

Article

A New Approach to Predicting Cryptocurrency Returns Based on the Gold Prices with Support Vector Machines during the COVID-19 Pandemic Using Sensor-Related Data

Esam Mahdi ^{1,*} , Víctor Leiva ² , Saed Mara'Beh ¹ , Carlos Martin-Barreiro ^{3,4} ¹ Department of Mathematics, Statistics and Physics, Qatar University, Doha 2713, Qatar; smarabeh@qu.edu.qa² School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362807, Chile; victor.leiva@pucv.cl³ Faculty of Natural Sciences and Mathematics, Universidad Politécnica ESPOL, Guayaquil 090902, Ecuador; cmmartin@espol.edu.ec⁴ Faculty of Engineering, Universidad Espíritu Santo, Samborondón 0901952, Ecuador

* Correspondence: emahdi@qu.edu.qa

Abstract: In a real-world situation produced under COVID-19 scenarios, predicting cryptocurrency returns accurately can be challenging. Such a prediction may be helpful to the daily economic and financial market. Unlike forecasting the cryptocurrency returns, we propose a new approach to predict whether the return classification would be in the first, second, third quartile, or any quantile of the gold price the next day. In this paper, we employ the support vector machine (SVM) algorithm for exploring the predictability of financial returns for the six major digital currencies selected from the list of top ten cryptocurrencies based on data collected through sensors. These currencies are Binance Coin, Bitcoin, Cardano, Dogecoin, Ethereum, and Ripple. Our study considers the pre-COVID-19 and ongoing COVID-19 periods. An algorithm that allows updated data analysis, based on the use of a sensor in the database, is also proposed. The results show strong evidence that the SVM is a robust technique for devising profitable trading strategies and can provide accurate results before and during the current pandemic. Our findings may be helpful for different stakeholders in understanding the cryptocurrency dynamics and in making better investment decisions, especially under adverse conditions and during times of uncertain environments such as in the COVID-19 pandemic.

Keywords: artificial intelligence; data science; digital currency; gold; SARS-CoV-2; machine learning; sensing and data extraction



Citation: Mahdi, E.; Leiva, V.; Mara'Beh, S.; Martin-Barreiro, C. A New Approach to Predicting Cryptocurrency Returns Based on the Gold Prices with Support Vector Machines during the COVID-19 Pandemic Using Sensor-Related Data. *Sensors* **2021**, *21*, 6319. <https://doi.org/10.3390/s21186319>

Academic Editor: Wai Lok Woo

Received: 30 August 2021

Accepted: 17 September 2021

Published: 21 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the last two years, the impact of the 2019 Coronavirus Disease (COVID-19) on the trading strategies between different types of cryptocurrencies, gold, exchange market, portfolio diversification, and macroeconomic policy has attracted a great deal of interest in the financial market [1]. The unexpected recent exponential increase in cryptocurrency prices has exceeded the investors' imaginations. However, this does not indicate that cryptocurrencies are a substitute for gold [2]. For more than 2000 years, gold has been known as a high-quality liquid asset that has served as a hedge and safe haven for exchanging commodities. Several robust statistical models have been proposed to explore the predictability of the gold price [2,3].

Cryptocurrencies are digital currencies that belong to a decentralized system still classified in the nascent asset class [4]. Their prices are very sensitive to good and bad news with significant instability leading to high volatility and making the predictions hard and challenging, especially under adverse market conditions [2,4]. Thus, investments in cryptocurrencies are, often, equated to investments in gold, where cryptocurrencies may benefit from higher allocations to gold and vice-versa [2].

The emergence of blockchain [5] and decentralized cryptocurrencies has been a catalyzed innovation in the financial market. Before the global outbreak of COVID-19, research focused on using the Granger-causality and quantile-on-quantile regression (QQR) [6,7] to model the relationship between price change of crude oil, gold, denominated currencies, and other financial markets, on the one hand, and price change of top-ranked cryptocurrencies, on the other hand [8–12].

In the wake of the COVID-19 pandemic, several studies were conducted to explore the influence of the pandemic intensity on the gold and cryptocurrency market. For example, the asymmetric impacts of COVID-19 on the multifractality of oil and gold prices (and vice-versa) based on upward and downward trends were examined in [13]. They showed that the gold and oil markets during the pre-COVID-19 period were more inefficient during downward (upward) trends, but this has changed during the outbreak period [14].

A type of multivariate GARCH model, named VAR-AGARCH, was used in [15] to examine the return and volatility spillover of Bitcoin (BTC), Ethereum (ETH), and Litecoin (LTC) cryptocurrencies before and during the COVID-19 period. The authors [15] did not compare the performance of the VAR-AGARCH model with modern techniques as machine learning algorithms [16]. The QQR was employed in [17] to explore the asymmetric impact between the changing intensity in the quantiles of COVID-19 and the quantiles of the daily returns of the top-ranked cryptocurrencies. Their numerical results are based on data collected over the first six months of the outbreak of the pandemic. By utilizing these results, contrary to what was indicated in [18], the QQR suggested that BTC, Cardano (ADA), and ETH performed better in comparison with other currencies.

The GARCH-MIDAS model adopted in [4] examined the impact of news, during COVID-19, on the predictability of return volatility of cryptocurrencies. They showed that the pandemic increases the volatility of digital currencies compared with the period before the outbreak of COVID-19. A significant causal relationship was shown in [19], in quantiles, between the US dollar and the top five cryptocurrencies during the pandemic, which had a large effect on the digital currencies' relationship.

The impact of the COVID-19 lockdowns on the BTC trading volume was analyzed in [20] by using Apple mobility volume trends and autoregressive-moving-average models [21]. The authors [20] reported that the investors traded more BTCs during the days with low mobility associated with lockdown mandates, even after controlling the stocks and gold prices.

In many applications, the support vector machine (SVM) technique [22], a machine learning tool, has shown a superior forecasting performance in linear and non-linear models, especially when the data have outliers, compared with most evolutionary algorithms [23]. SVM can be classified into two algorithms: support vector classification (SVC) and support vector regression (SVR) [24]. Both algorithms are computationally intensive and require a large amount of memory, but with the age of big data [25] and quantum computing, this can be affordable. Interest in these algorithms has increased in the cryptocurrency market and it deserves to continue. The discriminant analysis (DA) and SVM algorithms were utilized in [26] to model the subsurface gold mineralization in the northwestern part of Iran. They compared the performance of four classification functions of DA and SVM methods and showed that the SVM algorithm worked better than the DA method.

SVM algorithms were applied in [27] to investigate the profitability of the cryptocurrency market and explore how high they would be. The authors showed that the SVM technique could outperform the benchmark strategies in terms of return-risk relation. They showed that the SVR algorithm provides better results in comparison to ARIMA models. An SVM algorithm based on particle swarm optimization (SVM-PSO) algorithm [28] was proposed in [29] to forecast the future price of cryptocurrency. They reported that the performance of the SVM-PSO technique is better than the SVM algorithm. The ARIMA and SVR models were employed in [3] to predict the daily price of gold, whereas in [30],

the SVC model was utilized to predict whether the price of three cryptocurrencies (BTC, ETH, and LTC) would increase or decrease in the next day.

Note that traditional econometric and time-series models, such as the ARMA and GARCH models, might not provide helpful results in predicting cryptocurrency returns due to the impact of risk and uncertainty on return volatility. Although several SVM techniques have already been used in numerous applications, very few have been done in the cryptocurrency market. To the best of our knowledge, the SVM has never been employed to predict the daily classification of cryptocurrency returns according to the quantiles of the gold price the next day.

Our objective is to propose a new approach for predicting cryptocurrency returns by using the SVC technique. Specifically, instead of forecasting the movement of the cryptocurrency return indicating whether it would go up or down, we forecast the type of this return reporting whether it would be in the first, second, third quartile, or any quantile of the gold price the next day. Therefore, our investigation fills this gap by proposing a new methodology for predicting the daily classification of cryptocurrency returns according to the quantiles of the gold price the next day. We classify the returns of the six major digital currencies (ADA, Binance Coin –BNB–, BTC, Dogecoin –DOGE–, ETH, and Ripple –XRP–) selected from the list of top ten cryptocurrencies into four categories according to the quartiles of the gold price values. In the current COVID-19 contingency scenario, data of cryptocurrency returns and COVID-19 are often collected through sensors [31–33].

