

Previsibilidade da Direção do Preço Intradiário do Bitcoin com Modelos de *Random Forest*

Paulo de Castro Rubio Poli, Leandro dos Santos Maciel

School of Economics, Business, Accounting and Actuary, University of Sao Paulo, Brazil

E-mails: paulo.poli@usp.br, leandromaciel@usp.br

RESUMO

O mercado de negociação do Bitcoin apresentou um acelerado crescimento a partir do ano de 2019 e atraiu a atenção de investidores individuais e institucionais. Dada a elevada volatilidade das cotações da moeda digital, previsões acuradas sobre a direção futura de seus preços são de grande importância para os participantes deste mercado. A dinâmica complexa dos preços da criptomoeda demanda o uso de técnicas sofisticadas de aprendizado de máquinas para a realização de previsões. Neste contexto, este trabalho objetivou avaliar a previsibilidade da direção do preço intradiário do Bitcoin durante o período de janeiro de 2020 a dezembro de 2022. Para este propósito, modelos de *random forest* foram utilizados para prever se a criptomoeda seria valorizada ou desvalorizada nos horizontes de previsão de um minuto, cinco minutos, quinze minutos, uma hora, seis horas e um dia. As variáveis explicativas se referem a defasagens de retornos de preços de fechamento, máximos e mínimos da própria criptomoeda. As previsões da moeda digital por modelos de *random forest* foram comparadas com as previsões dos modelos ARIMA e de regressão logística, em termos de medidas de poder preditivo. Os resultados encontrados apontam violações na verificação empírica da hipótese de eficiência de mercado em sua forma fraca para os preços intradiários do Bitcoin. Os movimentos intradiários da moeda puderam ser previstos por modelos de *random forest* com acurácia superior a observada por um modelo de passeio aleatório e pelos modelos competidores para diferentes horizontes de previsão.

Palavras-chave: Bitcoin. Previsão. Dados intradiários. *Random Forest*.

1. Introdução

O mercado de criptomoedas passou por um vertiginoso crescimento nos últimos anos. De acordo com a CoinMarketCap¹, a capitalização total do mercado passou de US\$ 1,3 bilhão, em 01 de maio de 2013, atingindo US\$ 3,0 trilhões em 10 de novembro de 2021, um aumento de 2.237%. O

¹ Disponível em: < <https://coinmarketcap.com/>>. Acesso em: 30 jan. 2023.

Bitcoin, principal criptomoeda negociada neste mercado, com maior capitalização de mercado, obteve uma valorização particularmente expressiva: nessas mesmas datas, sua cotação passou de US\$ 105 para US\$ 68.798. Em 31 de dezembro de 2022, a cotação e o valor de mercado do Bitcoin já haviam recuado para US\$ 16.547 e US\$ 318,52 bilhões, respectivamente. Em um contexto de alta volatilidade de preços das criptomoedas, o desenvolvimento de um modelo de previsão acurado e confiável é muito importante para investidores e participantes do mercado (GYAMERAH, 2019).

A forma mais intuitiva de se abordar a tarefa de previsão em séries temporais financeiras é pela previsão numérica de seus preços, o que consiste em um problema de regressão. Entretanto, uma alternativa utilizada na literatura, e que traz alguns benefícios como resultados mais interpretáveis, por exemplo, é a predição do movimento, ou da direção, dos preços dos ativos. Em outras palavras, estrutura-se um problema de classificação, ou seja, uma tarefa de prever se um ativo irá valorizar ou desvalorizar em algum determinado horizonte futuro de previsão (LEUNG ET AL., 2000). Enke e Thawornwong (2005) afirmam preferir o problema de classificação ao problema de regressão, pois a literatura sugere que resultados mais acurados são observados quando utilizados dados de mercados financeiros para antecipação da direção de preços.

As características complexas e o considerável volume de dados de séries temporais financeiras levaram a adoção de métodos, modelos e técnicas de simulação mais sofisticados. Procedimentos de aprendizado de máquina, particularmente, são capazes de analisar grandes quantidades de dados, que apresentam ruídos, dinâmicas não lineares e variantes no tempo, com o objetivo de detectar padrões e/ou prever movimentos futuros (BORGES & NEVES, 2020). Analogamente à predição de preços de ações e de taxas de câmbio utilizando algoritmos de aprendizado de máquina, o preço de criptomoedas também pode ser previsto utilizando tais técnicas (GYAMERAH, 2019).

Hastie et al. (2009) afirmam, por exemplo, que modelos de florestas aleatórias (*Random Forest* - *RF*) conseguem modelar adequadamente dados ruidosos, são altamente robustos ao sobreajuste (*overfitting*²), e requerem menos parâmetros que modelos de reforço de gradiente (*Gradiente Boosting*³ - *GB*).

Dentre as aplicações do estudo da modelagem de séries temporais financeiras, o chamado *High-Frequency Trading* (HFT), ou negociação em alta frequência, permite que investidores obtenham consideráveis ganhos ao acumular pequenos ganhos em um alto número de transações em

² *Overfitting* refere-se a situações em que modelos de previsão/classificação se ajustam excessivamente aos dados utilizados na sua modelagem, apresentando desempenho inferior em demais amostras, ou seja, sem capacidade de generalização.

³ Modelos de *gradiente boosting* são uma técnica de aprendizado de máquina que combina diversos previsores de forma sequencial com o objetivo de melhorar a acurácia geral de qualquer algoritmo, ou seja, um mecanismo de combinação (*ensemble*).

curto espaço de tempo. O uso do HFT é especialmente atraente no contexto das criptomoedas devido à alta volatilidade intradiária (ALONSO ET AL., 2020).

Neste contexto, o objetivo deste trabalho é analisar a previsibilidade da direção do preço da criptomoeda mais negociada no mercado de moedas digitais, o Bitcoin, na frequência intradiária, nos horizontes de previsão de um minuto, cinco minutos, 15 minutos, 60 minutos (uma hora), 360 minutos (seis horas) e 1.440 minutos (um dia). Para tal tarefa, serão aplicados modelos de *random forest* utilizando como atributos (variáveis explicativas) defasagens dos retornos intradiários das séries. Os resultados são comparados em termos de poder preditivo (acurácia, sensibilidade, especificidade e área sob a curva) com *benchmarks*, como é o caso de modelos de regressão logística e de modelos autorregressivos integrados e de médias móveis (ARIMA).

A questão da previsibilidade dos preços de criptomoedas na frequência intradiária ainda é escassa na literatura, sendo os resultados ainda não conclusivos (AKYILDIRIM ET AL., 2022; GHOSH ET AL., 2022; GIL, 2022; GRADOJEVIC ET AL., 2023; ORTE ET AL., 2023). Em adição, não é abrangente a pesquisa existente em relação ao uso de modelos de *random forest* na previsão de movimentos de preços de criptomoedas. Assim, este trabalho busca contribuir com a literatura do tema com novas evidências empíricas sobre a aplicação de modelos de aprendizado de máquina, em especial, de *random forest*, na tarefa de previsão de movimentos futuros do Bitcoin na frequência intradiária. Contribui-se também para prover os agentes do mercado financeiro de novas ferramentas de apoio à tomada de decisão, sobretudo para operações de HFT.

Após essa breve introdução, o artigo está organizado como segue. A seção 2 apresenta a revisão de literatura sobre previsibilidade de preços e movimentos de ativos financeiros, com ênfase nos trabalhos que tratam de modelos de *random forest* aplicados a criptomoedas, e na frequência intradiária. A estrutura dos modelos de floresta aleatória é descrita na seção 3, em conjunto com as métricas de avaliação dos resultados. A seção 4 apresenta os resultados e as discussões. Por fim, as conclusões e indicações de tópicos para pesquisa futura compreendem a seção 5.

2. Revisão da Literatura

Diferentes autores investigaram a previsibilidade de preços de ativos financeiros na frequência intradiária com base em modelos de aprendizado de máquina (NELSON ET AL., 2017; GYAMERAH, 2019; LAHMIRI & BEKIROU, 2020; ALONSO ET AL., 2020; BORGES & NEVES, 2020; AKYILDIRIM ET AL., 2021; IBRAHIM ET AL., 2021; JAQUART ET AL., 2021; AKYILDIRIM ET AL., 2022; GHOSH ET AL., 2022; GIL, 2022; GRADOJEVIC ET AL., 2023; ORTE ET AL., 2023).

Gyamerah (2019), por exemplo, estudou a modelagem e a predição da variação do preço do Bitcoin utilizando um modelo chamado VMD-GAM - modelo combinado dos algoritmos método de decomposição variacional (VMD - *Variational Mode Decomposition*) e modelos aditivos generalizados (GAM - *Generalized Additive Models*). Com base em dados de 1 de janeiro de 2019 a 24 de janeiro de 2019, com periodicidade de um minuto, o autor demonstrou que o modelo proposto performou melhor, no horizonte de previsão de um minuto, que o modelo utilizado como referência EMD-GAM - modelo combinado dos algoritmos método de decomposição empírico (EMD) e modelos aditivos generalizados (GAM) - em termos de medidas de erro.

Lahmiri e Bekiros (2020) realizaram um estudo comparativo, por meio da aplicação de sete diferentes modelos de aprendizado de máquina, sobre a previsibilidade intradiária do preço do Bitcoin. A pesquisa utilizou dados de janeiro de 2016 a março de 2018, com frequência de cinco minutos, e as previsões da criptomoeda foram realizadas para um passo à frente (próximos cinco minutos). Os autores apontam que todos os modelos proporcionaram previsões acuradas, sendo efetivos em capturar a característica ruidosa, não linear e complexa da série temporal em questão.

Modelos baseados em *random forest* também foram explorados para previsão em finanças. Um modelo baseado em *random forest* foi proposto por Pradeepkumar e Ravi (2016) na tarefa de previsão dos valores diários das moedas USD/JPY (dólar americano / iene japonês), USD/GBP (dólar americano / libra esterlina) e USD/EUR (dólar americano / euro). Com dados diários e previsões de um passo à frente, os autores propuseram o modelo QRRF-CM, abordagem híbrida dos modelos regressão quantílica de floresta aleatória (QRRF - *Quantile Regression Random Forest*) e modelo de teoria do caos (CM - *Chaos Modeling*), e compararam seus resultados com os de dois outros modelos, denominados regressão quantílica caótica (CQR - *Chaos Quantile Regression*) e floresta aleatória caótica (CRF - *Chaos Random Forest*). Os resultados apontam que o modelo proposto rendeu previsões mais acuradas que os de referência.

Ghosh et al. (2022), de outra forma, estudaram a previsibilidade da direção intradiária dos preços das ações participantes do índice S&P 500. O período coberto foi o de janeiro de 1993 a dezembro de 2018, totalizando 29 anos de base de dados. Os autores aplicaram os modelos *random forest* e rede neural LSTM em uma base de dados com 93 indicadores técnicos e, inicialmente, calcularam a probabilidade, para cada dia, de o preço de cada ação valorizar ou desvalorizar. Em seguida, simularam uma estratégia de investimento em que, na abertura diária do mercado, as 10 ações que apresentassem maior probabilidade de subida eram compradas e as 10 ações com maior probabilidade de queda eram vendidas, e essas operações eram realizadas (finalizadas) no fechamento do mercado. As previsões dos modelos RF e LSTM renderam maiores retornos diários e maior taxa

de retornos positivos, e as estratégias de investimento, que consideraram custos de transação compatíveis com os praticados pelo mercado, obtiveram retornos superiores e menor risco.

