

UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E ATUÁRIA  
DEPARTAMENTO DE ADMINISTRAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO

Paulo de Castro Rubio Poli

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

São Paulo  
2023

Prof. Dr. Carlos Gilberto Carlotti Júnior  
Reitor da Universidade de São Paulo

Profa. Dra. Maria Dolores Montoya Diaz  
Diretora da Faculdade de Economia, Administração, Contabilidade e Atuária

Prof. Dr. João Maurício Gama Boaventura  
Chefe do Departamento de Administração

Prof. Dr. Felipe Mendes Borini  
Coordenador do Programa de Pós-Graduação em Administração

Paulo de Castro Rubio Poli

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

Dissertação apresentada ao Programa de Pós-Graduação em Administração do Departamento de Administração da Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo, como requisito parcial para obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Leandro dos Santos Maciel

**Versão Corrigida**

(versão original disponível na Biblioteca da Faculdade de Economia, Administração e Contabilidade)

São Paulo  
2023

Catlogação na Publicação (CIP)  
Ficha Catalográfica com dados inseridos pelo autor

Poli, Paulo.

Previsibilidade da Direção do Preço Intradiário do Bitcoin com Modelos de Random Forest / Paulo Poli. - São Paulo, 2023.

94 p.

Dissertação (Mestrado) - Universidade de São Paulo, 2023.

Orientador: Leandro dos Santos Maciel.

1. Bitcoin. 2. Previsão. 3. Dados Intradiários. 4. Random Forest. I. Universidade de São Paulo. Faculdade de Economia, Administração, Contabilidade e Atuária. II. Título.

Nome: Paulo de Castro Rubio Poli

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

Dissertação apresentada ao Programa de Pós-Graduação em Administração do Departamento de Administração da Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo, como requisito para obtenção do título de Mestre em Ciências.

Aprovado em:

Banca Examinadora

Professor Dr. Leandro dos Santos Maciel (Presidente)

Instituição: Faculdade de Economia, Administração, Contabilidade e Atuária, Universidade de São Paulo

Professora Dra. Rosangela Ballini

Instituição: Instituto de Economia, Universidade Estadual de Campinas

Julgamento:

Professor Dr. Reinaldo Augusto da Costa Bianchi

Instituição: Centro Universitário da FEI

Julgamento:

Professor Dr. João Vinicius de França Carvalho

Instituição: Faculdade de Economia, Administração, Contabilidade e Atuária, Universidade de São Paulo

Julgamento:

À Bianca e Alice.

## RESUMO

Poli, P. C. R. (2023). *Previsibilidade da Direção do Preço Intradiário do Bitcoin com Modelos de Random Forest* (Dissertação de Mestrado). Faculdade de Economia, Administração, Contabilidade e Atuária, Universidade de São Paulo, São Paulo.

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*.

## ABSTRACT

Poli, P. C. R. (2023). *Predicting Bitcoin Intraday Price Direction using Random Forest Models* (Dissertação de Mestrado). Faculdade de Economia, Administração, Contabilidade e Atuária, Universidade de São Paulo, São Paulo.

The Bitcoin trading market has experienced a fast growth since 2019 and draw the attention of institutional and individual investors. Given the high volatility of the digital coin, accurate forecasting about the future direction of its prices has become more important to market's investors. The complex dynamics of the cryptocurrency prices requires the use of sophisticated machine learning techniques to perform the predictions. In this context, this work aims to evaluate the predictability of Bitcoin's intraday price direction, during the time period ranging from January 2020 to December 2022. For this purpose, random forest models are used to predict whether the cryptocurrency price would increase or decrease in the predictive horizons of one minute, five minutes, fifteen minutes, one hour, six hours and one day. The model features refer to returns of the own series using the cryptocurrency's close, high and low trading prices. The random forest predictions for the digital coin were compared with ARIMA and logistic regression predictions, in terms of predictive power metrics. The results point to violations in the empirical verification of the weak form efficient market hypothesis for Bitcoin's intraday prices. The coin's intraday movements were predicted by random forest models with higher accuracy compared to a random walk model and the competitors in different intraday predictive horizons.

**Keywords:** Bitcoin. Forecast. Intraday data. Random Forest.

## LISTA DE FIGURAS

Figura 1 - Exemplo de uma árvore de decisão .....	35
Figura 2 - Matriz de Confusão.....	53
Figura 3 - Divisão da base de dados nos 12 distintos conjuntos de treinamento/teste para aferição da capacidade preditiva dos modelos de previsão .....	59
Figura 4 - Evolução do preço de fechamento diário do Bitcoin no período de janeiro de 2020 a 2022 .....	60

## LISTA DE TABELAS

Tabela 1 - Resumo da revisão da literatura que considera modelos de aprendizado de máquinas na previsão de séries temporais financeiras .....	28
Tabela 2 - Parâmetros de controle da função “ <i>randomForest()</i> ” do pacote “ <i>randomForest</i> ” do <i>software</i> RStudio.....	43
Tabela 3 - Estatísticas descritivas dos retornos intradiários do Bitcoin no período de janeiro de 2020 e dezembro de 2022 .....	60
Tabela 4 - Resultados das simulações dos modelos RF com diferentes configurações de parâmetros de controle.....	63
Tabela 5 - Importância média de cada atributo - RF.....	64
Tabela 6 - 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 nos anos de 2020 a 2022 .....	66
Tabela 7 - Valores-p ( <i>p-values</i> ) dos testes de proporção para a acurácia dos modelos RF, realizados para cada conjunto de teste nos diferentes horizontes de previsão .....	68
Tabela 8 - Porcentagem de valorização do Bitcoin por faixas de previsão das saídas dos modelos RF nos diferentes horizontes de previsão .....	70
Tabela 9 - Comparativo dos valores de acurácia entre os trabalhos presentes na literatura sobre previsão de direção do preço do Bitcoin no horizonte intradiário utilizando modelos <i>Random Forest</i> .....	71
Tabela 10 - Comparativo do poder preditivo dos modelos RF, regressão logística e ARIMA.....	74
Tabela 11 - Ordenação do poder preditivo entre os modelos RF, regressão logística e ARIMA por horizonte de previsão .....	76
Tabela 12 - Comparativo da porcentagem de valorização do Bitcoin por faixas de previsão dos modelos RF e regressão logística no horizonte de previsão de 15 minutos.....	76