The rest of this article is organized as follows. In Section 2, we propose an algorithm that uses a sensor in the database to predict the return of cryptocurrencies considering COVID-19 data and the price of gold. Section 3 defines the methodology based on the SVM models employed in this study. In Section 4, we present the data and results of our investigation, whereas in Section 5 our conclusions are reported.

2. Sensing Data of Cryptocurrencies, Gold, and COVID-19

To perform the empirical analysis proposed in this work (see Section 4), a software has been implemented with a local database that uses the SVM algorithm considering two aspects. In addition, we design an algorithm that allows us to sense changes in the data and show updated results. The database management system permits the configuration of a digital sensor. This sensor can detect if new data are entered into the database, if the data are modified, or if data are deleted.

In this work, a sensor that detects the presence of new data is utilized. As a response to new data, a trigger is implemented, which enables instructions to be registered and executed when the new data arrive. This sensor permits the software that performs the calculations to be notified that new records were entered in the database and therefore that the computations must be updated.

In summary, the local database is updated daily with external data (websites). A sensor in the local database sends a signal to the developed software to perform the calculations considering the new data. The software receives the signal and executes the corresponding processing. Lastly, the software delivers the results.

Next, we describe Algorithm 1 that senses changes in the data and shows updated results. For the sake of better understanding, we relate some steps of this algorithm to our case study in Section 4.

Algorithm 1 Approach to predict the daily classification of cryptocurrencies returns relating them to gold price and COVID-19 data using sensors.

- 1: Update table of deaths and infected cases due to COVID-19 with the data extracted from the website <https://github.com/CSSEGISandData/COVID-19> (accessed on 29 August 2021) and use this table for constructing the data warehouse.
 - 2: Update table of gold prices with the data extracted from the website <https://www.gold.org/goldhub/data/gold-prices> (accessed on 29 August 2021) and employ this table for building the data warehouse.
 - 3: Update table of cryptocurrency returns taking the daily closing price and the market capitalization of the six studied digital currencies with the data extracted from the website <https://coinmarketcap.com> (accessed on 29 August 2021) and utilize this table for creating the data warehouse.
 - 4: Send a warning to the software that performs a data analysis when new records are detected by means of the trigger. Note that the database has a digital sensor with the ability to detect when new data are entered.
 - 5: Communicate the trigger (in the database) with the software through a middleware so that the data analysis may be performed by executing the following steps:
 - 5.1 Plot Figures 2–4, and construct Table 1.
 - 5.2 Use the probability density function to state the type of skewness and kurtosis of the data empirical distribution and describe the shape of their distribution.
 - 5.3 Employ the Jarque-Bera (JB) test and elaborate Table 2.
 - 5.4 Execute the SVM algorithm considering the one-to-one method (see Section 3), that is, separate the observations into training and testing data sets, and select the best model based on the indicators defined in (11).
 - 5.5 Apply the SVM algorithm considering the one-to-rest method (see Section 3) and build a dynamic model based on the available data.
 - 5.6 Construct Figures 5–7.
 - 6: End the algorithm reporting the results of the updated data analysis in the graphical user interface of the software.
-

3. Methodology

SVM with binary classification is a powerful learning algorithm that aims to differentiate between two classes by finding an optimal hyperplane that maximizes the separation between the two-dimensional space points [23].

An algorithm that maximizes the margin between the training patterns and the decision boundary is a helpful tool that can be applied to a wide variety of classifiers. In this tool, the number of parameters must be fit automatically to match the complexity of the problem, with the solution being formulated as a linear combination of support patterns, which are a subset of training patterns that are closest to the decision boundary.

Next, we provide a background of the method used in the present investigation. For detailed mathematical derivations of the results presented in this section, we refer the readers to [34].

For m -dimensional space points, with $2 < m \in \mathbb{Z}$, we cannot utilize the binary SVM to obtain the separation hyperplane (decision surface) directly. To overcome this issue, the multi-classification problem breaks down into multiple binary classification problems. Then, the binary SVM algorithm is applied to these binary classes. This may be done in a one-to-one or one-to-rest manner explained as follows:

- One-to-one method: Here, we break down the m classes into $m(m - 1)/2$ mutual binary classes, where a binary SVM is employed to differentiate between every two

binary classes. Each binary SVM finds a hyperplane that separates between every two classes, neglecting the data points of the other classes. For example, consider $m = 4$, say, with classes 1, 2, 3, and 4. In this case, we have six different SVMs applied to the binary classes $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$, $\{2, 4\}$, and $\{3, 4\}$.

- One-to-rest method: Here, the classifier can use m binary SVMs by obtaining an optimal hyperplane that separates between a class and all others at once. For example, consider $m = 3$, say, with classes 1, 2, and 3. In this case, we have three different SVMs applied to the binary classes $\{1, \text{not } 1\}$, $\{2, \text{not } 2\}$, and $\{3, \text{not } 3\}$.

The data points classification depends on the type of the kernel function to be employed, which aims to reduce the computational time and effort by transforming a non-linear decision surface to a linear equation in a higher number of dimension spaces. For any two vectors \mathbf{u} and \mathbf{v} , the most popular kernels, utilized with SVM algorithms, are linear, polynomial, radial basis, and sigmoid functions defined as:

- Linear function:

$$K(\mathbf{u}, \mathbf{v}) = \mathbf{u}^\top \mathbf{v}, \quad (1)$$

where K is the kernel function and $^\top$ denotes the transpose of a vector.

- Polynomial function:

$$K(\mathbf{u}, \mathbf{v}) = (\gamma \mathbf{u}^\top \mathbf{v} + 1)^d, \quad (2)$$

where $\gamma > 0$ and $d > 0$ are the parameters of scale and degree, respectively.

- Radial basis function:

$$K(\mathbf{u}, \mathbf{v}) = \exp(-\gamma \|\mathbf{u} - \mathbf{v}\|^2), \quad (3)$$

where $\gamma > 0$ is the inverse influence radius of points selected as support vectors.

- Sigmoid function:

$$K(\mathbf{u}, \mathbf{v}) = \tanh(\gamma \mathbf{u}^\top \mathbf{v} + 1), \quad (4)$$

where $\gamma > 0$ is a tuning meta-parameter that defines how far the influence of a training point can reach.

Let $(x_1, y_1), \dots, (x_n, y_n)$ be the training data set, where each pair belongs to one of two classes that we name A or B, according to

$$y_k = \begin{cases} 1, & \text{if } x_k \in \text{class A;} \\ -1, & \text{if } x_k \in \text{class B.} \end{cases}$$

Each x_k , with $k = 1, \dots, n$, is a p -dimensional vector of real numbers. From the training data set, the maximum margin hyperplane algorithm is used to estimate the parameters of the decision function $D(\mathbf{x})$. Hence, the classification of unknown patterns is predicted based on the following criteria: (i) $\mathbf{x} \in \text{class A}$ if $D(\mathbf{x}) > 0$; or (ii) $\mathbf{x} \in \text{class B}$, otherwise. In the direct space, $D(\mathbf{x})$ has the form $D(\mathbf{x}) = \mathbf{w}^\top \boldsymbol{\varphi}(\mathbf{x}) + \mathbf{b}$, where $\boldsymbol{\varphi}(\mathbf{x})$ is a predefined function, which separates the hyperplane in n -dimensional spaces, whereas $\mathbf{w} = (w_1, \dots, w_n)^\top$ and \mathbf{b} are vectors of weights and noise coefficients, respectively, with both \mathbf{w} and \mathbf{b} being adjustable parameters computed along with minimizing the empirical risk. The distance between the pattern \mathbf{x} and the separated hyperplane is $D(\mathbf{x}) / \|\mathbf{w}\|$ (see Figure 1). Note that the axes in this figure (left) are labeled as x_1 and x_2 . Nonetheless, the figure on the right-hand side is the same as that on the left-hand side but as an image (photography).

