Journal of Applied Finance & Banking, Vol. 14, No. 6, 2024, 95-119
ISSN: 1792-6580 (print version), 1792-6599(online)
https://doi.org/10.47260/jafb/1466
Scientific Press International Limited

A Darwinian Approach via ML to the Analysis of Cryptocurrencies' Returns

Federico Cini[1] and Annalisa Ferrari[2]

Abstract

This study adopts a Darwinian approach leveraging machine learning (ML) to analyze cryptocurrency returns and their interactions with traditional financial markets. Using a daily dataset from 2018 to 2023, the Random Forest model proved particularly effective in identifying key factors influencing cryptocurrency returns, including technology stock indices (NASDAQ), global equity indices (S&P500, Eurostoxx600), commodity prices (gold, crude oil), and market sentiment (Google Trends). The analysis reveals consistent positive relationships between market sentiment and cryptocurrency returns, highlighting the crucial role of public interest in shaping long-term outcomes. Cryptocurrencies emerge as a distinct asset class with specific correlations to traditional markets and investor sentiment. The study provides strategic insights into understanding cryptocurrency behavior and integrating these dynamics into informed portfolio strategies. It emphasizes the importance of monitoring both traditional financial indices and market sentiment for investment decisions across various time horizons.

JEL classification numbers: C58, G11, G15.
Keywords: Crypto Assets, Bitcoin, Machine Learning, Investor Decisions.

---

[1] Sapienza University of Rome, PhD School in Social and Economic Sciences, Roma, Italy.
[2] University of Rome UnitelmaSapienza, Department of Law and Digital Society, Roma, Italy.

Article Info: Received: October 29, 2024. Revised: November 22, 2024.
Published online: December 12, 2024.

1. Introduction

Bitcoin was conceived as a decentralized system for digital transactions, operating autonomously and without intermediaries. Transaction validation requires substantial computational power, incentivizing participants to contribute to exchange for rewards. Miners validating transactions receive fees and Bitcoins, which are used within the system. This makes Bitcoins digital currencies with unique identifiers, functioning akin to traditional currencies but limited to exchanges within the system.

The profitability of mining hinges on the validation method and the revenue gained from participating. This relationship underscores how network computational power influences cryptocurrency price trends, system stability, and functionality. (Dutta, Kumar, & Basu, 2020). After Bitcoin consolidated its technology and initially spread as a cryptocurrency, many other projects emerged, each based on different systems and representing various cryptocurrencies. According to the latest data available in June 2024, over 24,000 different tokens are listed in CoinMarketCap (2024), and the total market capitalization is 1.18 trillion USD. An overview of major cryptocurrencies in order of capitalization is provided in Table 1.

Table 1: Top 10 Most Capitalized Cryptocurrencies - CoinMarketCap (2024)

Name	Ticker	Price (USD)	Market Cap (USD)	Circulating Supply
Bitcoin	BTC	59789.07	1.18T	19,719,378 BTC
Ethereum	ETH	3277.72	393.48B	120,193,480 ETH
Tether	USDT	0.9989	112.52B	112,644,773,126 USDT
BNB	BNB	551.98	81.47B	147,583,003 BNB
Solana	SOL	140.29	64.92B	462,770,578 SOL
USD Coin	USDC	0.9999	32.64B	32,641,286,638 USDC
XRP	XRP	0.4633	25.80B	55,688,327,582 XRP
Toncoin	TON	7.77	19.11B	2,460,373,617 TON
Dogecoin	DOGE	0.1176	17.05B	144,967,156,384 DOGE
Cardano	ADA	0.4034	14.43B	35,760,987,560 ADA

This proliferation of systems, cryptocurrencies, and tokens (Howell, Niessner, & Yermack, 2020) has spawned a complete financial sector. This sector encompasses not only cryptocurrencies but also various derivative instruments, which are attracting growing interest from both the public and professionals, prompting market reactions and shaping governmental policies (Augustin, Rubtsov, & Shin, 2020).

The exponential growth of the cryptocurrency market in recent years has caused increasingly pressing questions about the true nature of these assets, their behavior, and the stylized aspects that can be attributed to them. The answers to these questions identify the possibility for investors to take informed decisions regarding their portfolio composition strategies. In fact, investing in cryptocurrencies presents unique challenges due to their distinctive risk and volatility profiles compared to traditional assets.

Assuming cryptocurrencies constitute a distinct and separate ecosystem, in this study, by adopting a Machine Learning (ML) approach, we analyze a novel dataset from 2018 to 2023 to investigate how cryptocurrencies interact with other asset classes and identify factors influencing their returns. The importance of the topic analyzed in this study can be defined by three points of interest, depending on the reader's perspective: (i) exploratory interest in the emerging asset class of cryptocurrencies, as they behave significantly differently from the rest of the financial ecosystem depending on the observation time window; (ii) methodological interest in applying innovative models to a novel dataset; (iii) strategic interest in highlighting the main behaviors of this new asset class.

This paper is subsequently organized as follows: Section 2 presents the theoretical framework; Section 4 outlines the methodology used; Section 3 describes the datasets; Section 5 discusses the results; and Section 6 concludes.

2. Literature Review

Since their introduction in the context of exchanges on decentralized infrastructures, cryptocurrencies have undoubtedly been seen to carry out transactions on the network (Nakamoto, 2008). However, the most significant evolution of the blockchain came with the introduction of the Ethereum platform (Buterin et al., 2014). This confirmed the transactional use of cryptocurrencies and provided users with the ability to use ETH as a currency for automated exchanges governed by innovative smart contracts.

Classifying cryptocurrencies as mere transaction tools is inadequate due to their exponential market growth and minimal transactional use: hedge funds and asset managers have begun to include cryptocurrency-related assets in their portfolios and trading strategies (Fang et al., 2022). Various schools of thought have attempted to compare cryptocurrencies to traditional asset classes, but inconsistent results led scholars to initially focus on what cryptocurrencies were not. Eventually, the hypothesis emerged that cryptocurrencies represent a new, entirely distinct asset class.

The term “cryptocurrency” highlights its initial characteristics: usage as currency and origins in cryptographic algorithms. Early blockchain scholars questioned cryptocurrencies’ ability to function as currencies, leading to the development of diverse and often conflicting viewpoints. Viewpoints (Franco, 2014).
However, all analyses, including those conducted later, share a common research approach fundamentally based on verifying the three functions that generally characterize money (Yermack, 2015).
For each of these functions, cryptocurrencies align or diverge from fiat currencies since the latter are issued by authorized central banks and are backed by both a real economy and reserves of value. Regarding the function as a medium of exchange, it is necessary to consider that cryptocurrencies originate within blockchains, which inherently imply trading on markets and transaction recording in dedicated ledgers. Therefore, cryptocurrencies are intended as a means of payment. However, this assertion is not without criticism from scholars. Challenges regarding the classification of cryptocurrencies as a means of payment primarily concern their liquidity. Estimating the liquidity of cryptocurrencies is complicated by the lack of regulated data feeds in their markets, making it challenging to apply for standard metrics such as bid-ask spreads (Brauneis, Mestel, Riordan, & Theissen, 2021).
In a study focusing on the top six cryptocurrencies by market capitalization, Phillip, Chan, and Peiris (2019) found that cryptocurrencies with slower transactions, such as Bitcoin, are associated with lower volatility and liquidity. In contrast, cryptocurrencies with faster transactions exhibit oscillatory characteristics and lower liquidity risk during transactions, making them preferable purely as a medium of exchange. Critiques regarding the actual function of cryptocurrencies as a medium of payment also concern the weight of transactions for the purchase of goods and services compared to total transactions (Baur, Hong, & Lee, 2018), asserting that the number of transactions for goods and services on major blockchains is marginal compared to the total (Yermack, 2015), with the majority of transactions aimed at currency exchanges for speculative purposes (Alfieri, Burlacu, & Enjolras, 2019).
Regarding the store of value function, cryptocurrencies face a significant hurdle due to their high volatility (Liang, Li, Chen, & Zeng, 2019; Panagiotidis, Papapanagiotou, & Stengos, 2022; Peng, Albuquerque, de Sa, Padula, & Montenegro, 2018). This volatility instills a sense of distrust among users regarding the reliability of cryp- tocurrencies as a store of value, often leading them to hold these assets for speculative purposes.

