Expert Systems With Applications 184 (2021) 115480

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Predicting Ethereum prices with machine learning based on Blockchain information

Han-Min Kim [a], Gee-Woo Bock [a,*], Gunwoong Lee [b]

[a] Business School, Sungkyunkwan University, Seoul, South Korea
[b] Business School, Korea University, Seoul, South Korea

ARTICLE INFO

Keywords:
Ethereum
Predictions
Machine learning
Blockchain information

ABSTRACT

With the growing interest in cryptocurrency and its fundamental algorithm, studies of cryptocurrency price predictions have been actively conducted in various academic disciplines. Since cryptocurrency is generated and consumed by Blockchain systems, Blockchain-specific information can be considered as the main component in forecasting the values of cryptocurrency. This perspective has been widely adopted in studies of Bitcoin price predictions. However, we find that Ethereum, a popular and leading cryptocurrency in the market, has Blockchain information that differs from that of Bitcoin. Hence, this study investigates the relationship between inherent Ethereum Blockchain information and Ethereum prices. Furthermore, we investigate how Blockchain information concerning other publicly available coins on the market is associated with Ethereum prices. Our key findings reveal that macro-economy factors, Ethereum-specific Blockchain information, and the Blockchain information of other cryptocurrency play important roles in the prediction of Ethereum prices.

1. Introduction

Blockchain has attracted significant attention as a technology that can resolve the issue of trust between business participants as it allows sharing and verification of a decentralized ledger with network participants (Antonopoulos & Wood, 2018). Blockchain employs a special system component called cryptocurrency to ensure the stable operation of its systems. Cryptocurrency has created new economic value beyond its purpose of operating Blockchain systems. Cryptocurrency has been widely used for purchasing products and services, as well exchanging legal currencies with coins. With the growing interest in cryptocurrency, its fundamental algorithm (i.e., Blockchain) has received increasing attention from academic researchers. Cryptocurrency is a compensation made when a new block is created and registered formally in Blockchain systems (Antonopoulos, 2014). Bitcoin and Ethereum, which are the most representative cryptocurrencies, have recorded the largest trading volumes and possess the highest market capitalizations. As of April 2021, Bitcoin has recorded a trading volume of USD 64 billion and a market capitalization of USD 1,304 billion, and Ethereum has reached a trading volume of USD 38 billion and a market capitalization of USD 265 billion (CoinMarketCap, 2021).

As cryptocurrency has attracted unprecedented attention, researchers have attempted to predict and analyze the future values of cryptocurrency. Unlike existing currencies or gold, the values of these representative cryptocurrencies (i.e., Bitcoin and Ethereum) cannot be predicted based on previously well-accepted general economic indicators (Ciaian et al., 2016). In line with this, many prior studies have attempted to predict the price of Bitcoin, a cryptocurrency associated with the first-generation Blockchain system. A few recent studies have found that information such as Google Trends for keyword searches of "Bitcoin" and Wikipedia views of Bitcoin is positively associated with the price fluctuations of Bitcoin (Kristoufek, 2013; Abraham et al., 2018). Kristoufek (2013) found that Google Trends predicted Bitcoin prices more accurately than Wikipedia views. Further, Ciaian et al. (2016) found that market-level indicators such as transaction volumes and user activities on Bitcoin community sites (e.g., numbers of postings, discussions, and new members), are significantly associated with future Bitcoin prices. In particular, the increase in new postings on Bitcoin community sites is closely related to the price rise of Bitcoin. Abraham et al. (2018) also showed that Google Trends search queries and tweet volumes about Bitcoin are positively associated with Bitcoin price fluctuations. They found that tweet volumes had a slightly higher positive correlation with Bitcoin price fluctuations than Google Trends.

However, cryptocurrency is issued and consumed by the Blockchain

* Corresponding author.
E-mail addresses: hanmin8809@gmail.com (H.-M. Kim), gwbock@skku.edu (G.-W. Bock), leegw@korea.ac.kr (G. Lee).

https://doi.org/10.1016/j.eswa.2021.115480
Received 9 February 2020; Received in revised form 9 May 2021; Accepted 23 June 2021
Available online 27 June 2021

system. Moreover, records related to cryptocurrency are stored in the system as Blockchain information, which includes different types of data such as the number of issued coins and transaction fees paid in cryptocurrency. Therefore, Blockchain information can be more deeply involved in cryptocurrency price from the perspective of a supply and demand relationship.

As such, recent literature on Bitcoin price predictions using machine-learning techniques has found that Bitcoin's Blockchain information on delivering transaction fees, difficulty levels in generating blocks, and the amount of generated coins are closely related to future Bitcoin prices (Saad et al., 2019). Other studies have reported that macro-economic factors, such as stock indices (e.g., S&P 500, DOW30, and NASDAQ), changes in crude oil and gold prices, global currency ratios, and Bitcoin's Blockchain information, are the critical factors in forecasting the fluctuation of Bitcoin prices (Jang & Lee, 2017; Mallqui & Fernandes, 2019). The studies argue that Blockchain information coupled with advanced machine-learning techniques can improve the accuracy of the prediction of Bitcoin prices and fluctuations (Jang & Lee, 2017; Saad et al., 2019).