Alonso et al. (2020), por exemplo, estudaram a previsão da direção do preço de seis populares criptomoedas, Bitcoin, Dash, Ether, Litecoin, Monero e Ripple. Com dados de 1 de julho de 2018 a 30 de junho de 2019 na frequência de 1 minuto, os autores utilizaram como variáveis explicativas 18 indicadores técnicos e compararam quatro diferentes arquiteturas de redes neurais. Os resultados mostraram que as séries podem ser previstas durante todo o período analisado por meio de redes neurais e que arquiteturas com memória proporcionam os melhores resultados em termos de acurácia.

Também considerando dados intradiários, Akyildirim et al. (2021) estudaram a previsibilidade da direção de preços de 12 populares criptomoedas nos horizontes de previsão de 15, 30, 60 minutos e 1 (um) dia. Os autores construíram um conjunto de 40 variáveis explicativas contendo variáveis de retornos passados e indicadores técnicos em uma base de dados de abril de 2013 a junho de 2018 com as mesmas periodicidades dos horizontes de previsão. A aplicação dos diferentes modelos de classificação apontou que os modelos considerados são hábeis em prever a direção do preço de criptomoedas em diferentes horizontes de previsão com acurácia maior que 50%, especialmente modelos de vetores suporte.

Uma comparação de modelos estado-da-arte sobre previsão do movimento do Bitcoin com frequência de 5 minutos foi realizada por Ibrahim et al. (2021). Além de indicadores populares na indústria do mercado de ações, os autores criaram preditores de medida de distância e de tendência de curto prazo. A saída dos modelos representava a probabilidade de o preço do Bitcoin valorizar a cada 5 minutos e os resultados foram comparados com as estratégias palpite aleatório (RG – *Random Guess*) e momento (repetição do movimento anterior). O modelo de rede neural MLP apresentou a melhor capacidade preditiva.

Recentemente, uma avaliação sobre o poder de predição da direção dos preços do Bitcoin futuro com modelos de RF em diferentes horizontes de previsão intradiários foi realizada por Orte et al. (2023). Os autores aplicaram o modelo RF em um base de dados de janeiro de 2013 a dezembro de 2020, com uma ampla gama de 135 indicadores técnicos, e realizaram previsões de direção dos preços nos horizontes de 15 e 30 minutos, 1 (uma), 2 (duas), 4 (quatro), 6 (seis), 12 horas e 1 (um) dia. Os resultados, mensurados pela métrica acurácia e por medidas de desempenho de carteira, mostraram a possibilidade da criação de estratégias lucrativas no mercado futuro de Bitcoin com o auxílio de modelos RF, principalmente com a utilização de horizontes de previsão mais longos.

Os estudos, em geral, apontam que os modelos baseados em *random forest* podem ser acurados para a tarefa da previsão de preços e movimentos do Bitcoin e de outros ativos financeiros. Os resultados destas pesquisas, avaliados por diferentes métricas de desempenho, exibem resultados

superiores aos clássicos de econometria de séries de tempo. A Tabela 1 resume os principais estudos que avaliam a previsibilidade de ativos financeiros utilizando modelos de aprendizado de máquinas, com destaque para dados intradiários.

Tabela 1. Resumo da literatura que considera modelos de aprendizado de máquinas na previsão de séries temporais financeiras.

Autores	Ativos Utilizados	Periodicidade dos Dados	Horizonte de Previsão	Tipo de Previsão	Métodos	Métodos Comparados	Melhor Modelo	Período	Variáveis Explicativas	Métricas de Avaliação
Pradeepkumar e Ravi (2016)	moedas USD/JPY, USD/GBP e USD/EUR	1 dia	1 dia	valor do preço	CM-QRRF	CQR e CRF	CM - QRRF	jan00 - dez13	-	MSE, MAPE
Nelson et al. (2017)	5 ações Ibovespa	15 minutos	15 minutos	direção do preço	LSTMNN	MLPNN, RF	LSTMNN	jan18 - dez15	180 indicadores técnicos	métricas de previsão
Gyamerah (2019)	Bitcoin	1 minuto	1 minuto	VMD-GAM	EMD-GAM	VMD-GAM	01jan19 - 24jan19	-	-	RMSE, MAPE, Bias
Alonso et al. (2020)	6 criptomoedas	1min	1min	direção do preço	CNN, CLSTM, MLPNN, RBFNN	RW, categoria mais frequente	CLSTM	jul18 - jun19	18 indicadores técnicos	acurácia
Borges e Neves (2020)	100 criptomoedas	1 min	1 min	direção do preço	RF, LR, SVC, GTB e EV	-	EV e RF	jul17 - out18	12 indicadores técnicos	acurácia e medidas de desempenho de carteiras
Lahmiri e Bekiros (2020)	Bitcoin	5 minutos	5 minutos	valor do preço	SVR, GRP, RT, kNN, FFNN, BRNN, RBFNN	-	Em ordem: RNN, BPNN, GRP, SVR, RT, RBFNN e kNN	jan16 - mar18	últimas 5 observações	RMSE e métricas não lineares
Akyildirim et al. (2021)	12 criptomoedas	15, 30, 60 minutos e 1 dia	15, 30, 60 minutos, 1 dia	direção do preço	SVM, LR, ANN, RF	ARIMA	SVM	abr13 - jun18	40 retornos passados e indicadores técnicos	acurácia

Tabela 1. Resumo da literatura que considera modelos de aprendizado de máquinas na previsão de séries temporais financeiras (continuação).

Autores	Ativos Utilizados	Periodicidade dos Dados	Horizonte de Previsão	Tipo de Previsão	Métodos	Métodos Comparados	Melhor Modelo	Período	Variáveis Explicativas	Métricas de Avaliação
Ibrahim et al. (2021)	Bitcoin	5 minutos	5min	direção do preço	ARIMA, Facebook Prophet, RF, LARFR e MLPNN	RG, estratégia momento	MLPNN	2014 - não revelado	9 indicadores técnicos	acurácia, teste McNemar
Jaquart et al. (2021)	Bitcoin	1 minuto	1, 5, 15, 60 minutos	direção do preço	GRUNN, LSTMNN, FNN, GBC, RF)	LR	GRUNN, LSTMNN e GBC	mar19 - dez19	11 indicadores técnicos	acurácia, teste Diebold Mariano
Akyildirim et al. (2022)	Bitcoin futuro	1 minuto	5, 15, 30, 60 minutos	direção do preço	kNN, NB, LR, RF, SVM, Boosting	ARIMA, <i>random walk</i>	RF e kNN	jan20- set20	5 retornos passados	acurácia, medidas de desempenho de carteira
Gil (2022)	5 futuros de metais preciosos	1 minuto	valor do preço	estratégia RSI, estratégia CK	-	estratégia RSI	jan20 - set21	2 indicadores técnicos - RSI e CK	medidas de retorno e risco, medidas de desempenho de carteiras	
Gradojevic et al. (2023)	Bitcoin	1 hora, 1 dia	1 hora, 1 dia	valor e direção do preço	ANN, FNN, SVM, RF	ARIMA, <i>random walk</i> , ridge, lasso, elastic	RF	abr13- jun19	46 indicadores técnicos	acurácia e MSPE
Orte et al. (2023)	Bitcoin	15, 30 minutos; 1, 2, 4, 6, 12 horas; 1 dia	15, 30 minutos; 1, 2, 4, 6, 12 horas; 1 dia	direção do preço	RF	-	-	jan13- jan21	135 indicadores técnicos	precisão e medidas de desempenho de carteira

3. Metodologia

O objetivo deste trabalho é estudar a previsibilidade da direção do preço da criptomoeda mais negociada do mercado, o Bitcoin, na frequência intradiária, utilizando modelos de *random forest*. As variáveis explicativas, ou independentes, dos modelos são referidas neste trabalho como “atributos”, enquanto a variável de interesse, ou dependente, é referida como “rótulo”.

A base de dados é simbolizada pelo conjunto $D = (X, \mathbf{y})$, em que $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]$ é a matriz que contém os P atributos (variáveis explicativas), com $\mathbf{x}_p = [x_{1p}, x_{2p}, \dots, x_{Np}]^T$, e $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$ contém o vetor dos rótulos (variável dependente), com $p = 1, 2, \dots, P$ e $n = 1, 2, \dots, N$, sendo N o número de observações na amostra. Cada observação da base de dados pode ser simbolizada por (\mathbf{x}_n, y_n) , e contém os valores observados dos atributos $\mathbf{x}_n = [x_{n1}, x_{n2}, \dots, x_{nP}]$ e o valor observado do rótulo y_n .

Um problema de classificação consiste em associar um conjunto de atributos a uma determinada categoria, ou classe. É uma regra que atribui uma das K classes de \mathbf{y} para cada observação do conjunto de atributos $\mathbf{x}_n = [x_{n1}, x_{n2}, \dots, x_{nP}]$ em X . Isto é, dado qualquer $\mathbf{x}_n \in X$, a regra atribui uma das classes $\{1, 2, \dots, K\}$ para \mathbf{x}_n na tarefa de prever y_n , ou seja, associa a adequada classe ou categoria para a n -ésima observação. Neste trabalho, $K = 2$, pois o rótulo contém apenas duas classes, “valorizar” e “desvalorizar”.

Os modelos de *random forest*, abordagem avaliada neste trabalho, possibilitam, além de associar os atributos a uma classe ou categoria, calcular a probabilidade de esse conjunto de atributos pertencer à classe associada. Neste trabalho, a função de previsão para cada rótulo y_n modela a probabilidade do Bitcoin se valorizar em determinado horizonte de previsão, dado a observação de atributos $\mathbf{x}_n \in X$, da seguinte forma:

$$\hat{f}_{RF}(\mathbf{x}_n) = \text{Prob}(\mathbf{y}_n = \text{“valorizar”} / \mathbf{x}_n) \quad (1)$$

Portanto, a função de previsão $\hat{f}_{RF}(\mathbf{x}_n)$ abrange valores no intervalo $[0, 1]$, em que um valor próximo de um indica maior probabilidade de valorização, enquanto um valor próximo de zero sugere maior probabilidade de desvalorização (ou menor probabilidade de valorização). Com isso, a escolha de um ponto de corte ou limiar irá definir se a resposta do modelo (saída) será “valorizar” ou “desvalorizar”. Este trabalho utiliza, como regra geral, do valor de 0,5 como limiar, ou seja, valores da função preditora maiores que 0,5 são associados a categoria/rótulo “valorizar”, enquanto valores menores ou iguais a 0,5 são associados ao rótulo “desvalorizar”.

O processo de classificação pode ser realizado utilizando diferentes classificadores/modelos. Na literatura atual de aprendizagem de máquinas, os principais classificadores são os modelos

baseados em redes neurais artificiais, máquinas de suporte vetorial, *boosting* e *random forest*. Neste trabalho, será abordado o modelo de *random forest* (RF), que se trata de um modelo do tipo *ensemble* de árvores de decisão. A seção a seguir descreve as árvores de decisão, constructo básico dos modelos de RF.

3.1 Árvores de Decisão

O modelo de árvores de decisão é creditado a Breiman et al. (1984), que desenvolveram a conhecida abordagem CART – *Classification and Regression Trees* (árvores de regressão e classificação), utilizada neste trabalho.

Métodos baseados em árvores envolvem estratificar ou segmentar um espaço preditor em diferentes regiões não sobrepostas e, em cada uma dessas regiões, todas as observações pertencentes recebem a mesma previsão/classificação para a variável de interesse. Como o conjunto de regras utilizado para segmentar esse espaço preditor é realizado em forma de árvore, essa abordagem é conhecida como árvore de decisão. As regiões finais encontradas são conhecidas como nós finais ou folhas, e os pontos ao longo da árvore, em que o espaço preditor é dividido, são conhecidos como nós internos (JAMES ET AL., 2013).