Concerning the unit of account function, it requires the currency's ability to facilitate easy comparison of the values of different goods. While it's feasible to compare prices of various goods using any cryptocurrency at any given time, this task becomes increasingly challenging over longer intervals for two main reasons.
Firstly, merchants could theoretically display prices in any cryptocurrency, but the varying unit values (ranging from fractions of a cent to tens of thousands of dollars) make practical implementation difficult. Secondly, the significant price volatility of cryptocurrencies (especially in relation to fiat currencies) poses substantial challenges in accurately comparing the value of goods (Figuet, 2016).

Based on the reasons mentioned above and the current literature, it can be concluded that currently only the first of the three functions that characterize traditional currencies can be partially attributed to cryptocurrencies.

It is important to highlight that for those arguing against cryptocurrencies being considered as currencies, a prominent argument in the literature concerns the protective aspects of decentralized systems from an economic perspective. Cryptocurrencies notably lack an entity empowered to set monetary policy, meaning there is no central bank capable of issuing the currency (Alfieri et al., 2019). Without a central bank able to adjust interest rates to safeguard the market, and with cryptocurrency values determined solely by market equilibrium, they fail to exhibit key characteristics of traditional currencies (Glaser, Zimmermann, Haferkorn, Weber, & Siering, 2014).

In the context of this discussion, it is pertinent to delve deeper into considerations specifically concerning the primary cryptocurrency, Bitcoin. Bitcoin's infrastructure is entirely predicated on the premise that at most exactly 21 million bitcoins can be mined. Consequently, there is no doubt that over time, the quantity of coins available on the market will tend to stabilize. If Bitcoin were indeed integrated into a traditional currency, it would lead to a unique scenario where, at some point, it would become impossible for any entity to mint new money (Yermack, 2015).

An alternative strand of literature suggests classifying cryptocurrencies as commodities. Due to specific characteristics such as scarcity, indestructibility, homogeneity, standardization, and divisibility, cryptocurrencies are likened to gold and other precious metals. They meet the commodity definition per the US Commodity Exchange Act (CEA) and can be used as underlying assets for futures contracts (Prentis, 2015). Some tokens, like Bitcoin, are finite by design, though mining continues, leading to potential future market behaviors like gold. Hence, investors consider tokens as safe havens, akin to gold (Alfieri et al., 2019). However, this safe-haven status is more limited compared to gold, influenced by the observation period of the returns. (Feng, Wang, & Zhang, 2018).

Recent studies argue that tokens lack intrinsic value and cannot function as a safe-heaven, classifying them as speculative assets. This view is supported by observing the behavior of cryptocurrency users, showing a growing prevalence of speculative investors and a decreasing minority of users for other purposes (Baur et al., 2018; Glaser et al., 2014; Nunez, Contreras-Valdez, & Franco-Ruiz, 2019).

With a comparative analysis of cryptocurrency, forex, and stock markets, Baur et al. (2018) highlight similarities and differences based on volatility, central assets, robustness, and risk. The study finds that the cryptocurrency market is most akin to the stock market, though they do not entirely overlap. The analysis conducted so far demonstrates that cryptocurrencies cannot be strictly classified within traditional asset classes but should rather be considered a new type of financial instrument, still within the speculative asset family but distinct due to their modern technological and economic foundations (Ankenbrand & Bieri, 2018; Glaser et al., 2014).

By looking at the asset's universe as a complex ecosystem, Pele, Wesselhofft, Hardle, Kolossiatis, and Yatracos (2023) provide empirical evidence that cryptocurrencies exhibit a synchronic evolution, i.e. individual cryptocurrencies develop similar statistical characteristics over time, allowing them to differentiate from classical assets. A related analysis can be found in ElBahrawy, Alessandretti, Kandler, Pastor-Satorras, and Baronchelli (2017), where the cryptocurrency market is seen as an evolutive system with several characteristics which are preserved over time: while new cryptocurrencies appear and disappear continuously and their market capitalization is increasing exponentially, the number of active cryptocurrencies, market share distribution and the turnover of cryptocurrencies have been stable for years.

This article aims to contribute to the debate on the real nature of cryptocurrencies to identify precise indications for investors. The proposed analysis adopts a Darwinian approach (Davison, 2020), whereby the characteristics of cryptocurrencies are identified basing on observable behavior over time.

3. Data

The dataset on which this work focuses consists of 2185 daily observations and is composed of historical series of daily prices for the period 2018 - 2023 available on Yahoo Finance. For each historical series in the data set, log returns were computed. The dataset can be divided into four macro-classes: cryptocurrencies and sentiment analysis index, commodities, equity index and currencies. Table 2 summarizes all the features in our dataset and their characteristics.

Table 2: Features

	Feature	Description
Crypto	BTC	Bitcoin to US Dollar exchange rate
	TREND_BTC[3]	Google trend with keyword 'Bitcoin'
	TREND_CRYPTO[3]	Google trend with keyword 'Cryptocurrencies'
	TREND_ETH[3]	Google trend with keyword 'Ethereum'
Currencies	EUR	Euro to US Dollar exchange rate
	JPY	Japanese Yen to US Dollar exchange rate
	GBP	British Pound to US Dollar exchange rate
	AUD	Australian Dollar to US Dollar exchange rate
	NZD	New Zealand Dollar to US Dollar exchange rate
	RUB	Russian Ruble to US Dollar exchange rate
Commodities	GOLD	Gold Futures, standard for trading gold
	SILVER	Silver Futures, standard for trading silver
	COPPER	Copper Futures, standard for trading copper
	CRUDEOIL	Crude Oil Futures, standard for trading crude oil
	CORN	Corn Futures, standard for trading corn
	OAT	Oats Futures, standard for trading oats
	HEATOIL	Heating Oil Futures, standard for trading heating oil
	NATGAS	Natural Gas Futures, standard for trading natural gas
	SOYOIL	Soybean Oil Futures, standard for trading soybean oil
	SOY	Soybean Futures, standard for trading soybeans
	SUGAR	Sugar Futures, standard for trading sugar
Market indices	SP500	S&P 500 Index, a major U.S. stock market index
	DOWJONES	Dow Jones Industrial Average, a major U.S. stock market index
	NASDAQ	NASDAQ Composite Index, focused on technology stocks
	FTSE	FTSE 100 Index, major UK stock market index
	SENSEX	BSE Sensex, benchmark index of the Bombay Stock Exchange
	SSE	SSE Composite Index, major stock market index of China
	NIKKEI	Nikkei 225, leading stock market index of Japan
	STOXX	STOXX Europe 600, European stock market index
	VIX	Volatility Index, measures market volatility expectations[3]

[3] Available on a weekly basis

To observe the role that media attention plays in the cryptocurrency dynamics (Tandon, Revankar, & Parihar, 2021), the global Google trend index was included in the dataset. Google Trends is a public web facility of Google Inc. that shows how often a particular search term is entered relative to the total search volume across various regions of the world and in various languages. This index assigns a score from 0 to 100 to the keywords related to cryptocurrencies based on the frequency of Google searches (Urquhart, 2018). This index is freely available in the form of open data and can be obtained on a weekly basis. It was therefore assumed that the score remains constant for each day of the week.

As already mentioned in Section 2, the similarity between cryptocurrencies and commodities has been a debated and still open issue in the literature (Ankenbrand & Bieri, 2018). For this reason, we included in the dataset numerous features belonging to the commodities class by collecting their 1-month futures prices.

As regards features relating to stock markets, differently from what was done in other works, we considered a selection of the most representative indices at a global level, as we observe that the phenomenon under investigation manifests itself with a wide dimension (Baur et al., 2018). In addition, we included also the VIX volatility indicator (Chen, Liu, & Zhao, 2020; Dutta et al., 2020).

Since both stocks and futures markets are not open 24/7, we assumed that the last available price remains constant.