Although most recent research has focused on Bitcoin prices, there has been less academic interest in understanding the price fluctuations of Ethereum, another popular and leading cryptocurrency on the market. This new but high-profile cryptocurrency has unique and interesting attributes that should be studied in academic disciplines. Specifically, Ethereum has four distinctive characteristics of a Blockchain system that Bitcoin does not have (Antonopoulos & Wood, 2018). First, unlike Bitcoin, which contains only transaction information in a block, Ethereum can include various types of information, such as user identity and product attributes, along with transaction information in a block. Second, unlike Bitcoin, Ethereum offers compensation coins for an uncle block[1]. Several new blocks can be simultaneously created in the Blockchain system. Among them, blocks with low mining difficulty cannot be registered as a formal block in the Blockchain network, and they eventually remain as uncle blocks without rewards. These uncle blocks can cause security problems and transaction delays in the Blockchain system. However, the Ethereum Blockchain system provides Ethereum rewards to uncle block miners to ensure stable operation of the Blockchain. Ethereum's uncle block reward system can be viewed as the supply of Ethereum, and it may be of significant relevance to the price of Ethereum. Therefore, our attention should be focused on the uncle block reward system of Ethereum. Third, Ethereum introduces the concept of "gas," which is not available in Bitcoin, to adjust its block size and to speed up the system. Ethereum gas is the running cost required to register various types of information in a block. The gas can be purchased with Ethereum. Information can be registered in the Ethereum block using the gas purchased with Ethereum. However, an Ethereum block has a limit on the amount of gas that can be used. In other words, the amount of registered information in a block is limited. Therefore, the Ethereum block size cannot be increased in the same manner that a Bitcoin block can be. Finally, the average block-generation time for Ethereum is much shorter than that for Bitcoin (approximately 15 s vs. 10 min to generate a block) as an Ethereum block is relatively small and does not handle much more than approximately 30 kB of information.

Although Ethereum is the second-generation Blockchain and has the above four distinct structural factors that deserve to be scrutinized in both academia and practice, there are few studies investigating the relationships between Ethereum Blockchain information[2] and Ethereum prices. Extant studies have not considered distinct Ethereum Blockchain information (gas limit, gas usage, gas price, and uncle block) and other coins' information for predicting cryptocurrency prices. Therefore, we highlight the need to use Ethereum Blockchain information to predict the value of Ethereum as it has a structure that is issued and consumed in a Blockchain system. In addition, we consider the Blockchain information[3] of other coins that can potentially affect the price of Ethereum, such as Bitcoin, Dashcoin, and Litecoin, which are also actively traded on the market (Songmuang, 2018).

Based on these inherent Ethereum attributes, this research attempts to identify relevant variables concerning Ethereum prices to achieve more accurate price predictions. Specifically, this study answers the following salient research questions:

(1) Are macro-economic factors related to predicting Ethereum prices?
(2) Is generic Blockchain information and Ethereum-specific Blockchain information, such as compensation for uncle blocks, gas limits, gas price, and gas consumption, associated with the successful prediction of Ethereum prices?
(3) Is the Blockchain information of other coins (i.e., Bitcoin, Dashcoin, and Litecoin) related to the prediction of Ethereum prices?

To answer these research questions, we used a set of advanced machine-learning techniques to forecast Ethereum prices using various factors. We employed generic Blockchain information, Ethereum-specific Blockchain information, and other coins' Blockchain information to predict Ethereum prices. We empirically identified a new but more valid set of variables that are closely related to Ethereum future prices. We expect that the key findings of this research will offer a new perspective to Blockchain research in various academic disciplines. In practice, the findings allow faster and more accurate prediction of Ethereum prices. Practitioners could conduct proper investment and sale of Ethereum. In addition, we provide Blockchain information that can aid the design of the structure of the Blockchain.

2. Literature review

2.1. Blockchain

Blockchain is "a record-keeping technology that is nearly impossible to tamper with" (Ferguson, 2018; pp.2). Since it is generally difficult to alter transaction-related information after a user has entered it into the system, Blockchain is considered to provide transaction trust and transparency to all participants. A block in a Blockchain has a hash value, a so-called "unique identity." Every block consists of a header and a body portion. The block header includes a nonce value (i.e., a number used once) used to generate a new block, a difficulty of block creation, a hash value of the previous block, a time stamp, the number of blocks, and a Merkle root, which refers to the hash value of the transaction information in the block body. The hash value of the block is created by a hash function with the block header information. The block body contains transaction information. Blockchain has a structure in which blocks are connected to each other in a chain. Blockchain system must employ a unique hash value of the previous block when creating a new block. The hash value of the block provides a completely different hash value even if there is a small change in the input values. Therefore, it is almost impossible for an attacker to manipulate the information registered in the Blockchain unless the attacker (or attackers) has more than 51% of the total computational power of the Blockchain network (Jang & Lee, 2017; Antonopoulos & Wood, 2018).

In general, coins in Blockchain are paid as compensation to miners

[1] "Uncle block" refers to a block that was not registered as a formal block in the Blockchain network.

[2] **Ethereum Blockchain information refers to data records such as transaction volume, transaction count, generated coins, uncle block, gas price, and gas limit that occur while operating the Ethereum Blockchain system.**

[3] **Blockchain information means general data records such as transaction volume, transaction count, and generated coins that occur while operating the Blockchain system.**

who formally created new blocks and registered them in the Blockchain network. This process is called Proof-of-Work. Coins can be spent by adding transaction information to a Blockchain network and can be exchanged for other coins.

2.2. Research related to prediction of cryptocurrency price and Blockchain applications

Previous literature on Blockchain has focused heavily on cryptocurrency prices and their applications to various industrial domains, such as finance, health care, or supply chain management (SCM) (Guo & Liang, 2016; Yue et al., 2016; Wu et al., 2017; Yli-Huumo et al., 2016). For example, health care studies have suggested a method to securely manage medical data and provide data usage rights to users based on Blockchain (Yue et al., 2016; Kuo et al., 2017). Wu et al. (2017) studied how to build transaction trust between participants and improve the visibility of SCM using Blockchain. Li et al. (2017) examined how Blockchain contributes to security issues faced in the field of the Internet of Things (IOT).

In another stream of research, many studies have focused on Bitcoin prices. Some studies have examined the relationships among social media data, search queries, and Bitcoin prices. Kristoufek (2013) found that search queries within Google Trends and Wikipedia on Bitcoin had positive associations with Bitcoin prices. Abraham et al. (2018) also found that the volume of Google Trends and Twitter messages regarding Bitcoin were closely related to the fluctuation of Bitcoin prices. In addition, Ciaian et al. (2016) showed that the numbers of new members, new postings, Bitcoin transactions, and Bitcoin addresses in a Bitcoin community site can be a set of credible indicators that affect the changes in Bitcoin prices.