Considere a matriz X dos P atributos, $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]$. O espaço preditor, ou seja, o conjunto de possíveis combinações de valores de $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ é dividido em J regiões distintas e não sobrepostas, R_1, R_2, \dots, R_J . Uma vez que estas regiões são encontradas, a estimação do rótulo y_n de cada observação em dada região R_j , $j = 1, \dots, J$, é feita pela categoria mais frequente dentre essas observações. Assim, essas J caixas (regiões ou espaços preditores) R_1, R_2, \dots, R_J precisam ser encontradas de forma a minimizar o erro total de classificação do modelo (JAMES ET AL., 2013).

Dentro de um nó j , referente à uma região R_j , contendo N_j observações, considere que \hat{p}_{jk} representa a proporção das observações de uma determinada classe k dentro desta região, com $k = 1, \dots, K$, ou, ainda, representa a frequência relativa de cada classe em cada nó, da seguinte forma (JAMES ET AL., 2013):

$$\hat{p}_{jk} = \frac{1}{N_j} \sum_{\mathbf{x}_n \in R_j} I(y_n = k) \quad (2)$$

Uma métrica de erro passível de medir o erro dentro de uma região é o erro de classificação (*classification error rate* – E_{class}). Em uma árvore de decisão, essa métrica representa a fração das observações de treinamento em determinada região que não pertencem à classe mais frequente (JAMES ET AL., 2013):

$$E_{class} = 1 - \max_k(\hat{p}_{jk}) \quad (3)$$

O erro de classificação não é suficientemente sensível ao crescimento da árvore, ou seja, mensura adequadamente a soma do erro dos nós resultantes de uma quebra, mas não o erro de cada nó isoladamente. Então, na prática, outra métrica considerada mais apropriada é o índice de Gini⁴ (JAMES ET AL., 2013). O Índice de Gini dentro de uma região j é definido por:

$$G_j = \sum_{k=1}^K \hat{p}_{jk}(1 - \hat{p}_{jk}) \quad (4)$$

O índice de Gini resulta em valores baixos quando todos os \hat{p}_{jk} se aproximam de zero ou um. Por essa razão, esse índice é referido como uma medida de pureza de nó. Um valor baixo para o índice de Gini indica que o nó é puro e contém predominantemente observações de uma única classe (JAMES ET AL., 2013). Em um caso de classificação binária (apenas duas categorias no rótulo), o índice de Gini pode ser simplificado (TAN, 2018):

$$G_j = 2\hat{p}_{jk}(1 - \hat{p}_{jk}) \quad (5)$$

Assim, em um modelo de árvore de decisão, o objetivo é encontrar os espaços preditores R_1, \dots, R_j de forma a minimizar o erro total de classificação do modelo. Nas quebras sequenciais da árvore de decisão, qualquer valor pertencente a um dos atributos $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ pode ser utilizado para repartir o espaço preditor anterior (nó-pai), R , em duas novas regiões, R_1 e R_2 (nós filhos). Então, para realizar a quebra binária do nó j , primeiramente são selecionados o atributo \mathbf{x}_p e o ponto de corte, simbolizado por t_j , que dividem o espaço preditivo nas regiões $R_1(p, t_j) = \{X | \mathbf{x}_p < t_j\}$ e $R_2(p, t_j) = \{X | \mathbf{x}_p \geq t_j\}$ (JAMES ET AL., 2013).

Simbolizando como $\mathbb{I}(j)$ o grau de impureza induzido pela divisão de um nó j em sub-regiões, o decrescimento da pureza deste nó pode ser expresso como (DE ALVARENGA JÚNIOR, 2018):

$$\Delta\mathbb{I}(j) = \mathbb{I}(N_j) - \hat{p}_{jL} \cdot \mathbb{I}(N_{jL}) - \hat{p}_{jR} \cdot \mathbb{I}(N_{jR}) \quad (6)$$

em que $\mathbb{I}(\cdot)$ se refere à medida de impureza do nó, \hat{p}_{jL} é a proporção das amostras do nó j alocada ao nó filho da esquerda ($\frac{N_{jL}}{N_j}$) após a partição e, similarmente, \hat{p}_{jR} é a proporção das amostras do nó j alocado ao nó filho da direita ($\frac{N_{jR}}{N_j}$) após a partição. N_{jL} e N_{jR} são os nós filhos da esquerda e da direita, respectivamente, e N_j se refere ao total de observações do nó j .

⁴ Não confundir o índice de Gini utilizado neste trabalho, que mensura grau de pureza dentro de uma região, com o índice de mesmo nome utilizado em economia e em outras áreas da ciência para mensurar concentração de renda ou de outra medida.

O processo de escolha do melhor preditor e do melhor ponto de corte para a divisão das novas regiões é repetido de forma a maximizar a redução de impureza dessas novas quebras (JAMES ET AL., 2013).

O crescimento indefinido da árvore resultará em nós finais com apenas uma observação. Estes nós apresentariam pureza máxima, mas o modelo seria demasiadamente complexo, o que geraria sobreajuste aos dados, reduzindo a capacidade de generalização. Por outro lado, uma árvore pequena pode não conseguir capturar toda a estrutura presente no conjunto de dados (DE ALVARENGA JÚNIOR, 2018).

Uma solução para esta questão consiste em escolher um critério de parada para o crescimento da árvore, como, por exemplo, escolher um número mínimo de observações em cada região para evitar sobreajustes. Entretanto, uma solução mais eficiente consiste em crescer uma árvore para, em seguida, cortar ou podar (*prune*) essa árvore em alguns pontos para obter uma sub-árvore. O objetivo aqui é selecionar algum ponto de corte na árvore, criada utilizando a base de treinamento, que gera a sub-árvore na base de validação com o menor valor de erro (HASTIE ET AL., 2009).

Finalmente, uma vez que as regiões R_1, \dots, R_j são encontradas, a predição do rótulo y_n dentro de cada região é feita pela classe mais frequente entre as observações de treinamento que pertencem a essa região. O modelo de árvore de decisão associa as observações da região j à classe $k(j) = \underset{k}{\operatorname{argmax}} \hat{p}_{jk}$, a classe majoritária dentro da região j (HASTIE ET AL., 2009).

Assim, a indução (crescimento) de uma árvore de decisão é um problema de otimização combinatória⁵ que envolve a escolha do melhor par de parâmetros (p, t_j) pra cada nó j do modelo. Para isto, os parâmetros da árvore, $\Theta = \{R_j, Y_j\}$, em que Y_j é a saída do modelo em cada região, são encontrados através da minimização da função de risco, $\mathcal{L}(y_j, Y_j)$, da seguinte forma (DE ALVARENGA JÚNIOR, 2018):

$$\Theta = \underset{\Theta}{\operatorname{argmin}} \sum_{j=1}^J \sum_{x_n \in R_j} \mathcal{L}(y_j, Y_j) \quad (7)$$

A resolução deste problema de otimização é dividida em duas etapas. Na primeira, dada uma região R_j , estima-se o valor para Y_j utilizando alguma medida de impureza (neste trabalho, a métrica Gini). Na segunda etapa, em que região R_j é determinada, alguma aproximação deve ser realizada. Uma estratégia usual neste caso é o crescimento recursivo (sequencial) da árvore, a partir de sua raiz, atuando em um nó de cada vez, sem um planejamento antecipado (HASTIE ET AL., 2009).

⁵ Em um problema de otimização combinatória, procura-se pela solução ótima entre um conjunto finito de Possibilidades (DE ALVARENGA JÚNIOR, 2018).

3.2 Modelos de *random forest*

Os modelos de *random forest* visam solucionar o problema de alta variância das árvores de decisão por meio do treinamento de múltiplas árvores de decisão em diferentes regiões da base de dados. Dessa forma, nenhuma árvore de decisão se ajusta à totalidade dos dados (TAN, 2018). *Random forest* são modelos do tipo *ensemble*, que combinam a predição de um conjunto de árvores de decisão para se obter uma única resposta como saída, e que tendem a apresentar melhor desempenho que cada árvore do modelo em separado, devido à redução da variância de suas previsões (DE ALVARENGA JÚNIOR, 2018). Hastie et al. (2009) demonstram que o viés das *random forest* é o mesmo que o das árvores de decisão, e que a melhora nas previsões das *random forest* é resultado apenas da redução da variância das previsões.

O método encontrado para solucionar o problema de alta variância das árvores de decisão foi o *bagging* ou *bootstrap aggregating* (Breiman, 1996), cuja ideia principal é a geração de múltiplas versões de um preditor para, em seguida, utilizá-las como um preditor agregado (*ensemble*). No caso das *random forest*, são criadas B diferentes bases de treinamento, desenvolvidas as funções de previsão $\hat{f}_T^1(X), \hat{f}_T^2(X), \dots, \hat{f}_T^B(X)$ pela aplicação de um modelo de árvore de decisão $\hat{f}_T^b(X)$ em cada uma dessas diferentes bases, e calculada a média de suas previsões, com o objetivo de se obter um modelo final de baixa variância. Assim, para cada observação de atributos \mathbf{x}_n (HASTIE ET AL., 2009):

$$\hat{f}_{bagging}(\mathbf{x}_n) = \frac{1}{B} \sum_{b=1}^B \hat{f}_T^b(\mathbf{x}_n) \quad (8)$$

Hastie et al. (2009) explicam que a média de B variáveis aleatórias i.i.d. (independentes e identicamente distribuídas), cada uma com variância σ^2 , possui variância de $\frac{\sigma^2}{n}$. Assim, as previsões realizadas pela técnica *bagging* apresentariam menor variância por se tratar de uma média de previsões. Breiman (1996) informa que o método *bagging* pode levar a ganhos substanciais em termos de poder preditivo, principalmente na presença de alto grau de ruído na base de dados.

Entretanto, não existem diferentes bases de dados, apenas uma. A solução para este aparente entrave está na consideração da técnica *bootstrap*. Essa técnica de amostragem permite gerar novos conjuntos de observações da base de dados de forma aleatória e com reposição para formar novas bases distintas e de mesmo tamanho que a base original (EFRON, 1979). Assim, as *random forest* utilizam a técnica *bagging* para realizar as classificações, empregando a técnica de *bootstrap* na geração das novas bases de treinamento.

Embora o método *bagging* reduza a variância do modelo, a aplicação desse método apresenta um problema adicional: as árvores resultantes da aplicação desse método são consideravelmente correlacionadas entre si, pois o método de formação de cada árvore é semelhante (JAMES ET AL., 2013). Hastie et al. (2009) apontam que, se as amostras forem apenas identicamente distribuídas (não independentes), com correlação positiva par a par ρ , a variância da média amostral se torna $\rho\sigma^2 + \frac{(1-\rho)}{B}\sigma^2$. Dessa forma, nota-se que a variância final do modelo possui relação direta com a correlação entre as amostras, e que há uma redução na variância com o aumento do número de amostras (o segundo termo se reduz).

Dietterich (1998) tratou desta questão da correlação criando a ferramenta *random split selection* (seleção aleatória de quebras) – uma evolução do método de randomização (aleatorização) de Kwok (1990). Nesta ferramenta, em vez de se utilizarem todos os atributos da base de dados como candidatos para cada quebra das árvores, apenas algumas dessas variáveis são selecionadas aleatoriamente. Dessa forma, árvores mais distintas são formadas, a correlação entre as árvores resultantes do modelo diminui e a variância total do modelo é reduzida.