## LISTA DE SIGLAS

ANN	<i>Artificial Neural Network</i>
ARIMA	<i>Autoregressive Integrated Moving Average</i>
AUC	<i>Area Under the Curve</i>
BR	<i>Bayesian Regularization</i>
CLSTM	<i>Convolutional Long Short-Term Memory</i>
CM	<i>Chaos Modeling</i>
CNN	<i>Convolutional Neural Network</i>
CQR	<i>Chaos Quantile Regression</i>
CRF	<i>Chaos Random Forest</i>
EMH	<i>Efficient Market Hypothesis</i>
EV	<i>Ensemble Voting</i>
FF	<i>Feedforward</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
FPR	<i>False Positive Rate</i>
GAM	<i>Generalized Additive Models</i>
GB	<i>Gradient Boosting</i>
GRP	<i>Gaussian Poisson Regressions</i>
GRU	<i>Gated Recurrent Unit</i>
KC	<i>Keltner Channels</i>
kNN	<i>k-Nearest Neighbors</i>
LARFR	<i>Random Forest Lagged-Auto-Regression</i>
LR	<i>Logistic Regression</i>
LSTM	<i>Long Short-Term Memory</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MLP	<i>Multilayer Perceptron</i>
MSE	<i>Mean Squared Error</i>
MSPE	<i>Mean Squared Prediction Error</i>
NB	<i>Naive Bayes</i>
OHLC	<i>Open, High, Low, Close</i>
PSO	<i>Particle Swarm Optimization</i>

QRRF	<i>Quantile Regression Random Forest</i>
RBF	<i>Radial Basis Function</i>
RF	<i>Random Forest</i>
RSI	<i>Relative Strength Index</i>
RT	<i>Regression Trees</i>
SVC	<i>Support Vector Classifier</i>
SVM	<i>Support Vector Machine</i>
SVR	<i>Support Vector Regression</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
TPR	<i>True Positive Rate</i>
VMD	<i>Variational Mode Decomposition</i>

## SUMÁRIO

1 INTRODUÇÃO.....	13
1.1 Objetivos.....	14
1.2 Justificativas.....	14
1.3 Organização do trabalho .....	15
2 FUNDAMENTAÇÃO TEÓRICA .....	16
2.1 Bitcoin.....	16
2.2 Relevância dos Modelos de <i>Random Forest</i> .....	18
2.3 Hipóteses de Pesquisa.....	20
3 REVISÃO DE LITERATURA .....	21
4 METODOLOGIA.....	30
4.1 Formulação do Problema .....	30
4.2 Árvores de Decisão .....	32
4.2.1 Viés e Variância.....	32
4.2.2 Formulação de Árvores de Decisão.....	34
4.3 <i>Random Forest</i> .....	39
4.4 Modelos Competidores .....	45
4.4.1 ARIMA.....	45
4.4.2 Regressão Logística.....	51
4.5 Métricas de Desempenho.....	53
5 RESULTADOS .....	56
5.1 Análise Descritiva dos Dados .....	56
5.2 Modelagem e Resultados – <i>Random Forest</i> .....	61
5.3 Comparativo dos Resultados dos Modelos.....	73
6 CONCLUSÃO.....	78
REFERÊNCIAS .....	80
APÊNDICES .....	84



## 1 INTRODUÇÃO

O mercado de criptomoedas passou por um vertiginoso crescimento nos últimos anos. De acordo com a CoinMarketCap<sup>1</sup>, a capitalização total do mercado passou de US\$ 1,3 bilhão, em 1 de maio de 2013, atingindo US\$ 3,0 trilhões em 10 de novembro de 2021. O Bitcoin, principal criptomoeda negociada, com maior capitalização de mercado, obteve uma valorização particularmente expressiva: nessas mesmas datas, sua cotação passou de US\$ 105 para US\$ 68.789. Em 31 de dezembro de 2022, a cotação e o valor de mercado do Bitcoin já haviam recuado para US\$ 16.547 e US\$ 318,5 bilhões, respectivamente.

Analisar séries temporais financeiras, como o preço de ativos tradicionais e criptomoedas, é extremamente desafiador devido à natureza dinâmica, não linear, não estacionária, ruidosa e caótica do mercado financeiro (Borges & Neves, 2020). Em um contexto de alta volatilidade do preço de criptomoedas, um modelo de previsão acurado e confiável é fundamental para investidores e participantes do mercado (Gyamerah, 2019).

A forma mais intuitiva de abordar a tarefa de previsão em séries temporais financeiras é pela previsão numérica de seus preços, 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, Daouk, & Chen, 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 são utilizados dados de mercado financeiro para antecipação da direção de preços.

As características complexas e o considerável volume de dados de séries temporais financeiras levaram à 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

---

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

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, Tibshirani, Friedman e Friedman (2009) afirmam, por exemplo, que modelos de florestas aleatórias (*Random Forest* – RF) conseguem modelar adequadamente dados ruidosos, são altamente robustos ao sobreajuste (*overfitting*<sup>2</sup>) e requerem menos parâmetros que modelos de reforço de gradiente (*Gradient Boosting*<sup>3</sup> – GB).

## 1.1 Objetivos

Considerando o contexto anteriormente apresentado, o objetivo desta dissertação é analisar a previsibilidade da direção do preço da criptomoeda mais negociada no mercado, o Bitcoin, na frequência intradiária, nos horizontes de previsão de 1 (um) minuto, 5 (cinco) minutos, 15 minutos, 60 minutos (uma hora), 360 minutos (seis horas) e 1440 minutos (um dia). Para atingir o objetivo, foram aplicados modelos de *random forest* utilizando como atributos (variáveis explicativas) defasagens dos retornos calculados com base nas séries de preços de fechamento, máximos e mínimos, estratégia tradicionalmente utilizada nos trabalhos de previsão de séries temporais. Os resultados foram comparados em termos de poder preditivo com *benchmarks*: os modelos regressão logística e autorregressivo integrado e de médias móveis (ARIMA).

## 1.2 Justificativas

A revisão de literatura realizada neste trabalho<sup>4</sup> aponta que a questão da previsibilidade dos preços de criptomoedas na frequência intradiária ainda não foi

---

<sup>2</sup> *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.

<sup>3</sup> Modelos de *gradient 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*).

<sup>4</sup> Revisão de literatura realizada pela busca de artigos, dissertações, teses e livros nas principais bases de dados de trabalhos científicos, como Science Direct, Web of Science, Scielo, Scopus, Dedalus USP e Google Acadêmico, contendo combinações das expressões mais relevantes desta pesquisa - “bitcoin”, “previsão”, “intradário”, “*random forest*”, “hipótese de eficiência de mercado” - e outras expressões

analisada de forma abrangente e que os resultados destes trabalhos não são conclusivos. Mais especificamente, não é abrangente a pesquisa existente em relação ao uso do modelo *random forest* na previsão de movimentos de preços de criptomoedas.