Other recent studies have employed Bitcoin's Blockchain information to predict Bitcoin prices. Saad et al. (2019) found that Bitcoin's Blockchain information, such as miner revenue, fees, hash rate, and network difficulty, were significantly associated with Bitcoin prices. Jang and Lee (2017) also reported that Bitcoin's Blockchain information and macro-economic factors were key predictors for forecasting Bitcoin prices. Further, they found that an artificial neural network (ANN) machine-learning technique provided the best predictions of Bitcoin prices. In addition, Mallqui and Fernandes (2019) found that Bitcoin's Blockchain information, macro-economic factors, Google popularity index, and Wikipedia searches could influence Bitcoin future prices.

However, very few studies of Ethereum have been conducted. Kim et al. (2016) found that the numbers of user replies and user comments for Ethereum are positively associated with the fluctuations of Ethereum prices. Abraham et al. (2018) showed that Google Trends and Twitter messages regarding Ethereum are significantly related to the fluctuation of Ethereum prices. Moreover, Valencia et al. (2019) found that users' sentiments regarding Ethereum affected Ethereum prices. Table 1 summarizes the key research variables used in previous studies.

Table 1
Prior Studies Related to Cryptocurrency.

Study	Dependent Variable	Key Factors
Kristoufek (2013)	Fluctuation of Bitcoin Prices	- Google Trends about Bitcoin - Wikipedia views for Bitcoin
Ciaian et al. (2016)	Bitcoin Prices	- **Market forces of Bitcoin**: number of Bitcoin transactions, number of Bitcoin addresses - **Bitcoin attractiveness for investors and users**: new posts in Bitcoin community site, new members in Bitcoin community site
Kim et al. (2016)	Fluctuation of Bitcoin, Fluctuation of Ethereum, Fluctuation of Ripple	- User replies - User comments for cryptocurrency
Abraham et al. (2018)	Fluctuation of Bitcoin and Ethereum Prices	- Google Trends on cryptocurrency - Tweet volumes about cryptocurrency
Saad et al. (2019)	Bitcoin Prices	- **Bitcoin's Blockchain information**: miner's revenue, fee, hash rate, network difficulty
Jang and Lee (2017)	Bitcoin Prices, Fluctuation of Bitcoin Prices	- **Bitcoin's Blockchain information**: trading volume, block size, transactions/block, median confirmation time, hash rate, difficulty, cost percentage of transactions, miner's revenue, confirmed transaction, total number of Bitcoin - **Macro-economic factors:** S&P 500, Euro Stoxx 50, DOW30, NASDAQ 100, Crude oil, SSE, Gold, VIX, Nikkei225, FTSE100 - Global currency ratio: GBP/USD, JPY/USD, CHF/USD, CNY/USD, EUR/USD
Mallqui and Fernandes (2019)	Bitcoin Prices, Fluctuation of Bitcoin Prices	- **Bitcoin's Blockchain information**: volume of trades, total transaction fees, cost per transaction, number of transactions, hash rate - **Macro-economic factors**: Crude oil, Gold, S&P 500 NASDAQ 100, DAX index - Google popularity index - Wikipedia search volume
Valencia et al. (2019)	Prices of Bitcoin, Ethereum, Ripple, and Litecoin	- Sentiment of tweets

2.3. Ethereum and the research focus

Ethereum is the second-generation Blockchain system and is designed to contain various types of information in addition to transaction information (Antonopoulos & Wood, 2018). Ethereum has made a significant contribution to the practical application of Blockchain technology thanks to its extensibility. Unlike Bitcoin, Ethereum can contain various types of information, such as identity, health care, and product information as well as transaction information. In addition, the smart contract of Ethereum, which is a function of conducting trade under certain conditions, improves the usage value of Ethereum.

However, studies on Ethereum prices have been scarce. Compared to studies of other cryptocurrencies, there may be additional variables related to Ethereum price. First, Bitcoin does not pay mining compensation for uncle blocks, whereas Ethereum does. The mining compensation for uncle blocks can be seen in terms of the issuance of Ethereum, and it might be related to the price of Ethereum. Second, Ethereum introduces the concept of "gas" to adjust the size of the block and the speed of the Blockchain network. The user needs a memory cost to register information in the Ethereum Blockchain. This memory cost should be paid by gas, which can be purchased with Ethereum. Thus, gas consumption, gas price, and gas limits may be related to Ethereum prices because gas consumes Ethereum. Third, Ethereum is exchanged with other coins in the cryptocurrency market. Considering that transactions can be made between different coins, the Blockchain information of other coins can be associated with the price of Ethereum.

Therefore, for Ethereum price predictions, this study considers Ethereum-specific Blockchain information variables such as gas used, gas limits, gas price, uncle block, Blockchain information of other coins, macro-economic factors, and generic Blockchain information used in previous Bitcoin studies.

3. Methodology

Here, we investigate and compare various research methods and machine-learning algorithms employed in the extant literature on cryptocurrency price predictions.

3.1. Data Description

We gathered data regarding Ethereum from various sources in the period of August 11, 2015 to November 28, 2018. Ethereum prices and Blockchain information were recorded on a daily basis from Etherscan, which provides daily Ethereum-related statistics. Macro-economic development and global currency ratio data were obtained from DataStream of the university database, a professional macro-economic and financial data platform. Other cryptocurrencies (coins) having Blockchain information were chosen based on the following two criteria: (1) the coins had high transaction volumes in the market and (2) the Blockchain information of other coins overlapped temporally with that of Ethereum. Hence, we selected the Blockchain information of Bitcoin, Litecoin, and Dashcoin.

We employed a set of predictor variables commonly used for Bitcoin price prediction in the previous literature (e.g., Jang & Lee, 2017; Poyser, 2019). Furthermore, we gathered gas and uncle block information closely related to the issuance and consumption of Ethereum (Antonopoulos & Wood, 2018). The values of products and currency are significantly related to their supply and demand relationship. A previous study argued that Blockchain information related to supply and demand is closely associated with the price of cryptocurrency (Jang & Lee, 2017). The Ethereum Blockchain system pays Ethereum coins to miners whenever an uncle block is created. In addition, Ethereum's gas system consumes Ethereum to input information into the Ethereum Blockchain. This can be regarded as the demand for Ethereum. Within this study, variables such as gas limits, gas price, gas used, and uncle block variables were selected appropriately as Ethereum-specific Blockchain information to predict Ethereum prices. The research variables used in this study are summarized in Table 2.