Em Breiman (2001), com base na literatura sobre o tema, é apresentado o modelo de *random forest* cuja base de funcionamento é a que se utiliza atualmente em aprendizagem de máquinas. A ideia principal foi unir habilmente o modelo de árvores de decisão (BREIMAN ET AL., 1984) com o método *bagging* (BREIMAN, 1996) e o método *random split selection* (DIETTERICH, 1998).

Em suma, as *random forest* são modelos do tipo *ensemble*, criadas pela construção de múltiplas árvores de decisão (*bagging*), em que cada árvore é aplicada em uma base de dados amostrada aleatoriamente da base de treinamento (*bootstrap*). No processo de crescimento de cada árvore, as quebras binárias do espaço preditor se utilizam de apenas uma parcela das variáveis explicativas disponíveis (*random split selection*).

Podemos definir, então, uma floresta aleatória como um preditor agregado formado por B classificadores $\{\hat{f}_T^1(X; \Theta_1), \hat{f}_T^2(X; \Theta_2), \dots, \hat{f}_T^B(X; \Theta_B)\}$, baseados nos diferentes classificadores $\hat{f}_T^b(X; \Theta_b)$, $b = 1, 2, \dots, B$, das B árvores de decisão, representadas por T_1, T_2, \dots, T_B , em que X representa a entrada de dados (conjunto dos atributos) e Θ_b simboliza o conjunto de parâmetros que define a b -ésima árvore de decisão.

A classificação final de uma floresta aleatória, $\hat{f}_{rf}(X; \Theta)$, pode ser determinada diretamente pela categoria que obteve a maioria dos B votos ou pela porcentagem de votos da categoria mais frequente (BREIMAN ET AL., 1984). Neste trabalho, a previsão para cada observação y_n é realizada pela porcentagem de votos ou classificações “valorizar” dentro das B árvores de decisão, ou seja:

$$\hat{f}_{rf}(\mathbf{x}_n; \Theta) = \frac{\sum_{b=1}^B I(\hat{f}_T^b(\mathbf{x}_n; \Theta_b) = \text{"valorizar"})}{B} \quad (9)$$

Considere, como exemplo, um modelo de *random forest* composto por 300 árvores de decisão, com a tarefa de prever o movimento do preço do Bitcoin nos próximos 30 minutos, ou seja, y_n . Se essa determinada observação y_n da variável resposta recebeu 210 classificações “valorizar” e 90 classificações “desvalorizar”, o resultado da função $\hat{f}_{rf}(\mathbf{x}_n; \Theta)$ para essa observação seria de 0,7 (210/300), ou seja, o modelo indica que existe 70% de probabilidade da criptomoeda se valorizar nos próximos 30 minutos. Como o valor de corte utilizado neste trabalho é 50% (ou 0,5), a classificação final para essa observação seria “valorizar”.

Akyldirim et al. (2021) elencam uma variedade de vantagens dos modelos RF. Os autores apontam que o modelo, devido à estrutura em que está baseado, reduz o risco de sobreajuste, possui um alto grau de acurácia, funciona eficientemente em extensas bases de dados e pode ser utilizado tanto em problemas de classificação como de regressão. Os autores adicionam, ainda, que as RF resolvem os problemas de alta variância observado em técnicas de árvores de decisão, e são menos custosos computacionalmente que modelos de máquina de vetores suporte (*Support Vector Machine - SVM*) e de redes neurais artificiais.

Neste trabalho, os experimentos computacionais são realizados com a utilização do pacote *randomForest (Classification and Regression with Random Forest)*, que implementa modelos de florestas aleatórias no *software R*. Os desenvolvedores do pacote se basearam no manual criado por Breiman (2002)⁶. O modelo de *random forest* depende de diferentes parâmetros, que são definidos pelo usuário.

3.3 Métricas de Avaliação de Desempenho

A avaliação do desempenho das previsões dos modelos nos horizontes de previsão considerados é realizada pela análise dos valores das seguintes métricas: acurácia, sensibilidade, especificidade e área sob a curva (*area under the curve - AUC*).

Em problemas de classificação, a matriz de confusão (*confusion matrix*) é a ferramenta-base utilizada para determinação nas métricas de poder preditivo. Essa matriz contém a contagem conjunta

⁶ Breiman, L (2002). Manual On Setting Up, Using, And Understanding Random Forest V3.1. https://www.stat.berkeley.edu/~breiman/Using_random_forest_V3.1.pdf.

das classificações dos modelos e dos valores reais da variável de interesse, conforme apresentado na Figura 1.

		Real	
		Positivo (Valorizar)	Negativo (Desvalorizar)
Previsão	Positivo (Valorizar)	TP	FP
	Negativo (Desvalorizar)	FN	TN

Figura 1. Matriz de Confusão.

O cálculo dessas métricas é a seguinte:

$$Acurácia = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$Sensibilidade = \frac{TP}{TP + FN} \quad (11)$$

$$Especificidade = \frac{TN}{TN + FP} \quad (12)$$

em que, TP é o número de verdadeiros positivos, TN número de verdadeiros negativos, FP o número de falsos positivos, e FN o número de falsos negativos.

Além disso, considerou-se a métrica AUC (HANLEY & MCNEIL, 1982) neste trabalho. Essa métrica apresenta valores entre 0,5 a 1, em que um valor de 0,5 revela que as previsões não são melhores que uma previsão aleatória, enquanto um valor de 1 denota que as previsões são perfeitas. A equação matemática da métrica AUC contém basicamente dois parâmetros: taxa de verdadeiros positivos (*true positive rate* - TPR), que é dada por $\frac{TP}{TP+FN}$; e taxa de falsos positivos (*false positive rate* - FPR), que é dada por $\frac{FP}{FP+TN}$. A curva chamada ROC (*receiver operating characteristic*) (FAWCETT, 2006) relaciona graficamente TPR e FPR em diferentes pontos de corte ou limiares de classificação. Por fim, a métrica AUC resume matematicamente essa representação gráfica por meio do cálculo da área sob a curva ROC, ou seja:

$$AUC = \int_0^1 TPR(FPR)dFPR \quad (13)$$

4. Resultados e Discussões

4.1 Dados

Os dados do Bitcoin foram coletados por intermédio do *website* da corretora Binance⁷. O formato em que os dados se encontram originalmente é conhecido como OHLC (*Open, High, Low, Close*). Os valores se referem aos preços de negociação de abertura, máximo, mínimo e fechamento das criptomoedas na frequência de um minuto. O período abrangido é de janeiro de 2019 a dezembro de 2022, totalizando 2.099.831 observações intradiárias de 1 (um) minuto. A data inicial foi escolhida por motivos de liquidez (a criptomoeda começou a apresentar maiores níveis de liquidez em 2019). O período final foi definido por tratar-se do último prazo disponível quando da condução da pesquisa.

O rótulo (variável resposta) da pesquisa em cada instante t , y_t , é do tipo numérico binário, representando as direções futuras dos preços de fechamento da criptomoeda nos horizontes de previsão (h) de um minuto, cinco minutos, quinze minutos, uma hora, seis horas e um dia. Assim, $y_t = \text{"valorizar"}$, se $P_{t+h} > P_t$ e $y = \text{"desvalorizar"}$, se $P_{t+h} \leq P_t$, em que P_t é o preço da criptomoeda no instante t , P_{t+h} é o preço de fechamento do Bitcoin no instante $t + h$, e h é o horizonte de previsão em minutos, com $h = \{1; 5; 15; 60; 360; 1440\}$. A variável binária é igual a unidade quando representa a categoria “valorizar” (aumento do preço no horizonte futuro selecionado) e é igual a zero quando representa a categoria “desvalorizar” (diminuição ou manutenção do preço no horizonte futuro selecionado). Esse conjunto de defasagens do rótulo é semelhante ao utilizado por Jaquart et al. (2021), acrescido de dois horizontes mais longos, de seis horas e um dia, pelo fato de a pesquisa tratar do horizonte intradiário.

Os atributos (variáveis explicativas) utilizados neste trabalho são do tipo numérico e referem-se a 15 diferentes retornos passados (defasados), calculados a partir dos preços de fechamento (*Close*), máximos (*High*) e mínimos (*Low*) da série temporal. Os retornos são calculados como $r_t = \ln(P_t) - \ln(P_{t-x})$, em que r_t é o retorno no instante t , P_t é o preço da criptomoeda no instante t , e P_{t-x} é o preço defasado no instante $t - x$, com $0 \leq x \leq 1400$ minutos.

O preço utilizado no cálculo dos retornos para o instante t , P_t , é sempre o último preço de fechamento (*Close*) do Bitcoin para o mesmo instante, ou seja, $P_t = \text{Close}_t$. Em relação aos preços defasados na fórmula de cálculo, P_{t-x} , foram considerados três subgrupos de valores.

⁷ Fonte: <https://data.binance.vision/>.

Os primeiros cinco atributos utilizam como preço defasado na fórmula de cálculo P_{t-x} , os preços de fechamento (*Close*) pontuais da criptomoeda nas defasagens de um minuto, cinco minutos, quinze minutos, uma hora, seis horas e um dia, ou seja, os valores

$\left\{ \underset{x=1}{Close_{t-x}}; \underset{x=5}{Close_{t-x}}; \underset{x=5}{Close_{t-x}}; \underset{x=15}{Close_{t-x}}; \underset{x=60}{Close_{t-x}}; \underset{x=1440}{Close_{t-x}} \right\}$. Os seguintes cinco atributos são

calculados utilizando como preço defasado, P_{t-x} , o maior valor dos preços máximos (*High*) da criptomoeda nos seguintes horizontes de cálculo: até um minuto, de um a cinco minutos, de cinco a quinze minutos, de quinze minutos a uma hora, de uma hora a seis horas, e de seis horas a um dia,

formando o conjunto de valores: $\left\{ \begin{array}{l} \max_{0 \leq x < 1} High_{t-x}; \max_{1 \leq x < 5} High_{t-x}; \max_{5 \leq x < 15} High_{t-x}; \\ \max_{15 \leq x < 60} High_{t-x}; \max_{60 \leq x < 360} High_{t-x}; \max_{360 \leq x < 1440} High_{t-x} \end{array} \right\}$. Os

últimos cinco atributos são calculados utilizando como preço defasado, P_{t-x} , o menor valor dos preços mínimos (*Low*), com os mesmos horizontes de cálculo dos atributos que utilizam os preços máximos (até um minuto, de um a cinco minutos, de cinco a quinze minutos, de quinze minutos a uma hora, de uma hora a seis horas, e de seis horas a um dia), ou seja,

$\left\{ \begin{array}{l} \min_{0 \leq x < 1} Low_{t-x}; \min_{1 \leq x < 5} Low_{t-x}; \min_{5 \leq x < 15} Low_{t-x}; \\ \min_{15 \leq x < 60} Low_{t-x}; \min_{60 \leq x < 360} Low_{t-x}; \min_{360 \leq x < 1440} Low_{t-x} \end{array} \right\}$.

Os dados são divididos em conjuntos de treinamento, para a realização do ajuste (estimação) dos modelos (*in-sample*), e de teste, para aferição da capacidade preditiva dos modelos (*out-of-sample*). O método escolhido para realizar as divisões é o de janelas de tempo móveis ou deslizantes (*sliding windows*), em que diferentes conjuntos treinamento-teste são criados respeitando a evolução temporal das observações. Essa metodologia de validação cruzada permite que os resultados dos modelos sejam mensurados para períodos que apresentam distintas dinâmicas, como de alta, baixa ou estabilidade.

