PARTITION BY hour ORDER BY year, month ROWS BETWEEN current row AND unbounded following) AS F5 FROM bit ORDER BY date_mm, year, month, hour;</pre>
<b>F6</b>	<p>Mean of BTC price from the same month between which there is a <math>\pm 5</math> difference relative to the current value</p> <pre>SELECT date_mm, year, month, hour, BTC_USD, AVG(BTC_USD) OVER (PARTITION BY year, month, hour ORDER BY month range BETWEEN 5 preceding AND 5 following) AS F6 FROM bit ORDER BY date_mm, year, month, hour;</pre>
<b>F7</b>	<p>Mean of BTC price from the same month between which there is a <math>\pm 10</math> difference relative to the current value</p> <pre>SELECT date_mm, year, month, hour, BTC_USD, AVG(BTC_USD) OVER (PARTITION BY year, month, hour ORDER BY month range BETWEEN 10 preceding AND 10 following) AS F7 FROM bit ORDER BY date_mm, year, month, hour;</pre>
<b>F8, F9</b>	<p>Min BTC_USD value, max BTC_USD value from the same hour, with less or equal values relative to the current value</p> <pre>SELECT date_mm, year, month, hour, BTC_USD, MIN(BTC_USD) OVER (PARTITION BY hour ORDER BY year, month, hour rows BETWEEN unbounded preceding AND current row) AS F8, MAX(BTC_USD) OVER (PARTITION BY year, month, hour ORDER BY year, month, hour rows BETWEEN unbounded preceding AND current row) AS F9 FROM bit ORDER BY date_mm, year, month, hour;</pre>
<b>F10, F11</b>	<p>Min BTC_USD value, max BTC_USD value from the same year, month hour, having equal or greater values relative to the current value</p> <pre>SELECT date_mm, year, month, hour, BTC_USD, MIN(BTC_USD) OVER (PARTITION BY year, month, hour ORDER BY year, month, hour rows BETWEEN current row AND unbounded following) AS F10, MAX(BTC_USD) OVER (PARTITION BY year, month, hour ORDER BY year, month, hour rows BETWEEN current row AND unbounded following) AS F11 FROM bit order by date_mm, year, month, hour;</pre>
<b>F12</b>	<p>Standard deviation of the BTC price from the same year and month (column in ORDER BY) with the current value from the same hour (column in PARTITION BY)</p> <pre>SELECT date_mm, year, month, hour, BTC_USD, stddev(BTC_USD) OVER (PARTITION BY hour ORDER BY year, month ROWS BETWEEN 1 PRECEDING AND 1 FOLLOWING) AS F12 FROM bit ORDER BY date_mm, year, month, hour;</pre>

**Table 2** (continued)

Analytical Feature	Depiction
<b>F13, F14</b>	F13 calculates BTC_USD percentile and returns a value that represents the percent of values that are less or equal than the current value. F14 calculates the percent rank of the current value within a specific hour  <pre>SELECT date_mm, year, month, hour, BTC_USD, CUME_DIST() OVER (PARTITION BY hour ORDER BY BTC_USD)*100 AS F13, PERCENT_RANK() OVER (PARTITION BY hour ORDER BY BTC_USD)*100 AS F14 FROM bit ORDER BY date_mm, year, month, hour;</pre>
<b>F15</b>	It computes the ratio of a value to the sum of a set of values  <pre>SELECT date_mm, year, month, hour, BTC_USD, RATIO_TO_REPORT(BTC_USD) OVER (PARTITION BY year, month, hour)*100 AS F15 FROM bit ORDER BY date_mm, year, month, hour;</pre>

last semester of 2021 were followed at the beginning of 2022 by the maximum values of the oil price.

The basic statistics applied on the fundamental features are presented in Appendix A in Tables A2–A6. The standard deviation was around 2,643 in 2019, almost doubled in 2020 and then was more than double in 2021 when the Bitcoin price reached a maximum value. However, the standard deviation increased 4 times from 2020 to 2021 in the case of the electricity price, whereas for the gas price, it increased more than 10 times. Impressive increases in standard deviation occurred for other input variables as illustrated in Appendix A.

The first considerable peak price was recorded in March 2021 – also known as the first bull market, but it was followed by the first bear market in May 2021 (as in Fig. 9). In August 2021, the Bitcoin price started its second upward slope that continued with the third upward slope in October 2021. However, in November 2021, the price reached the highest value and then went back to the average price level in September. Therefore, four intervals (entire months) are studied in the current paper: May and November—bear markets and August and October—bull markets.

## 4.2 Simulations

The training interval spanned from 1st of January 2019 until the day before the prediction horizon that is set for 168 h ahead or a week. For simulations, we focus on two upward and two downward slopes. Apart from the statistical variables, the analytical functions are implemented in Oracle 12c. Their implementation is briefly described in Table 2.

For the input variables selection, we tested SelectKBest method from `sklearn.feature_selection`, but no significant improvement was recorded, thus, the entire input variables set (that consists of 41 variables: 15 fundamental, 11 statistical and 15 SQL analytical functions) was used in simulations. The standard 80:20 split was employed for training and testing. Although alternative ratios such as