In the prediction, Ethereum price is the dependent variable. Macro-economic development indices, global currency ratios, generic Blockchain information (on Ethereum, Bitcoin, Litecoin, and Dashcoin), and Ethereum-specific Blockchain information are used as the predictors. The descriptive statistics of the variables are shown in Tables 3 and 4.

Studies have reported that the macro-economic development indices and global currency ratios are significantly associated with changes in Bitcoin prices. Specifically, Jang and Lee (2017) found that the key macro-economic factors of the S&P 500, DOW30, Eurostoxx, NASDAQ, Crude Oil, SSE, Gold, VIX, Nikkei225, and FTSE100 contributed to the prediction of Bitcoin price. Poyser (2019) showed that British Currency Sterling/US Dollar (GBP/USD), Japanese Yen/US Dollar (JPY/USD), Swiss Franc/US Dollar (CHF/USD), China Yuan Renminbi/US Dollar (CNY/USD), and Euro/US Dollar (EUR/USD) were also associated with Bitcoin price.

Table 4 provides a description of generic Blockchain information (Ethereum, Bitcoin, Litecoin, and Dashcoin) and Ethereum-specific Blockchain information. From the left side of Table 4, we divide generic Blockchain information and Ethereum-specific Blockchain information by category and describe the Blockchain information variables. We report the mean and standard deviation values for the Blockchain information variables in Table 5.

3.2. Research design

We employed Rapid Miner for the main analyses and applied a 10-fold cross-validation to ensure the reliability of the analysis results. To investigate the relationship between the predictor variables and the dependent variable, we conducted a stepwise analysis from Models I-1 to I-6. In addition, we analyzed Models II-1, II-2, and II-3 to better evaluate the impact of other coins' Blockchain information on changes in Ethereum prices. We conducted further analyses to examine the effects of Ethereum-specific Blockchain information in Models II-4, II-5, and II-6.

We used time-series analyses and advanced machine-learning techniques to predict Ethereum prices. To analyze time-series data, studies have applied the traditional regression model, autoregressive integrated moving average (ARIMA) (Kaytez et al., 2015; Pati, et al., 2017; Jang & Lee, 2017). Some studies have employed machine-learning techniques to analyze time-series data, concluding that machine learning performs better than conventional time-series analytical approaches (Wang, 2011; Kaytez et al., 2015; Pati et al., 2017). In addition, machine-learning methods provide better predictions of Blockchain coin prices

Table 2
Description of Key Research Variables.

Category	Research Variables	Reference
Dependent Variable	- Ethereum Price (in USD)	This Study
Macro-economic Development Index	- Standard & Poor's 500 index (S&P 500), - Dow Jones Industrial Average 30 (DOW30), - Stock Index of Eurozone (Euro Stoxx 50), - National Association of Securities Dealers Automated Quotations (NASDAQ), - Crude Oil, - Shanghai Stock Exchange (SSE), - Gold, - Volatility Index of S&P500 (VIX), - Nikkei Stock Average for the Tokyo Stock Exchange (Nikkei225), - Financial Times Stock Exchange 100 Index (FTSE100)	Jang & Lee (2017)
Global Currency Ratio	- British Currency Sterling (GBP)/ US Dollar (USD), - Japanese Yen (JPY)/ US Dollar (USD), - Swiss Franc (CHF)/ US Dollar (USD), - China Yuan Renminbi (CNY)/ US Dollar (USD), - Euro (EUR)/ US Dollar (USD)	Jang & Lee (2017)
Generic Blockchain Information (Ethereum, Bitcoin, Litecoin, Dashcoin)	- Transaction Volume (in USD), - Transaction Count, - Generated Coins, - Transaction Fee, - Active Address, - Block Size, - Block Count, - Difficulty	Jang & Lee (2017), Poyser (2019)
Generic Blockchain Information (Bitcoin, Litecoin, Dashcoin,)	- Price (in USD)	Jang & Lee (2017), Poyser (2019)
Ethereum-Specific Blockchain Information	- Uncle Block, - Gas Limit, - Gas Price, - Gas Used	This Study

Table 3
Descriptive Statistics of Macro-economic Factors and Global Currency Ratios.

Macro-Economy Factors		Global Currency Ratio	
Index	**Mean (S.D.)**	**Currency**	**Mean (S.D.)**
S&P 500	2375.978 (300.838)	GBP/USD	1.352 (0.090)
DOW30	20972.448 (3248.853)	JPY/USD	0.010 (0.024)
Eurostoxx	3299.973 (233.588)	CHF/USD	1.018 (0.021)
NASDAQ	72.294 (11.394)	CNY/USD	0.151 (0.005)
Crude oil	51.898 (10.887)	EUR/USD	1.135 (0.050)
SSE	1417.445 (120.925)	–	–
Gold	1241.827 (73.241)	–	–
VIX	14.851 (4.704)	–	–
Nikkei225	19967.100 (2353.307)	–	–
FTSE100	6979.030 (566.491)	–	–

Table 4
Description of Blockchain Information.

Category	Attribute	Description
Generic Blockchain Information	Price	Coin price per day (in USD)
	Transaction Volume	Total volume of transactions completed per day (in USD)
	Transaction Count	Number of transactions completed per day
	Generated Coins	Amount of coin issued per day
	Transaction Fee	Registration fee required to enter information in block per day
	Active Address	Number of accounts active in Blockchain system
	Block Size	Average memory size of blocks created
	Block Count	Number of blocks created per day
	Difficulty	Average difficulty of block creation
Ethereum-Specific Blockchain Information	Uncle Block	Number of blocks that are not registered as a formal block per day
	Gas Limit	Average amount of gas allowed per block
	Gas Price	Average memory cost required to enter information into the block
	Gas Used	Amount of gas used per day

than those of other analytical methods (Jang & Lee, 2017). Studies have used nonlinear methods, such as an ANN (Zhou et al., 2017; Jang & Lee, 2017) and support-vector machine (SVM) (Xiong et al., 2014; Wang, 2011; Kaytez et al., 2015), for time-series analyses. They found that these methods worked better for various types of time-series data (Zhou et al., 2017; Jang & Lee, 2017).