Consistently with what was done for the equity sector, we also included currencies prices (Baur et al., 2018; Chen et al., 2020). We excluded the US dollar as it is the reference currency of all the assets considered in this work.

4. Models and methods

In our study, we used random forest regression models (RF) due to their effectiveness. As part of the ensemble learning family, RF models are designed to handle complex datasets by minimizing prediction error through averaging the outputs of multiple decision trees. This approach not only captures intricate patterns in the data but also enhances model robustness and interpretability, making RF a valuable tool for both researchers and practitioners (Breiman, 2001).

More precisely, a Random Forest (RF) predictor is an ensemble of individual decision tree predictors that averages the results of each tree. Using the bagging technique and random feature selection, each tree is trained with a subset of the data and fea- tures, reducing overfitting and enhancing model performance. RF models are popular for their applicability to a wide range of prediction problems with minimal parameter tuning (Breiman, 2001).

Given a set (forest) F of decision tree regression predictors (f_{DT}), each trained on a bootstrap sample $d \in D$ of the training dataset D and an input vector x of features, we can write an RF regression predictor output $\hat{y}$ as:

$$\hat{y} = f_{RF}(x, D) = \frac{1}{|F|} \sum_{f_{DT} \in F} f_{DT}(x, D)$$

One of the key advantages of the Random Forest (RF) regression model is its interpretability through the use of feature importance. Feature importance provides a quantitative measure of the influence each input variable has on the model's predictions. In RF, this is typically calculated by assessing the average decrease in impurity or by permutation importance, which evaluates the increase in prediction error when the values of a feature are permuted. This interpretability helps understand the under- lying relationships between predictors and the target variable, facilitating insights into the factors that most significantly impact the model's output (Liaw, Wiener, et al., 2002). Consequently, feature importance not only aids in model validation but also enhances the decision-making process by highlighting key drivers of Bitcoin returns.

Additionally, we incorporated two control models: a linear regression model and an SVR (Support Vector Regression) model. These models serve as benchmarks to interpret the relationships between cryptocurrencies and the other financial assets discussed in Section 3. While we believe that the RF model is the most relevant to satisfy our needs, the control models help ensure robustness and validate our findings.

A linear regression model predicts the output $\hat{y}$ as a linear combination of input features. It is expressed as:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

where β_0 is the intercept, β_i are the coefficients of the features x_i, and ε is the error term (Seber & Lee, 2012).

A SVR model aims to find a function that approximates the data within a certain margin of tolerance (ε). For this execution, we set $\varepsilon = 0.01$ and use the Radial basis function (RBF) kernel, which is effective for capturing nonlinear relationships between cryptocurrencies and financial assets. It is expressed as:

$$\hat{y} = f_{SVR}(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x) + b$$

where $K(x_i, x) = e^{-\gamma \|x_i - x\|^2}$, α_i, α_i^* are Lagrange multipliers, and b is the bias term. The RBF kernel is chosen for its flexibility in modeling complex interactions in the data (Smola & Scholkopf, 2004).

To provide a comprehensive understanding of the performance of each model, we provide a brief description of the metrics that we used to compare the predicted values $\hat{y}$ with the actual ones y:

- *Explained Variance Score* measures the proportion of the target variable's variance Var that is accounted for by the model. It is calculated as:

$$EVScore = 1 - \frac{Var(y - \hat{y})}{Var(y)}$$

higher values indicate better performance.

- *Max Error* is a metric that captures the largest absolute error in the model's predictions, highlighting the worst-case prediction scenario. It is defined as:

$$MaxError = \max(|y_i - \hat{y}_i|)\,, \forall i$$

- *Mean Absolute Error* represents the average absolute difference between the predicted values and the actual values. It is computed as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where n is the number of observations. MAE provides a straightforward measure of prediction accuracy.

- *Mean Squared Error* is calculated as the average of the squared differences between predicted and actual values:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

This metric gives more weight to larger errors due to the squaring process.

- *R2 Score* assesses the proportion of the variance in the dependent variable that is predictable from the independent variables. It is computed as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

where $\bar{y}$ is the mean of the actual values. An R2 score closer to 1 indicates a model that better explains the variance of the target variable.

While metrics and feature importance provide insights into model performance, *Partial Dependence Plots* (PDPs) are needed to obtain a deeper understanding of the relationships between features and Bitcoin returns (Molnar et al., 2023). PDPs illustrate these relationships by averaging out the effects of other features, making complex models more interpretable. A PDP shows the marginal effect of a feature on the predicted outcome and can be calculated as follows:

$$f_P(x_j) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(x_j, x_{i-j})$$

where $\hat{f}$ is the prediction function, x_j is the feature of interest, and x_{i-j} represents all other features.

This visualization helps in understanding how changes in a specific feature influence the predictions while holding other features constant. For example, a positive relationship in a PDP would show that as the feature value increases, the predicted outcome also increases, indicating the feature positively contributes to the prediction. Conversely, a negative relationship would show that an increase in the feature value decreases the predicted outcome (Molnar et al., 2023; Petch, Di, & Nelson, 2022). PDPs are particularly useful in identifying nonlinear relationships and interactions between features. Moreover, It is important to note that PDPs assume independece among the features, thus we need to verify that a low level of correlation holds between the features.
For further details about models and metrics we refer the reader to Alpaydin (2020).

5. Results

This section presents a detailed analysis of the results obtained from the ML models. The performance metrics of the RF model, as well as those of the control algorithms - Linear Regression (Linear) and SVR - across different holding periods (1 day, 7 days, 15 days, and 30 days), are summarized in Table 3.

Table 3: ML Model Performance Across Different Holding Periods

Metric	**Model**	**1 Day**	**7 Days**	**15 Days**	**30 Days**
Explained Variance Score	RF	0.7343	0.9325	0.9613	0.9820
	Linear	0.1285	0.1923	0.2616	0.3835
	SVR	0.4618	0.7669	0.8288	0.8968
Max Error	RF	0.1472	0.1294	0.2399	0.1980
	Linear	0.3328	0.3942	0.5059	0.6559
	SVR	0.1582	0.3834	0.4139	0.5172
Mean Absolute Error	RF	0.0121	0.0183	0.0201	0.0200
	Linear	0.0234	0.0659	0.0962	0.1296
	SVR	0.0153	0.0278	0.0360	0.0405
Mean Squared Error	RF	0.0004	0.0007	0.0008	0.0008
	Linear	0.0012	0.0080	0.0161	0.0280
	SVR	0.0007	0.0023	0.0037	0.0047
R2 Score	RF	0.7343	0.9325	0.9613	0.9820
	Linear	0.1285	0.1923	0.2616	0.3835
	SVR	0.4618	0.7668	0.8288	0.8968

The technical analysis of the results shows that the RF model consistently outperforms both the Linear and SVR models across all holding periods based on the Explained Variance Score (EVS). It achieves scores of 0.7343 (1-day), 0.9325 (7-days), 0.9613 (15-days), and 0.9820 (30-days), indicating superior ability to explain variance in cryptocurrency returns. It also demonstrates the lowest Max Error, MAE, MSE values, and highest R2 scores among the models, showcasing robust performance in predicting cryptocurrency dynamics. Thus, we can conclude that the RF model's ensemble learning approach effectively handles the nonlinearities and interactions inherent in cryptocurrency data (Khedr, Arif, El-Bannany, Alhashmi, & Sreedharan, 2021).
Figures from 1 to 4 presents the significance of various features in analysing cryptocurrency returns over different holding periods (1 day, 7 days, 15 days, and 30 days). Feature importance in RF models highlights each variable's influence on predictions, revealing key factors affecting cryptocurrency returns (see Section 4).