Este trabalho busca enriquecer a literatura do tema principalmente pela apresentação de novas evidências sobre a aplicação de modelos de aprendizado de máquina, em especial *random forest*, na tarefa de previsão de movimentos futuros do Bitcoin e, em particular, na frequência intradiária.

Foi levada em consideração a busca por modelos parcimoniosos, ou seja, que fossem ao mesmo tempo menos complexos, com a utilização de um conjunto reduzido de atributos, mas que apresentassem também capacidade de poder preditivo compatível com os resultados dos trabalhos pesquisados na fundamentação teórica.

Adicionalmente, este trabalho propõe um nível maior de análise e compreensão das saídas que os modelos apresentam em relação aos trabalhos investigados sobre o tema, por meio da avaliação da acurácia segmentada pela probabilidade de alta ou baixa (saída do modelo de *random forest*). Ainda, houve notável atenção para a robustez dos resultados, com a consideração de um número maior de amostras teste, com diferentes dinâmicas, assim como com base em diferentes horizontes preditivos intradiários.

### 1.3 Organização do trabalho

Esta dissertação está organizada da seguinte forma. O Capítulo 2 apresenta o referencial teórico da pesquisa, contendo a discussão sobre o aumento do interesse do público pela criptomoeda Bitcoin, uma introdução aos modelos de *random forest*, e a relação da previsibilidade de preços de ativos financeiros com a hipótese de eficiência de mercado em sua forma fraca. 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 na frequência intradiária, é discutida no Capítulo 3. No Capítulo 4, são descritos os modelos utilizados neste trabalho (*random forest*, regressão logística e ARIMA), bem como as métricas de avaliação dos resultados. O Capítulo 5 apresenta os resultados e as discussões. Por fim, as conclusões e indicações de tópicos para pesquisas futuras compreendem o Capítulo 6.

---

derivadas, como “previsibilidade”, “criptomoedas”, “ativos financeiros”, “modelos de aprendizado de máquina” e “mercado eficiente”, tanto em língua portuguesa como em inglesa.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentadas as bases teóricas em que o trabalho está pautado, sendo abordados temas que vão desde a definição de Bitcoin e de criptomoedas, até os motivos que tornaram os modelos baseados em *random forest* potencialmente adequados à tarefa de previsão de preços e movimentos de séries temporais financeiras.

Primeiramente, são elencados alguns motivos que provocaram a atenção da academia e do público em geral pelas criptomoedas, tendo como destaque a valorização expressiva de seus preços e o aumento do volume negociado nas corretoras. Em seguida, são abordadas tentativas de pesquisadores em classificar as moedas digitais em alguma categoria de ativo financeiro, aspecto esse que ainda não apresenta consenso na literatura.

Posteriormente, é elaborada uma discussão sobre a dinâmica das séries temporais de preços de criptomoedas, com destaque para as séries intradiárias. A literatura aponta que as características próprias dessas séries inviabilizam a aplicação dos modelos tradicionais de previsão, o que abre caminho para a busca de técnicas mais complexas e adequadas para realização de tal tarefa com um maior desempenho. Na discussão acerca de previsibilidade é inserida sua associação com a hipótese de mercados eficientes, muito utilizada em modelos da teoria moderna de finanças. O capítulo segue com a indicação de estudos para a utilização de modelos de *random forest* na tarefa da previsão de movimentos futuros da moeda digital. Por fim, é tratada a importância prática da utilização de dados em alta frequência (prazos curtos de tempo) e suas aplicações como o *high-frequency trading*.

### 2.1 Bitcoin

Além do significativo crescimento, em termos de volume negociado, e da alta capitalização de mercado das criptomoedas, pesquisadores elencaram outros motivos pelos quais as moedas digitais atraíram a atenção de pesquisadores e investidores. Alguns motivos incluem: suas características inovadoras (Mallqui & Fernandes, 2019); a alta volatilidade de preços (Huang, Huang, & Ni, 2019); o alto potencial de ganhos financeiros com o aumento de seus preços (Maciel & Ballini, 2019); a capacidade de antecipação de choques negativos do mercado e seu uso como proteção (*hedge*) contra ativos de risco (Dyhrberg, 2016); o formato seguro de realizar transações digitais com anonimato (Huang et al., 2019); e o ecossistema do tipo *blockchain*, sistema par a par (*peer-to-peer*), de

natureza não regulamentada pelo governo e com baixos custos de transação (Alonso-Monsalve et al., 2020).

No ano de 2008, o artigo intitulado “*Bitcoin: A Peer-to-Peer Electronic Cash System*”, publicado por um autor de pseudônimo Satoshi Nakamoto (2008)<sup>5</sup>, descreveu, pela primeira vez, o Bitcoin e introduziu o conceito de criptomoeda descentralizada (Borges & Neves, 2020). Criptomoedas são um tipo de ativo digital baseado em tecnologias e protocolos criptografados, a *blockchain*, e são executadas em redes descentralizadas tornando as transações seguras e difíceis de serem falsificadas (Alonso-Monsalve et al., 2020).

Gyamerah (2019) tratou do assunto de classificar as moedas digitais em alguma categoria de ativo financeiro, afirmando que as grandes flutuações no preço do Bitcoin ameaçam seu potencial de funcionar como reserva de valor, em decorrência da sua alta volatilidade. Similarmente, Maciel e Ballini (2019) expõem que a dinâmica do preço do Bitcoin promove o incentivo de ações especulativas, com o objetivo de obter lucros com a variação de seus preços.

Em uma diferente análise, Ibrahim, Kashef e Corrigan (2021) mencionam que o Bitcoin e as criptomoedas parecem estar mais relacionados com ações do que com moedas tradicionais, utilizadas para troca de mercadorias. Os autores adicionam que, como essas moedas digitais possuem maior volatilidade que moedas tradicionais, são mais parecidas com pequenas ações (*penny stocks*<sup>6</sup>).

Alonso-Monsalve et al. (2020), por outro lado, defendem que as criptomoedas constituem uma nova classe de ativo, com mais elementos em comum com *commodities* especulativas do que com moedas de troca, pelo fato de o valor das moedas digitais não ser baseado em nenhum ativo tangível. Similarmente, Jaquart, Dann e Weinhardt (2021) indicam, com base na literatura, que o Bitcoin pode simplesmente representar uma nova classe de ativos.

Em relação à dinâmica temporal das séries, pode-se dizer que os preços de ativos financeiros são geralmente caracterizados pela presença de ruídos, de um alto grau de incertezas e de interrelações não identificadas. A previsão de tais séries temporais, na prática, é uma tarefa desafiadora, em vista de sua não estacionariedade,

---

<sup>5</sup> Não existe certeza na literatura sobre a autoria deste artigo.