ANNs have a multilayer perceptron (MLP) structure with multiple hidden layers between the input and output layers. The value entered in the input layer is processed by the hidden layers, and the estimated value is given to the output layer. If the estimated values significantly differ from the actual values, the ANN propagates the error values back to the input layer. Thereafter, the ANN performs a series of calculations by reflecting the errors in the input values. This process is called a back-propagation algorithm, through which the ANN performs a learning process that minimizes the errors between the estimated and actual values. ANN has one or more hidden layers and can handle various nonlinear problems, such as natural language processing, financial analysis, image recognition, and text classification (Murphy, 2012; Laboissiere et al., 2015; Kara et al., 2011). Studies have applied ANN to solve various problems, such as predicting economic indicators and coin prices. Pati et al. (2017) revealed that ANN works better than the ARIMA model with respect to software clone evolution predictions. ANN has also been applied in the study of coin price prediction. In the study of Jang and Lee (2017), an ANN based on Bayesian theory showed the best prediction of Bitcoin prices. They used the Blockchain information of Bitcoin and macro-economic factors to predict Bitcoin prices and found that these variables play important roles in forecasting Bitcoin prices.

SVM is a machine-learning method that can perform classification, regression, and time-series analysis. It is suitable for analyzing linear and nonlinear problems (Kara et al., 2011). Cortes and Vapnik (1995) proposed SVM using the principle of a hyperplane to solve the classification problem. The basic principle of SVM is to create a hyperplane by finding a margin that maximizes the distance between the positive and negative aspects. If the data are not linearly separated, SVM increases the dimensionality of the data by using kernels. In higher dimensions, SVM finds a hyperplane that can separate data. The representative kernels of SVM have a polynomial and radial basis. They can be expressed by Eqs. (1) and (2), in which d is the degree and σ is a gamma parameter. If a polynomial kernel is used, the parameter d, which is expressed in polynomial degrees as in (1), must be defined. In the radial kernel, the standard deviation σ in Eq. (2) must be defined.

$$\text{Polynomial} = (x_i \bullet x_j + 1)^d \tag{1}$$

$$\text{Radial} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{2}$$

Several studies have found that SVM provides improved analysis results for time-series analysis. Wang (2011) revealed that SVM works well for the prediction of market currencies, such as GBP/USD, USD/JPY, AUD/USD, and EUR/USD. Zhou et al. (2017) employed an ANN and SVM to predict the depth-averaged current velocities of underwater

Table 5
Descriptive Statistics of Blockchain Information.

Category	Attribute	Ethereum Mean (S.D.)	Bitcoin Mean (S.D.)	Litecoin Mean (S.D.)	Dashcoin Mean (S.D.)
Generic Blockchain Information	Price	212.996 (276.290)	3607.498 (3964.767)	47.904 (65.290)	177.967 (250.796)
	Transaction Volume	$161.797{*}10^7$ ($312.773{*}10^7$)	$443.010{*}10^7$ ($615.580{*}10^7$)	$32.201{*}10^7$ ($80.878{*}10^7$)	$4.740{*}10^7$ ($11.881{*}10^7$)
	Transaction Count	$290.392{*}10^3$ ($317.211{*}10^3$)	$232.869{*}10^3$ ($58.957{*}10^3$)	$18.454{*}10^3$ ($24.730{*}10^3$)	$8.576{*}10^3$ ($88.922{*}10^3$)
	Generated Coins	25811.197 (5172.817)	2418.076 (891.710)	14749.282 (1882.250)	2331.639 (1172.292)
	Transaction Fee	323.179 (527.904)	126.023 (160.171)	46.865 (57.657)	8.654 (5.405)
	Active Address	$13.493{*}10^4$ ($15.556{*}10^4$)	$61.329{*}10^4$ ($18.038{*}10^4$)	$5.807{*}10^4$ ($7.014{*}10^4$)	$2.681{*}10^4$ ($1.912{*}10^4$)
	Block Size	$9.745{*}10^3$ ($9.799{*}10^3$)	$125194.490{*}10^3$ ($27410.866{*}10^3$)	$10727.559{*}10^3$ ($12148.730{*}10^3$)	$5371.258{*}10^3$ ($17436.452{*}10^3$)
	Block Count	5575.780 (708.120)	151.396 (15.919)	582.940 (43.865)	548.430 (4.140)
	Difficulty	$1.198{*}10^3$ ($1.364{*}10^3$)	$1668317741.607{*}10^3$ ($2218942504.854{*}10^3$)	$2376.546{*}10^3$ ($3625.691{*}10^3$)	$22842.672{*}10^3$ ($33408.680{*}10^3$)
Ethereum-Specific Blockchain Information	Uncle Block	658.187 (399.256)	–	–	–
	Gas Limit	$5.414{*}10^6$ ($2.048{*}10^6$)	–	–	–
	Gas Price	$3.165{*}10^{10}$ ($3.832{*}10^{10}$)	–	–	–
	Gas Used	$1.541{*}10^{10}$ ($1.702{*}10^{10}$)	–	–	–

gliders and found that the SVM worked better than the ANN. Kaytez et al. (2015) reported that SVMs provided better predictions than those of ANNs and traditional regression models for electrical energy consumption. A study into the prediction of Bitcoin prices also showed that SVMs worked better than ANNs (Mallqui & Fernandes, 2019). This study uses ANN and SVM analyses to predict Ethereum prices, drawing from previous studies.

We employed the root mean square error (RMSE) and mean absolute percentage error (MAPE) to evaluate the results. The formulas of the evaluation criteria are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i - \widehat{y}_i\right)^2}{N}} \tag{3}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \tag{4}$$

where N is the number of samples, y_i is the actual value, and $\widehat{y}_i$ is the estimated value.