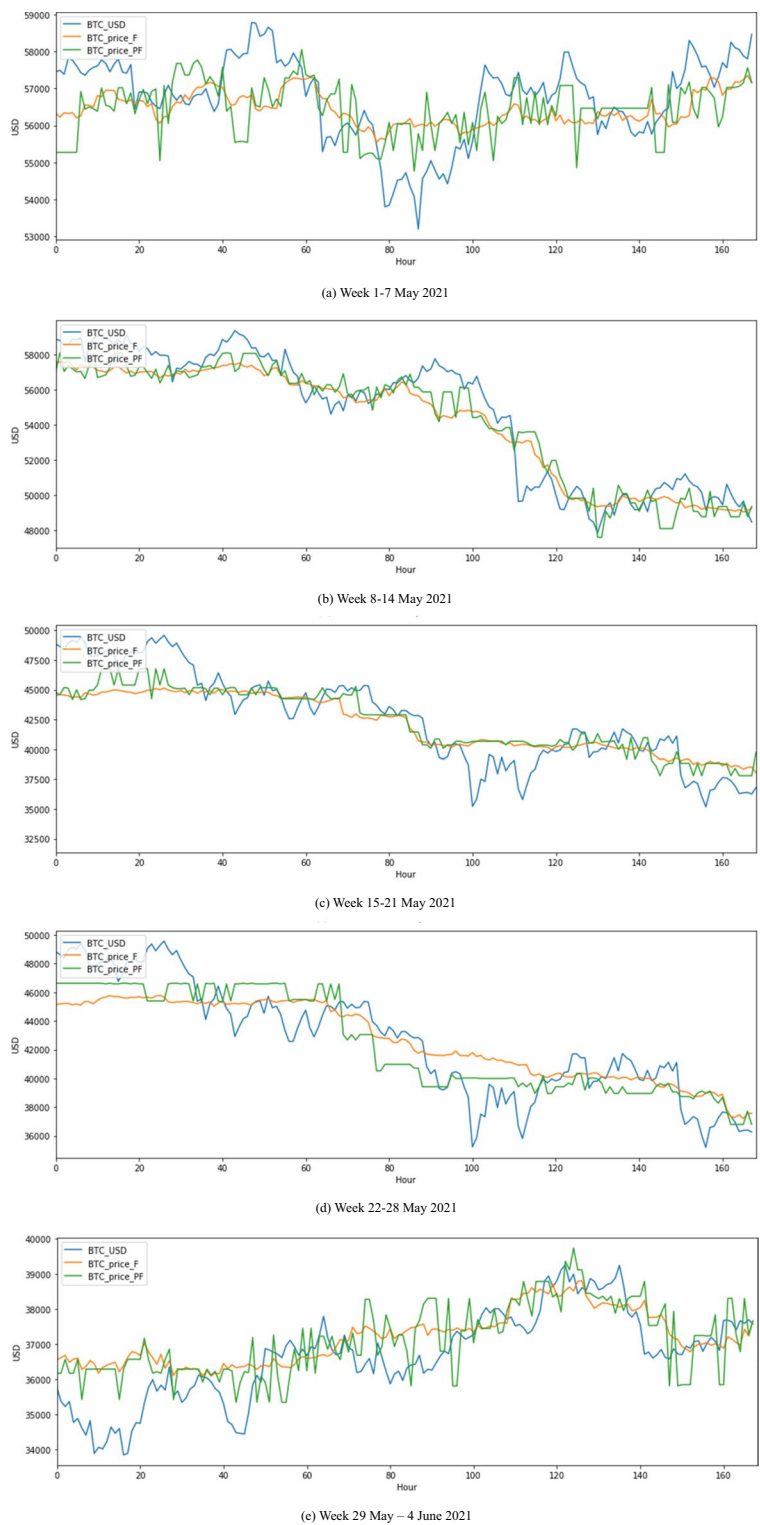
60:40, 70:30, 75:25, 85:15, and 90:10 were explored, the 80:20 ratio demonstrated the most effective results. One of the most interesting months in 2021 from the Bitcoin price evolution point of view were: May, August, October and November. The first significant downward slope (when the prices dropped from almost 59,000 USD to 34,000 USD) that is assimilated to a bearish market that is analyzed in the current paper refers to month May 2021. The first training interval spanned from January 2019 to 30 April 2021. The prediction is tested during the following weeks: 1–7, 8–14, 15–21, 22–28 May, 29 May – 4 June. The results are graphically presented on a weekly basis (Fig. 10). BTC\_USD represents the prices (blue curve), whereas BTC\_price\_F represents the average forecast with the results of the 5 ML algorithms (orange curve) and BTC\_price\_PF represents the weighted forecast (green curve) as described in Eqs. (19, 20).

By the middle of the first week of May, the price showed signs of decline. It went from almost 59,000 to 53,000 USD, but it recovered by the end of the week. In the next three weeks, the Bitcoin price continued to go down and it stabilized around 37,000 USD at the beginning of June. From the first analyzed month, one can notice that most of the time, the real curve is well followed by the prediction obtained by averaging the results of the five predictors or by stacking the results using the decision tree regressor (green) curve. However, abrupt slopes (drops or peaks) are difficult to predict. For instance, the sudden drop from the first day of the last week (29 May – 04 June) of almost 2,000 USD is not well predicted or the model fails to provide an accurate forecast.

The first upward slope analyzed in the current paper was recorded in August 2021. The first training interval spanned from January 2019 to 31 July 2021. The prediction is tested during the following weeks: 1–7, 8–14, 15–21, 22–28 August, 29 August – 4 September. The results are graphically presented on a weekly basis (Fig. 11).



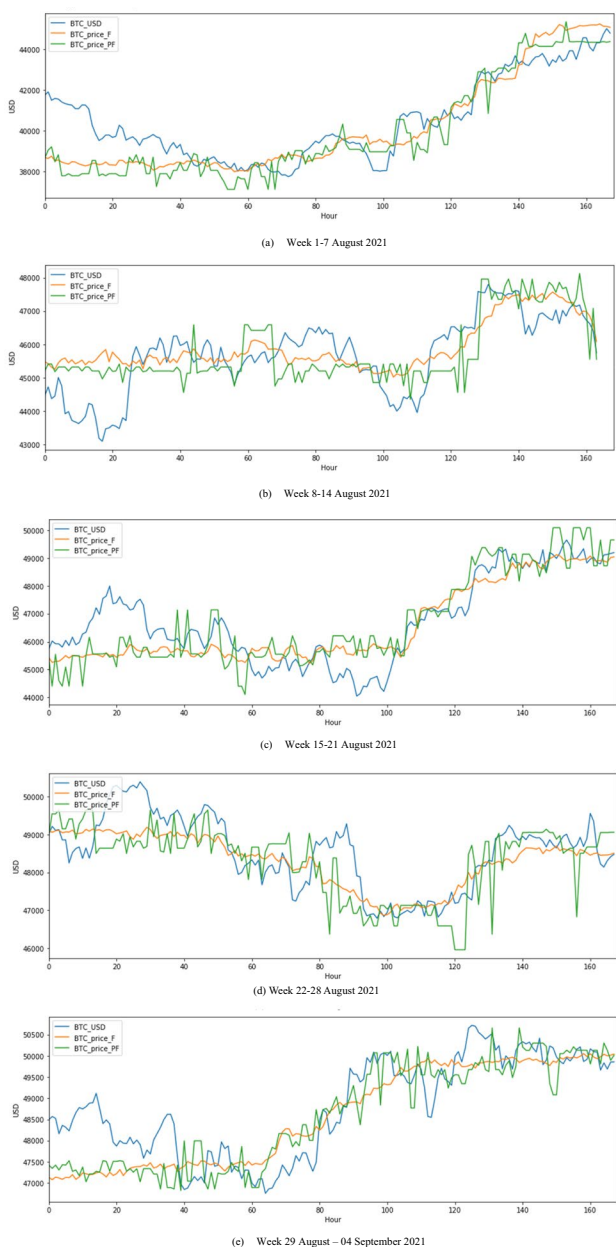
**Fig. 10** The first downward slope – May 2021



During August 2021, the Bitcoin price suffered significant oscillations. During the first two weeks it recorded ups and downs, but by the end of the third week, it reached again the 50,000 USD threshold. The next two weeks were characterized by ups and downs but at the beginning of September, the price was again around 50,000 USD. During August, the

prediction is also good. The green and orange curves most of the time followed the real price. Sudden drops like in the first day of the second week of almost 2,000 USD or the sudden increase in the last week were difficult to predict.

The second upward slope was recorded in Oct 2021. The first training interval spanned from January 2019 to



**Fig. 11** The first upward slope – August 2021

30 September 2021. The prediction is tested during the following weeks: 1–7, 8–14, 15–21, 22–28 October, 29 October – 4 November. The results are graphically presented on a weekly basis (Fig. 12). In the first week of October, the price increased from almost 45,500 to 55,000 USD clearly indicating a bullish market. After the second week, the Bitcoin price went up to more than 59,000 USD (reaching the value recorded at the beginning of May). The prediction in the first two weeks is very good, the real price was closely followed by the prediction curves.