<sup>6</sup> *Penny stocks* é um termo utilizado no mercado financeiro em referência a ações negociadas a um baixo preço de mercado.

heterocedasticidade<sup>7</sup>, descontinuidade e presença de valores atípicos (*outliers*). De forma geral, observa-se que o mercado de criptomoedas é conhecido por suas significativas flutuações de preços, em outras palavras, por sua alta volatilidade, em adição às comuns características das séries de preços de ativos tradicionais (Borges & Neves, 2020).

Nesse sentido, o desenvolvimento de modelos preditivos acurados para ativos financeiros assume papel de destaque, sobretudo sobre os ativos que apresentam volatilidade considerável, como é o caso das criptomoedas. Dentre as técnicas de previsão da literatura, os modelos de aprendizado de máquina (*machine learning* – ML) ganham destaque por apresentarem flexibilidade e bom desempenho para o tratamento de séries temporais complexas. Algoritmos de aprendizado de máquina ganharam o interesse de acadêmicos e atuantes do mercado devido à habilidade de capturarem dinâmicas não lineares, sobretudo na predição de movimentos futuros de preços, sem dependerem de premissas estatísticas tradicionais acerca dos dados (Alonso-Monsalve et al., 2020). Nesta pesquisa, dentre os modelos de *machine learning*, a técnica de *random forest* é adotada para análise, e as justificativas dessa escolha são apresentadas a seguir.

## 2.2 Relevância dos Modelos de *Random Forest*

Modelos de *random forest*, uma das técnicas de aprendizado de máquina, vêm recebendo uma maior atenção de pesquisadores (Krauss, Do, & Huck, 2017; Gyamerah, 2019; Borges & Neves, 2020; Akyildirim, Goncu, & Sensoy, 2021; Ibrahim et al., 2021; Jaquart et al., 2021; Ghosh, Neufeld, & Sahoo, 2022; Gradojevic, Kukolj, Adcock, & Djakovic, 2023; Orte, Mira, Sánchez, & Solana, 2023) e investidores, em virtude dos seus benefícios (que serão melhor explicados na seção de metodologia) na modelagem de séries temporais financeiras.

O termo *random forest* se refere a uma abordagem derivada dos modelos de árvores de decisão (*Decision Trees* – DT) e, por isso, herdou suas principais características. Huang et al. (2019) salientam que a árvore de decisão é robusta a dados ruidosos e capaz de aprender expressões disjuntivas. Jaquart et al. (2021) apontam que as técnicas baseadas em árvores se destacam pelo baixo custo computacional e pela geração de modelos interpretáveis.

---

<sup>7</sup> Variância não constante dos resíduos de um modelo.

Krauss et al. (2017) apresentam algumas qualidades da abordagem RF, declarando que esse método é mais simples de ser aplicado que o *boosting*, reduz substancialmente a variância das árvores de decisão, é robusto a ruídos nas variáveis e ainda reduz o problema do viés<sup>8</sup> (*bias*). Gyamerah (2019) reforça que, devido ao princípio da lei forte dos grandes números<sup>9</sup> (*strong law of large numbers*), o sobreajuste não é um problema para o RF.

Uma discussão clássica na literatura de finanças se dá em torno da hipótese de eficiência de mercado (*efficient market hypothesis* – EMH), que trata da possibilidade de preços de ativos financeiros serem previstos. Na definição clássica, Fama (1970) caracterizou um mercado eficiente como aquele cujos preços refletem totalmente as informações disponíveis aos interessados do mercado. Em relação ao comportamento dos preços, o autor afirma que estes seguem o chamado passeio aleatório, não sendo possível, portanto, que os agentes criem mecanismos para a obtenção de lucros anormais em períodos de tempo pré-determinados.

Fama (1970) debateu também a questão da medida em que os preços de mercado “refletem totalmente” conjuntos particulares de informações relevantes disponíveis. Ele propôs três formas de eficiência de mercado, que tratam do efeito da disponibilidade dessas informações na formação dos preços de mercado: (i) a forma forte de eficiência, que debate se algum investidor individual ou institucional tem acesso privativo a qualquer informação recente relevante à formação dos preços de mercado; (ii) a forma semi-forte de eficiência, que aborda a velocidade de ajuste de preços em relação a outras informações públicas, como anúncio de desdobramento de ações, reportes anuais e emissão de novos títulos; e (iii) a forma fraca de eficiência, em que o conjunto de informações de interesse consiste somente de retornos ou preços passados. Assim sendo, este trabalho busca constatar violações na hipótese fraca de mercado, demonstrando a possibilidade de prever a direção do preço de um ativo financeiro por meio de modelos que utilizam apenas preços e retornos passados como entradas.

---

<sup>8</sup> Viés, ou *bias*, é uma medida de erro calculada pela diferença entre os valores estimados por um modelo e os valores reais de uma variável prevista. É uma medida de poder preditivo de modelos.

<sup>9</sup> A lei forte dos grandes números é um teorema fundamental da teoria da probabilidade que descreve o resultado da realização de um mesmo experimento repetidas vezes. De acordo com o teorema, a média aritmética dos resultados da realização da mesma experiência repetidas vezes tende a se aproximar de seu valor esperado à medida em que mais tentativas são realizadas.

A literatura recente aponta que ao menos a forma fraca de eficiência pode ser rejeitada em determinadas situações. Akyildirim et al. (2021) estudaram e apresentaram diversos estudos sobre o Bitcoin em que a hipótese de forma fraca de eficiência foi violada. Os autores concluem que os preços desse ativo são ineficientes, mas que estão se alterando rumo à forma fraca de eficiência ao longo do tempo.

### 2.3 Hipóteses de Pesquisa

Este trabalho avalia a previsibilidade da direção futura dos preços de fechamento do Bitcoin, na frequência intradiária, com base em modelos de *random forest*. Desse modo, as seguintes hipóteses são avaliadas nesta pesquisa:

- *H1*: a direção futura do preço do Bitcoin, na frequência intradiária, pode ser prevista por meio de modelos de *random forest*, com resultados superiores aos obtidos por um modelo de passeio aleatório.
- *H2*: Os modelos de *random forest* apresentam capacidade preditiva superior a modelos tradicionais de econometria de séries temporais, mais especificamente, modelos de regressão logística e ARIMA.

Em relação à primeira hipótese, *H1*, foram realizadas previsões da direção do preço intradiário do Bitcoin utilizando modelos de *random forest* para os horizontes de previsão de 1 (um) minuto, 5 (cinco) minutos, 15 minutos, 60 minutos (uma hora), 360 minutos (seis horas) e 1440 minutos (um dia). São consideradas diferentes amostras teste para avaliação da robustez dos resultados.