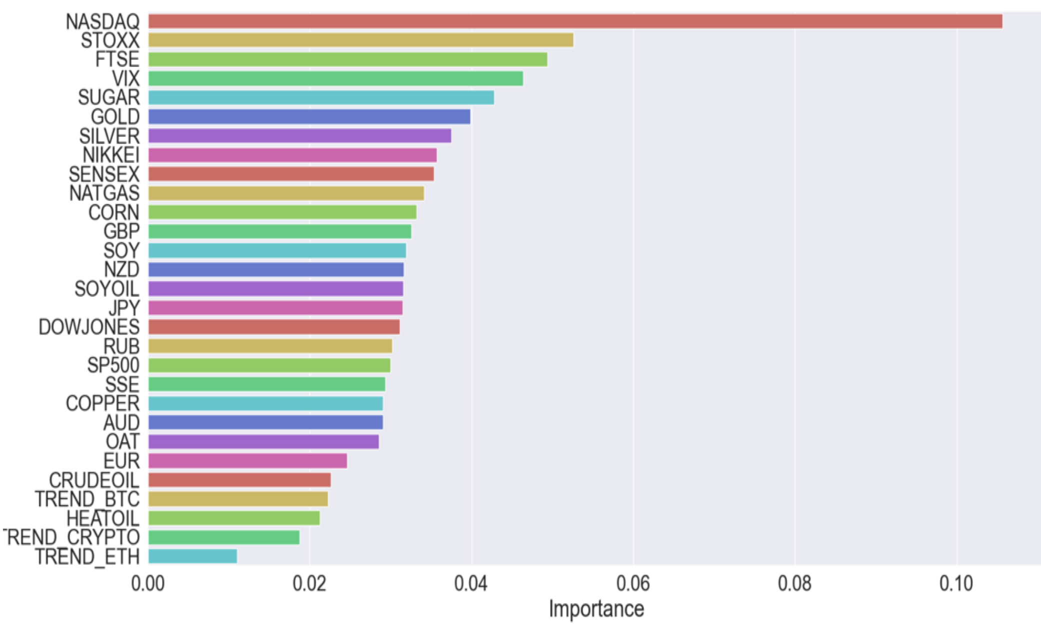

Figure 1: Feature importance of RF model with 1-day holding period

For the 1-day holding period (Figure 1), the NASDAQ index emerges as the most important feature, suggesting that short-term cryptocurrency returns are highly influenced by the performance of technology stocks, likely due to the tech-centric nature of the cryptocurrency market (Umar, Hung, Chen, Iqbal, & Jebran, 2020). Additionally, European stock indices such as STOXX and FTSE also show significant importance, indicating that global equity markets may play a role in shaping short-term cryptocurrency price movements with reference of the analysed holdind period (Aliu, Nuhiu, Krasniqi, & Jusufi, 2021). The VIX index, which measures market uncertainty/fear, is another important feature, reflecting the sensitivity of short-term cryptocurrency returns to market uncertainty (Ghorbel & Jeribi, 2021). The notable absentee for the 1-day holding period is Google trends for general cryptocurrency, Bitcoin, and Ethereum searches (TREND_CRYPTO, TREND_BTC, and TREND_ETH). As reported in Section 3, this is explained by the weekly basis availability of trends (Almeida & Goncalves, 2023).

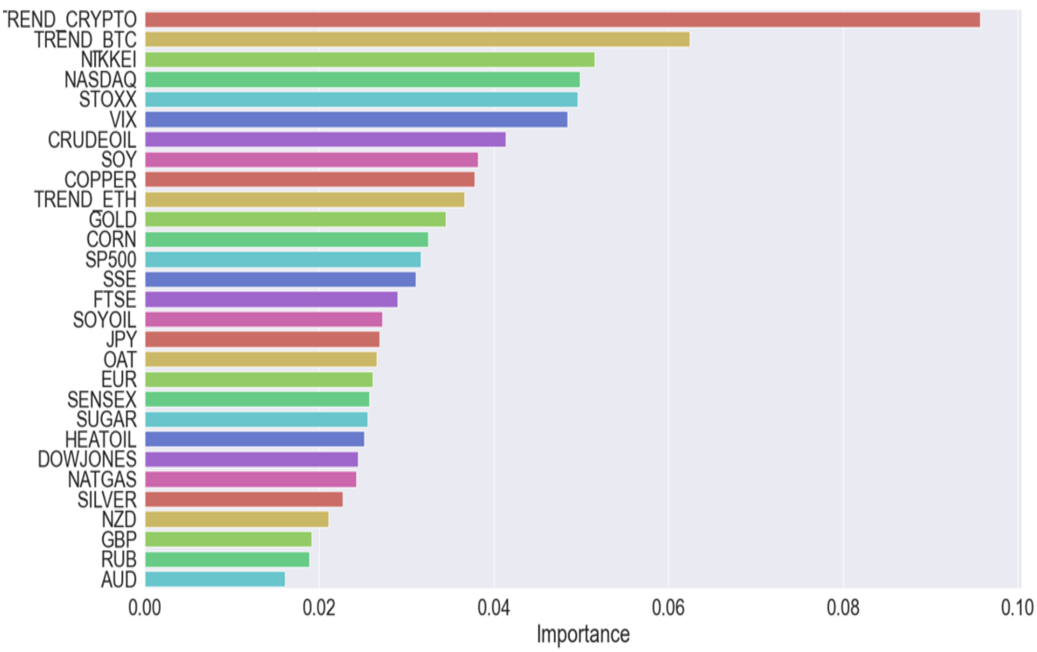

Figure 2: Feature importance of RF model with 7-days holding period

As far as the 7-days holding period (Figure 2) is concerned, TREND_CRYPTO becomes, as expected, the most important feature, indicating that market sentiment and public interest over a week have a significant impact on cryptocurrency returns. The continued importance of TREND_BTC and TREND_ETH suggests that specific interest in the corresponding cryptocurrencies drives weekly returns (Almeida & Gon¸calves, 2023; Bianchi, 2020).

The relevance of NASDAQ and STOXX indices remains high (Umar et al., 2020) while the importance of CRUDEOIL and GOLD, suggesting that both markets still mantain an inefficient component able to engage with cryptos dynamics (Mensi, Rehman, & Vo, 2020; Salisu, Akanni, & Raheem, 2020).

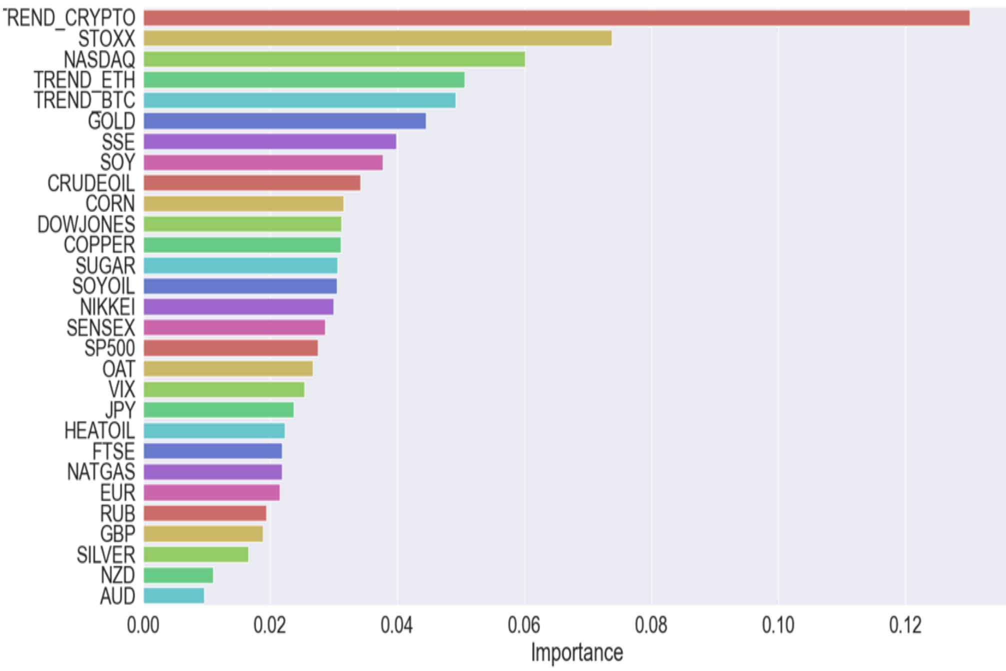

Figure 3: Feature importance of RF model with 15-days holding period

For the 15-days holding period (Figure 3), TREND_CRYPTO maintains its top position, underscoring the persistent impact of market sentiment over medium-term periods. The consistent importance of NASDAQ and STOXX indices highlights the correlation between cryptocurrency returns and broader financial markets. The specific interest in Ethereum and Bitcoin continues to be relevant, indicating that trends in these major cryptocurrencies are critical for medium-term predictions. GOLD and CRUDEOIL maintain their high importance, suggesting their influence on cryptocurrencies' returns (Ozturk, 2020).