4. Results

Our analyses show that the ANN works better than the SVM across all models. Among them, Models I-4 and II-4 with the ANN presented the best performance, as shown in Tables 6 and 7 (RMSE = 0.068, MAPE = 0.048). This study conducted a stepwise analysis for Models I-1 to I-6. Model I-1 includes only macro-economic factors (RMSE = 0.131, MAPE = 0.067). Model I-2 adds generic Blockchain information, and we found that RMSE and MAPE were improved (RMSE = 0.086, MAPE = 0.054). Conversely, Model I-3 with Ethereum-specific Blockchain information did not significantly improve the results of the analysis (RMSE = 0.107 MAPE = 0.052). However, we found that RMSE and MAPE were improved in Model I-4, which included Bitcoin's Blockchain information (RMSE = 0.068, MAPE = 0.048). Further, this study confirmed that in Model I-5, adding Litecoin's Blockchain information did not improve the analysis result (RMSE = 0.107, MAPE = 0.053). We also found that Model I-6, with Dashcoin's Blockchain information, did not improve the performance (RMSE = 0.099, MAPE = 0.053). These results reveal that generic Blockchain information includes information that is directly related to Ethereum prices. However, Ethereum-specific Blockchain information and the Blockchain information of Litecoin and Dashcoin did not contribute significantly to the prediction of the price of Ethereum. Therefore, it is possible that unnecessary variables are included in the prediction models, according to the stepwise analysis of Models I-1 to I-6.

We conducted further analyses to better identify the influence of other coins. Models II-1 to II-3 were evaluated by adding the attributes of other coins independently into Model I-2. We found that the performances of Models II-1, II-2, and II-3 were worse than that of Model I-4. These results may have been caused by excluding Ethereum-specific Blockchain information.

To evaluate the influence of Ethereum's special Blockchain information more accurately, we considered Models II-4, II-5, and II-6, which added variables of Ethereum-specific Blockchain information into Model I-3. Model II-4 showed the best performance in the prediction of Ethereum prices (RMSE = 0.068, MAPE = 0.048). In this study, macro-economic factors, generic Blockchain information, Ethereum-specific Blockchain information, and Bitcoin's Blockchain information were identified as the most suitable variables for forecasting Ethereum prices.

We conducted additional analyses to evaluate whether there is a potential confounding effect on the results of the final model. First, this study conducted the same analysis by excluding macro-economic factors from Model II-4. As a result, we confirmed that the results were worse than those of Model II-4 (RMSE = 0.103, MAPE = 0.092). The second analysis was performed by excluding the generic Blockchain information of Ethereum from Model II-4. We also found that the results were worse (RMSE = 0.073, MAPE = 0.051). As a result, we conclude that macro-economic factors and generic Blockchain information of Ethereum contribute to the accurate prediction of Ethereum prices.

5. Discussion

Given that Blockchain and cryptocurrency are important information technologies that can affect various academic domains, information system researchers need to focus more attention on Blockchain. The novelty of this study is the discovery of new important variables related to Ethereum prices. Our findings provide new knowledge for understanding Blockchain and cryptocurrencies.

5.1. *Theoretical implications*

This study bears theoretical implications for related literature. First, this research identified whether the Blockchain information of Ethereum is valid or relevant for predicting Ethereum prices. Previous studies have found that Bitcoin Blockchain information is important for predicting Bitcoin prices (Jang & Lee, 2017; Mallqui & Fernandes, 2019; Saad et al., 2019). However, this perspective has not been considered in Ethereum price prediction studies. We found a significant association between Ethereum price and Blockchain information and suggested that Blockchain information should be employed for predicting future Ethereum prices. Considering the interest and future value of cryptocurrency, the prediction of cryptocurrency prices is expected to be an important and valuable topic for researchers as well as practitioners. The findings of this study not only contribute to achieving accurate Ethereum price predictions, but also provide rigorous evidence of the benefits of applying Blockchain information to these predictions.

Second, we found additional variables of the Ethereum Blockchain related to price prediction, in addition to the Blockchain variables related to Bitcoin price prediction. As mentioned earlier, the Blockchain structure of Ethereum is different than that of Bitcoin; it has a gas limit, gas usage, compensation for uncle blocks, and gas price. Despite these

Table 6
The Results of Data Analysis for Model I.

		Model I-1	Model I-2	Model I-3	Model I-4	Model I-5	Model I-6
Macro-Economy Factors (Number of Inputs: 15)		√	√	√	√	√	√
Generic Blockchain Information(Number of inputs: 8)			√	√	√	√	√
Ethereum-Specific Blockchain information (Number of inputs: 4)				√	√	√	√
Bitcoin's Blockchain Information (Number of inputs: 9)					√	√	√
Litecoin's Blockchain Information (Number of inputs: 9)						√	√
Dashcoin's Blockchain Information (Number of inputs: 9)							√
ANN	RMSE	0.131	0.086	0.107	**0.068**	0.107	0.099
	MAPE	0.067	0.054	0.052	**0.048**	0.053	0.053
SVM	RMSE	2.933	3.645	1.589	1.306	0.604	0.571
	MAPE	0.361	0.403	0.231	0.201	0.139	0.136

Notes: (√: Add to Variables), (Bold: The Best Value)

Table 7
The Results of Data Analysis for Model II.