*H1* avalia a hipótese de eficiência de mercado em sua forma fraca, como citado por Fama (1970), que atesta um comportamento de um passeio aleatório para os retornos do preço de ativos financeiros, isto é, a imprevisibilidade de preços. Assim, um teste de hipótese para proporções visa mensurar se o poder preditivo dos modelos é estatisticamente superior a 50%, ou seja, superior a um preditor ingênuo, como o caso do passeio aleatório.

A segunda hipótese de pesquisa, *H2*, trata da questão da superioridade do poder preditivo dos modelos de *random forest* quando comparados aos modelos de regressão logística e ARIMA, clássicos da literatura de econometria de séries temporais. Para tanto, a acurácia dos modelos foi comparada em termos estatísticos.

### 3 REVISÃO DE LITERATURA

Diferentes autores investigaram a previsibilidade de preços de ativos financeiros na frequência intradiária com base em modelos de aprendizado de máquina (Nelson, Pereira, & Oliveira, 2017; Gyamerah, 2019; Alonso-Monsalve et al., 2020; Borges & Neves, 2020; Lahmiri & Bekiros, 2020; Akyildirim et al., 2021; Ibrahim et al., 2021; Jaquart et al., 2021; Akyildirim, Çepni, Corbet, & Uddin, 2022; Ghosh et al., 2022; Gil, 2022; Gradojevic et al., 2023; Orte et al., 2023). Com maior intensidade, algumas pesquisas utilizaram modelos baseados em redes neurais (Nelson et al., 2017; Alonso-Monsalve et al., 2020; Lahmiri & Bekiros, 2020; Akyildirim et al., 2021; Ibrahim et al., 2021; Jaquart et al., 2021; Akyildirim et al., 2022; Gradojevic et al., 2023).

Nelson et al. (2017), por exemplo, investigaram a capacidade de previsão da direção do preço de cinco diferentes ações pertencentes à bolsa de valores brasileira (B3) no período de 2008 a 2015, utilizando dados com frequência de 15 minutos. Os autores consideraram como variáveis explicativas diferentes indicadores de análise técnica, enquanto a avaliação dos modelos se deu com base nas métricas de acurácia, precisão, sensibilidade e uma medida-F, calculada como a média harmônica entre previsão e sensibilidade. Os resultados indicaram que os modelos rede neural de memória de curto e longo prazo (*Long Short-Term Memory – LSTM*), rede neural do tipo perceptron (*Multilayer Perceptron – MLP*) e de *random forest* são acurados, com destaque para a LSTM.

Na mesma linha, a pesquisa de Gil (2022) demonstrou que modelos baseados em indicadores técnicos de dados com frequência de um minuto podem ser utilizados para propiciar operações lucrativas no mercado futuro de metais preciosos. O autor desenvolveu dois sistemas de negociação baseados, respectivamente, nos valores dos indicadores índice de força relativa (*Relative Strength Index – RSI*) e canais de Keltner (*Keltner Channels – KC*). O autor aplicou o método de otimização por enxame de partículas (*Particle Swarm Optimization – PSO*) para encontrar a melhor configuração de parâmetros de cálculo de cada indicador. Exemplos de parâmetros presentes nesses indicadores são o número de barras de negociação<sup>10</sup> e o período de tempo compreendido

---

<sup>10</sup> Uma barra de negociação representa graficamente o movimento de uma ação em certo período. Por exemplo, uma barra de negociação de prazo de um minuto retrata os preços de abertura, máximo, mínimo e de fechamento alcançados pela cotação de um ativo no horizonte de tempo de um minuto.

em cada uma dessas barras. Os resultados do trabalho, avaliados por meio de diferentes métricas de risco e retorno, exibiram operações lucrativas tanto em posições compradas como em vendidas, e demonstraram que esses modelos podem também ser utilizados na gestão de carteiras de outros ativos arriscados, como ações, moedas e outras *commodities*.

Com o passar do tempo, estudos sobre previsibilidade de preços específicos para o mercado de criptomoedas foram sendo publicados. O Bitcoin foi o principal alvo desses trabalhos.

Gyamerah (2019) 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 (*Variational Mode Decomposition – VMD*) e modelos aditivos generalizados (*Generalized Additive Models – GAM*). Com base de dados do período entre 1 de janeiro de 2019 e 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. Esse resultado foi superior ao 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 modelos foram categorizados em três tipos: (i) modelos estatísticos de aprendizado de máquina, contendo regressões de vetores suporte (*Support Vector Regression – SVR*) e regressões de Poisson gaussianas (*Gaussian Poisson Regressions – GRP*); (ii) modelos baseados em algoritmos, abrangendo árvores de regressão (*Regression Trees – RT*) e o método dos k-vizinhos mais próximos (*k-Nearest Neighbors – kNN*); e (iii) modelos de redes neurais, incluindo redes neurais do tipo *feedforward* (*Feedforward Neural Network – FFNN*), rede neural com regularização bayesiana (*Bayesian Regularization – BR*) e rede neural com função de base radial (*Radial Basis Function – RBF*). Os autores apontaram que todos os modelos proporcionaram previsões acuradas, sendo efetivos em capturar a característica ruidosa, não linear e complexa da série temporal em questão. O modelo BR apresentou os melhores resultados, seguido por FF, GRP, SVR, RT, RBF e kNN.

Os resultados das pesquisas mencionadas apontam para a adequação de diversos modelos de aprendizado de máquina na tarefa da previsão de preços e movimentos de diferentes ativos financeiros, incluindo o Bitcoin. Os modelos baseados em *random forest*, objeto de pesquisa deste trabalho, também foram explorados por alguns autores em pesquisas relativas ao tema das moedas digitais.

Por exemplo, 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 na frequência de um dia e previsões de um passo à frente (um dia), os autores propuseram o modelo QRRF-CM, composto pelos modelos de regressão quantílica de floresta aleatória (*Quantile Regression Random Forest – QRRF*) e de teoria do caos (*Chaos Modeling – CM*), e compararam seus resultados com os de outros dois modelos, denominados regressão quantílica caótica (*Chaos Quantile Regression – CQR*) e floresta aleatória caótica (*Chaos Random Forest – CRF*). A base de dados se referiu ao período de janeiro de 2000 a dezembro de 2013 e os resultados foram avaliados por duas métricas: erro quadrático médio (*Mean Squared Error – MSE*) e erro percentual absoluto médio (*Mean Absolute Percentage Error – MAPE*). O teste Diebold-Mariano (Diebold & Mariano, 1995) foi utilizado na comparação estatística da acurácia dos diferentes modelos. Os resultados apontaram que o modelo proposto rendeu previsões mais acuradas que os de referência e que o QRRF-CM pode ser explorado em outras séries temporais financeiras.