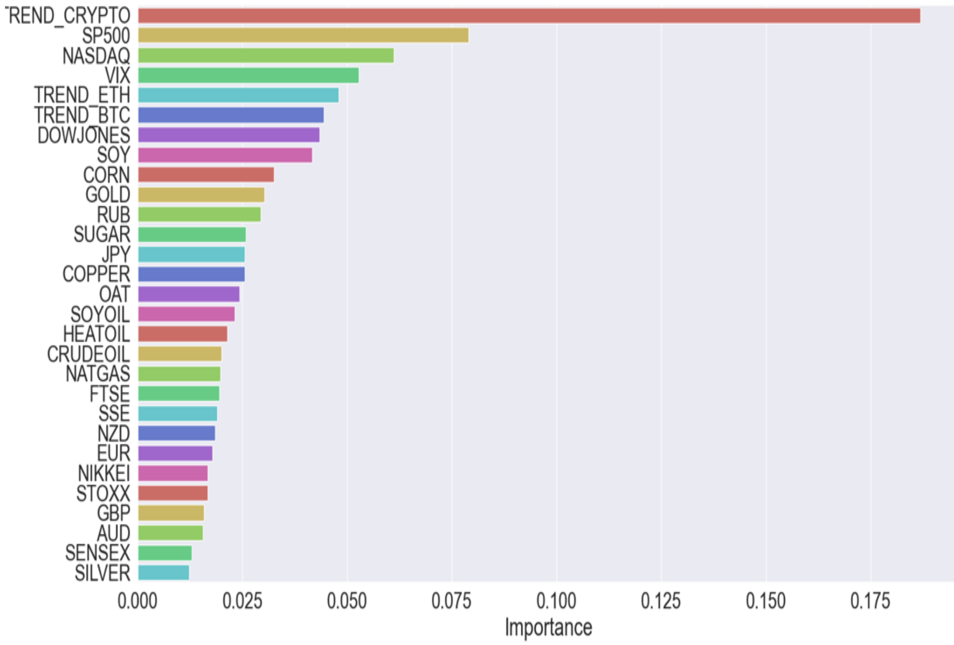

Figure 4: Feature importance of RF model with 30-days holding period

For the 30-days holding period (Figure 4), TREND_CRYPTO continues to dominate, emphasizing a stable and prominent long-term effect of sustained public interest on cryptocurrency returns. The significance of SP500 and NASDAQ indices peaks, indicating that long-term cryptocurrency returns are aligned with overall market performance and suggesting that, despite the lack of symmetry between positive and negative price changes documented in the literature, cryptocurrencies returns are affected bu tradi- tional stock market dynamics (Umar et al., 2020). The continued importance of the VIX highlights the impact of market uncertainty on long-term returns, that is also reflected by the fact that commodities indices such GOLD and CRUDEOIL remain significant also in this holding period. To better observe what is highlighted in Figures from 1 to 4, we report in Figure 5 in the form of a heatmap the trend of the importance of a subset of features as the holding period varies.

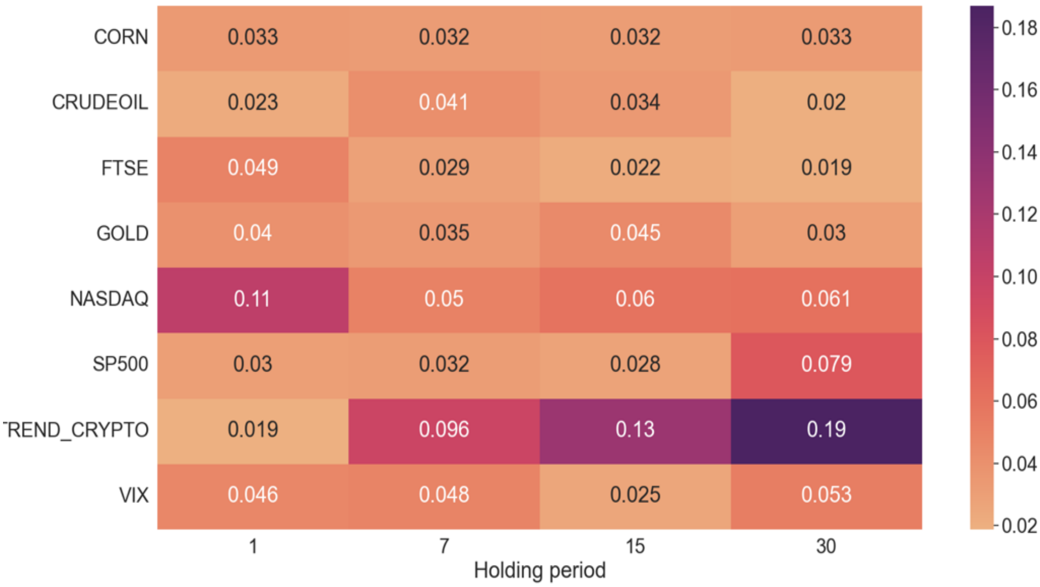

Figure 5: Feature importance comparison for different holding periods

The importance of TREND_CRYPTO, exhibits a notable grow across varying holding periods from short to long term except in the first holding period for the reasons already explained. This trend suggests that public interest and market sentiment, as gauged through social media and search trends, play a critical and enduring role in shaping cryptocurrency market dynamics (Liu & Tsyvinski, 2021). The consistent high importance highlights the need for continuous monitoring of social media and search trends by investors to assess market sentiment and anticipate price movements effectively.

The importance of the SP500 steadily increases across different holding peri- ods, emphasizing its growing influence over longer investment holding periods. For investors, this underscores the importance of monitoring major stock indices like the SP500 when making long-term investment decisions in cryptocurrencies (Aliu et al., 2021; Umar et al., 2020).

Gold's importance remains stable across different holding periods, with a slight increase noted for the 15-days period. As a traditional safe-haven asset, gold plays a consistent role in influencing investor behavior, reflecting broader economic trends and sentiment towards risk (Selmi, Mensi, Hammoudeh, & Bouoiyour, 2018; Tarchella, Khalfaoui, & Hammoudeh, 2024).

The VIX stays notable across different holding periods. Initially, it holds moderate significance for the 1-day period, which then increases notably for the 7-days period before stabilizing thereafter. This trend underscores the pivotal role of market volatility in influencing cryptocurrency returns, particularly over shorter timeframes, that are usually more sensitive (Ghorbel & Jeribi, 2021). For investors, monitoring the VIX may offer insights into market risk dynamics, aiding in the anticipation of potential price fluctuations in tokens' returns.

The influence of CORN on cryptocurrency returns remains moderate and consistent across varying holding periods, showing only negligible fluctuations without a clear trend. This stability indicates that corn prices may consistently affect cryptocurrency performance only in a moderate magnitude (Jareno, Gonzalez, & Belmonte, 2022).

The influence of CRUDEOIL futures peaks significantly for the 7-days and 15-days periods but shows a slight decrease by the 30-days mark. This trend underscores the potential impact of industrial commodity prices on short to medium-term cryptocurrency performance. Investors can leverage this information to gain a more comprehensive understanding of short to medium-term market movements (Okorie & Lin, 2020).

Figures 6 to 9 displays Partial Dependence Plots (PDPs) (See Section 4) for four critical features: TREND_CRYPTO, GOLD, SP500, and VIX, in holding periods 1-day, 7-days, 15-days, and 30-days. PDPs depict the isolated effect of each feature on outcomes while keeping all other variables constant. This analysis offers complementary information to that of feature importance, allowing in particular to measure not only the magnitude of the importance but also the sign of the relationship.