		Model II-1	Model II-2	Model II-3	Model II-4	Model II-5	Model II-6
Macro-Economy Factors (Number of Inputs: 15)		√	√	√	√	√	√
Generic Blockchain Information (Number of Inputs: 8)		√	√	√	√	√	√
Ethereum-Specific Blockchain Information (Number of Inputs: 4)					√	√	√
Bitcoin's Blockchain Information (Number of Inputs: 9)		√			√		
Litecoin's Blockchain Information (Number of Inputs: 9)			√			√	
Dashcoin's Blockchain Information (Number of Inputs: 9)				√			√
ANN	RMSE	0.102	0.097	0.135	**0.068**	0.082	0.071
	MAPE	0.052	0.053	0.055	**0.048**	0.050	0.049
SVM	RMSE	3.910	1.184	2.921	1.306	1.193	0.812
	MAPE	0.427	0.195	0.337	0.201	0.198	0.163

Notes: (√: Add to Variables), (Bold: The Best Value)

differences, studies related to cryptocurrency have not considered the specific Blockchain information of Ethereum for Ethereum price prediction (Jang & Lee, 2017; Mallqui & Fernandes, 2019; Saad et al., 2019). The unique Blockchain information of Ethereum is directly related to the issuance and consumption of coins. Therefore, we used Ethereum-specific Blockchain information to predict Ethereum prices and found additional Blockchain variables related to Ethereum prices. We suggest that future studies of Ethereum prices should employ Ethereum-specific Blockchain information. The structures of the Blockchains of various coins may differ slightly. In other words, there may be special variables associated with each Blockchain that are useful for predicting coin prices. Drawing from the findings of this study, the roles of these additional variables must be determined, considering Blockchain structural differences, to accurately predict coin prices. In particular, if the additional Blockchain information generated by the system structure is directly related to the supply and demand of cryptocurrency, the researcher should consider it more important in the prediction of cryptocurrency prices. The findings of our study offer supportive empirical evidence for this argument.

Third, in addition to the Ethereum Blockchain information, we found that the Blockchain information of Bitcoin contributes significantly toward predicting Ethereum prices. Since Ethereum is actively being traded with other coins (e.g., Bitcoin, Litecoin, and Dashcoin) in the cryptocurrency market, traders are likely to refer to the information of other coins when buying and selling Ethereum. Considering that Blockchain information and cryptocurrency price are related (Jang & Lee, 2017; Mallqui & Fernandes, 2019), there is also the possibility that the Blockchain information of other coins is related to Ethereum prices. That is, it is necessary to investigate whether the Blockchain information of other coins has an impact on future Ethereum prices. Although the Blockchain information of other coins may play a pivotal role in predicting Ethereum prices, research from this perspective has not yet been conducted. In this study, we substantiated the relationship between the Blockchain variables of other coins and Ethereum prices. In particular, Bitcoin's Blockchain information is more related to Ethereum prices than the Blockchain information of Litecoin and Dashcoin. Therefore, we recommend considering the Blockchain information of Bitcoin for the prediction of prices. Finally, our study predicted Ethereum prices with an accuracy of 95.2% when assessed with MAPE (see Model II-4 in Table 6; MAPE = 0.048). This study employed popular and well-accepted machine-learning techniques (i.e., ANN and SVM) and used publicly available Blockchain-related datasets. For these reasons, researchers and practitioners who aim to predict Ethereum prices in the future can easily replicate the methods and results of this study. This study provides the academic implications associated with the discovery of meaningful variables for predicting Ethereum prices and broadened theoretical perspectives.

5.2. Practical implications

The practical implications of this study are as follows. First, we identified and used a set of key variables to predict Ethereum prices. The Blockchain information of Ethereum and other coins can be identified and collected online in real time. Practitioners can carry out proper investment in and sales of Ethereum using the Blockchain information and ANN machine-learning techniques employed in this research.

Second, we demonstrated that macro-economic factors and the Blockchain information of Ethereum and Bitcoin are more important than the Blockchain information of Litecoin and Dashcoin for predicting Ethereum prices. Based on this finding, practitioners can predict Ethereum prices rapidly and with high accuracy. Using only the key variables related to Ethereum prices can reduce unnecessary computing power and time. Practitioners should use macro-economic factors and the Blockchain information of Ethereum and Bitcoin actively to accurate forecast Ethereum prices.

Third, the findings of this study suggest that one should consider the Blockchain information when adjusting Ethereum prices and issuing other coins. The Blockchain information related to the price may help to design the structure of the Blockchain. In addition, companies that attempt to build Blockchain platforms based on Ethereum may prepare for extreme fluctuations in coin prices by referring to the results of this study.

5.3. Limitations and future directions

The limitations and future direction of this study are as follows. First, we attempted to predict Ethereum prices using only an ANN and SVM machine-learning method. Studies have used ANN and SVM for time-series analysis and reported that these algorithms provide better results than those of other analytical methods. However, ensemble techniques or other analytical algorithms can be used to predict Ethereum prices. Considering this, future research should introduce various analysis algorithms, including ANN and SVM.

Second, this study only considered macro-economic factors and Blockchain information for predicting Ethereum prices. However, studies have reported that user opinions, sentiment, Google Trends, and information from cryptocurrency community sites are also related to cryptocurrency prices (Kristoufek, 2013; Ciaian et al., 2016; Kim et al., 2016). The variables extensively used in studies (e.g., user opinion, sentiment, Google Trends, and so on) can also be related to the price of Ethereum. These variables should be considered when predicting Ethereum prices. In future studies, social media data should be used for predicting cryptocurrency prices, along with the variables predicted in this study.

Third, this study does not aim to predict the volatility of future Ethereum prices. Although the study identified a set of influential factors related to the changes in Ethereum prices and future Ethereum prices, this may be insufficient for traders who actually buy and sell Ethereum. In practice, however, we believe that predicting Ethereum's volatility can provide important information for companies using Ethereum. Therefore, further research regarding the prediction of the volatility of Ethereum prices is necessary.

6. Conclusion

Ethereum, like Bitcoin, is actively traded in the cryptocurrency market, and its value is likely to attract more attention in the future. We started this research with the salient question "What variables can be related to Ethereum prices?" and identified them variables. Previous studies have not identified the benefits that can be obtained from the special variables of Blockchain information of other coins for predicting Ethereum prices because they have mostly focused on Bitcoin. The key findings of this study addressed these research questions. First, macroeconomic factors significantly improve the performance of predicting Ethereum prices. Second, the factors in Ethereum-specific Blockchain information (i.e., the uncle block, gas price, gas consumption, and gas limit) are significantly associated with Ethereum prices. In particular, Ethereum-specific Blockchain information contributed to the best performance when it was considered within the model along with macroeconomic factors, the generic Blockchain information of Ethereum, and Bitcoin's Blockchain information. Third, among the Blockchain information of other coins (i.e., Bitcoin, Litecoin, and Dashcoin), Bitcoin's Blockchain information is significantly related to Ethereum prices. Our findings not only provide a theoretical base for future cryptocurrency researchers to discover additional variables but also broaden the current view of Bitcoin research. Cryptocurrency can contribute to effectively implementing the Blockchain system in areas where Blockchain is needed. We hope that the findings of this study will contribute to the expansion of knowledge in the field of cryptocurrency research.