Ghosh et al. (2022), de outra forma, estudaram a previsibilidade da direção intradiária dos preços das ações participantes do índice S&P500, composto por 500 ativos cotados nas bolsas de NYSE ou NASDAQ. O período analisado 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. Por meio de métricas de desempenho de carteiras de investimento, os autores encontraram resultados superiores aos apresentados por Krauss et al. (2017). As previsões dos modelos RF e LSTM tiveram maiores retornos diários e maior taxa de retornos

positivos, enquanto as estratégias de investimento, que consideraram custos de transação compatíveis com os praticados pelo mercado, obtiveram maior retorno com menor risco.

O aumento do interesse de pesquisadores sobre o tema das criptomoedas desencadeou o surgimento de publicações de estudos específicos sobre previsão de preços e movimentos dessas moedas digitais com frequência intradiária.

Alonso-Monsalve et al. (2020) estudaram a previsão da direção do preço de seis populares criptomoedas (Bitcoin, Dash, Ether, Litecoin, Monero e Ripple) em cada minuto próximo. Com dados de 1 de julho de 2018 a 30 de junho de 2019 com frequência de um minuto, os autores utilizaram como variáveis explicativas 18 indicadores técnicos e compararam quatro diferentes arquiteturas de redes neurais – rede neural convolucional (*Convolutional Neural Network* – CNN), rede neural convolucional híbrida customizada (*Custom Hybrid Convolutional* – CLSTM), rede neural MLP e rede neural com função de base radial (*Radial Basis Function* – RBF). Os resultados mostraram que as séries podem ser previstas durante todo o período analisado por meio desses modelos e que a arquitetura CLSTM proporciona os melhores resultados em termos de acurácia.

Com o objetivo de gerar investimentos com maior retorno e menor risco que a estratégia simplista *Buy & Hold*<sup>11</sup>, Borges e Neves (2020) propuseram uma estratégia de compra e venda de criptomoedas com base em diferentes modelos de aprendizado de máquina. Em um conjunto de 12 indicadores técnicos calculados com dados de julho de 2017 a outubro de 2018, os autores aplicaram os modelos *random forest*, regressão logística, classificador de vetores de suporte (*Support Vector Classifier* – SVC), *boosting* e um modelo combinado de votação utilizando todos os outros modelos (*Ensemble Voting* – EV). A avaliação dos resultados, realizada por meio de diferentes métricas de avaliação de previsão e de desempenho de carteiras, apontou que o modelo combinado de votação apresentou os melhores resultados. RF obteve o melhor desempenho entre os modelos individuais.

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 desenvolveram 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 a mesma periodicidade dos horizontes de previsão. A aplicação dos

---

<sup>11</sup> A estratégia *Buy & Hold*, ou comprar e manter (em tradução livre), se refere à simples compra do ativo em referência e sua manutenção em carteira por todo o período analisado.

diferentes modelos de classificação dos tipos máquina de vetores suporte (*Support Vector Machine* – SVM), regressão logística (*Logistic Regression* – LR), rede neural artificial (*Artificial Neural Network* – ANN) e *random forest*, 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%. O modelo SVM apresentou os melhores resultados.

Uma comparação de modelos sofisticados 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 e testaram os modelos ARIMA, Prophet (modelo de previsão disponibilizado gratuitamente pela equipe de ciência de dados da empresa Facebook), *random forest*, regressão autorregressiva com floresta aleatória (LARFR - *Random Forest Lagged-Auto-Regression*) e rede neural MLP. No estudo, a saída dos modelos representava a probabilidade de o preço do Bitcoin valorizar a cada 5 (cinco) minutos e os resultados foram comparados com as estratégias palpite aleatório (RG – *Random Guess*) e momento (repetição do movimento anterior). O modelo MLP apresentou os melhores resultados. Os autores afirmam que há uma oportunidade de investimento com a adoção da estratégia proposta, por meio de um robô, em opções binárias<sup>12</sup>.

Jaquart et al. (2021) estudaram a previsibilidade da direção do preço do Bitcoin nos intervalos de 1 (um), 5 (cinco), 15 e 60 minutos. Para tal, os autores utilizaram um conjunto amplo de variáveis explicativas – incluindo indicadores técnicos, de cadeia de blocos (*blockchain*), de sentimento e econômicos – como entrada em diferentes modelos de aprendizado de máquina: rede neural unidade de portão recorrente (*Gated Recurrent Unit* – GRU), rede neural LSTM, rede neural FF, *boosting* e *random forest*. Com observações entre março e dezembro de 2019 e previsões com periodicidade de 1 (um) minuto, os resultados indicam que todos os modelos estudados superaram um classificador aleatório e que a acurácia nas previsões tende a aumentar para períodos mais longos. Os indicadores técnicos se apresentaram como as variáveis explicativas mais importantes.

---

<sup>12</sup> Opções binárias são plataformas de negociação digitais em que os investidores apostam na direção correta de cada ativo em um determinado horizonte de tempo. Possuem como vantagem o fato de não haver o problema da liquidez dos mercados tradicionais, em que o investidor não consegue comprar ou vender um ativo instantaneamente pelo último preço negociado. Assim, essas plataformas são uma alternativa de investimento para quem faz uso de modelos de previsão de direção.

Akyildirim et al. (2022) avaliaram a habilidade de diferentes modelos de aprendizado de máquina (RF, kNN, *Naive Bayes* – NB, regressão logística, *random forest*, SVM e *boosting*) de previsibilidade do movimento dos preços do Bitcoin no mercado futuro nos horizontes de previsão de 5 (cinco), 15, 30 e 60 minutos. A partir de uma base de dados de janeiro de 2020 a setembro de 2020, nas mesmas periodicidades dos horizontes de previsão, os resultados mostraram que os modelos estudados apresentam melhor desempenho que os competidores ARIMA e passeio aleatório em previsões de movimentos de preços do Bitcoin no mercado futuro, com destaque para os modelos RF e kNN. Os resultados foram avaliados com o cálculo da métrica acurácia e também de medidas de desempenho de carteiras.

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, contendo 135 indicadores técnicos como entradas. Foram realizadas 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 de 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.