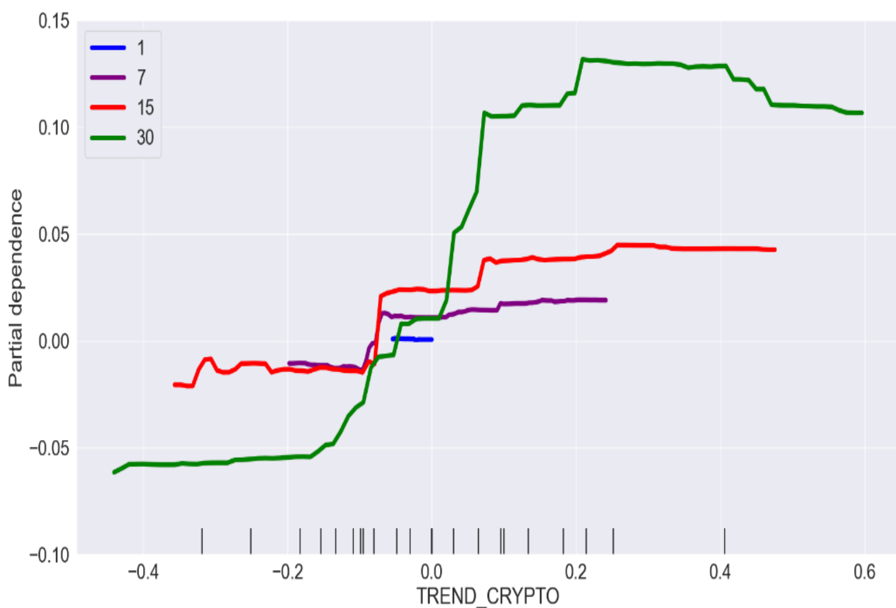

Figure 6: PDP for TREND_CRYPTO in different holding periods

Figure 6 illustrates the relationship between TREND_CRYPTO and Bitcoin returns across different holding periods. For the 1-day holding period, the PDP shows a positive but relatively flat relationship, likely due to the weekly availability of this feature, indicating anticipation of public interest. Over the 7-days holding period, the influence on returns becomes more positive. At 15 days, the relationship strengthens, with public interest having a cumulative effect on market performance. By the 30-days holding period, the trend peaks, showing the maximum impact of public interest over a month. The relationship remains positive, highlighting the lasting effect of market sentiment on long-term returns. In the short term, spikes in public interest can lead to modest price increases, offering opportunities for quick gains. For medium to long-term strategies, maintaining high levels of public interest is crucial (Burggraf, Huynh, Rudolf, & Wang, 2021; Da, Engelberg, & Gao, 2015; Jung, Lee, Lee, & Kim, 2023; Kristoufek, 2013).

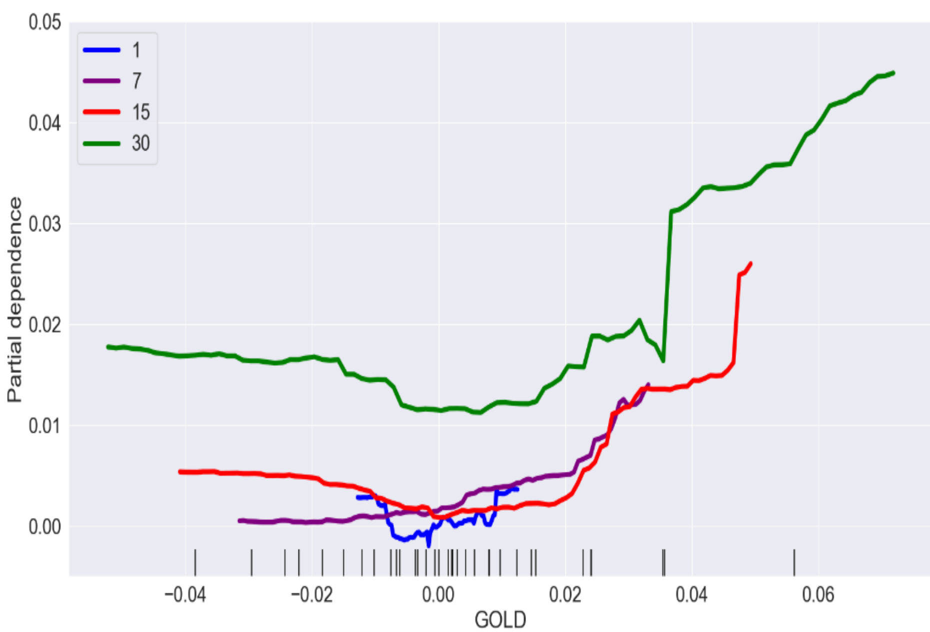

Figure 7: PDP for GOLD in different holding periods

Figure 7 explores the relationship between Gold Futures and Bitcoin returns. Over a 1-day holding period, the PDP shows a slight positive relationship, indicating that changes in gold prices have a modest impact on returns. Moving to the 7-days holding period, the relationship becomes more positive, suggesting that weekly fluctuations in gold prices start to more noticeably influence returns. Over the 30-days holding period, the PDP shows a clear and consistent positive trend. Higher gold prices consistently lead to higher BTC returns. When gold prices rise, signaling potential economic uncertainty or inflationary pressures, investors may turn to BTC as alternative investments (Selmi et al., 2018; Tarchella et al., 2024).

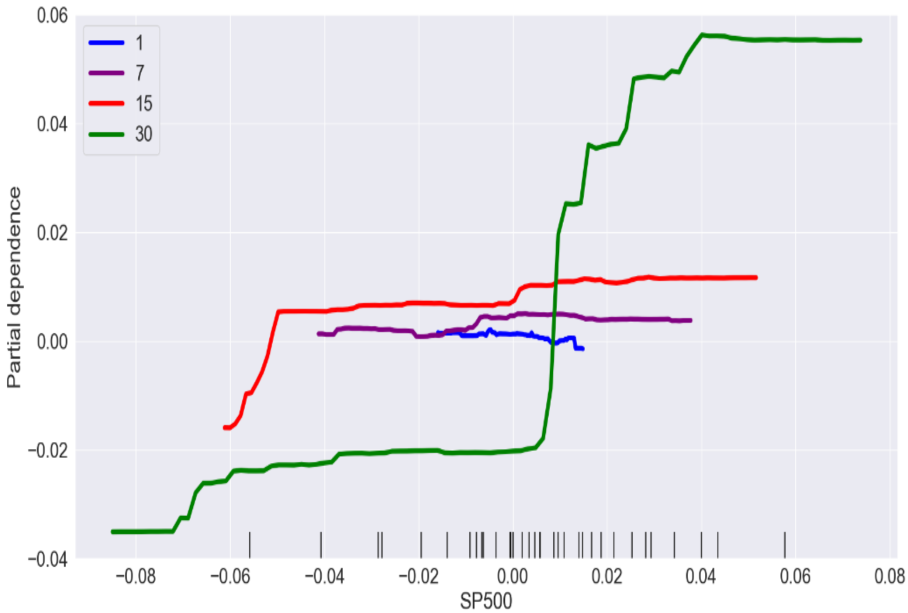

Figure 8: PDP for SP500 in different holding periods

Figure 8 examines the relationship between the SP500 and Bitcoin returns. The 1-day holding period shows a slightly negative relationship, indicating that higher SP500 values might lead to lower short-term cryptocurrency returns. This relationship turns slightly positive over the 7-days and 15-days holding periods, suggesting that medium-term increases in the SP500 are associated with positive variations in Bitcoin returns. Up to this holding period, the relationship is ambiguous and moderate (Zeng, Yang, & Shen, 2020). However, the PDP for the 30-days holding period, consistent with the feature importance (See Figure 4), shows a stable positive trend, indicating that over a month, higher SP500 values are associated with higher returns. The shift from an ambiguous to a positive relationship between the SP500 and BTC returns suggests that while short-term movements provide little insight, over longer periods, a thriving equity market complements the token's performance. When constructing diversified portfolios, investors should consider this signal of market integration (Lahiani, Jlassi, et al., 2021).