Serão considerados 12 diferentes conjuntos de amostra teste, referentes aos trimestres dos anos de 2020, 2021 e 2022. O conjunto de treinamento abrange sempre o semestre anterior a cada amostra teste, conforme ilustra a Figura 2.

A Figura 3 apresenta a evolução temporal do preço do Bitcoin no período de janeiro de 2020 a dezembro de 2022. As linhas horizontais indicam a divisão dos períodos por trimestres de amostras teste. É costume na literatura do tema a utilização de apenas um ou dois períodos de amostra teste. Com a finalidade de trazer maior robustez aos resultados, este trabalho se utilizou de uma diversidade maior nos períodos de amostras teste.

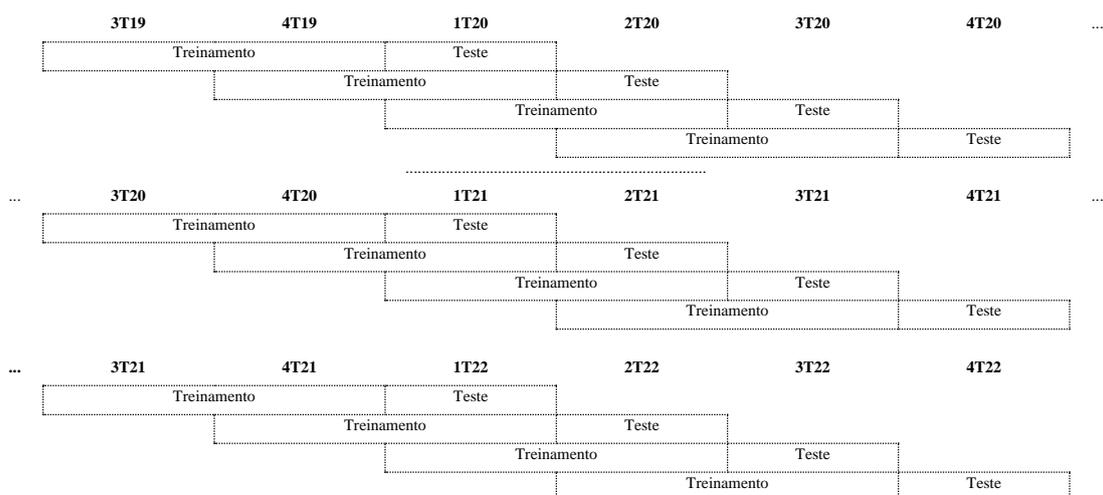


Figura 2. Divisão da base de dados nos 12 distintos conjuntos de treinamento/teste

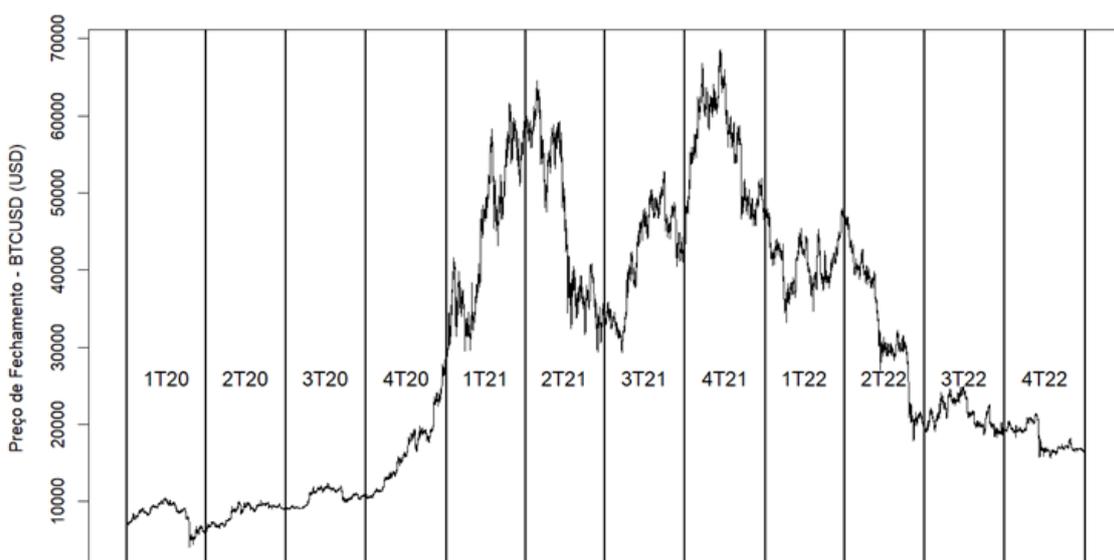


Figura 3. Evolução do preço de fechamento intradiário do Bitcoin no período de janeiro de 2020 a 2022. A linhas horizontais indicam a divisão dos períodos pelos 12 trimestres de amostras teste.

O preço do Bitcoin apresentou alta volatilidade no período analisado (ver Figura 4). A criptomoeda iniciou o ano de 2020 negociada a aproximadamente US\$ 3,7 mil, quase atingiu a marca de US\$ 70,0 mil em novembro de 2021 e, em abril de 2022, voltou a ser negociada a patamares similares aos do início de 2021. Por fim, a moeda encerrou o ano de 2022 negociada a US\$ 16,5 mil. A Tabela 2 exibe as principais estatísticas descritivas dos retornos intradiários, segmentados de acordo com a caracterização dos atributos a serem considerados nesse trabalho.

Tabela 2. Estatísticas descritivas dos retornos intradiários do Bitcoin no período de janeiro de 2020 e dezembro de 2022, segmentados de acordo com a periodicidade dos atributos a serem considerados nos modelos preditivos.

Prazo de Retorno	Estatísticas Descritivas								
	Média	Desvio Padrão	Assimetria	Curtose	Mínimo	Quartil 1	Mediana	Quartil 2	Máximo
1 min.	0,00%	0,11%	6,41%	155,35	-7,51%	-0,04%	0,00%	0,04%	7,23%
5 min.	0,00%	0,25%	-60,68%	136,56	-15,75%	-0,09%	0,00%	0,09%	16,91%
15 min.	0,00%	0,42%	-24,41%	106,61	-23,40%	-0,15%	0,00%	0,15%	28,25%
1 hora	0,00%	0,81%	-62,06%	55,81	-27,54%	-0,29%	0,01%	0,30%	29,31%
6 horas	0,02%	1,90%	-86,14%	17,76	-45,33%	-0,70%	0,03%	0,79%	33,85%
1 dia	0,08%	3,83%	-93,92%	12,04	-68,95%	-1,58%	0,12%	1,91%	36,20%

Os valores da Tabela 2 demonstram que a média dos retornos é próxima de zero, e que a volatilidade, mensurada pelo desvio-padrão dos retornos, aumenta de acordo com o horizonte de previsão, ou seja, a variabilidade intradiária é maior quanto maior for o espaço de tempo intradiário. A análise dos valores de desvio padrão e de medidas de posição, como máximo, mínimo e os quartis, indicam alta volatilidade do preço da criptomoeda (Tabela 2). No horizonte de um dia, por exemplo, em mais da metade dos valores dos retornos (valores entre o primeiro e o terceiro quartil), o retorno do Bitcoin variou mais de 1,5%. A maior desvalorização diária no período analisado foi de 68,95%, enquanto a maior valorização da moeda foi de 36,20%. Os retornos apresentaram assimetria negativa (com exceção do retorno de um minuto), crescente com o horizonte de cálculo dos retornos. Os retornos exibem um padrão leptocúrtico (curtose com valores acima de 3). Por outro lado, segundo a Tabela 2, a curtose decresce com o aumento do horizonte de cálculo dos retornos.

4.2 Resultados do Modelo de *Random Forest*

O primeiro passo do processo de modelagem consistiu na escolha da melhor configuração dos parâmetros de controle do modelo de RF. Esta tarefa foi realizada utilizando observações da base de dados referentes ao ano de 2019 (em um total de 523.836 observações), considerando o período de janeiro a agosto (348.419 observações) como amostra de treinamento, e o período de setembro a dezembro (175.417 observações) como amostra de validação.

Foram simulados os resultados de 36 diferentes configurações de escolha de parâmetros de controle, decorrentes da utilização dos seguintes valores para os parâmetros: *sampsiz* (0,25;0,50;0,75), a porcentagem de observações utilizada nas amostragens de cada base de

treinamento; *nodesize* (25;50;75;100), o número mínimo de observações em cada nó final; e *mtry* (2;4;6), o número de variáveis disponíveis no processo de quebra do nós. A qualidade das configurações de parâmetros de controle foi mensurada de acordo com a métrica AUC.

Nas modelagens, todos os atributos das bases de dados foram utilizados e foram realizadas previsões para todos os horizontes de previsão. A qualidade das configurações de parâmetros de controle foi mensurada de acordo com a métrica AUC. A Tabela 3 apresenta os 10 melhores resultados, ordenados pela média da métrica AUC, para a amostra de validação considerada, para diferentes horizontes de previsão.

Tabela 3. Resultados das simulações dos modelos de RF com diferentes configurações de parâmetros de controle. Para cada horizonte de predição, são apresentados os resultados da métrica AUC para a amostra de validação.

Parâmetros de Controle			AUC nos Horizontes de Previsão						
<i>Sampsize</i>	<i>Nodesize</i>	<i>Mtry</i>	1 min.	5 min.	15 min.	1 hora	6 horas	1 dia	Média
0,25	100	2	0,543	0,561	0,567	0,570	0,560	0,522	0,554
0,25	75	2	0,542	0,560	0,567	0,569	0,559	0,524	0,554
0,50	100	2	0,545	0,560	0,563	0,567	0,557	0,521	0,552
0,75	100	2	0,544	0,560	0,564	0,564	0,557	0,523	0,552
0,50	75	2	0,544	0,560	0,563	0,567	0,557	0,52	0,552
0,25	50	2	0,542	0,560	0,563	0,566	0,557	0,522	0,552
0,75	75	2	0,543	0,560	0,562	0,565	0,554	0,521	0,551
0,50	50	2	0,542	0,559	0,562	0,564	0,554	0,521	0,550
0,25	25	2	0,540	0,557	0,562	0,564	0,555	0,523	0,550
0,25	75	4	0,545	0,561	0,564	0,563	0,551	0,515	0,550

A configuração de parâmetros de controle com maior média AUC possui valores para *sampsiz*e de 0,25, *nodesize* de 100 e *mtry* de 2. Esse conjunto de valores foi selecionado para ser utilizado nos modelos de RF para os 12 conjuntos de treinamento e teste considerados.

A função de modelagem RF utilizada, do pacote *randomForest()* do software *R*, oferece automaticamente uma informação sobre a importância de cada atributo, ou seja, das entradas nos modelos RF. A importância de cada atributo é medida por meio do aumento médio da acurácia gerado nas diferentes quebras das árvores de decisão em cada floresta aleatória. A Tabela 4 exibe o valor médio de importância de cada um dos atributos nos diferentes horizontes de previsão, calculados nas modelagens em cada uma das 12 diferentes bases de treinamento. Com a finalidade de padronização, os valores estão ajustados de forma que a variável menos importante receba o valor de um, ou seja, os valores de importância, para cada horizonte de previsão, foram divididos pelo menor valor entre

eles. Os valores em negrito referem-se aos valores máximos de importância em cada horizonte de previsão.