Gradojevic et al. (2023) abordaram a questão da predição do valor e da direção do preço do Bitcoin pelos modelos não lineares ANN, FFNN, SVM e RF nos horizontes de previsão de uma hora e um dia para o período de abril de 2013 a junho de 2019. Como atributos, foi elaborado um conjunto completo de 46 indicadores técnicos que representam tendência (*trend*), momento (*momentum*), sentimento (*sentiment*) e volume. Pela análise dos valores de métricas de acurácia e erro médio quadrático de previsão (*Mean Squared Prediction Error* – MSPE) das predições realizadas, os autores concluíram que os preços do Bitcoin atendem à hipótese fraca de eficiência no horizonte de previsão de uma hora, mas não no caso do horizonte de um dia. O modelo RF apresentou a maior acurácia entre os modelos utilizados.

Os estudos mencionados apontam que os modelos baseados em *random forest* podem ser considerados na tarefa da previsão de preços e movimentos do Bitcoin e de outros ativos financeiros em geral. Os resultados dessas pesquisas, avaliados por diferentes métricas de desempenho, exibem resultados superiores aos considerados clássicos do tema, como o ARIMA e a regressão logística. A Tabela 1 resume os

principais estudos que avaliaram a previsibilidade de ativos financeiros utilizando modelos de aprendizado de máquinas.

A Tabela 1 mostra a escassez de trabalhos que tratam da previsibilidade dos preços de criptomoedas na frequência intradiária. Destaca-se que o modelo RF foi pouco examinado pelos autores. Jaquart et al. (2020) realizaram uma revisão bibliográfica da adequação de métodos de aprendizado de máquina para problemas de predição do Bitcoin e relataram que, dentre os 33 artigos selecionados para análise, apenas oito possuíam frequência menor que uma hora, e apenas sete deles utilizaram modelos baseados em árvore, como o RF. Dentre esses estudos, apenas três abordaram simultaneamente modelos baseados em árvores de decisão e prazos intradiários.

Além dos pontos discutidos, a leitura dos trabalhos anteriores indica também problemas com possibilidades de melhoria/avanço na literatura, tais como: a utilização de um conjunto excessivo de variáveis explicativas (Nelson et al., 2017; Gyamerah, 2019; Huang et al., 2019; Basher & Sadorsky, 2022; Ghosh et al., 2022), que, sem o tratamento correto, pode gerar sobreajuste; falta de atenção à evolução temporal das observações na divisão dos bancos de dados para treinamento e para teste (Alonso-Monsalve et al., 2020; Basher & Sadorsky, 2022); períodos de teste longos sem atualização nos modelos (Pradeepkumar & Ravi, 2016; Gyamerah, 2019; Huang et al., 2019; Mallqui & Fernandes, 2019; Lahmiri & Bekiros, 2020; Ibrahim et al., 2021; Jaquart et al., 2021; Gil, 2022); escassez de trabalhos com janelas móveis de dados (Krauss et al., 2017; Nelson et al., 2017; Ghosh et al., 2022); e presença de sobreajuste nos resultados (Alonso-Monsalve et al., 2020; Ibrahim et al., 2021; Gradojevic et al., 2023; Orte et al., 2023).

Tabela 1 - Resumo da revisão 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	variação do preço	VMD-GAM	EMD-GAM	VMD-GAM	01jan19 - 24jan19	-	RMSE, MAPE, Bias
Alonso-Monsalve et al. (2020)	6 criptomoedas	1 minuto	1 minuto	direção do preço	CNN, CLSTM, MLPNN, RBFNN	RW, categoria mais frequente	CLSTM	jul18 - jun19	18 indicadores técnicos	acurácia
Borges e Neves (2020)	100 criptomoedas	método novo de reagrupamento (base 1min)	método novo de reagrupamento (base 1min)	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
Ibrahim et al. (2021)	Bitcoin	5 minutos	5 minutos	direção do preço	ARIMA, Facebook Prophet, RF, LARFR e MLPNN	RG, estratégia momento	MLPNN	2014 - não revelado	9 indicadores técnicos	acurácia, teste McNemar

Continua

<b>Autores</b>	<b>Ativos Utilizados</b>	<b>Periodicidade dos Dados</b>	<b>Horizonte de Previsão</b>	<b>Tipo de Previsão</b>	<b>Métodos</b>	<b>Métodos Comparados</b>	<b>Melhor Modelo</b>	<b>Período</b>	<b>Variáveis Explicativas</b>	<b>Métricas de Avaliação</b>
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, random walk	RF e kNN	jan20 - set20	5 retornos passados	acurácia, medidas de desempenho de carteira
Ghosh et al. (2022)	500 ações - S&P 500	1 dia	1 dia	direção do preço	LSTMNN, RF	-	LSTMNN	jan93 - dez18	93 variáveis de retornos passados	métricas estatística, de risco, de desempenho de carteiras
Gil (2022)	5 futuros de metais preciosos	1 minuto	variável	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, random walk, 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

Fonte: elaborado pelo autor (2023).

Conclusão

## 4 METODOLOGIA

Este capítulo apresenta os procedimentos metodológicos. Inicia-se com a formulação do problema de pesquisa, contendo as notações utilizadas, a caracterização do problema de classificação e a função de previsão dos modelos de *random forest*. Em seguida, é realizada a explanação sobre os modelos de árvores de decisão, apontando suas principais características, vantagens e limitações. A próxima subseção trata dos modelos de *random forest*, abrangendo seus fundamentos, formulação matemática, um pseudo-código e a descrição dos parâmetros de controle. Em seguida, os modelos competidores – ARIMA e regressão logística – são descritos. Por fim, são apresentadas as métricas de desempenho utilizadas para a avaliação do poder preditivo dos modelos.

### 4.1 Formulação do Problema

O objetivo principal 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  $y_n$ .

Uma tarefa de previsão pode ser entendida como um problema de regressão ou um problema de classificação, dependendo se o atributo em questão for numérico (regressão) ou categórico (classificação). Este trabalho trata de um problema de classificação, pois o rótulo é categórico, representado por uma variável binária, ou seja,  $y_n = \{0, 1\}, \forall n$ , em que o valor da unidade está associado a uma valorização do ativo (retorno positivo), enquanto o valor zero representa uma desvalorização (retorno negativo).

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, especificamente, a função de previsão para cada rótulo  $y_n$  modela a probabilidade do Bitcoin se valorizar em determinado horizonte de previsão, dada uma observação de atributos  $\mathbf{x}_n \in X$ , da seguinte forma:

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

Portanto, a função de previsão  $\hat{f}_{RF}(\mathbf{x}_n)$  abrange valores no intervalo  $[0, 1]$ , em que um valor próximo de 1 (um) indica maior probabilidade de valorização, enquanto um valor próximo de 0 (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, o valor de 0,5 como limiar, ou seja, valores da função preditora maiores que 0,5 são associados à categoria/rótulo “valorizar”, enquanto valores menores ou iguais a 0,5 são associados ao rótulo “desvalorizar”.