In the third week of October, the price continued to increase up to almost 67,000 USD. However, the sudden

increase from 64,000 to 67,000 USD is difficult to predict. In the next week (the fourth week in October), only one noticeable drop to almost 58,000 USD took place, but the price recovered in a day or two and it stabilized at 62,000 USD at the beginning of November.

The second downward slope was recorded in November 2021 after the second week when the maximum value was recorded (around 69,000 USD) and then decreased by more than 20,000 USD and at the beginning of December it valued 48,000 USD. The first training interval spanned from January 2019 to 31 October 2021. The prediction is tested during the following weeks: 1–7, 8–14, 15–21, 22–28 November, 29 November – 5 December. The results are graphically presented on a weekly basis (Fig. 13).

The prediction for November is again very good. Only several days of sudden ups and downs (very limited intervals) raise prediction issues when price variations were on a steep slope. In any case, upon the return from these sudden price movements, the prediction continued to maintain the trend.

The numerical results are illustrated in Tables 3, 4, 5, 6.

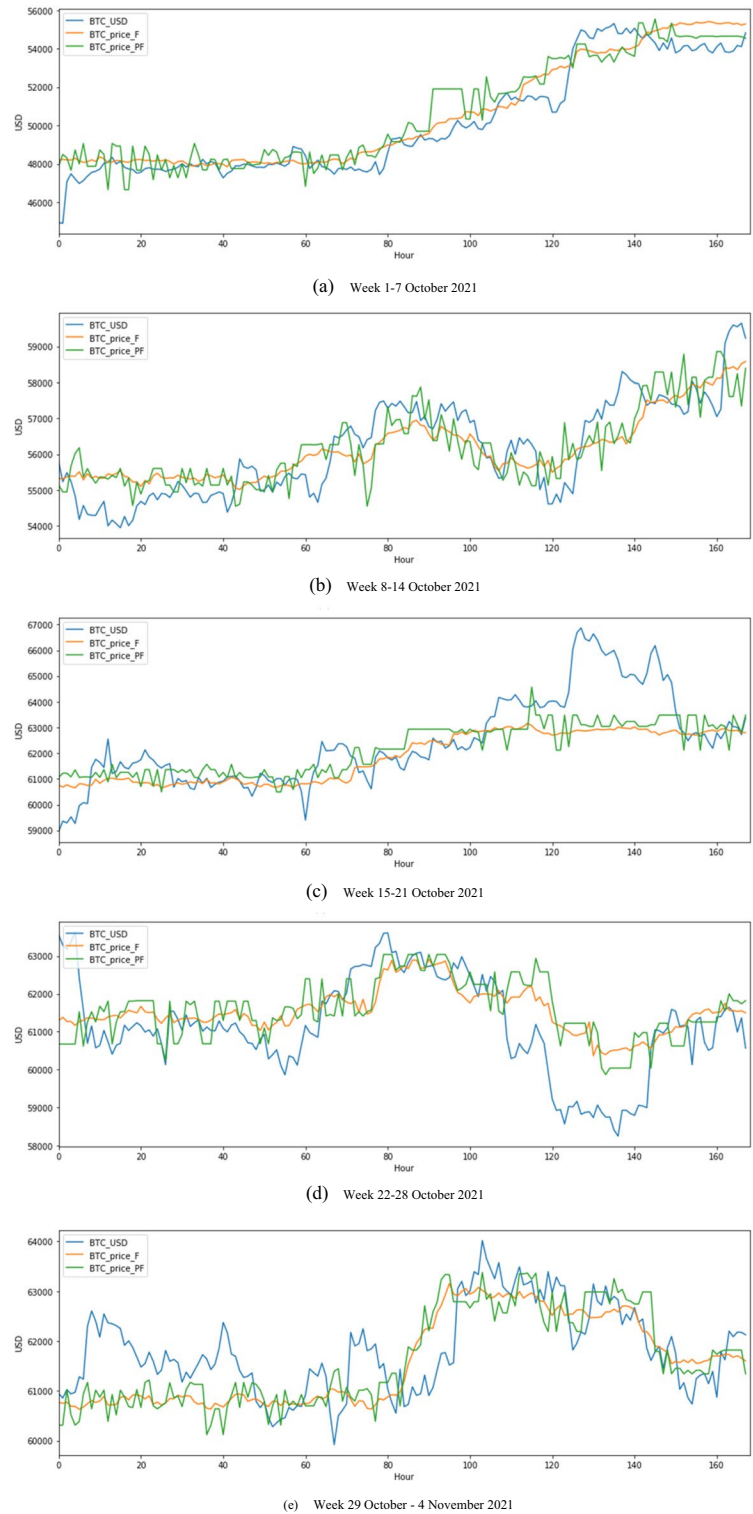
In Table 7, a comparison between the proposed Meta-Model (MM) and the five individual ML algorithms (RF, LGB, HGB, XGB and VR) is provided. The feature engineering is identical and on average, MAE improves by 7.42% and RMSE improves by 3.25% using the MM. However, without the proposed feature engineering (statistical variables and SQL analytical functions), MAE improves by 31.98% and RMSE improves by 35.55% using the MM.

The following key parameters were considered in simulations (as in Table 8):

## 5 Conclusions

This study focuses on the importance of Bitcoin in the world economy, from the perspective of the relationship with energy consumption. The interest of investors, researchers, public authorities for Bitcoin and in general for cryptocurrencies is growing considering the multiple values that Bitcoin has and its role as a star on the financial market despite its volatility and lack of regulation. The innovation process on the financial market is very intense, but by far, cryptocurrencies are considered the most significant financial innovation that brings together new elements both from an economic and technical point of view. Blockchain technology, which is the basis of cryptocurrencies, is of great interest for other fields as well, given its potential to transform several industries in the context of the twin transition.