The maximum margin hyperplane algorithm aims to find the vector \mathbf{w} that maximizes the margin M between the training set and class boundary, which is formulated as

$$\begin{aligned} M^* &= \max_{\mathbf{w}, \|\mathbf{w}\|=1} M \\ \text{subject to } y_k D(\mathbf{x}) &\geq M, \quad k = 1, \dots, p, \end{aligned} \quad (5)$$

where the bound M^* for supporting patterns attains to satisfy the condition $M^* = \min_k y_k D(x_k)$. The decision function, $D(x)$ namely, with maximum margin M^* , is shown in Figure 1. Therefore, the hyperplane with maximum margin can be obtained by solving the minimax problem established as $\max_{w, \|w\|=1} \min_{k,b} y_k D(x_k)$. Usually, the product of the norm of w and the margin M is scaled to be fixed, that is, $M\|w\| = 1$, to avoid the problem of getting an infinite number of possible solutions. Therefore, maximizing M as in (5) is equivalent to minimizing $\|w\|$. Thus, finding the hyperplane with maximum margin ($M^* = 1/\|w^*\|$) reduces to solve the quadratic form given by

$$\min_{w,b} \|w\|^2, \tag{6}$$

subject to $y_k D(x_k) \geq 1, \quad k = 1, \dots, p.$

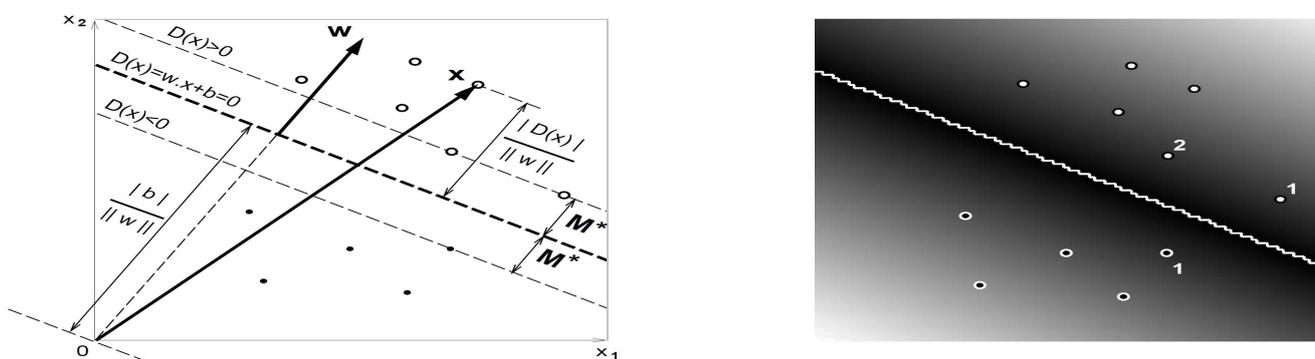


Figure 1. Plot of the maximum margin linear decision function $D(x) = w^T x + b$ ($\varphi(x) = x$) (left) and an image of it (right). The gray levels encode $|D(x)|$, solid black corresponds to $D(x) = 0$, and the numbers indicate the supporting vectors (Source: [34]).

Often, we may tolerate some cost errors $\xi \geq 0$ by allowing for some observations to violate the margin. The cost C is defined by the number of observations beyond the margin. In this case, instead of minimizing the form stated in (6), we need to minimize

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{k=1}^p \xi_k, \tag{7}$$

subject to $y_k D(x_k) \geq 1 - \xi_k, \quad k = 1, \dots, p,$

where $\xi_k \geq 0$ and $\sum_{k=1}^p \xi_k \leq C$. However, although the problem stated in (7) can be solved by numerical methods in direct space, these methods are impractical for high dimensionality of the dual space φ . In this dual space, the decision function is defined as

$$D(x) = \sum_{k=1}^p \alpha_k K(x_k, x) + b, \tag{8}$$

where $0 \leq \alpha_k \leq c$ are adjustable parameters computed along with minimizing the empirical risk, C is the cost given by the number of observations outside the margin, and K is a predefined kernel function. The values of the kernel function stated in (8) may be calculated by the inner product of $\varphi(x_i)$ and $\varphi(x_j)$, that is, $K(x_i, x_j) = \varphi(x_i) \circ \varphi(x_j)$, where the operator \circ denotes the inner product of two vectors. We aforementioned the most popular kernels in the expressions defined in (1) to (4). The problem stated in (7) can be transformed into a dual space using a Lagrangian algorithm [31] by means of the representation formulated as

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{k=1}^p \alpha_k (y_k D(x_k) - 1), \tag{9}$$

under the condition that the Lagrangian multipliers $\alpha_k \geq 1$, for $k = 1, \dots, p$ [35].

Note that the tolerance errors are again introduced, defining ζ as the value outside the error margins and ζ^* to limit the value to the dependent variable. Thus, the Lagrangian dual of the problem stated in (9) may be solved by minimizing the expression given by

$$\frac{1}{2} \|\mathbf{w}\|^2 + c \sum_{k=1}^p (\zeta_k - \zeta_k^*), \quad (10)$$

under the conditions $(\mathbf{w}^\top \mathbf{x}_k + \mathbf{b}) - y_k \leq 1 + \zeta_k$ and $y_k - (\mathbf{w}^\top \mathbf{x}_k + \mathbf{b}) \leq 1 + \zeta_k^*$.

4. Empirical Analysis with Real-World Data

To conduct our study with real-world data, we employ the SVC algorithm proposed in [23]. We use the R software with the `e1071` [36] and `ggplot2` [37] packages to carry out the data analysis [38]. As mentioned, we perform our analysis using three type of data: (i) daily confirmed COVID-19 infections and deaths; (ii) daily gold price—in US dollars—; and (iii) daily closing price of major cryptocurrencies—in US dollar—.

- Daily confirmed COVID-19 infections and deaths:
First, we collect the daily confirmed COVID-19 infections and deaths from the online repository <https://github.com/CSSEGISandData/COVID-19> (accessed on 29 August 2021). This web-hosting has been operated by the Johns Hopkins University Center for Systems Science and Engineering from 23 January 2020 until the present, and it is updated twice daily. We consider the period from 23 January 2020 to 14 July 2021. This period essentially accounts for all available data officially published [39] up to the date where we started this research.

- Daily gold price in US dollars:
Second, to compare the impact of the pre-COVID-19 period with the COVID-19 period on the classification of cryptocurrency returns according to the quantile values of the gold price (in US dollars), we collected the gold price data from the website <https://www.gold.org/goldhub/data/gold-prices> (accessed on 29 August 2021) over the period 04 December 2018 to 14 July 2021. In this article, the word return is used synonymously with log returns.

The four graphs in the upper and middle of Figure 2 present the distribution and trend of the daily infections and deaths due to COVID-19. The graphs show that, for more than 100 days, the number of daily deaths ranged from 5000 to 7500 cases. The plots in Figure 2 show that the daily confirmed COVID-19 infections and deaths and the daily gold prices followed a similar trend during the period of the ongoing COVID-19 pandemic. The gold prices report high volatility clustering in the pandemic compared with the pre-COVID-19 period. For the entire period of study, the probability density function of the gold prices is slightly skewed to the left with a high excess kurtosis.

Remark. Note that the shaded region in Figures 2–4 describes the period after the beginning of the pandemic for better visualization of the impact of COVID-19 on the gold price/cryptocurrency return. Also, in these figures and in Figures 5–7, the vertical dotted line represents the threshold between the pre-COVID-19 period and ongoing COVID-19 period.

- Daily closing price of major cryptocurrencies in US dollars:
Third, we collect the daily closing price and market capitalization for some major cryptocurrencies available from the online source <https://coinmarketcap.com> (accessed on 29 August 2021). For the current research, as mentioned, we study the six cryptocurrency assets, that is, ADA, BNB, BTC, DOGE, ETH, and XRP, selected from the top 10 ranked currencies.

Table 1 provides the main characteristics of the top-ranked cryptocurrencies and Figure 3 presents the market capitalization of used currencies over the study period. Again, the vertical dotted lines represent the threshold between the two periods of time (before and during the pandemic). The trends show an exponential increase in

the market capitalization of BTC and ETH and less increase in the other currencies during the period of the outbreak of the pandemic.