Tabela 4. Importância média de cada atributo - RF, medida por meio do aumento médio da acurácia gerado nas diferentes quebras das árvores de decisão em cada floresta aleatória. Padronização realizada pela divisão de cada valor pelo menor valor em cada horizonte de previsão. Os valores em negrito referem-se aos valores máximos de importância em cada horizonte de previsão.

Atributo	Horizontes de Previsão					
	1 min.	5 min.	15 min.	1 hora	6 horas	1 dia
Fechamento 1 min.	3,76	1,51	1,24	1	1	1
Fechamento 5 min.	4,10	4,06	3,14	2,58	1,95	1,52
Fechamento 15 min.	3,44	4,26	5,75	4,45	3,12	2,38
Fechamento 1 hora	1,94	4,44	7,78	8,93	6,64	4,94
Fechamento 6 horas	1,18	3,31	7,14	12,14	14,85	11,47
Fechamento 1 dia	1	3,30	6,70	12,19	18,99	17,73
Max. 1 min.	4,40	1,15	1,09	1,10	1,28	1,47
Max. 5 min.	4,40	2,89	2,76	2,53	2,37	2,17
Max. 15 min.	4,49	5,09	5,40	4,93	3,98	3,37
Max. 1 hora	3,18	6,11	9,53	9,32	7,67	6,41
Max. 6 horas	1,26	3,96	8,53	13,44	16,76	12,91
Max. 1 dia	1,24	3,52	7,41	13,44	20,21	19,18
Min. 1 min.	3,73	1	1	1	1,15	1,34
Min. 5 min.	3,81	2,92	2,69	2,32	2,10	1,97
Min. 15 min.	4,84	5,11	5,33	4,59	3,65	3,10
Min. 1 hora	3,17	5,70	8,41	8,42	7,10	6,14
Min. 6 horas	1,42	4,05	8,52	13,09	14,48	12,69
Min. 1 dia	1,04	3,36	7,32	12,83	18,78	19,41

Os resultados da Tabela 4 apontam que, com o aumento do horizonte de previsão, os atributos referentes a maiores defasagens se tornam mais relevantes. Para o horizonte de previsão de um minuto, por exemplo, os retornos passados de até quinze minutos se apresentaram como os atributos mais importantes. De outra forma, para os horizontes de previsão mais longos, a partir de seis horas, os retornos passados com horizontes a partir de seis horas se mostraram como os atributos mais importantes.

A Tabela 5 ilustra o poder preditivo dos modelos de RF, nos diferentes horizontes de previsão, para a antecipação da direção futura do preço do Bitcoin nas diferentes amostras teste consideradas. A capacidade preditiva do modelo é medida pelas métricas: acurácia, sensibilidade, especificidade e AUC.

Tabela 5. Métricas de avaliação preditiva dos modelos RF para a antecipação futura da direção do preço intradiário do Bitcoin para o conjunto de amostras teste trimestrais no dos anos de 2020 a 2022. A última coluna (“Total”) computa os valores considerando o prazo total de testes. Os valores em negrito demarcam os maiores valores das métricas para cada período de teste. O termo “1T20” refere-se ao primeiro trimestre de 2021.

Horizonte de Previsão	Período do Conjunto Teste												Total
	1T20	2T20	3T20	4T20	1T21	2T21	3T21	4T21	1T22	2T22	3T22	4T22	
Painel A: Acurácia													
1 min.	0,535	0,540	0,534	0,532	0,522	0,526	0,522	0,515	0,518	0,520	0,516	0,522	0,525
5 min.	0,542	0,550	0,545	0,534	0,530	0,534	0,530	0,521	0,523	0,522	0,523	0,528	0,532
15 min.	0,538	0,546	0,543	0,544	0,532	0,537	0,534	0,527	0,531	0,525	0,531	0,538	0,536
1 hora	0,54	0,547	0,540	0,533	0,514	0,530	0,528	0,528	0,535	0,535	0,544	0,54	0,534
6 horas	0,515	0,542	0,510	0,536	0,506	0,526	0,514	0,533	0,549	0,515	0,530	0,537	0,526
1 dia	0,484	0,524	0,520	0,521	0,502	0,521	0,500	0,524	0,518	0,481	0,533	0,531	0,513
Painel B: Sensibilidade													
1 min.	0,507	0,487	0,483	0,503	0,498	0,506	0,521	0,510	0,500	0,521	0,505	0,475	0,501
5 min.	0,488	0,499	0,503	0,499	0,552	0,546	0,531	0,539	0,503	0,514	0,517	0,471	0,513
15 min.	0,496	0,516	0,528	0,514	0,582	0,567	0,528	0,540	0,510	0,516	0,510	0,482	0,524
1 hora	0,513	0,512	0,474	0,501	0,573	0,579	0,528	0,518	0,502	0,543	0,528	0,505	0,523
6 horas	0,484	0,492	0,446	0,542	0,482	0,533	0,491	0,505	0,528	0,554	0,491	0,403	0,496
1 dia	0,456	0,435	0,529	0,486	0,495	0,527	0,426	0,529	0,504	0,560	0,428	0,543	0,492
Painel C: Especificidade													
1 min.	0,563	0,594	0,585	0,561	0,546	0,545	0,523	0,521	0,536	0,519	0,528	0,569	0,549
5 min.	0,596	0,602	0,588	0,571	0,507	0,521	0,530	0,504	0,544	0,531	0,528	0,585	0,550
15 min.	0,582	0,576	0,559	0,576	0,482	0,506	0,541	0,513	0,552	0,533	0,552	0,595	0,547
1 hora	0,567	0,584	0,607	0,569	0,449	0,482	0,529	0,536	0,570	0,527	0,561	0,575	0,546
6 horas	0,548	0,598	0,583	0,528	0,533	0,520	0,539	0,562	0,570	0,479	0,569	0,658	0,558
1 dia	0,513	0,637	0,508	0,588	0,512	0,517	0,584	0,519	0,532	0,423	0,628	0,520	0,537
Painel D: AUC													
1 min.	0,551	0,558	0,547	0,544	0,533	0,538	0,531	0,523	0,527	0,529	0,525	0,531	0,536
5 min.	0,563	0,572	0,564	0,549	0,542	0,548	0,545	0,530	0,535	0,532	0,531	0,54	0,546
15 min.	0,555	0,567	0,562	0,565	0,547	0,553	0,549	0,540	0,544	0,533	0,543	0,552	0,551
1 hora	0,558	0,564	0,556	0,551	0,523	0,54	0,539	0,541	0,549	0,545	0,56	0,552	0,548
6 horas	0,520	0,563	0,530	0,546	0,514	0,544	0,521	0,546	0,565	0,525	0,54	0,539	0,538
1 dia	0,477	0,544	0,525	0,541	0,498	0,530	0,508	0,528	0,520	0,486	0,543	0,535	0,520

Os resultados exibidos na Tabela 5 apontam que os modelos RF possuem poder preditivo superior ao de um passeio aleatório ou de um modelo ingênuo - valores de 0,5 para todas as métricas de avaliação - na tarefa de previsão da direção do preço do Bitcoin nos horizontes intradiários considerados.

A métrica acurácia, exibida no Painel A da Tabela 5, representa a capacidade de acerto total de classificação dos modelos. Considerando o resultado para o período total de testes (anos completos de 2020 a 2022) – exibido na última coluna da Tabela 5 – os valores de acurácia (Painel A) foram de

0,525; 0,532; 0,536; 0,534; 0,526; 0,513, para os horizontes respectivos de um minuto, cinco minutos, quinze minutos, uma hora, seis horas e um dia. Todos os valores foram superiores a 0,5. A acurácia foi maior para os horizontes de previsão intermediários, de cinco minutos (0,532), quinze minutos (0,536) e uma hora (0,534). O melhor resultado se refere ao horizonte de quinze minutos (0,536), enquanto o pior resultado está associado ao horizonte mais longo de previsão, de um dia (0,513). Em relação aos conjuntos trimestrais de teste (Painel A da Tabela 5), os valores de acurácia encontrados foram superiores a 0,5 em 71 dos 72 períodos de teste e horizontes de previsão considerados, com exceção de um resultado (conjunto teste “2T22” no horizonte de previsão de um dia). O maior valor de acurácia encontrado (0,550) refere-se ao segundo trimestre de 2020, no horizonte de previsão de cinco minutos.

A métrica sensibilidade, exibida no Painel B da Tabela 5, indica a capacidade do modelo de realizar previsões corretas de valores positivos do rótulo, no caso deste trabalho, de valorizações da criptomoeda. Os resultados apontam que os modelos RF apresentaram menor habilidade em detectar as valorizações do preço do Bitcoin, principalmente nos horizontes de previsão mais longos, de seis horas e um dia, com valores encontrados inferiores a 0,5. Considerando o período total de teste (última coluna Painel B da Tabela 5), os valores de sensibilidade foram de 0,501 (um minuto); 0,513 (cinco minutos); 0,524 (quinze minutos); 0,523 (uma hora); 0,496 (seis horas); e 0,492 (um dia). Tal como os resultados de acurácia, os valores maiores de sensibilidade ocorreram nos horizontes intermediários (cinco minutos, quinze minutos de uma hora), com maior valor para o horizonte de quinze minutos (0,524) e pior valor no horizonte de uma hora (0,492). Em 47 dos 72 testes de base trimestral, os modelos não conseguiram prever as valorizações do Bitcoin de forma superior a de um passeio aleatório (0,5). O ano de 2021 apresentou os maiores valores de sensibilidade no período de teste considerado.

A métrica especificidade que, por outro lado, mensura neste trabalho a capacidade de realizar previsões corretas entre as desvalorizações da criptomoeda. Considerando o período total de testes (última coluna Painel C da Tabela 5), exibiu valores de 0,549; 0,550; 0,547; 0,546; 0,558; 0,537 para os horizontes respectivos de um minuto, cinco minutos, quinze minutos, uma hora, seis horas e um dia. O horizonte de previsão de seis horas apresentou o melhor resultado (0,558), enquanto o pior resultado esteve associado ao horizonte de previsão mais longo, de um dia (0,537). Quanto aos trimestres de teste, houve resultados superior a 0,5 em 67 dos 72 conjuntos considerados, com maior valor encontrado (0,658) no quarto trimestre de 2022, com horizonte de previsão de seis horas.

A análise dos valores de sensibilidade e especificidade revela que os modelos de RF apresentaram maior habilidade em identificar desvalorizações da criptomoeda. Os valores de sensibilidade foram superiores aos de especificidade em todos os horizontes de previsão quando considerado o prazo completo de teste (últimas colunas dos Painéis B e C da Tabela 5). Uma possível razão para a maior capacidade dos modelos de detectar desvalorizações é a de que os modelos RF realizaram com mais frequência as previsões do rótulo “desvalorizar” que do rótulo “valorizar”, principalmente para os horizontes de previsão mais longos.

A métrica de poder preditivo AUC é a mais completa das utilizadas neste trabalho, pois abrange em seu cálculo (de medição de erro de classificação) todos os valores de corte na tarefa de transformar uma previsão do modelo, medida em probabilidade de valorização entre zero e um, em uma classificação, de rótulo “valorizar” ou “desvalorizar”. As demais métricas utilizam apenas o valor de corte de 0,5 como referência. Os resultados encontrados de AUC, presentes no Painel D da Tabela 5, indicam que as previsões são superiores que uma previsão aleatória, com valores de 0,536 (um minuto); 0,546 (cinco minutos); 0,551 (quinze minutos); 0,548 (uma hora); 0,538 (seis horas); 0,520 (um dia), considerando o período total de teste (última coluna do Painel D da Tabela 5). De mesma forma que nos resultados da métrica acurácia (Painel A da Tabela 5), o maior poder preditivo se associou ao horizonte de quinze minutos (0,551) e o pior resultado ao horizonte de um dia (0,520). Os maiores valores de AUC encontrados para os trimestres de teste (Painel D da Tabela 5) decorrem do ano de 2020, atingindo o valor máximo de 0,572, relativo ao segundo trimestre de 2020 (“2T20”) e horizonte de previsão de cinco minutos. Em 69 dos 72 conjuntos de teste, o valor de AUC foi superior a 0,5.