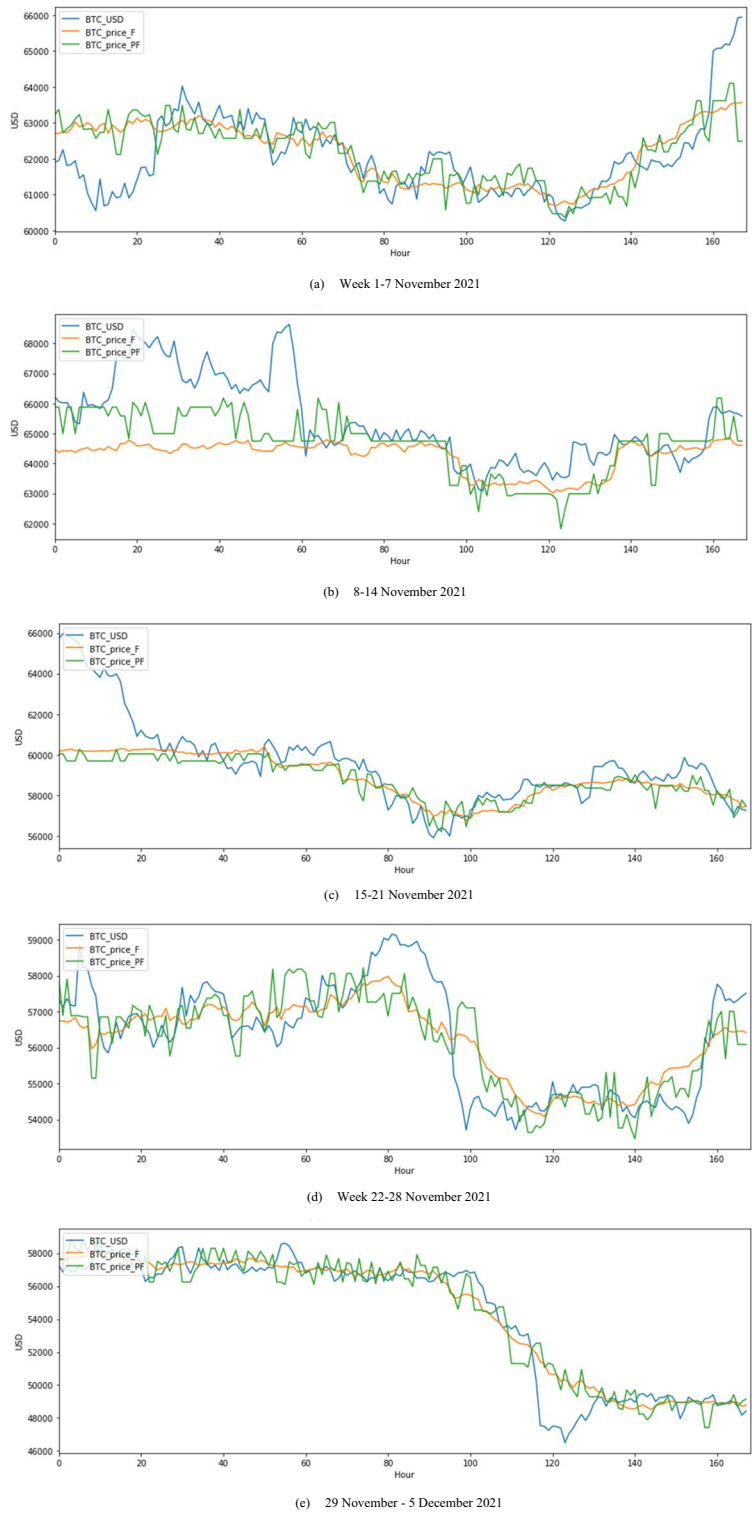
**Fig. 12** The second upward—  
October 2021



Bitcoin price prediction is essential from the perspective of investors who are looking for solutions to optimize the structure of portfolios, but also to reduce the herd behavior that is specific to markets under the sign of the bear. Researchers' attention has gradually shifted from purely financial aspects

related to Bitcoin trading, namely the growing volatility, interdependence with other segments of the financial market, the need to regulate the issuance and trading of these assets to sustainability issues like energy consumption and carbon emissions that generate cryptocurrency mining.

**Fig. 13** The second downward slope – November 2021



**Table 3** Metrics for the first downward slope (May 2021)

Metric	01–07 May	08–14 May	15–21 May	22–28 May	29 May-04 Jun
MAE	184.11	182.44	175.46	191.48	196.52
MAPE	0.00781	0.00809	0.00735	0.00822	0.00815
RMSE	380.00	348.92	351.55	380.78	366.72
R <sup>2</sup>	0.9994	0.9995	0.9995	0.9995	0.9995

**Table 4** Metrics for the first upward slope (August 2021)

Metric	01–07 Aug	08–14 Aug	15–21 Aug	22–28 Aug	29 Aug.-04 Sep
MAE	240.58	259.98	241.62	230.70	230.09
MAPE	0.00854	0.00898	0.00860	0.00828	0.00833
RMSE	448.97	487.33	453.05	432.45	423.49
R <sup>2</sup>	0.9993	0.9992	0.9993	0.9993	0.9994

**Table 5** Metrics for the second upward slope (October 2021)

Metric	01–07 Oct	08–14 Oct	15–21 Oct	22–28 Oct	29 Oct.-04 Nov
MAE	246.74	259.62	235.98	268.02	240.22
MAPE	0.00870	0.00854	0.00818	0.00878	0.00772
RMSE	440.18	458.10	436.41	471.66	434.85
R <sup>2</sup>	0.9993	0.9993	0.9994	0.9993	0.9994

The Bitcoin-energy relationship is important both from the perspective of investors who, based on the evolution of energy prices, can predict the price of this cryptocurrency, but also of public authorities who are increasingly concerned about twin transition. On the one hand, the mining process specific to cryptocurrencies has implications on the energy transition through the reconsideration of nuclear energy and the orientation towards renewable energy to reduce the negative externalities on the environment. On the other hand, blockchain technology specific to cryptocurrencies can be a solution for the digitization of the energy market that faces the problems generated by the production of renewable energy that comes with specific challenges related to intermittent production and the need to store surplus energy.

The proposed prediction using the Meta-Model (MM) for Bitcoin prices was rigorously tested during periods of high volatility and strong market fluctuations. These testing intervals were chosen based on intervals when Bitcoin experienced substantial price shifts. Specifically, the MM was assessed during the months when Bitcoin's value saw notable declines, such as in May and November, where prices fell from approximately \$59,000 to \$34,000 and from \$68,000 to \$46,000, respectively. Additionally, periods of significant price increases were also considered, like in August and October, where Bitcoin's value rose from around \$38,000 to over \$50,000 and from \$46,000 to \$67,000. To ensure the model's robustness and to mitigate the influence of chance, its forecasting ability was evaluated across a 16-week timeframe. Moreover, to compare the results of the proposed MM and the individual models (RF, LGB, XGB, HGB and VR), the testing interval span over the entire year.

By stacking the ensemble ML models and with the newly added features, on average, the RMSE improved by 35.55% and MAE improved by 31.98% compared to the case with individual models and only fundamental variables. Nevertheless, when the feature engineering is identical, MAE improves by 7.42% and RMSE improves by 3.25% proving the superiority of the proposed MM.