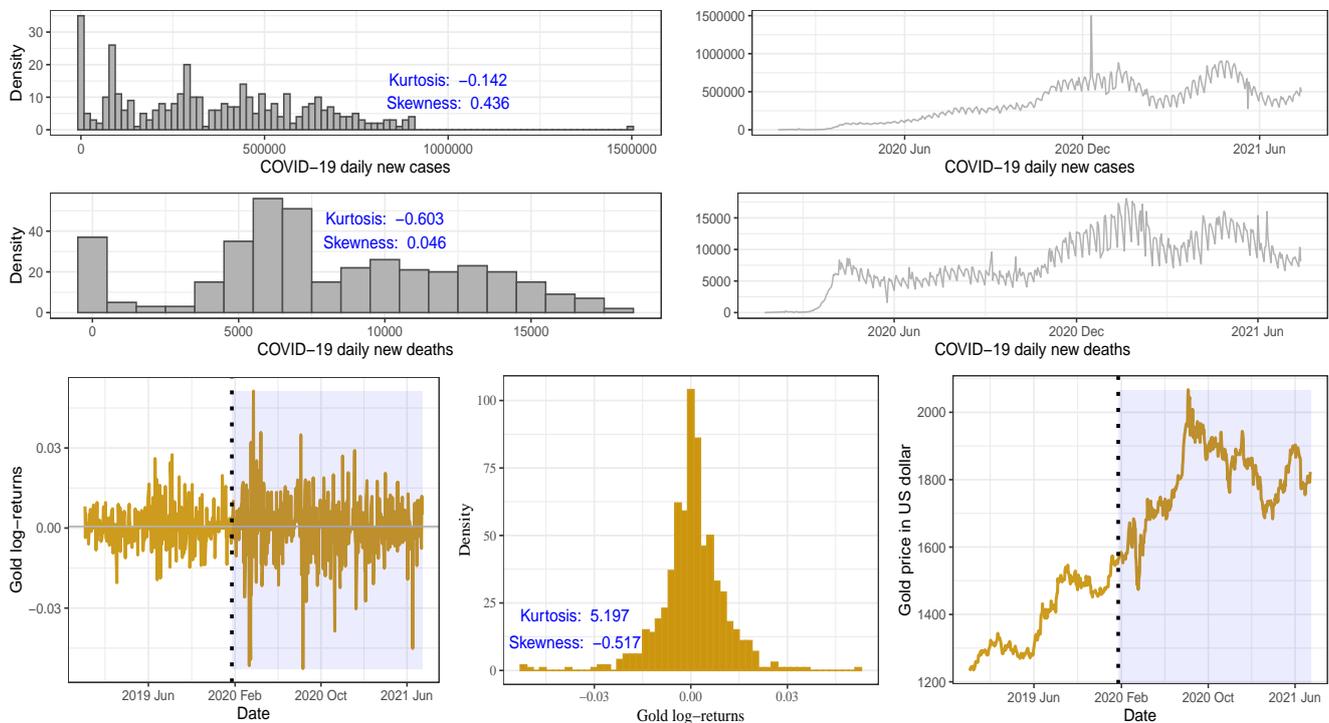


Figure 2. Probability density function and trend of daily (from 23 January 2020 to 01 July 2021) confirmed COVID-19 infections and deaths (**upper and middle**) and daily (from 04 December 2018 to 14 July 2021) returns (**lower left**), probability density function (**lower middle**), and prices (**lower right**) of the gold. The vertical dotted line represents the threshold between the pre-COVID-19 period (**left**) and during the COVID-19 period (**right**) so that the shaded zone represents the period after the pandemic begins to visualize its impact on the gold price.

Almost all currencies followed a similar trend during the period of study. The graphical plots also indicate that the market capitalization substantially plummeted, from mid-May 2021 until the end of the study period, when Tesla (<https://www.bbc.com/news/business-57096305>, accessed on 29 August 2021) announced that it would not accept or sell the cryptocurrencies due to pollution concerns. Such bad news often impacts the cryptocurrency prices leading to a substantial increase in the volatility of the returns. Thus, the traditional time-series models might not provide accurate cryptocurrency returns forecasting results, especially during adverse market conditions. Then, the SVM model might be a good option to achieve forecasting accuracy as we will show next.

We split the gold price into four classes according to their quartiles and labeled the values as Q_1 , Q_2 , Q_3 , and Q_4 . Note that, although our approach can work for all quartiles (from 0 to 100), we consider only four quartiles for simplicity's sake. Readers may use the supplementary source code for any subset of gold-price quartiles. We combine the quartiles of the gold prices one day in advance with the daily log-returns of the employed currencies and the COVID-19 cases and deaths. The resulting date corresponds to a time series comprised of 682 observations from 04 December 2018 to 22 January 2020 (before COVID-19 with 297 observations) and from 23 January 2020 to 14 July 2021 (during COVID-19 with 385 observations).

Figures 4a,b illustrate the daily returns, distribution, box-plot, and prices of the studied currencies. The box-plots illustrate that the cryptocurrency returns have many outliers in both periods and the volatility of the returns during the COVID-19 period is considerably higher than in the pre-COVID-19 pandemic period.

Table 1. Characteristics of the top ten cryptocurrencies (14 July 2021). Source: <https://coinmarketcap.com> (accessed on 29 August 2021).

Rank	Name	Symbol	Price (in US \$)	Market Capitalization (in US \$)
1	Bitcoin	 (BTC)	32,822.35	615,635,042,692
2	Ethereum	 (ETH)	1,994.33	232,727,182,406
3	Tether	 (USDT)	1.00	62,093,784,456
4	Binance Coin	 (BNB)	309.41	47,473,117,237
5	Cardano	 (ADA)	1.26	40,444,089,747
6	Ripple	 (XRP)	0.62	28,669,334,327
7	USD Coin	 (USDC)	1.00	26,668,074,913
8	Dogecoin	 (DOGE)	0.20	25,738,852,158
9	Polkadot	 (DOT)	13.67	13,310,832,721
10	Binance USD	 (BUSD)	1.00	11,579,221,385

Table 2 summarizes the statistics of the daily gold prices and returns of the studied currencies. The averages of the gold price and all cryptocurrencies returns have substantially increased over time, especially during the ongoing COVID-19 pandemic. The p -values based on the JB test strongly reject the hypothesis that the returns are normally distributed, which is coherent with the graphical analysis. During the entire period of the study, gold and currencies have a significant excess kurtosis with negative skewness (except the probability density function of the DOGE returns which has a positive skewness), as noted in Figures 2–4.

To examine for a bi-directional causal relationship between the daily number of newly infected cases (deaths) of COVID-19 and the daily gold price (cryptocurrencies returns), we apply the Granger causality test at a lag equal to three. The results are reported in Table 3, which reports that no Granger causality exists. In contrast, we find a significant casualty between the quartiles of the gold price and almost all of the cryptocurrencies returns the next day. According to the findings, contrary to the BNB and DOGE, the returns of the ADA, BTC, and ETH can affect the quartile value of the gold price, but the opposite is not true. In addition, we do not find enough evidence for a causal relationship between the XRP returns and the quartiles of the gold prices.



Figure 3. The market capitalization of the indicated cryptocurrencies. The vertical dotted line represents the threshold between the pre-COVID-19 period (left) and during the COVID-19 period (right) so that the shaded zone represents the period after the pandemic's beginning to visualize its impact on the cryptocurrency return.