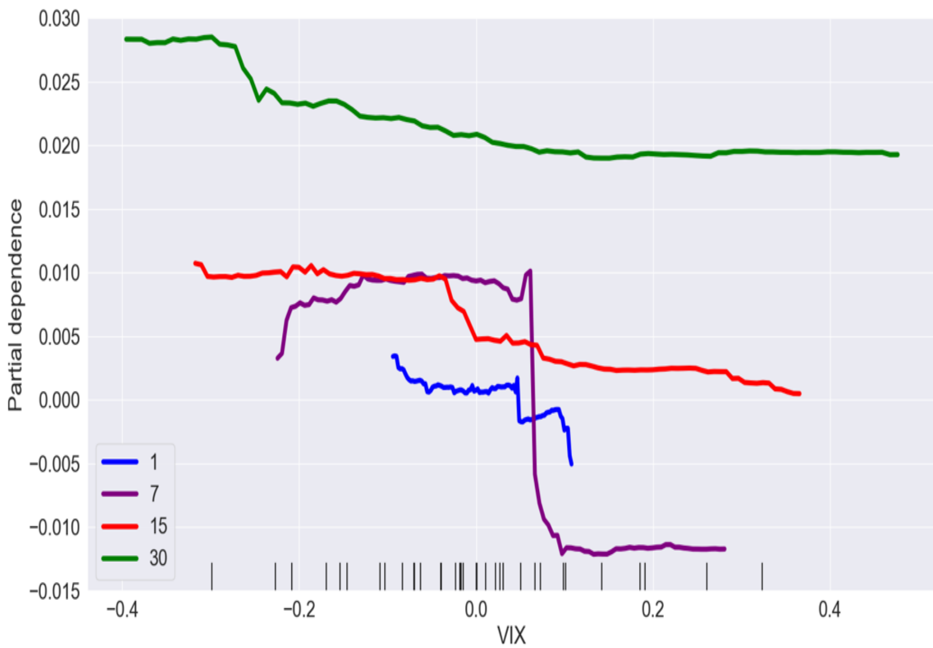

Figure 9: PDP for VIX in different holding periods

Figure 9 explores the relationship between the VIX and Bitcoin returns. For the 1-day holding period, the relationship is relatively flat with a slight downward trend. This indicates that, in the very short term, an increase in market volatility does not significantly impact Bitcoin returns, suggesting that BTC might not be immediately sensitive to daily fluctuations (Bouri, Molnar, Azzi, Roubaud, & Hagfors, 2017). Moving to the 7-days holding period, the relationship becomes more pronounced, with a notable drop in returns as the VIX rises. This suggests that over a week, higher market volatility tends to negatively impact returns, reflecting investors' risk aversion during volatile periods. For the 15-days holding period, the PDP continues to show a negative relationship, although it is less pronounced than the 7-days period. The partial dependence steadily decreases as the VIX increases, indicating that bi-weekly market volatility still negatively affects cryptocurrency returns, but the effect is some- what stabilized compared to the shorter period. For the 30-days holding period, the relationship is slightly positive but overall relatively flat. This suggests that over a month, the impact of market volatility on cryptocurrency returns diminishes. The positive slope at higher VIX levels might indicate that long-term investors are less concerned with short-term volatility and might even see it as an opportunity, leading to a stabilization or slight increase in returns (Da et al., 2015).

6. Conclusions

In this study, we explore the relationships between Bitcoin returns and a selected set of features, aiming to understand cryptocurrencies' nature and behaviors using Bitcoin as a market benchmark. Adopting a Darwinian approach, we examine whether cryptocurrencies, as significant literature suggests, represent a distinct asset class, effectively identifying a new species.

We innovate this research field by analyzing a novel dataset spanning from 2018 to 2023, confirming the Random Forest model's superiority in revealing the informative content of each selected feature concerning return distributions. The robust and consistent relationships identified by the model are further explored through Partial dependece plots, providing investors with precise strategic insights for portfolio com- position decisions across 1-day, 7-days, 15-days, and 30-days holding periods. The RF model's effectiveness is high and improves with longer holding periods.

For investments with holding periods less than 7 days, the significant importance of the NASDAQ index and other global equity indices (S&P500 and FTSE) suggests that short-term traders should closely monitor the performance of the stock market, especially in the technology sector. Market uncertainty/fear, as measured by VIX, is equally crucial, indicating that short-term trading strategies should consider broader market uncertainty. The relevance of commodity prices (gold, crude oil, and corn) reflects their role in shaping investor behavior and market dynamics, suggesting that short-term investors should incorporate these indicators into their analysis.

For investments with holding periods greater than 7 days up to 30 days, The Google trend assumes a central role, emphasizing the enduring impact of public interest on long-term cryptocurrency returns. The significance of VIX indices (to a lesser extent for the 15-days holding period), S&P500 (more intensely for the 30-days holding period), NASDAQ, and (with decreasing relevance) FTSE suggests that long-term investors should integrate traditional market performance into their analysis, as cryptocurrencies appear increasingly correlated with overall market trends. The ongoing relevance of commodities prices indicates their influence as long-term economic indicators, suggesting that investors should monitor them as part of their broader economic analysis.

The results suggest that cryptocurrencies are increasingly identifying as separate species compared to other asset classes, yet maintain precise and consistent relationships with traditional markets and market sentiment. As a result of these relationships, cryptocurrency returns appear to benefit parasitically from investor sentiment (fear, euphoria, confidence, etc.) and from the value created or destroyed by traditional markets.

Investment strategy in cryptocurrencies presents unique challenges compared to traditional assets, particularly due to their distinct risk and volatility characteristics and momentum dynamics. For that this information becomes relevant for investors who can adjust their strategies based on the selected holding period, using signals provided by features to integrate cryptocurrencies into their portfolios in an informed and coherent manner aligned with their objectives in terms of both returns and diversification. These findings will be further developed to define a precise portfolio composition model.

ACKNOWLEDGEMENTS
The authors affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

References

[1] Alfieri, E., Burlacu, R., Enjolras, G. (2019). On the nature and financial performance of bitcoin. The Journal of Risk Finance.
[2] Aliu, F., Nuhiu, A., Krasniqi, B.A., Jusufi, G. (2021). Modeling the optimal diversification opportunities: the case of crypto portfolios and equity portfolios. Studies in Economics and Finance, 38(1), 50–66.
[3] Almeida, J., & Goncalves, T.C. (2023). A systematic literature review of investor behavior in the cryptocurrency markets. Journal of Behavioral and Experimental Finance, 37, 100785.
[4] Alpaydin, E. (2020). Introduction to machine learning. MIT press.
[5] Ankenbrand, T., & Bieri, D. (2018). Assessment of cryptocurrencies as an asset class by their characteristics. Investment Management & Financial Innovations, 15(3), 169.
[6] Augustin, P., Rubtsov, A., Shin, D. (2020). The impact of derivatives on cash markets: Evidence from the introduction of bitcoin futures contracts. Available at SSRN.
[7] Baur, D.G., Hong, K., Lee, A.D. (2018). Bitcoin: Medium of exchange or speculative assets? Journal of International Financial Markets, Institutions and Money, 54, 177–189.
[8] Bianchi, D. (2020). Cryptocurrencies as an asset class? an empirical assessment. The Journal of Alternative Investments, 23(2), 162–179.
[9] Bouri, E., Molnar, P., Azzi, G., Roubaud, D., Hagfors, L.I. (2017). On the hedge and safe haven properties of bitcoin: Is it really more than a diversifier? Finance Research Letters, 20, 192–198.
[10] Brauneis, A., Mestel, R., Riordan, R., Theissen, E. (2021). How to measure the liquidity of cryptocurrency markets? Journal of Banking & Finance, 124, 106041.
[11] Breiman, L. (2001). Random forests. Machine Learning , 45 (1), 5–32, https://doi.org/10.1023/a:1010933404324.
[12] Burggraf, T., Huynh, T.L.D., Rudolf, M., Wang, M. (2021). Do fears drive bitcoin? Review of Behavioral Finance, 13(3), 229–258.
[13] Buterin, V., et al. (2014). A next-generation smart contract and decentralized application platform. white paper, 3(37), 2–1.