**Table 6** Metrics for the second downward slope (November 2021)

Metric	01–07 Nov	08–14 Nov	15–21 Nov	22–28 Nov	29 Nov.-05 Dec
MAE	253.73	264.23	260.31	264.597	263.75
MAPE	0.00823	0.00839	0.00837	0.00827	0.00819
RMSE	444.35	467.17	463.62	473.93	473.12
R <sup>2</sup>	0.9994	0.9994	0.9994	0.9994	0.9994



**Table 7** Comparison between MM and the five individual ML algorithms

Metrics	Models	May 2021	August 2021	October 2021	November 2021
MAE	RF	201.76	255.32	264.45	272.62
	LGB	196.87	251.25	266.89	271.87
	HGB	253.53	275.29	278.99	289.23
	XGB	192.24	252.43	265.24	270.02
	VR	198.29	249.99	263.19	268.76
	MM	186.00	240.59	250.12	261.32
MAPE	RF	0.009812	0.009888	0.010218	0.010168
	LGB	0.009321	0.011982	0.012658	0.011857
	HGB	0.015634	0.012753	0.019825	0.018823
	XGB	0.008973	0.013932	0.011231	0.010634
	VR	0.009832	0.010175	0.009782	0.009989
	MM	0.007924	0.008546	0.008384	0.008290
RMSE	RF	378.61	462.96	461.82	479.96
	LGB	373.92	463.27	462.37	477.27
	HGB	380.58	478.89	475.29	486.89
	XGB	375.98	461.52	460.36	478.53
	VR	371.28	459.29	457.82	479.29
	MM	365.59	449.06	448.24	464.44
R <sup>2</sup>	RF	0.9988	0.9986	0.9982	0.9984
	LGB	0.9989	0.9985	0.9981	0.9983
	HGB	0.9987	0.9988	0.9979	0.9980
	XGB	0.9989	0.9989	0.9982	0.9984
	VR	0.9989	0.9988	0.9989	0.9988
	MM	0.9994	0.9993	0.9993	0.9994

**Table 8** Key parameters

Model	Parameters
RF	<ul style="list-style-type: none"> <li>- n_estimators: 100 (number of trees in the forest)</li> <li>- criterion: 'gini' for classification, 'mse' for regression (function to measure the quality of a split)</li> <li>- max_depth: 10 (maximum depth of a tree)</li> <li>- min_samples_split: 2 (minimum number of samples required to split an internal node)</li> <li>- min_samples_leaf: 1 (minimum number of samples required to be at a leaf node)</li> <li>- max_features: 'auto' (number of features to consider when looking for the best split)</li> </ul>
XGB	<ul style="list-style-type: none"> <li>- booster: 'gbtree' (type of booster used, typically tree-based)</li> <li>- eta: 0.3 (learning rate)</li> <li>- max_depth: 10 (maximum depth of a tree)</li> <li>- min_child_weight: 1 (minimum sum of instance weight needed in a child)</li> <li>- subsample: 1 (subsample ratio of the training instances)</li> <li>- colsample_bytree: 1 (subsample ratio of columns when constructing each tree)</li> </ul>
LGB	<ul style="list-style-type: none"> <li>- boosting_type: 'gbdt' (type of algorithm to use, gradient boosting decision tree)</li> <li>- num_leaves: 31 (maximum number of leaves in one tree)</li> <li>- max_depth: 10 (maximum depth of a tree)</li> <li>- learning_rate: 0.001 (learning rate)</li> <li>- n_estimators: 100 (number of boosting iterations)</li> <li>- subsample_for_bin: 200,000 (number of samples for constructing bins)</li> </ul>
HGB	<ul style="list-style-type: none"> <li>- loss: 'auto' (loss function to use, auto depends on whether it's a regression or classification task)</li> <li>- learning_rate: 0.001 (learning rate)</li> <li>- max_iter: 100 (maximum number of iterations)</li> <li>- max_depth: 10 (maximum depth of each tree)</li> <li>- min_samples_leaf: 20 (minimum number of samples per leaf)</li> </ul>

Ensemble models are valuable because they take advantage of the diversity of models' predictions, leading to improved generalization and better performance. However, it is important to note that while the proposed ensemble model significantly enhances accuracy, it further increases complexity and requires more computational resources. For instance, RF trains multiple trees and storing them can be memory-intensive, and thus training large ensembles can require significant computation. Additionally, finding the right mix of models and hyper-parameters for an ensemble involves trial and error steps.

Predicting the future price of Bitcoin, the world's most famous cryptocurrency, remains a challenging endeavor, particularly within a short forecasting horizon of 1 to 7 days. Its highly volatile nature, coupled with a multitude of influencing factors, poses significant challenges in accurately forecasting its trajectory. One of the primary obstacles lies in Bitcoin's inherent volatility. Its price is prone to sharp swings, often defying conventional market patterns. For instance, while the situation remained relatively stable for an extended period, from March to October 2023, a significant change occurred in November. During this month, there was a rapid Bitcoin price escalation, leading to the current bullish phase. This unpredictability stems from its decentralized nature, lack of regulatory oversight and the involvement of various global factors, making it difficult to establish a consistent relationship between price and its underlying factors. Another challenge is the vast array of factors that influence Bitcoin's price. The complexity of data analysis further compounds the difficulty of predicting Bitcoin's price. The sheer volume and diversity of data from various sources necessitate sophisticated data collection and pre-processing techniques to extract meaningful insights.

Moreover, the dynamic nature of the cryptocurrency market introduces additional challenges. As Bitcoin's adoption and usage grow, the market becomes more interconnected, with events in one region potentially influencing prices globally. This interconnectedness can amplify the impact of unexpected events, making it difficult to predict their precise effect on Bitcoin's price. Despite these challenges, researchers and analysts are constantly exploring new methods and techniques to improve Bitcoin price prediction accuracy. ML algorithms, with their ability to analyze vast amounts of data and identify complex patterns, are gaining traction in this field.

In conclusion, predicting Bitcoin's price within a short forecasting horizon of 1 to 7 days is a challenging task due to its volatility, the multitude of influencing factors and the complexity of data analysis. In our upcoming work, we aim to broaden the scope of our initial dataset by incorporating additional variables. These include sentiment analysis of news and specialized platforms, along with the identification of speculative trends using natural language processing techniques, to provide a more comprehensive view of Bitcoin's price dynamics.

## Appendix A

**Table 9** Data sample

Date	Y	m	h	O	H	L	BTC_ USD	V	QAV	Num_ trades	Taker_ BV	Taker_QV	EL_ price_ DAM	Gas_ price_ DAM	Infla- tion_ EU	Oil_ price	Price_ EUETS	EL_ quan- tity	Gas_ quan- tity_ DAM
1/5/2020	2020	1	0	7348.04	7365	7330	7430.05	1272.743	9,349,986	10,447	754.2173	5,541,046	159.87	61.1	1.4	70.25	25.5	3259.3	1365
1/5/2020	2020	1	1	7346.4	7369.18	7339.28	7427.74	1084.111	7,969,047	10,286	632.4431	4,649,101	151.87	61.1	1.4	70.25	25.5	3278	1365
1/5/2020	2020	1	2	7364.32	7369.63	7304.36	7415.12	1807.552	13,285,643	13,744	901.6336	6,628,279	145.18	61.1	1.4	70.25	25.5	3245.1	1365
1/5/2020	2020	1	3	7318.15	7339.27	7302.1	7422.33	1520.455	11,137,558	12,378	845.0462	6,190,647	140	61.1	1.4	70.25	25.5	3225.1	1365
1/5/2020	2020	1	4	7336.75	7339.58	7305	7457.01	1131.423	8,280,307	10,870	695.2298	5,087,716	148	61.1	1.4	70.25	25.5	3214.3	1365