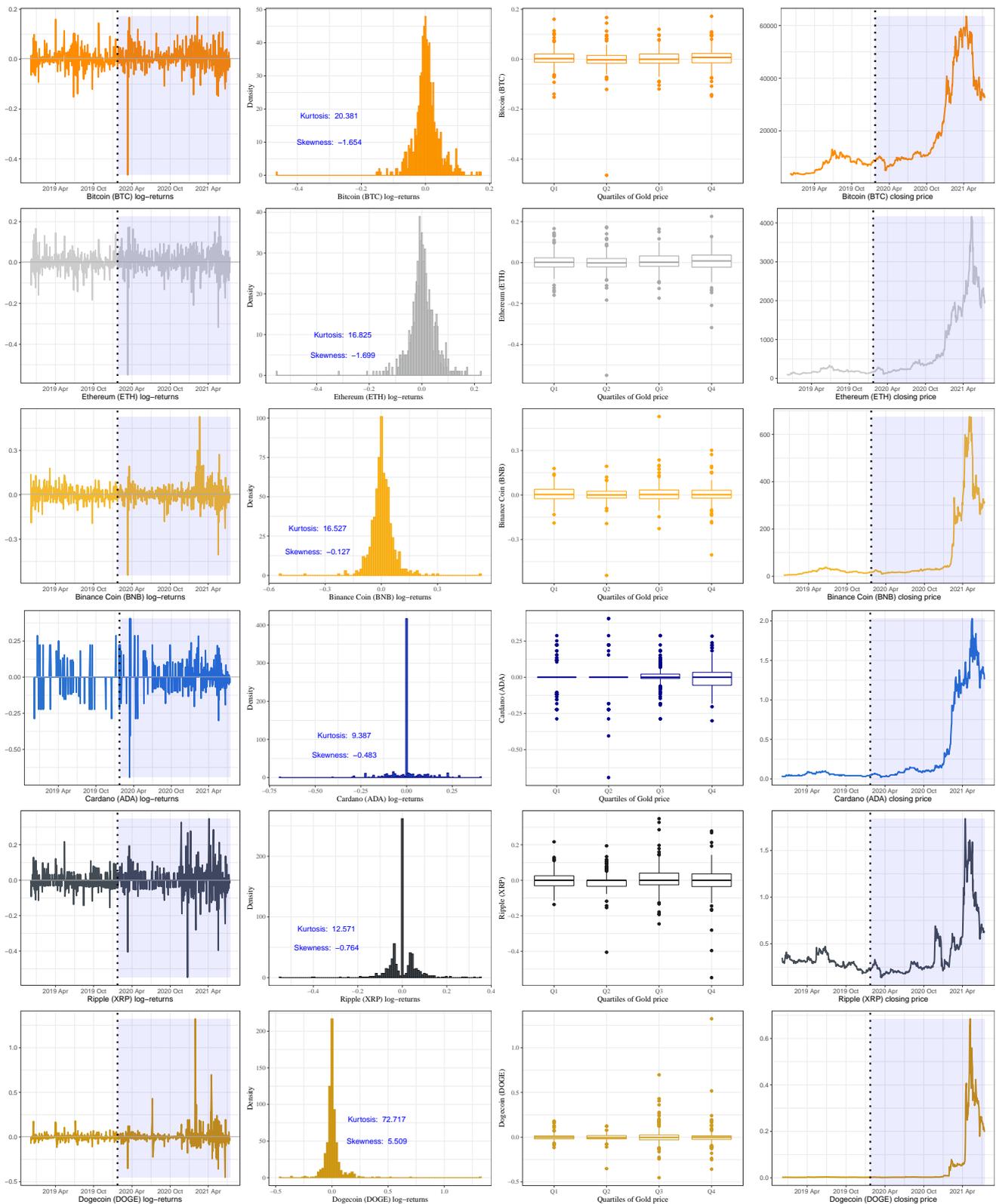


Figure 4. Cryptocurrency daily returns and box-plots of these returns based on the quartiles of the gold prices. The vertical dotted line represents the threshold between the pre-COVID-19 period (left) and during the COVID-19 period (right) so that the shaded zone represents the period after the pandemic’s beginning to visualize its impact on the cryptocurrency return.

Table 2. Descriptive statistics and preliminary tests for data of gold price, cryptocurrency return, and COVID-19.

Variable	Period	Min	Q ₁	Mean	Median	Q ₃	Max	SD *	JB Test **
Gold	Pre-COVID-19	1235	1292	1391	1400	1490	1573	101.29	31
	COVID-19	1474	1720	1792	1800	1880	2067	113	12
	Both periods	1235	1429	1617	1617	1817	2067	226.07	51
BTC	Pre-COVID-19	−0.1518	−0.0156	0.0018	0.0009	0.0188	0.1600	0.0376	137
	COVID-19	−0.4647	−0.0147	0.0036	0.0033	0.0232	0.1718	0.0477	9767.9
	Both periods	−0.4647	−0.0151	0.0028	0.0018	0.0210	0.1718	0.0436	12196
ETH	Pre-COVID-19	−0.1833	−0.0212	0.0003	−0.0012	0.0183	0.1661	0.0453	96
	COVID-19	−0.5507	−0.0203	0.0043	0.0052	0.0354	0.2256	0.0618	5793
	Both periods	−0.5507	−0.0204	0.0026	0.0016	0.0293	0.2256	0.0552	8429
BNB	Pre-COVID-19	−0.1929	−0.0235	0.0027	0.0005	0.0310	0.1787	0.0489	43
	COVID-19	−0.5428	−0.0216	0.0058	0.0038	0.0327	0.5292	0.0752	4002.3
	Both periods	−0.5428	−0.0232	0.0045	0.002	0.0323	0.5292	0.0650	7818
ADA	Pre-COVID-19	−0.2877	0.0000	−0.0018	0.0000	0.0000	0.2877	0.0777	317
	COVID-19	−0.6931	−0.0082	0.0042	0.0000	0.0217	0.4055	0.0940	1817
	Both periods	−0.6931	0.0000	0.0016	0.0000	0.0000	0.4055	0.0873	2550
XRP	Pre-COVID-19	−0.1542	−0.0317	−0.0018	0.0000	0.0000	0.2171	0.0445	109
	COVID-19	−0.5486	−0.0339	0.0011	0.0000	0.0392	0.3474	0.0814	1657
	Both periods	−0.5486	−0.0328	−0.0002	0.0000	0.0317	0.3474	0.0678	4590
DOGE	Pre-COVID-19	−0.1156	−0.0182	0.0003	−0.0024	0.0133	0.1803	0.0377	452
	COVID-19	−0.4523	−0.0244	0.0091	−0.0006	0.0218	1.3235	0.1163	36,238
	Both periods	−0.4523	−0.0208	0.0053	−0.0017	0.0179	1.3235	0.0909	154,646

* SD stands for standard deviation and ** all JB *p*-values are significant at 1%.

Table 3. Granger causality test results for data of gold prices, cryptocurrency return, and COVID-19.

Null Hypothesis	COVID-19 Infected Cases			COVID-19 Death Cases		
	Wald	<i>p</i> -Value		Wald	<i>p</i> -Value	
COVID-19 \nRightarrow Gold	0.4376	0.7262		COVID-19 \nRightarrow Gold	0.1968	0.8985
COVID-19 \nRightarrow BTC	0.9753	0.4044		COVID-19 \nRightarrow BTC	0.6815	0.5638
COVID-19 \nRightarrow ETH	1.9525	0.1207		COVID-19 \nRightarrow ETH	0.8400	0.4726
COVID-19 \nRightarrow BNB	0.7258	0.5371		COVID-19 \nRightarrow BNB	0.2029	0.8944
COVID-19 \nRightarrow ADA	1.0662	0.3634		COVID-19 \nRightarrow ADA	2.0982	0.1000
COVID-19 \nRightarrow XRP	2.4939	0.0597		COVID-19 \nRightarrow XRP	2.2387	0.0834
COVID-19 \nRightarrow DOGE	0.8942	0.4442		COVID-19 \nRightarrow DOGE	2.2739	0.0796

The results in Table 2 and Figures 2–4 also show a significant increase in the volatility of the returns of gold and currencies after the outbreak of the COVID-19, compared with the pre-pandemic period. Note that, in these figures, as mentioned, the vertical dotted line represents a threshold between the pre-COVID-19 period and during the pandemic. In addition, the circled points are the support vectors, and misclassification appears when the color of the support vectors does not match the color of the decision boundary. Thus, the traditional econometric and time-series models might not provide helpful results in such cases. Hence, due to its superior forecasting performance, we implement the SVM in our analysis.

The SVM optimization strategy is usually done in two stages. First, we separate the observations into training and testing data sets. A rule of thumb indicates to consider the first 75% contiguous days as a training set to fit the model and the remaining 25% days as a testing set, which is used to validate the prediction accuracy of this model [40]. The best model is selected based on minimizing the mean absolute percentage error (MAPE), root mean square error (RMSE), or normalized root mean square error (NRMSE) defined as

$$\begin{aligned}
 \text{MAPE} &= \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%, \\
 \text{RMSE} &= \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \\
 \text{NRMSE} &= \sqrt{\frac{N \sum_{i=1}^N (y_i - \hat{y}_i)^2}{(\sum_{i=1}^N y_i)^2}},
 \end{aligned} \tag{11}$$

where N is the total number of observations, \hat{y}_i is the predictive value of y_i , for $i = 1, \dots, N$, and $y_i - \hat{y}_i$ denotes the prediction error in y_i .