Para avaliar se os modelos RF são capazes de prever a direção do preço de criptomoedas com poder preditivo maior que de um passeio aleatório, ou um simples lance de uma moeda, realizou-se um teste de hipótese paramétrico para proporções (Morettin & Bussab, 2017), em que é verificado se o poder preditivo dos modelos, medido pela acurácia, é estatisticamente igual a 0,5, como hipótese nula. A hipótese alternativa atesta acurácia maior que 0,5. A Tabela 6 exibe os valores-p (*p-values*) dos testes de proporção realizados para cada conjunto de teste nos diferentes horizontes de previsão. Valores encontrados abaixo de 0,05, apresentados em negrito, indicam que a hipótese nula de igualdade de proporções do teste é rejeitada, ou seja, a acurácia é estatisticamente superior a 0,5.

Tabela 6. P-valores (*p-values*) dos testes de proporção para a acurácia dos modelos RF, realizados para cada conjunto de teste nos diferentes horizontes de previsão. Valores encontrados abaixo de 0,05, demarcados em negrito, indicam situações em que a acurácia é estatisticamente superior a 0,5.

A última coluna (“Total”) computa os valores considerando o prazo total de testes.

Horizonte de Previsão	Período do Conjunto Teste													
	1T20	2T20	3T20	4T20	1T21	2T21	3T21	4T21	1T22	2T22	3T22	4T22	Total	
1 min.	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
5 min.	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
15 min.	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
1 hora	0,000	0,000	0,000	0,002	0,212	0,005	0,008	0,010	0,001	0,001	0,000	0,000	0,000	0,000
6 horas	0,563	0,115	0,715	0,175	0,874	0,345	0,602	0,230	0,065	0,637	0,274	0,192	0,001	0,001
1 dia	0,834	0,675	0,755	0,755	1,000	0,834	1,000	0,755	0,752	0,834	0,602	0,602	0,381	0,381

Os resultados da Tabela 6, quando considerado o período completo de teste (última coluna), revelam que os modelos RF apresentaram acurácia estatisticamente superior a 0,5 em quase todos os horizontes de previsão, com exceção do horizonte de um dia. Quanto aos trimestres de teste, foi atestada acurácia superior a 0,5 em todos os períodos de teste nos horizontes mais curtos, de um minuto, cinco minutos e 15 minutos, e em 11 dos 12 períodos de teste relativos ao horizonte de uma hora. Em relação ao horizonte mais longo considerado, de um dia, em todos os casos não houve superioridade estatística dos modelos RF.

Os resultados encontrados nos testes de proporção apontam para a rejeição da hipótese de eficiência de mercado de Fama (1970) em sua forma fraca, nos horizontes de previsão mais curtos utilizados, de um minuto, cinco minutos, quinze minutos e uma hora. Dessa forma, há indícios de que os movimentos do preço do Bitcoin na frequência intradiária não seguem o comportamento de um passeio aleatório e podem ser previstos por modelos de aprendizado de máquina, como o *random forest*, em determinados horizontes de previsão.

4.3 Comparação dos Modelos

Foram avaliados, também, como *benckmarks*, os modelos de regressão logística e ARIMA. As variáveis independentes do modelo de regressão logística foram as mesmas do modelo de RF. A estruturas consideradas dos modelos ARIMA, de acordo com os horizontes de previsão, são apresentadas no Apêndice A. A Tabela 7 apresenta conjuntamente o poder preditivo dos diferentes modelos, mensurado pelas métricas: AUC, acurácia, sensibilidade e especificidade. Os valores de cada métrica foram calculados considerando o período completo de teste, de janeiro de 2020 a

dezembro de 2020. Os resultados abrangentes dos modelos de regressão logística e ARIMA são apresentados no Apêndice B.

Tabela 7. Comparação do poder preditivo dos modelos *random forest*, regressão logística e ARIMA por meio do conjunto de métricas: AUC, acurácia, sensibilidade e especificidade, para cada horizonte de previsão considerando o período completo de teste (janeiro de 2020 a dezembro de 2020). Os valores em negrito destacam o maior valor para a métrica de cada horizonte de previsão.

Horizonte de Previsão	Modelo		
	<i>Random Forest</i>	Regressão Logística	ARIMA
Painel A: Acurácia			
1 min.	0.516	0.514	0.504
5 min.	0.524	0.526	0.514
15 min.	0.527	0.531	0.509
1 hora	0.526	0.532	0.516
6 horas	0.519	0.526	0.518
1 dia	0.510	0.529	0.498
Painel B: Sensibilidade			
1 min.	0.493	0.510	0.459
5 min.	0.510	0.521	0.511
15 min.	0.518	0.518	0.410
1 hora	0.515	0.512	0.431
6 horas	0.493	0.503	0.421
1 dia	0.475	0.502	0.208
Painel C: Especificidade			
1 min.	0.538	0.518	0.549
5 min.	0.537	0.531	0.518
15 min.	0.536	0.545	0.609
1 hora	0.538	0.552	0.603
6 horas	0.546	0.551	0.621
1 dia	0.547	0.559	0.802
Painel D: AUC			
1 min.	0.523	0.522	0.504
5 min.	0.535	0.537	0.515
15 min.	0.538	0.541	0.509
1 hora	0.536	0.541	0.513
6 horas	0.528	0.536	0.519
1 dia	0.513	0.541	0.506

A leitura da Tabela 7 demonstra a superioridade de poder preditivo do modelo de *random forest*, seguido pela regressão logística, com resultados próximos, e pelo ARIMA, com resultados inferiores. Com exceção do horizonte de um dia, o modelo *random forest* apresentou valores superiores de acurácia e de AUC (Painel A e Painel B da Tabela 7) em relação aos gerados pelos modelos regressão logística e ARIMA em todos os demais horizontes de previsão. Os modelos ARIMA apresentaram os piores resultados de acurácia e AUC para todos os horizontes de previsão.

Todos os modelos apresentaram maior capacidade de detectar desvalorizações da criptomoeda (ver Painel B e Painel C da Tabela 7), pois apresentaram, em média, maiores valores para a métrica especificidade (acerto entre as desvalorizações da criptomoeda) em relação à métrica sensibilidade (acerto entre as valorizações), fato ainda mais evidente nos horizontes de previsão mais longos.

Os modelos RF e regressão logística apresentaram maior poder preditivo nos horizontes de previsão intermediários, de cinco minutos, quinze minutos e uma hora, com valores de acurácia de 0,532, 0,536 e 0,534, respectivamente para os modelos de RF, e de 0,528, 0,531 e 0,528 para os modelos de regressão logística. Os maiores valores de acurácia se referem ao horizonte de previsão de quinze minutos. Aparentemente, o horizonte de um minuto é curto para realizar previsões de direção de preços e o horizonte de um dia descola do contexto intradiário do trabalho.

5. Conclusão

Este artigo avaliou a previsibilidade da direção do preço da criptomoeda mais negociada do mercado, o Bitcoin, na frequência intradiária, em diferentes horizontes de previsão, com o uso de modelos de *random forest*. Os resultados foram comparados com os modelos ARIMA e regressão logística de acordo com métricas de acurácia de classificação.

Os resultados apontam para a rejeição da hipótese de eficiência de mercado em sua forma fraca de Fama (1970) para a maioria dos horizontes de previsão considerados na pesquisa, com exceção apenas do horizonte mais longo, de um dia. Dessa forma, atestou-se a previsibilidade da direção do preço intradiário do Bitcoin. A análise dos valores de sensibilidade e especificidade mostrou que os modelos RF apresentaram maior habilidade em detectar desvalorizações do Bitcoin. Os melhores resultados dos modelos RF ocorreram nos horizontes de previsão intermediários, de 15 minutos e de uma hora.

A acurácia dos modelos RF se apresentou estatisticamente superior quando comparada com os modelos regressão logística e ARIMA nos horizontes de previsão de um minuto, quinze minutos e uma hora. Os melhores resultados dos modelos RF ocorreram no horizonte de previsão de quinze minutos.

Pesquisas futuras podem abordar modelos com uma maior gama de atributos (variáveis explicativas), com a inclusão de diferentes indicadores técnicos e índices econômicos, como já ocorre em pesquisas na literatura do tema. Outras criptomoedas podem ser incluídas como objeto de pesquisa. O potencial econômico das previsões dos modelos também pode ser explorado.

Referências Bibliográficas

Akyildirim, E., Çepni, O., Corbet, S., & Uddin, G. S. (2022). Forecasting Mid-Price Movement of Bitcoin Futures Using Machine Learning. E Akyildirim, O Cepni, S Corbet, GS Uddin, Forecasting

mid-price movement of Bitcoin futures using machine learning, **Annals of Operations Research**, 1-32.

Akyildirim, E., Goncu, A., & Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. **Annals of Operations Research**, 297(1), 3-36.

Alonso-Monsalve, S., Suárez-Cetrulo, A. L., Cervantes, A., & Quintana, D. (2020). Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. **Expert Systems with Applications**, 149, 113250.

Borges, T. A., & Neves, R. F. (2020). Ensemble of machine learning algorithms for cryptocurrency investment with different data resampling methods. **Applied Soft Computing**, 90, 106187.

Breiman, L. (1996). Bagging predictors. **Machine learning**, 24(2), 123-140.

Breiman, L. (2001). Random forest. **Machine learning**, 45(1), 5-32.

Breiman, L (2002). Manual On Setting Up, Using, And Understanding Random Forest V3.1. https://www.stat.berkeley.edu/~breiman/Using_random_forest_V3.1.pdf.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. **Wadsworth Statistics/Probability (1st ed.,)**. New York, N.Y: Chapman and Hall/CRC.

de Alvarenga Júnior, W. J. (2018). Métodos de otimização hiperparamétrica: um estudo comparativo utilizando árvores de decisão e florestas aleatórias na classificação binária. **Universidade Federal de Minas Gerais**.

Dietterich, T. G. (1998). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization. **Machine learning**, 32, 1-22.

Efron, B. (1979). Computers and the theory of statistics: thinking the unthinkable. **SIAM review**, 21(4):460–480.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. **The journal of Finance**, 25(2), 383-417.

Enke, D., Thawornwong, S., 2005. The use of data mining and neural networks for forecasting stock market returns. **Expert Systems with Applications**, 29 (4), 927–940.

Fawcett, T. (2006). An introduction to ROC analysis. **Pattern recognition letters**, 27(8), 861-874.

Ghosh, P., Neufeld, A., & Sahoo, J. K. (2022). Forecasting directional movements of stock prices for intraday trading using LSTM and random forest. **Finance Research Letters**, 46, 102280.

Gil, C. (2022). Intraday Trading of Precious Metals Futures Using Algorithmic Systems. **Chaos, Solitons & Fractals**, 154, 111676.

Gradojevic, N., Kukolj, D., Adcock, R., & Djakovic, V. (2023). Forecasting Bitcoin with technical analysis: A not-so-random forest? **International Journal of Forecasting**, 39(1), 1-17.