**Table 10** Descriptive statistics of the input dataset for 2019

2019	<i>BTC_USD</i>	<i>El_price_DAM</i>	<i>Gas_price_DAM</i>	<i>Inflation_EU</i>	<i>Oil_price</i>	<i>Price_EUETS</i>	<i>El_quantity</i>	<i>Gas_quantity_DAM</i>
Mean	7362.522	239.5852	28.43903	1.189089	64.3559	26.09134	2642.024	1,896,357
Standard Error	28.32383	1.073505	0.425821	0.003044	0.047417	0.022064	5.854789	29,995.04
Median	7816.72	229	0.11	1.2	64	26.15	2635.5	890,000
Mode	3962	229	0.1	1	74.39	25.68	2757.1	376,800
Standard Deviation	2642.934	100.1702	39.73387	0.284035	4.42457	2.058852	546.318	2,798,877
Sample Variance	6,985,098	10,034.07	1578.78	0.080676	19.57682	4.23887	298,463.4	7.83E+12
Kurtosis	-1.21493	2.915351	-1.11916	-0.87527	-0.47807	0.313743	-0.35182	12.01878
Skewness	-0.08057	1.112342	0.810354	-0.06112	0.294883	-0.09585	0.257978	2.94787
Range	10,594.35	749.9	113.51	1	21.71	11.01	3247.8	19,879,960
Minimum	3366.41	0.1	0.07	0.7	53.23	20.32	1401.7	56
Maximum	13,960.76	750	113.58	1.7	74.94	31.33	4649.5	19,880,016
Sum	64,105,476	2,086,068	247,618.6	10,353.4	560,346.8	227,177.3	23,004,099	1.65E+10
Count	8707	8707	8707	8707	8707	8707	8707	8707

**Table 11** Descriptive statistics of the input dataset for 2020

2020	<i>BTC_USD</i>	<i>El_price_DAM</i>	<i>Gas_price_DAM</i>	<i>Inflation_EU</i>	<i>Oil_price</i>	<i>Price_EUETS</i>	<i>El_quantity</i>	<i>Gas_quantity_DAM</i>
Mean	11,100.55	191.3291	55.79414	0.246923	41.80641	25.43888	2837.304	8278.028
Standard Error	45.74536	0.911039	0.125088	0.006084	0.128772	0.038154	6.421505	65.72077
Median	9699.615	183	55.95	0.1	42.35	25.55	2768.05	6740
Mode	11,332	229	67.22	-0.3	21.74	24.02	2686	1404
Standard Deviation	4277.13	85.18096	11.69552	0.568871	12.03996	3.56736	600.4021	6144.804
Sample Variance	18,293,837	7255.796	136.7851	0.323614	144.9607	12.72606	360,482.7	37,758,618
Kurtosis	3.39173	4.080447	-1.32311	-0.58754	0.114655	-0.20201	0.064902	3.164561
Skewness	1.806442	1.274487	-0.13184	0.785001	-0.20441	-0.2676	0.620056	1.716363
Range	25,279.35	730.28	48.47	1.7	61.13	17.54	3482.4	33,474
Minimum	4130.64	0.1	31.61	-0.3	9.12	16.12	1570.5	100
Maximum	29,409.99	730.38	80.08	1.4	70.25	33.66	5052.9	33,574
Sum	97,041,033	1,672,599	487,752.3	2158.6	365,471.6	222,386.7	24,803,711	72,366,520
Count	8742	8742	8742	8742	8742	8742	8742	8742

**Table 12** Descriptive statistics of the input dataset for 2021

2021	<i>BTC_USD</i>	<i>El_price_DAM</i>	<i>Gas_price_DAM</i>	<i>Inflation_EU</i>	<i>Oil_price</i>	<i>Price_EUETS</i>	<i>El_quantity</i>	<i>Gas_quantity_DAM</i>
Mean	47,374.51	549.7568	219.5546	2.610851	70.7203	53.67466	2952.918	11,432.98
Standard Error	104.8115	4.200483	1.635339	0.014971	0.086748	0.13439	6.148384	93.59948
Median	47,895.5	397.49	148.14	2.2	70.9	53.76	2923.65	8966
Mode	38,700	359	85.15	0.9	61.47	42.81	3055.3	2399
Standard Deviation	9801.977	392.8294	152.937	1.40012	8.112699	12.56819	574.9972	8753.428
Sample Variance	96,078,761	154,314.9	23,389.73	1.960336	65.81588	157.9595	330,621.8	76,622,497
Kurtosis	-1.13998	2.783413	0.514849	-1.11513	-0.34408	-0.14	-0.35343	1.234839
Skewness	-0.04305	1.620409	1.115896	0.485931	-0.27645	0.384239	0.244697	1.24588
Range	39,633.68	2684.62	788.06	4.1	35.39	57.57	3477.8	46,649.81
Minimum	29,000.01	0.1	60.52	0.9	50.37	31.84	1561.3	47
Maximum	68,633.69	2684.72	848.58	5	85.76	89.41	5039.1	46,696.81
Sum	4.14E+08	4,808,173	1,920,225	22,834.5	618,519.7	469,438.6	25,826,224	99,992,862
Count	8746	8746	8746	8746	8746	8746	8746	8746

**Table 13** Descriptive statistics of the input dataset for 2022

2022	<i>BTC_USD</i>	<i>El_price_DAM</i>	<i>Gas_price_DAM</i>	<i>Inflation_EU</i>	<i>Oil_price</i>	<i>Price_EUETS</i>	<i>El_quantity</i>	<i>Gas_quantity_DAM</i>
Mean	39,402.77	1022.676	494.7669	6.784496	103.823	83.48569	3054.492	23,478.78
Standard Error	81.39496	7.169671	2.036329	0.018511	0.205903	0.112741	10.24772	184.386
Median	40,074.94	1009.26	469.9	7.4	104.94	83.46	3061	21,395
Mode	43,200	599	495.72	7.4	105.49	79.97	3032.1	15,171.29
Standard Deviation	4883.019	430.1205	122.1628	1.110501	12.35248	6.763535	614.7778	11,061.63
Sample Variance	23,843,876	185,003.6	14,923.74	1.233211	152.5838	45.7454	377,951.8	1.22E+08
Kurtosis	-0.12061	2.668024	8.518703	-1.39747	-0.55103	1.10425	-0.13502	0.679328
Skewness	-0.7231	0.817834	2.370524	-0.41443	-0.04455	-0.60691	-0.12594	0.916533
Range	20,856.15	3283.76	887.62	3	54.93	38.63	3432.2	56,327.71
Minimum	27,114.84	2.24	295.11	5.1	78.25	58.3	1244.1	3422.24
Maximum	47,970.99	3286	1182.73	8.1	133.18	96.93	4676.3	59,749.95
Sum	1.42E+08	3,680,612	1,780,666	24,417.4	373,659.1	300,465	10,993,118	84,500,117
Count	3599	3599	3599	3599	3599	3599	3599	3599