Second, we utilize the SVM optimization strategy as a dynamic model that involves all available data. In [41], it is showed that the dynamic model can achieve more accurate forecasting results compared with the one based on a fixed training set. Therefore, in this article, we adopt the dynamic SVM algorithm.

The optimized estimation of the coefficient c stated in (10) and the parameters γ, d defined in (2) and (3), respectively, may be obtained based on the k -fold cross validation method, where the index expressed in (11) has the minimum value. In this article, we consider a 10-fold cross validation and a set of tuning parameters $c \in \{0.001, 0.01, 1, 2, 4, 8, 16, 32, 64, 128, 256\}$, $d \in \{1, 2, 3\}$, and $\gamma \in \{0.3, 0.4, 0.5\}$. The estimated parameters of the linear, polynomial, and radial kernel models for the entire sample are shown in Table 4. Applying the optimized parameters in Table 4, the SVM models with each kernel are run on a daily series and the forecasting performance accuracy (in %) for each currency is calculated. The results are indicative of the SVM's superior predictive power when using a radial kernel (between 89.4% and 91.8% performance accuracy) compared with linear (between 71.8% and 74.5% performance accuracy) and polynomial (between 74.8% and 79.3% performance accuracy) kernels.

To illustrate the SVM classification prediction capabilities, using the radial classifier, during both periods, pre- and during the COVID-19 outbreak, Figures 5–7 show the decision rules for a different set of patterns in the hyperplane. The four classes of gold price quartiles are represented by dotted points with four different colors. In the figures, we indicate the support vectors (patterns) with dotted points and double circles, and the errors (misclassifications) with dotted points and double circles but in different colors. Figures 5–7 display that the majority of the errors occurred during the COVID-19 period compared with the pre-period of COVID-19, where some of the cryptocurrencies returns from Q_3 class of the gold price are misclassified as Q_4 . This is justifiable as the uncertainty about the COVID-19 has resulted in significant effects that continue to drive perceptions of risk and volatility in cryptocurrencies and gold prices.

Table 4. Optimal parameters and performance accuracy (in %) for the indicated currency and kernel.

Currency	Linear Kernel		Polynomial Kernel			Radial Kernel		
	Accuracy	c	Accuracy	c	d	Accuracy	c	γ
Binance Coin	73.2	64	76.8	512	3	91.5	512	0.4
Bitcoin	74.5	256	77.3	256	3	91.8	512	0.3
Cardano	72.6	256	74.8	512	3	91.1	128	0.4
Dogecoin	71.8	256	76.5	512	3	89.4	16	0.4
Ethereum	73.0	64	79.3	128	3	90.9	128	0.3
Ripple	72.3	128	75.8	512	3	90.8	256	0.3

Where the parameter c is used for all kernels, d is the degree parameter of the polynomial kernel, and γ is the radial kernel parameter.

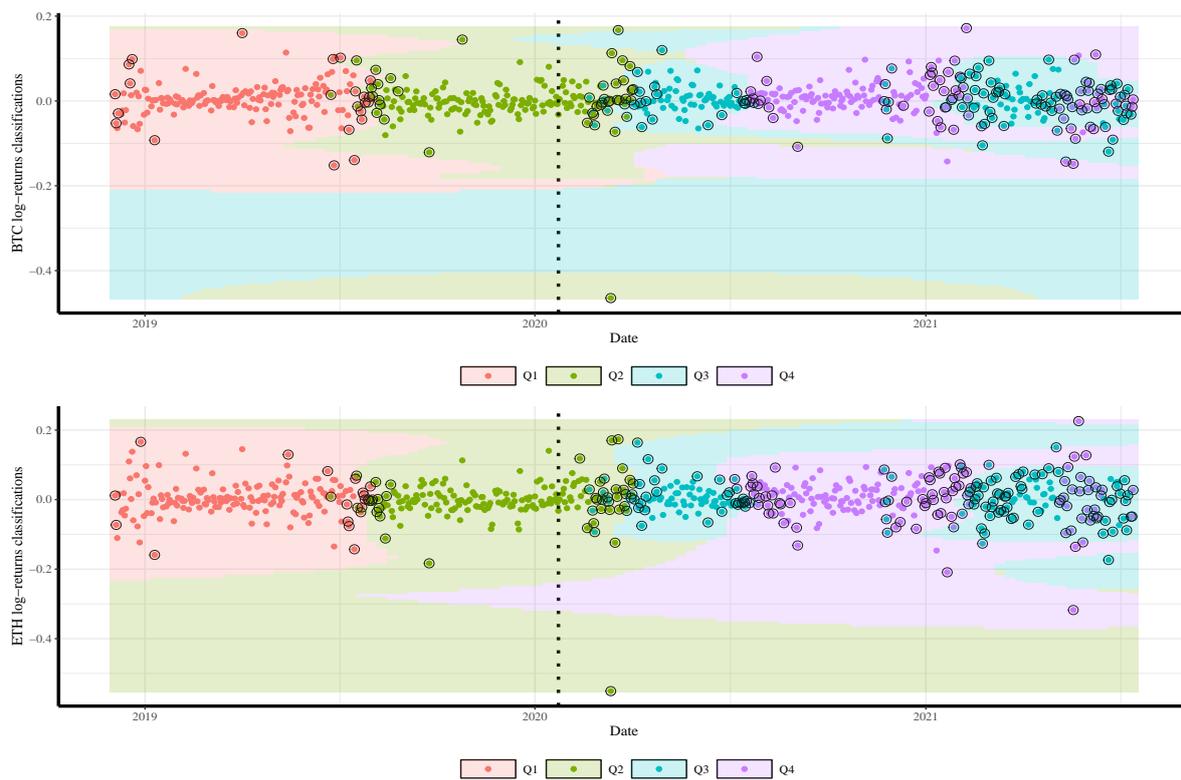


Figure 5. Classifications of BTC (top) and ETH (bottom) daily returns based on quartiles of daily gold prices predicted by the SVM model. The vertical dotted line represents the threshold between the pre-COVID-19 period (left) and ongoing COVID-19 period (right) to visualize its impact on the cryptocurrency return.