CRediT authorship contribution statement

Han-Min Kim: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Investigation. **Gee-Woo Bock:** Visualization, Supervision. **Gunwoong Lee:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abraham, J., Higdon, D., Nelson, J., & Ibarra, J. (2018). Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Science Review, 1*(3), 1–23.

Antonopoulos, A. M. (2014). Mastering Bitcoin: unlocking digital cryptocurrencies, O'Reilly Media, Inc.

Antonopoulos, A. M., Wood, G. (2018). Mastering Ethereum: building smart contracts and dapps, O'Reilly Media.

Ciaian, P., Rajcaniova, M., & Kancs, D. A. (2016). The economics of BitCoin price formation. *Applied Economics, 48*(19), 1799–1815. https://doi.org/10.1080/00036846.2015.1109038

CoinMarketCap. (2021). Today's Cryptocurrency Prices by Market Cap. Retrieved from https://coinmarketcap.com/.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

Ferguson, M. (2018). Preparing for a Blockchain Future. *MIT Sloan Management Review*.

Guo, Y., & Liang, C. (2016). Blockchain application and outlook in the banking industry. *Financial Innovation, 2*(1), 24. https://doi.org/10.1186/s40854-016-0034-9

Jang, H., & Lee, J. (2017). An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. *IEEE Access, 6*, 5427–5437. https://ieeexplore.ieee.org/document/8125674.

Kim, Y. B., Kim, J. G., Kim, W., Im, J. H., Kim, T. H., Kang, S. J., … Zhou, W.-X. (2016). Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PloS one, 11*(8), e0161197. https://doi.org/10.1371/journal.pone.0161197

Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert systems with Applications, 38*(5), 5311–5319. https://doi.org/10.1016/j.eswa.2010.10.027

Kaytez, F., Taplamacioglu, M. C., Cam, E., & Hardalac, F. (2015). Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *International Journal of Electrical Power & Energy Systems, 67*, 431–438. https://doi.org/10.1016/j.ijepes.2014.12.036

Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports, 3*, 1–7.

Kuo, T. T., Kim, H. E., & Ohno-Machado, L. (2017). Blockchain distributed ledger technologies for biomedical and health care applications. *Journal of the American Medical Informatics Association, 24*(6), 1211–1220. https://doi.org/10.1093/jamia/ocx068

Laboissiere, L. A., Fernandes, R. A., & Lage, G. G. (2015). Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. *Applied Soft Computing, 35*, 66–74. https://doi.org/10.1016/j.asoc.2015.06.005

Li, Z., Kang, J., Yu, R., Ye, D., Deng, Q., & Zhang, Y. (2017). Consortium blockchain for secure energy trading in industrial internet of things. *IEEE Transactions on Industrial Informatics, 14*(8), 3690–3700. https://ieeexplore.ieee.org/document/8234700/.

Mallqui, D. C., & Fernandes, R. A. (2019). Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques. *Applied Soft Computing, 75*, 596–606. https://doi.org/10.1016/j.asoc.2018.11.038

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press.

Pati, J., Kumar, B., Manjhi, D., & Shukla, K. K. (2017). A Comparison Among ARIMA, BP-NN, and MOGA-NN for Software Clone Evolution Prediction. *IEEE Access, 5*, 11841–11851. https://ieeexplore.ieee.org/abstract/document/7938609.

Poyser, O. (2019). Exploring the dynamics of Bitcoin's price: A Bayesian structural time series approach. *Eurasian Economic Review, 9*(1), 29–60. https://link.springer.com/article/10.1007/s40822-018-0108-2.

Saad, M., Choi, J., Nyang, D., Kim, J., & Mohaisen, A. (2019). Toward Characterizing Blockchain-Based Cryptocurrencies for Highly Accurate Predictions. *IEEE Systems Journal.*, 1–12. https://ieeexplore.ieee.org/document/8840919.

Songmuang, K. (2018). The forecasting of cryptocurrency price by correlation and regression analysis. *Kasem Bundit Journal, 19*, 287–296. https://www.tci-thaijo.org/index.php/jkbu/article/view/130214.

Valencia, F., Gómez-Espinosa, A., & Valdés-Aguirre, B. (2019). Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. *Entropy, 21*(6), 589. https://doi.org/10.3390/e21060589

Wang, P. (2011). Pricing currency options with support vector regression and stochastic volatility model with jumps. *Expert Systems with Applications, 38*(1), 1–7. https://doi.org/10.1016/j.eswa.2010.05.037

Wu, H., Li, Z., King, B., Ben Miled, Z., Wassick, J., & Tazelaar, J. (2017). A distributed ledger for supply chain physical distribution visibility. *Information, 8*(4), 1–18. https://doi.org/10.3390/info8040137

Xiong, T., Bao, Y., & Hu, Z. (2014). Multiple-output support vector regression with a firefly algorithm for interval-valued stock price index forecasting. *Knowledge-Based Systems, 55*, 87–100. https://doi.org/10.1016/j.knosys.2013.10.012

Yli-Huumo, J., Ko, D., Choi, S., Park, S., Smolander, K., & Song, H. (2016). Where is current research on blockchain technology?—a systematic review. *PloS one, 11*(10), e0163477. https://doi.org/10.1371/journal.pone.0163477

Yue, X., Wang, H., Jin, D., Li, M., & Jiang, W. (2016). Healthcare data gateways: Found healthcare intelligence on blockchain with novel privacy risk control. *Journal of Medical Systems, 40*(10), 218. https://doi.org/10.1007/s10916-016-0574-6

Zhou, Y., Yu, J., & Wang, X. (2017). Time series prediction methods for depth-averaged current velocities of underwater gliders. *IEEE Access, 5*, 5773–5784. https://doi.org/10.1109/ACCESS.2017.2689037