**Table 14** Descriptive statistics of the input dataset for 2019–2022

2019–2022	<i>BTC_USD</i>	<i>El_price_DAM</i>	<i>Gas_price_DAM</i>	<i>Inflation_EU</i>	<i>Oil_price</i>	<i>Price_EUETS</i>	<i>El_quantity</i>	<i>Gas_quantity_DAM</i>
Mean	24,075.13	411.0711	148.8978	2.005904	64.37528	40.92997	2840.409	562,812.7
Standard Error	109.5027	2.215894	1.034248	0.012708	0.120867	0.123995	3.458756	10,073.77
Median	11,605.09	263.35	70.85	1.3	64.31	28.59	2808	10,764
Mode	3962	229	0.1	-0.3	59.46	25.68	2686	376,800
Standard Deviation	18,901.2	382.4842	178.5209	2.193536	20.8627	21.40269	597.0137	1,738,828
Sample Variance	3.57E+08	146,294.1	31,869.71	4.8116	435.2523	458.0751	356,425.4	3.02E+12
Kurtosis	-1.18936	5.099674	2.200848	0.94547	0.760388	-0.24928	-0.21198	40.24929
Skewness	0.585637	2.108941	1.618773	1.37163	0.400356	1.052557	0.330264	5.44328
Range	65,267.28	3285.9	1182.66	8.4	124.06	80.81	3808.8	19,879,969
Minimum	3366.41	0.1	0.07	-0.3	9.12	16.12	1244.1	47
Maximum	68,633.69	3286	1182.73	8.1	133.18	96.93	5052.9	19,880,016
Sum	7.17E+08	12,247,452	4,436,262	59,763.9	1,917,997	1,219,468	84,627,152	1.68E+10
Count	29,794	29,794	29,794	29,794	29,794	29,794	29,794	29,794

**Acknowledgements** This work was supported by a grant from the Ministry of Research, Innovation and Digitization, CNCS- UEFIS-CDI, project number PN-III-P4-PCE-2021-0334, within PNCDI III.

**Author contributions** Adela Bâra: Conceptualization, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualization, Supervision. Simona-Vasilica Oprea: Method, Validation, Formal analysis, Investigation, Writing – Original Draft, Writing – Review and Editing, Visualization, Project administration, Mirela Panait: Writing on Introduction, Literature survey, Conclusion.

**Data availability** The data will be made available upon reasonable request.

## Declarations

**Ethical approval** Not applicable.

**Informed consent** Not applicable.

**Competing interest** The authors have no relevant financial or non-financial interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agur I, Lavyssière X, Bauer GV, Deodoro J, Peria SM, Sandri D, Tourpe H (2023) Lessons from crypto assets for the design of energy efficient digital currencies. *Ecol Econ* 212:107888
- Akbulaev N, Abdulhasanov T (2023) Analyzing the Connection between Energy Prices and Cryptocurrency throughout the Pandemic Period. *International Journal of Energy Economics and Policy* 13(1):227–234
- Alexandros N (2021) Cryptocurrency analysis: Benefits, dangers and price prediction using neural networks. *Romanian Journal of Economics* 52(1):61
- Al-Shboul M, Assaf A, Mokni K (2022) When bitcoin lost its position: Cryptocurrency uncertainty and the dynamic spillover among cryptocurrencies before and during the COVID-19 pandemic. *Int Rev Financ Anal* 83:102309
- Aivaz KA, Munteanu IF, Jakubowicz FV (2023) Bitcoin in Conventional Markets: A Study on Blockchain-Induced Reliability, Investment Slopes. *Financial and Accounting Aspects Mathematics* 11(21):4508
- Bastian-Pinto CL, Araujo FVDS, Brandão LE, Gomes LL (2021) Hedging renewable energy investments with Bitcoin mining. *Renew Sustain Energy Rev* 138:110520
- Biju AVN, Thomas AS (2023) Uncertainties and ambivalence in the crypto market: an urgent need for a regional crypto regulation. *SN Business & Economics* 3(8):136
- Bjerg O (2016) How is bitcoin money? *Theory Cult Soc* 33(1):53–72
- Bibi S (2023) Money in the time of crypto. *Res Int Bus Financ* 65:101964
- Bouri E, Gabauer D, Gupta R, Tiwari AK (2021) Volatility connectedness of major cryptocurrencies: The role of investor happiness. *J Behav Exp Financ* 30:100463
- Corbet S, Lucey B, Yarovaya L (2021) Bitcoin-energy markets interrelationships-New evidence. *Resour Policy* 70:101916
- De Vries A (2018) Bitcoin's growing energy problem. *Joule* 2(5):801–805
- Dennin T (2023) Digital Gold and Gold-Backed Cryptocurrencies: The Return of the Gold Standard. *Financial Innovation and Value Creation: The Impact of Disruptive Technologies on the Digital World*. Springer International Publishing, Cham, pp 3–20
- Derbali A, Jamel L, Ben Ltaifa M, Elnagar AK, Lamouchi A (2020) Fed and ECB: which is informative in determining the DCC between bitcoin and energy commodities? *J Cap Markets Stud* 4(1):77–102
- Dogan E, Majeed MT, Luni T (2022) Are clean energy and carbon emission allowances caused by bitcoin? A novel time-varying method. *J Clean Prod* 347:131089
- Dzidzikashvili D, Kheladze M (2022) The future of blockchain tech in transactional business. *Econ Insights-Trends & Challenges XI(LXXIV)(4):77–97*. <https://www.doi.org/10.51865/EITC.2022.04.05>
- Dulisse BC, Connealy N, Logan MW (2024) The influence and role of cryptoculture on target congruence in cryptocurrency investment behavior: a theoretical model. *Crime Law Soc Change* 81:421–441. <https://doi.org/10.1007/s10611-023-10126-6>
- Erdogan S, Ahmed MY, Sarkodie SA (2022) Analyzing asymmetric effects of cryptocurrency demand on environmental sustainability. *Environ Sci Pollut Res Int* 29(21):31723
- Fang F, Ventre C, Basios M, Kanthan L, Martinez-Rego D, Wu F, Li L (2022) Cryptocurrency trading: a comprehensive survey. *Financial Innovation* 8(1):1–59
- Foteinis S (2018) Bitcoin's alarming carbon footprint. *Nature* 554(7690):169–170
- García-Corral FJ, Cordero-García JA, de Pablo-Valenciano J, Uribe-Toril J (2022) A bibliometric review of cryptocurrencies: how have they grown? *Financial Innovation* 8(1):1–31
- Goodkind AL, Jones BA, Berrens RP (2020) Cryptodamages: Monetary value estimates of the air pollution and human health impacts of cryptocurrency mining. *Energy Res Soc Sci* 59:101281
- Gurrib I, Nourani M, Bhaskaran RK (2022) Energy crypto currencies and leading US energy stock prices: are Fibonacci retracements profitable? *Financial Innovation* 8:1–27
- Hashemi Joo M, Nishikawa Y, Dandapani K (2020) Cryptocurrency, a successful application of blockchain technology. *Manag Financ* 46(6):715–733. <https://doi.org/10.1108/MF-09-2018-0451>
- Huynh TLD, Shahbaz M, Nasir MA, Ullah S (2022) Financial modelling, risk management of energy instruments and the role of cryptocurrencies. *Ann Oper Res* 313(1):47–75
- Kohli V, Chakravarty S, Chamola V, Sangwan KS, Zeadally S (2023) An analysis of energy consumption and carbon footprints of cryptocurrencies and possible solutions. *Digital Communications and Networks* 9(1):79–89
- Küfeoğlu S, Özkuran M (2019) Bitcoin mining: A global review of energy and power demand. *Energy Res Soc Sci* 58:101273
- Kumar S (2021) Review of geothermal energy as an alternate energy source for Bitcoin mining. *Journal of Economics and Economic Education Research* 23(1):1–12