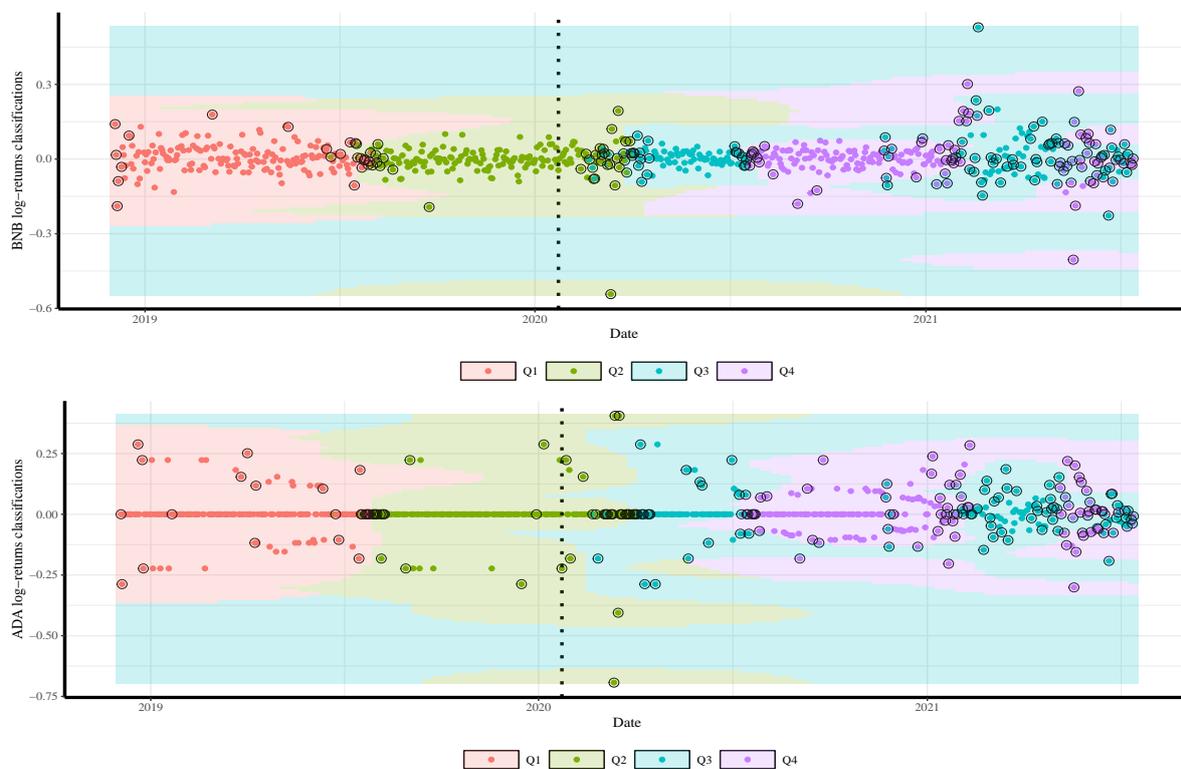


Figure 6. Classifications of BNB (top) and ADA (bottom) daily returns based on quartiles of daily gold prices predicted by SVM model. The vertical dotted line represents the threshold between the pre-COVID-19 period (left) and ongoing COVID-19 period (right) to visualize its impact on the cryptocurrency return.

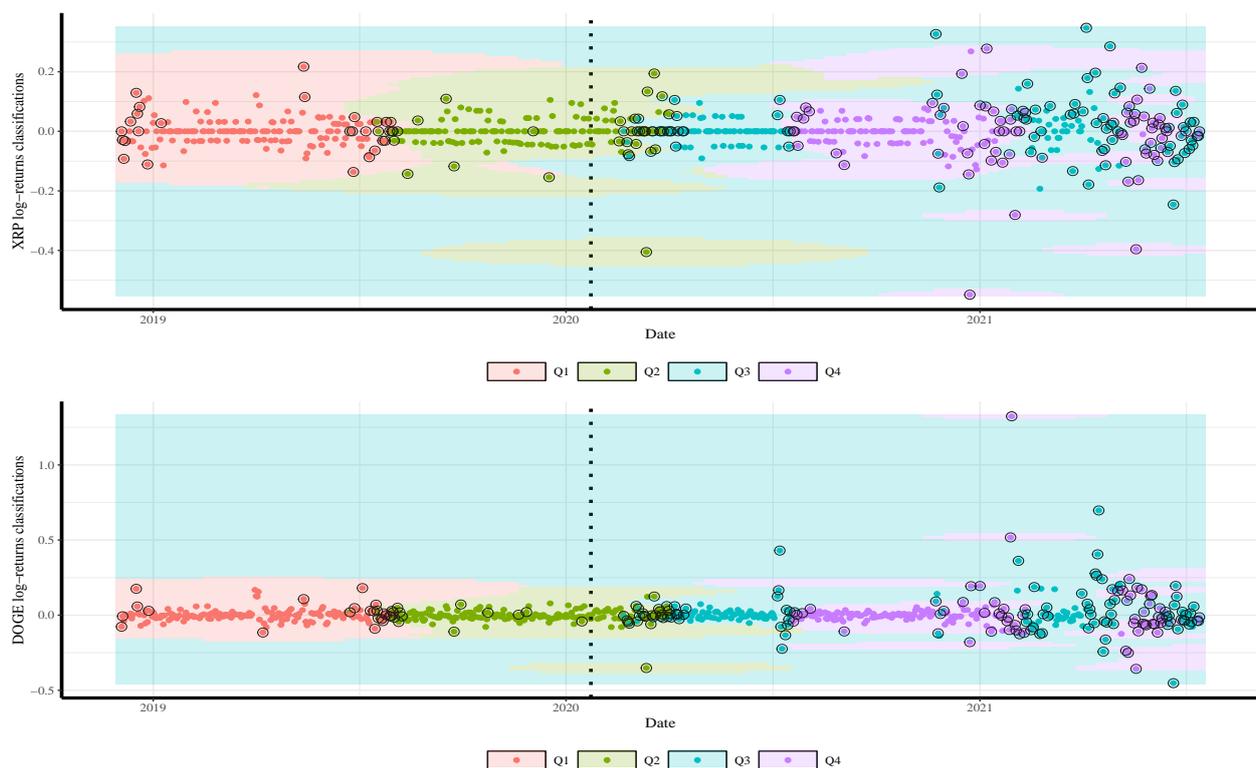


Figure 7. Classifications of XRP (top) and DOGE (bottom) daily returns based on quartiles of daily gold prices predicted by SVM model. The vertical dotted line represents the threshold between the pre-COVID-19 period (left) and ongoing COVID-19 period (right) to visualize its impact on the cryptocurrency return.

5. Conclusions

In this study, we have applied the support vector machine model to daily returns of six major cryptocurrencies: Binance Coin, Bitcoin, Cardano, Dogecoin, Ethereum, and Ripple, before and during the COVID-19 pandemic. The daily returns and volatility spillover substantially increased during the period of the pandemic, due to the high fluctuation in prices in comparison with the pre-pandemic period, making accurate forecasting a complex duty. Instead of using a traditional time series model to predict the return of the cryptocurrencies, we have proposed a new easy approach based on an advanced machine learning algorithm to achieve accurate forecasting results for the classification of the returns.

We have designed an algorithm that allows us to use sensors when submitting requests for data of cryptocurrency returns, gold prices, and COVID-19. After collecting the price series of the gold and a cryptocurrency, one can classify the returns of the cryptocurrency in the previous day based on the quartile of the gold price the next day. Also, we may employ the proposed model to predict the classification of the return, whether it will be in the first, second, third or fourth quartile of the gold price.

We have shown that the SVM model can provide significant accurate forecasting, especially during crises. Overall, our model is not only valuable for predicting the classification of the returns of the aforementioned cryptocurrencies, but it is also helpful to be used for other cryptocurrencies.

Our findings support the claim that the SVM technique provides a robust algorithm for the predictability of cryptocurrencies under bad or good news. We believe that the findings of this research are not only helpful for exploring the relationship between the classification returns of the cryptocurrencies and the quantiles of the gold price, but they are also of great interest to macroeconomic policymakers, portfolio managers, and investors who are actively trading in the cryptocurrency market, especially under adverse market conditions and during times of uncertainty. In summary, our new approach can help

to devise trading strategies for cryptocurrencies and help investors and macroeconomic policymakers making better investment decisions.

This article provides avenues for further research in digital currency portfolios by comparing our model with other existing evolutionary machine learning algorithms to achieve more accurate results.

Author Contributions: Data curation, E.M. and S.M.; investigation, E.M. and V.L.; formal analysis and methodology, E.M., V.L., S.M., and C.M.-B.; writing—original draft, E.M.; writing—review and editing, V.L. and C.M.-B. All authors have read and agreed to the submitted version of the manuscript.

Funding: This research was partially supported by FONDECYT, project grant number 1200525 (V.L.), from the National Agency for Research and Development (ANID) of the Chilean government under the Ministry of Science, Technology, Knowledge and Innovation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The analyzed data and used codes are available under request.