[14] Chen, C., Liu, L., Zhao, N. (2020). Fear sentiment, uncertainty, and bitcoin price dynamics: The case of covid-19. Emerging Markets Finance and Trade, 56(10), 2298–2309.
[15] CoinMarketCap (2024). Cryptocurrency prices, charts and data. Retrieved from https://coinmarketcap.com/ (Accessed: 2024-06-30).
[16] Da, Z., Engelberg, J., Gao, P. (2015). The sum of all fears investor sentiment and asset prices. The Review of Financial Studies, 28(1), 1–32.
[17] Davison, A. (2020). A darwinian approach to postmodern critical theory: Or, how did bad ideas colonise the academy? Society, 57(4), 417–424.
[18] Dutta, A., Kumar, S., Basu, M. (2020). A gated recurrent unit approach to bitcoin price prediction. Journal of Risk and Financial Management, 13(2), 23.
[19] ElBahrawy, A., Alessandretti, L., Kandler, A., Pastor-Satorras, R., Baronchelli, A. (2017). Evolutionary dynamics of the cryptocurrency market. Royal Society open science, 4(11), 170623.
[20] Fang, F., Ventre, C., Basios, M., Kanthan, L., Martinez-Rego, D., Wu, F., Li, L. (2022). Cryptocurrency trading: a comprehensive survey. Financial Innovation, 8(1), 1–59.
[21] Feng, W., Wang, Y., Zhang, Z. (2018). Can cryptocurrencies be a safe haven: a tail risk perspective analysis. Applied Economics, 50(44), 4745–4762.
[22] Figuet, J.-M. (2016). Bitcoin et blockchain: quelles opportunit´es? Revue d'´economie financiere(3), 325–338.
[23] Franco, P. (2014). Understanding bitcoin: Cryptography, engineering and economics. John Wiley & Sons.
[24] Ghorbel, A., & Jeribi, A. (2021). Investigating the relationship between volatili- ties of cryptocurrencies and other financial assets. Decisions in Economics and Finance, 44(2), 817–843.
[25] Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M.C., Siering, M. (2014). Bitcoin- asset or currency? revealing users' hidden intentions. Revealing Users' Hidden Intentions (April 15, 2014). ECIS.
[26] Howell, S.T., Niessner, M., Yermack, D. (2020). Initial coin offerings: Financing growth with cryptocurrency token sales. The Review of Financial Studies, 33(9), 3925–3974.
[27] Jareno, F., Gonzalez, M., Belmonte, P. (2022). Asymmetric interdependencies between cryptocurrency and commodity markets: The covid-19 pandemic impact. Quant. Financ. Econ, 6, 83–112.
[28] Jung, H.S., Lee, S.H., Lee, H., Kim, J.H. (2023). Predicting bitcoin trends through machine learning using sentiment analysis with technical indicators. Computer Systems Science & Engineering, 46(2).
[29] Khedr, A.M., Arif, I., El-Bannany, M., Alhashmi, S.M., Sreedharan, M. (2021). Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. Intelligent Systems in Accounting, Finance and Management, 28(1), 3–34.

[30] Kristoufek, L. (2013). Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. Scientific reports, 3(1), 3415.

[31] Lahiani, A., Jlassi, N.B., et al. (2021). Nonlinear tail dependence in cryptocurrency-stock market returns: The role of bitcoin futures. Research in International Business and Finance, 56, 101351.

[32] Liang, J., Li, L., Chen, W., Zeng, D. (2019). Towards an understanding of cryptocurrency: a comparative analysis of cryptocurrency, foreign exchange, and stock. 2019 ieee international conference on intelligence and security informatics (isi) (pp. 137–139).

[33] Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomforest. R news, 2(3), 18–22.

[34] Liu, Y., & Tsyvinski, A. (2021). Risks and returns of cryptocurrency. The Review of Financial Studies, 34(6), 2689–2727.

[35] Mensi, W., Rehman, M.U., Vo, X.V. (2020). Spillovers and co-movements between precious metals and energy markets: Implications on portfolio management. Resources Policy, 69, 101836.

[36] Molnar, C., Freiesleben, T., König, G., Herbinger, J., Reisinger, T., Casalicchio, G., . . . Bischl, B. (2023). Relating the partial dependence plot and permutation feature importance to the data generating process. World conference on explainable artificial intelligence (pp. 456–479).

[37] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Decentralized business review, 21260.

[38] Nunez, J.A., Contreras-Valdez, M.I., Franco-Ruiz, C.A. (2019). Statistical analysis of bitcoin during explosive behavior periods. PloS one, 14(3), e0213919.

[39] Okorie, D.I., & Lin, B. (2020). Crude oil price and cryptocurrencies: evidence of volatility connectedness and hedging strategy. Energy economics, 87, 104703.

[40] Ozturk, S.S. (2020). Dynamic connectedness between bitcoin, gold, and crude oil volatilities and returns. Journal of Risk and Financial Management, 13 (11), 275.

[41] Panagiotidis, T., Papapanagiotou, G., Stengos, T. (2022). On the volatility of cryptocurrencies. Research in International Business and Finance, 62, 101724.

[42] Pele, D.T., Wesselhofft, N., Hardle, W.K., Kolossiatis, M., Yatracos, Y.G. (2023). Are cryptos becoming alternative assets? The European Journal of Finance, 29(10), 1064–1105.

[43] Peng, Y., Albuquerque, P.H.M., de Sa, J.M.C., Padula, A.J.A., Montenegro, M.R. (2018). The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression. Expert Systems with Applications, 97, 177–192.

[44] Petch, J., Di, S., Nelson, W. (2022). Opening the black box: the promise and lim- itations of explainable machine learning in cardiology. Canadian Journal of Cardiology, 38(2), 204–213.

[45] Phillip, A., Chan, J., Peiris, S. (2019). On long memory effects in the volatility measure of cryptocurrencies. Finance Research Letters, 28, 95-100.
[46] Prentis, M. (2015). Digital metal: Regulating bitcoin as a commodity. Case W. Res. L. Rev., 66, 609.
[47] Salisu, A.A., Akanni, L., Raheem, I. (2020). The covid-19 global fear index and the predictability of commodity price returns. Journal of behavioral and experimental finance, 27, 100383.
[48] Seber, G.A., & Lee, A.J. (2012). Linear regression analysis. John Wiley & Sons.
[49] Selmi, R., Mensi, W., Hammoudeh, S., Bouoiyour, J. (2018). Is bitcoin a hedge, a safe haven or a diversifier for oil price movements? a comparison with gold. Energy Economics, 74, 787–801.
[50] Smola, A.J., & Scholkopf, B. (2004). A tutorial on support vector regression. Statistics and computing, 14, 199–222.
[51] Tandon, C., Revankar, S., Parihar, S.S. (2021). How can we predict the impact of the social media messages on the value of cryptocurrency? insights from big data analytics. International Journal of Information Management Data Insights, 1(2), 100035.
[52] Tarchella, S., Khalfaoui, R., Hammoudeh, S. (2024). The safe haven, hedging, and diversification properties of oil, gold, and cryptocurrency for the g7 equity markets: Evidence from the pre-and post-covid-19 periods. Research in international business and finance, 67, 102125.
[53] Umar, M., Hung, N.T., Chen, S., Iqbal, A., Jebran, K. (2020). Are stock markets and cryptocurrencies connected? The Singapore Economic Review, 1–16.
[54] Urquhart, A. (2018). What causes the attention of bitcoin? Economics Letters, 166, 40–44.
[55] Yermack, D. (2015). Is bitcoin a real currency? an economic appraisal. Handbook of digital currency (pp. 31–43). Elsevier.
[56] Zeng, T., Yang, M., Shen, Y. (2020). Fancy bitcoin and conventional financial assets: Measuring market integration based on connectedness networks. Economic Modelling, 90, 209–220.