- Gyamerah, S. A. (2019). Are Bitcoins price predictable? Evidence from machine learning techniques using technical indicators. **arXiv preprint arXiv:1909.01268**.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. **Radiology**, 143(1), 29-36.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). **New York: Springer**.
- Ibrahim, A., Kashef, R., & Corrigan, L. (2021). Predicting market movement direction for bitcoin: A comparison of time series modeling methods. **Computers & Electrical Engineering**, 89, 106905.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). **New York: Springer**.
- Jaquart, P., Dann, D., & Weinhardt, C. (2021). Short-term bitcoin market prediction via machine learning. **The Journal of Finance and Data Science**, 7, 45-66.
- Kwok, S. W., & Carter, C. (1990). Multiple decision trees. In Machine intelligence and pattern recognition (Vol. 9, pp. 327-335). **North-Holland**.
- Lahmiri, S., & Bekiros, S. (2020). Intelligent forecasting with machine learning trading systems in chaotic intraday Bitcoin market. **Chaos, Solitons & Fractals**, 133, 109641.
- Leung, M. T., Daouk, H., & Chen, A. S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. **International Journal of forecasting**, 16(2), 173-190.
- Morettin, P. A., & Bussab, W. O. (2017). *Estatística Básica*. São Paulo: Saraiva.
- Nelson, D. M., Pereira, A. C., & De Oliveira, R. A. (2017, May). Stock market's price movement prediction with LSTM neural networks. **In: 2017 International joint conference on neural networks (IJCNN)**, pp. 1419-1426.
- Orte, F., Mira, J., Sánchez, M. J., & Solana, P. (2023). A random forest-based model for crypto asset forecasts in futures markets with out-of-sample prediction. **Research in International Business and Finance**, 64, 101829.
- Pradeepkumar, D., & Ravi, V. (2016, March). FOREX rate prediction using chaos and quantile regression random forest. **In: 2016 3rd International Conference on Recent Advances in Information Technology (RAIT)**, pp. 517-522.

Apêndice A

Tabela A1. Estrutura ARMA(p,q) ótima encontrada em cada conjunto treinamento-teste para cada horizonte de previsão. Foram testados valores para p e q de zero a cinco em cada situação e selecionada a melhor estrutura pelo critério de informação bayesiano (*bayesian information criterion* – BIC). O termo “1T20” refere-se ao primeiro trimestre de 2021.

Horizonte de Previsão	Período do Conjunto Teste											
	1T20	2T20	3T20	4T20	1T21	2T21	3T21	4T21	1T22	2T22	3T22	4T22
1 min.	(1,1)	(1,5)	(1,5)	(1,5)	(5,0)	(1,3)	(1,2)	(2,1)	(1,5)	(1,5)	(1,2)	(1,0)
5 min.	(5,0)	(1,3)	(2,2)	(5,0)	(5,0)	(1,2)	(1,0)	(2,2)	(2,2)	(4,0)	(1,5)	(5,0)
15 min.	(3,2)	(2,2)	(5,2)	(2,1)	(5,0)	(1,1)	(1,2)	(1,2)	(1,0)	(1,0)	(1,5)	(1,3)
1 hora	(3,3)	(4,5)	(1,1)	(1,1)	(5,0)	(2,2)	(1,1)	(1,1)	(3,3)	(1,0)	(1,0)	(2,0)
6 horas	(2,2)	(1,4)	(2,2)	(4,4)	(5,0)	(2,3)	(2,3)	(1,1)	(1,0)	(1,0)	(1,0)	(1,0)
1 dia	(2,2)	(1,0)	(1,0)	(1,0)	(1,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)

Apêndice B

Tabela B1. Métricas de avaliação preditiva dos modelos de regressão logística para a antecipação futura da direção do preço intradiário do Bitcoin para o conjunto de amostras teste trimestrais no dos anos de 2020 a 2022. A última coluna (“Total”) computa os valores considerando o prazo total de testes. Os valores em negrito demarcam os maiores valores das métricas para cada período de teste. O termo “1T20” refere-se ao primeiro trimestre de 2021.

Horizonte de Previsão	Período do Conjunto Teste												Total
	1T20	2T20	3T20	4T20	1T21	2T21	3T21	4T21	1T22	2T22	3T22	4T22	
Panel A: Acurácia													
1 min.	0,524	0,528	0,506	0,511	0,520	0,519	0,514	0,510	0,513	0,514	0,518	0,519	0,516
5 min.	0,537	0,541	0,534	0,527	0,533	0,534	0,529	0,516	0,523	0,517	0,518	0,523	0,528
15 min.	0,533	0,540	0,542	0,539	0,538	0,529	0,525	0,525	0,529	0,515	0,527	0,527	0,531
1 hora	0,526	0,540	0,537	0,524	0,523	0,522	0,521	0,522	0,529	0,526	0,542	0,520	0,528
6 horas	0,518	0,530	0,523	0,514	0,520	0,522	0,511	0,533	0,514	0,532	0,519	0,514	0,521
1 dia	0,467	0,533	0,620	0,490	0,534	0,502	0,500	0,504	0,543	0,506	0,530	0,545	0,523
Panel B: Sensibilidade													
1 min.	0,503	0,541	0,435	0,476	0,533	0,516	0,557	0,55	0,508	0,550	0,541	0,671	0,532
5 min.	0,519	0,514	0,49	0,502	0,573	0,541	0,463	0,529	0,493	0,496	0,537	0,720	0,532
15 min.	0,494	0,525	0,484	0,486	0,583	0,531	0,439	0,509	0,501	0,459	0,533	0,767	0,526
1 hora	0,428	0,517	0,554	0,442	0,595	0,590	0,403	0,393	0,505	0,505	0,590	0,783	0,525
6 horas	0,433	0,480	0,454	0,425	0,502	0,66	0,323	0,361	0,476	0,517	0,475	0,583	0,472
1 dia	0,366	0,359	0,667	0,433	0,509	0,649	0,224	0,243	0,453	0,400	0,358	0,375	0,421
Panel C: Especificidade													
1 min.	0,544	0,514	0,578	0,547	0,508	0,521	0,471	0,471	0,517	0,48	0,495	0,370	0,501
5 min.	0,555	0,568	0,579	0,555	0,493	0,527	0,596	0,503	0,552	0,537	0,498	0,324	0,524
15 min.	0,574	0,555	0,600	0,597	0,492	0,528	0,612	0,540	0,557	0,569	0,520	0,287	0,536
1 hora	0,627	0,565	0,519	0,620	0,445	0,454	0,648	0,646	0,553	0,546	0,494	0,26	0,531
6 horas	0,609	0,586	0,601	0,640	0,541	0,387	0,707	0,703	0,555	0,546	0,563	0,451	0,572
1 dia	0,576	0,756	0,554	0,599	0,569	0,377	0,815	0,78	0,636	0,585	0,684	0,695	0,635
Panel D: AUC													
1 min.	0,534	0,543	0,514	0,518	0,527	0,527	0,522	0,515	0,520	0,523	0,525	0,530	0,525
5 min.	0,549	0,558	0,550	0,537	0,546	0,548	0,540	0,522	0,531	0,523	0,522	0,538	0,539
15 min.	0,542	0,553	0,554	0,552	0,553	0,544	0,534	0,536	0,539	0,520	0,533	0,548	0,542
1 hora	0,539	0,549	0,547	0,542	0,531	0,527	0,534	0,526	0,540	0,534	0,555	0,551	0,540
6 horas	0,531	0,539	0,535	0,551	0,539	0,538	0,523	0,545	0,526	0,543	0,523	0,513	0,534
1 dia	0,486	0,569	0,621	0,528	0,557	0,523	0,531	0,526	0,563	0,487	0,520	0,534	0,537

Tabela B2. Métricas de avaliação preditiva dos modelos ARIMA para a antecipação futura da direção do preço intradiário do Bitcoin para o conjunto de amostras teste trimestrais no dos anos de 2020 a 2022. A última coluna (“Total”) computa os valores considerando o prazo total de testes. Os valores em negrito demarcam os maiores valores das métricas para cada período de teste. O termo “1T20” refere-se ao primeiro trimestre de 2021.

Horizonte de Previsão	Período do Conjunto Teste												Total
	1T20	2T20	3T20	4T20	1T21	2T21	3T21	4T21	1T22	2T22	3T22	4T22	
Painel A: Acurácia													
1 min.	0,512	0,524	0,501	0,503	0,486	0,513	0,505	0,501	0,513	0,500	0,492	0,504	0,504
5 min.	0,531	0,535	0,523	0,528	0,458	0,516	0,519	0,503	0,514	0,523	0,503	0,518	0,514
15 min.	0,497	0,514	0,526	0,544	0,472	0,519	0,530	0,517	0,499	0,501	0,495	0,489	0,509
1 hora	0,474	0,522	0,549	0,555	0,479	0,501	0,532	0,531	0,535	0,517	0,498	0,493	0,516
6 horas	0,518	0,575	0,520	0,544	0,508	0,526	0,518	0,515	0,496	0,488	0,501	0,512	0,518
1 dia	0,584	0,416	0,440	0,344	0,523	0,494	0,500	0,505	0,494	0,556	0,549	0,571	0,498
Painel B: Sensibilidade													
1 min.	0,515	0,519	0,498	0,492	0,488	0,597	0,505	0,487	0,513	0,500	0,406	0,000	0,459
5 min.	0,533	0,534	0,522	0,565	0,470	0,787	0,516	0,483	0,516	0,527	0,393	0,294	0,511
15 min.	0,502	0,514	0,519	0,637	0,500	0,744	0,515	0,524	0,000	0,000	0,309	0,150	0,410
1 hora	0,477	0,521	0,551	0,727	0,516	0,707	0,497	0,540	0,533	0,000	0,000	0,060	0,431
6 horas	0,505	0,587	0,536	0,486	0,578	0,823	0,532	0,514	0,000	0,469	0,000	0,000	0,421
1 dia	0,568	0,000	0,000	0,000	0,979	1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,208
Painel C: Especificidade													
1 min.	0,508	0,530	0,504	0,514	0,483	0,430	0,504	0,515	0,513	0,499	0,579	1,000	0,549
5 min.	0,529	0,536	0,524	0,489	0,447	0,252	0,522	0,523	0,513	0,520	0,616	0,743	0,518
15 min.	0,492	0,514	0,533	0,444	0,444	0,298	0,544	0,511	1,000	1,000	0,686	0,829	0,609
1 hora	0,470	0,524	0,547	0,354	0,438	0,299	0,569	0,524	0,537	1,000	1,000	0,922	0,603
6 horas	0,532	0,562	0,503	0,622	0,433	0,247	0,503	0,516	1,000	0,505	1,000	1,000	0,621
1 dia	0,600	1,000	1,000	1,000	0,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	0,802
Painel D: AUC													
1 min.	0,512	0,524	0,501	0,503	0,486	0,514	0,505	0,501	0,513	0,500	0,493	0,500	0,504
5 min.	0,531	0,535	0,523	0,527	0,458	0,519	0,519	0,503	0,514	0,524	0,504	0,519	0,515
15 min.	0,497	0,514	0,526	0,541	0,472	0,521	0,530	0,517	0,500	0,500	0,498	0,489	0,509
1 hora	0,473	0,522	0,549	0,541	0,477	0,503	0,533	0,532	0,535	0,500	0,500	0,491	0,513
6 horas	0,519	0,574	0,519	0,554	0,505	0,535	0,517	0,515	0,500	0,487	0,500	0,500	0,519
1 dia	0,584	0,500	0,500	0,500	0,489	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,506