Acknowledgments: The authors would also like to thank the editor and two reviewers for their constructive comments which led to improving the presentation of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Chahuan-Jimenez, K.; Rubilar, R.; de la Fuente-Mella, H.; Leiva, V. Breakpoint analysis for the COVID-19 pandemic and its effect on the stock markets. *Entropy* **2021**, *23*, 100. [\[CrossRef\]](#)
- Dyhrberg, A.H. Bitcoin, gold and the dollar—A GARCH volatility analysis. *Financ. Res. Lett.* **2016**, *16*, 85–92. [\[CrossRef\]](#)
- Makala, D.; Li, Z. Prediction of gold price with ARIMA and SVM. *J. Phys. Conf. Ser.* **2021**, *1767*, 012022. [\[CrossRef\]](#)
- Salisu, A.A.; Ogbonna, A.E. The return volatility of cryptocurrencies during the COVID-19 pandemic: Assessing the news effect. *Glob. Financ. J.* **2021**. [\[CrossRef\]](#)
- Shen, M.; Zhu, L.; Xu, K. *Blockchain: Empowering Secure Data Sharing*; Springer: Singapore, 2020.
- Sanchez, L.; Leiva, V.; Galea, M.; Saulo, H. Birnbaum-Saunders quantile regression and its diagnostics with application to economic data. *Appl. Stoch. Model. Bus. Ind.* **2021**, *37*, 53–73. [\[CrossRef\]](#)
- Saulo, H.; Dasilva, A.; Leiva, V.; Sanchez, L.; de la Fuente-Mella, H. Log-symmetric quantile regression models. *Stat. Neerl.* **2021**. [\[CrossRef\]](#)
- Bouoiyour, J.; Selmi, R.; Wohar, M.E. Measuring the response of gold prices to uncertainty: An analysis beyond the mean. *Econ. Model.* **2018**, *75*, 105–116. [\[CrossRef\]](#)
- Bouri, E.; Gupta, R.; Tiwari, A.K.; Roubaud, D. Does bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. *Financ. Res. Lett.* **2017**, *23*, 87–95. [\[CrossRef\]](#)
- Dastgir, S.; Demir, E.; Downing, G.; Gozgor, G.; Lau, C.K.M. The causal relationship between bitcoin attention and bitcoin returns: Evidence from the copula-based granger causality test. *Financ. Res. Lett.* **2019**, *28*, 160–164. [\[CrossRef\]](#)
- Kim, M.J.; Canh, N.P.; Park, S.Y. Causal relationship among cryptocurrencies: A conditional quantile approach. *Financ. Res. Lett.* **2021**. [\[CrossRef\]](#)
- Sim, N.; Zhou, H. Oil prices, us stock return, and the dependence between their quantiles. *J. Bank. Financ.* **2015**, *55*, 1–8. [\[CrossRef\]](#)
- Mensi, W.; Sensoy, A.; Vo, X.V.; Kang, S.H. Impact of covid-19 outbreak on asymmetric multifractality of gold and oil prices. *Resour. Policy* **2020**, *69*, 101829. [\[CrossRef\]](#)
- de la Fuente-Mella, H.; Rubilar, R.; Chahuan-Jimenez, K.; Leiva, V. Modeling COVID-19 cases statistically and evaluating their effect on the economy of countries. *Mathematics* **2021**, *9*, 1558. [\[CrossRef\]](#)
- Yousaf, I.; Ali, S. Discovering interlinkages between major cryptocurrencies using high-frequency data: new evidence from COVID-19 pandemic. *Financ. Innov.* **2020**, *6*, 1. [\[CrossRef\]](#)
- Bustos, N.; Tello, M.; Droppelmann, G.; Garcia, N.; Feijoo, F.; Leiva, V. Machine learning techniques as an efficient alternative diagnostic tool for COVID-19 cases. *Signa Vitae* **2021**. [\[CrossRef\]](#)
- Iqbal, N.; Fareed, Z.; Wan, G.; Shahzad, F. Asymmetric nexus between covid-19 outbreak in the world and cryptocurrency market. *Int. Rev. Financ. Anal.* **2021**, *73*, 101613. [\[CrossRef\]](#)
- Conlon, T.; Corbet, S.; McGee, R.J. Are cryptocurrencies a safe haven for equity markets? An international perspective from the covid-19 pandemic. *Res. Int. Bus. Financ.* **2020**, *54*, 101248. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mokni, K.; Ajmi, A.N. Cryptocurrencies vs. US dollar: Evidence from causality in quantiles analysis. *Econ. Anal. Policy* **2021**, *69*, 238–252. [\[CrossRef\]](#)

20. Guzman, A.; Pinto-Gutierrez, C.; Trujillo, M.A. Trading cryptocurrencies as a pandemic pastime: COVID-19 lockdowns and bitcoin volume. *Mathematics* **2021**, *9*, 1771. [[CrossRef](#)]
21. Leiva, V.; Saulo, H.; Souza, R.; Aykroyd, R.G.; Vila, R. A new BISARMA time series model for forecasting mortality using weather and particulate matter data. *J. Forecast.* **2021**, *40*, 346–364. [[CrossRef](#)]
22. Palacios, C.A.; Reyes-Suarez, J.A.; Bearzotti, L.A.; Leiva, V.; Marchant, C. Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. *Entropy* **2021**, *23*, 485. [[CrossRef](#)]
23. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
24. Razi, M.A.; Athappilly, K. A comparative predictive analysis of neural networks (NNS), nonlinear regression and classification and regression tree (CART) models. *Expert Syst. Appl.* **2005**, *29*, 65–74. [[CrossRef](#)]
25. Aykroyd, R.G.; Leiva, V.; Ruggeri, F. Recent developments of control charts, identification of big data sources and future trends of current research. *Technol. Forecast. Soc. Chang.* **2019**, *144*, 221–232. [[CrossRef](#)]
26. Geranian, H.; Tabatabaei, S.H.; Asadi, H.H.; Carranza, E.J.M. Application of discriminant analysis and support vector machine in mapping gold potential areas for further drilling in the Sari-Gunay gold deposit, NW Iran. *Nat. Resour. Res.* **2016**, *25*, 145–159. [[CrossRef](#)]
27. Ślepaczuk, R.; Zenkova, M. Robustness of support vector machines in algorithmic trading on cryptocurrency market. *Cent. Eur. Econ. J.* **2018**, *5*, 186–205. [[CrossRef](#)]
28. Ramirez-Figueroa, J.A.; Martin-Barreiro, C.; Nieto, A.B.; Leiva, V.; Galindo-Villardón, M.P. A new principal component analysis by particle swarm optimization with an environmental application for data science. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 1969–1984. [[CrossRef](#)]
29. Hitam, N.A.; Ismail, A.R.; Saeed, F. An optimized support vector machine based on particle swarm optimization (PSO) for cryptocurrency forecasting. *Procedia Comput. Sci.* **2019**, *163*, 427–433. [[CrossRef](#)]
30. Sebastião, H.; Godinho, P. Forecasting and trading cryptocurrencies with machine learning under changing market conditions. *Financ. Innov.* **2021**, *7*, 3. [[CrossRef](#)]
31. Cabezas, X.; García, S.; Martin-Barreiro, C.; Delgado, E.; Leiva, V. A two-stage location problem with order solved using a Lagrangian algorithm and stochastic programming for a potential use in COVID-19 vaccination based on sensor-related data. *Sensors* **2021**, *21*, 5352. [[CrossRef](#)]
32. Martin-Barreiro, C.; Ramirez-Figueroa, J.; Cabezas, X.; Leiva, V.; Galindo-Villardón, M. Disjoint and functional principal component analysis for infected cases and deaths due to COVID-19 in South American countries with sensor-related data. *Sensors* **2021**, *21*, 4094. [[CrossRef](#)] [[PubMed](#)]
33. Rojas, F.; Leiva, V.; Huerta, M.; Martin-Barreiro, C. Lot-size models with uncertain demand considering its skewness/kurtosis and stochastic programming applied to hospital pharmacy with sensor-related COVID-19 data. *Sensors* **2021**, *21*, 5198. [[CrossRef](#)] [[PubMed](#)]
34. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. *A Training Algorithm for Optimal Margin Classifiers*; Association for Computing Machinery: New York, NY, USA, 1992. [[CrossRef](#)]
35. Luenberger, D.G.; Ye, Y. *Linear and Nonlinear Programming*; Springer: New York, NY, USA, 2008.
36. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F.; Chang, C.C.; Lin, C.C. e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. R Package Version V.1.7-9. 2021. Available online: <https://cran.r-project.org/web/packages/e1071> (accessed on 16 September 2021).
37. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016. Available online: <https://ggplot2.tidyverse.org> (accessed on 15 September 2021).
38. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
39. World Health Organization. *Coronavirus Disease (COVID-19) Outbreak*; WHO: Geneva, Switzerland, 2021. Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed on 20 September 2021).
40. Stone, M. Cross-validation choice and assessment of statistical predictions *J. R. Stat. Soc. B* **1974**, *36*, 111–147.
41. Henrique, B.M.; Sobreiro, V.A.; Kimura, H. Stock price prediction using support vector regression on daily and up to the minute prices. *J. Financ. Data Sci.* **2018**, *4*, 183–201. [[CrossRef](#)]