29. Li J, Li N, Peng J, Cui H, Wu Z (2019) Energy consumption of cryptocurrency mining: A study of electricity consumption in mining cryptocurrencies. *Energy* 168:160–168
30. Long SC, Lucey B, Zhang D, Zhang Z (2023) Negative elements of cryptocurrencies: exploring the drivers of bitcoin carbon footprints. *Finance Res Lett* 58:104031. <https://doi.org/10.1016/j.frl.2023.104031>
31. Maiti M (2022) Dynamics of bitcoin prices and energy consumption. *Chaos, Solitons & Fractals*: X 9:100086
32. Miśkiewicz R, Matan K, Karnowski J (2022) The role of crypto trading in the economy, renewable energy consumption and ecological degradation. *Energies* 15(10):3805
33. Mohsin M, Naseem S, Ivaşcu L, Cioca LI, Sarfraz M, Stănică NC (2021) Gauging the effect of investor sentiment on Cryptocurrency market: an analysis of Bitcoin currency. *Romanian Journal of Economic Forecasting* 24(4):87
34. Maurushat A, Halpin D (2022) Investigation of Cryptocurrency Enabled and Dependent Crimes. *Financial Technology and the Law: Combating Financial Crime*. Springer International Publishing, Cham, pp 235–267
35. Nández Alonso SL, Jorge-Vázquez J, EcharteFernández MÁ, Reier-Forradellas RF (2021) Cryptocurrency mining from an economic and environmental perspective. Analysis of the most and least sustainable countries. *Energy* 14(14):4254
36. Pagone E, Hart A, Salonitis K (2023) Carbon Footprint Comparison of Bitcoin and Conventional Currencies in a Life Cycle Analysis Perspective. *Procedia CIRP* 116:468–473
37. Rao A, Gupta M, Sharma GD, Mahendru M, Agrawal A (2022) Revisiting the financial market interdependence during COVID-19 times: a study of green bonds, cryptocurrency, commodities and other financial markets. *International Journal of Managerial Finance* 18(4):725–755
38. Salisu AA, Obiora K (2021) COVID-19 pandemic and the crude oil market risk: hedging options with non-energy financial innovations. *Financial Innovation* 7(1):1–19
39. Sapra N, Shaikh I (2023) Impact of Bitcoin mining and crypto market determinants on Bitcoin-based energy consumption. *Manag Financ* 49(11):1828–1846. <https://doi.org/10.1108/MF-03-2023-0179>
40. Sarker PK, Lau CKM, Pradhan AK (2023) Asymmetric effects of climate policy uncertainty and energy prices on bitcoin prices. *Innovation and Green Development* 2(2):100048
41. Sarkodie SA, Owusu PA (2022) Dataset on bitcoin carbon footprint and energy consumption. *Data Brief* 42:108252
42. Savona P (2022) Prospects for Reforming the Money and Financial System. *Open Econ Rev* 33(1):187–195
43. Şcheau MC, Achim MV (2022) TRENDS IN COMBATING MONEY LAUNDERING IN THE EUROPEAN CONTEXT. In *DIEM: Dubrovnik Int Econ Meet* 7(1):142–152 (**Sveučilište u Dubrovniku**)
44. Şcheau MC, Crăciunescu SL, Brici I, Achim MV (2020) A cryptocurrency spectrum short analysis. *Journal of Risk and Financial Management* 13(8):184
45. Sundaram D (2023) Advancing the Environmental, Social, and Governance (ESG) with Blockchain: A PRISMA Review. In *Blockchain and Applications, 5th International Congress* 778:103
46. Symitsi E, Chalvatzis KJ (2018) Return, volatility and shock spillovers of Bitcoin with energy and technology companies. *Econ Lett* 170:127–130
47. Su CW, Qin M, Tao R, Umar M (2020) Financial implications of fourth industrial revolution: Can bitcoin improve prospects of energy investment? *Technol Forecast Soc Chang* 158:120178
48. Thakur A (2023) Blockchain and Cryptocurrency Frauds: Emerging Concern. *Financial Crimes: A Guide to Financial Exploitation in a Digital Age*. Springer International Publishing, Cham, pp 131–145
49. Tiwari AK, Raheem ID, Kang SH (2019) Time-varying dynamic conditional correlation between stock and cryptocurrency markets using the copula-ADCC-EGARCH model. *Physica A* 535:122295
50. Toumi A, Najaf K, Dhiab MM, Li NS, Kanagasabapathy S (2023) The role of Fintech firms' sustainability during the COVID-19 period. *Environ Sci Pollut Res* 30(20):58855–58865
51. Yan X, Yan WJ, Xu Y, Yuen KV (2023) Machinery multi-sensor fault diagnosis based on adaptive multivariate mode decomposition and multi-attention fusion residual convolutional neural network. *Mech Syst Signal Process* 202:110664
52. Yan X, She D, Xu Y (2023) Deep order-wavelet convolutional variational autoencoder for fault identification of rolling bearing under fluctuating speed conditions. *Expert Syst Appl* 216:119479
53. Yuan X, Su CW, Peculea AD (2022) Dynamic linkage of the bitcoin market and energy consumption: An analysis across time. *Energy Strat Rev* 44:100976
54. Yüksel S, Dinçer H, Çağlayan Ç, Uluer GS, Lisin A (2022) Bitcoin mining with nuclear energy. In: Dinçer H, Yüksel S (eds) *Multidimensional strategic outlook on global competitive energy economics and finance*. Emerald Publishing Limited, Leeds, pp 165–177. <https://doi.org/10.1108/978-1-80117-898-320221019>
55. Zheng M, Feng GF, Zhao X, Chang CP (2023) The transaction behavior of cryptocurrency and electricity consumption. *Financial Innovation* 9(1):1–18
56. Zimmer Z (2017) Bitcoin and potosí silver: historical perspectives on cryptocurrency. *Technol Culture* 58(2):307–334. <https://doi.org/10.1353/tech.2017.0038>
57. Wang Q, Wei Y, Zhang Y, Liu Y (2023) Evaluating the safe-haven abilities of bitcoin and gold for crude oil market: Evidence during the COVID-19 pandemic. *Eval Rev* 47(3):391–432
58. Wen F, Tong X, Ren X (2022) Gold or Bitcoin, which is the safe haven during the COVID-19 pandemic? *Int Rev Financ Anal* 81:102121
59. Wu W, Tiwari AK, Gozgor G, Leping H (2021) Does economic policy uncertainty affect cryptocurrency markets? Evidence from Twitter-based uncertainty measures. *Res Int Bus Financ* 58:101478

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.