A avaliação do desempenho de modelos de classificação demanda alguma forma de medição da correspondência entre os valores reais e os valores preditos. A qualidade da previsão  $\hat{f}(\mathbf{x}_n)$  em relação ao valor observado  $y_n$  pode ser medida pelo erro de classificação (*ErroClass*), que mensura a proporção de erros cometidos na aplicação das estimações feitas pelo classificador, ou seja (Hastie et al., 2009):

$$\text{ErroClass} = \frac{1}{N} \sum_{n=1}^N I(y_n \neq \hat{f}(\mathbf{x}_n)) \quad (2)$$

em que  $\hat{f}(\mathbf{x}_n)$  e  $y_n$  são as categorias predita e observada, respectivamente, para a  $n$ -ésima observação,  $n = 1, 2, \dots, N$ , e  $I(\cdot)$  é uma função indicadora com valor unitário se  $y_n \neq \hat{f}(\mathbf{x}_n)$  (uma classificação incorreta) ou zero se  $y_n = \hat{f}(\mathbf{x}_n)$  (classificação correta). Assim, a Equação (2) expressa o percentual das classificações incorretas dentro de uma amostra de tamanho  $N$ .

Por fim, o problema de pesquisa deste trabalho pode ser definido como a busca da função preditora  $\hat{f}_{RF}(\mathbf{x}_n)$  – Eq. (1), baseada em modelos de *random forest*, que minimiza o erro de classificação *ErroClass* – Eq. (2), dada a amostra considerada.

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

## 4.2 Árvores de Decisão

O modelo de árvores de decisão é atualmente creditado a Breiman, Friedman, Olshen, & Stone (1984), que desenvolveram a conhecida abordagem *Classification and Regression Trees* (CART, em português: árvores de regressão e classificação). Embora Breiman et al. (1984) recebam o crédito pela criação dos modelos de árvores de decisão, formas mais simples desse método já haviam sido desenvolvidas anteriormente por pesquisadores da Universidade de Michigan, EUA. As árvores de regressão remontam ao denominado *Automatic Interaction Detection* (AID), desenvolvido nos anos 1960 por Morgan e Sonquist (1963), enquanto as árvores de classificação remetem ao chamado *Chi-squared Automatic Interaction Detection* (THAID), desenvolvido no começo dos anos 1970 por Morgan e Messenger (1973). Além do CART (Breiman et al., 1984), existem outras abordagens para implementação de árvores de decisão, como o ID3 (Quinlan, 1986) e o C4.5 (Quinlan, 1993). O algoritmo das RF deste trabalho utiliza a abordagem CART.

Para melhor compreensão do modelo de árvores de decisão, suas vantagens e limitações e como esse modelo evoluiu para o modelo de *random forest*, os conceitos de viés e variância estão apresentados a seguir.

### 4.2.1 Viés e Variância

Considere um caso de classificação binária em que, para o rótulo  $\mathbf{y}$ , são atribuídos valores numéricos de 0 (zero) ou 1 (um), dependendo de sua classe. O erro de classificação do modelo, nesse caso, pode ser calculado pela métrica chamada erro quadrático médio (*mean squared error* – MSE), da seguinte forma (Hastie et al., 2009):

$$MSE = \sum_{n=1}^N (y_n - \hat{f}(\mathbf{x}_n))^2 \quad (3)$$

A Equação (4) mostra que o valor esperado do MSE pode ser decomposto nos termos erro redutível e erro não redutível ou, também, em três valores fundamentais: o quadrado do viés de  $\hat{f}(\mathbf{x}_n)$ , a variância de  $\hat{f}(\mathbf{x}_n)$  e a variância do termo de erro  $\epsilon$ . Assumindo  $y_n = f(\mathbf{x}_n) + \epsilon$ , em que  $E(\epsilon) = 0$  e  $Var(\epsilon) = \sigma_\epsilon^2$  (Hastie et al., 2009):

$$\begin{aligned} E(y_n - \hat{f}(\mathbf{x}_n))^2 &= E[f(\mathbf{x}_n) + \epsilon - \hat{f}(\mathbf{x}_n)]^2 \\ &= \underbrace{[f(\mathbf{x}_n) - \hat{f}(\mathbf{x}_n)]^2}_{\text{erro redutível}} + \underbrace{\sigma_\epsilon^2}_{\text{erro não redutível}} \\ &= [E(\hat{f}(\mathbf{x}_n) - f(\mathbf{x}_n))]^2 + E[\hat{f}(\mathbf{x}_n) - E(\hat{f}(\mathbf{x}_n))]^2 + \sigma_\epsilon^2 \\ &= \text{viés}[\hat{f}(\mathbf{x}_n)]^2 + \text{var}(\hat{f}(\mathbf{x}_n)) + \sigma_\epsilon^2 \\ &= \text{viés}^2 + \text{variância} + \text{erro irreduzível} \quad (4) \end{aligned}$$

em que  $E(\cdot)$  é o operador esperança e  $E(y_n - \hat{f}(\mathbf{x}_n))^2$  refere-se à média do MSE obtido em caso da estimação repetida de  $f$  utilizando uma grande quantidade de bases de dados para treinamento do modelo, e da aplicação de cada  $f$  em  $\mathbf{x}_n$ .

A Equação (4) indica que a tarefa de minimização do erro de classificação exige a escolha de um modelo estatístico que alcance, ao mesmo tempo, baixa variância e baixo viés.

O erro redutível (viés e variância) existe porque  $\hat{f}$  frequentemente não é um estimador perfeito para  $f$ . Esse erro é chamado de redutível porque a função pode ser potencialmente aperfeiçoada pelo uso de uma técnica de aprendizado mais adequada. Entretanto, ainda que fosse possível formar um estimador perfeito para  $f$ , ainda existiria erro nas previsões, pois  $y$  também é uma função de  $\epsilon$ , que não pode ser previsto utilizando  $X$ . Assim, a variabilidade associada a  $\epsilon$  (erro irreduzível) também afeta o poder preditivo dos modelos (Hastie et al., 2009). Este trabalho trata apenas da questão do erro redutível.

A variância de um modelo refere-se ao montante pelo qual  $\hat{f}$  mudaria se fosse estimada utilizando uma base de dados diferente como entrada. Bases diferentes geram  $\hat{f}$  diferentes. Um método que possui alta variância é aquele em que pequenas mudanças na base de dados acarreta

uma grande mudança em  $\hat{f}$ . Viés, por outro lado, ocorre pela aproximação de um problema real, que pode ser complexo, em modelo mais simplificado. Por exemplo, a aproximação de uma relação não linear por um modelo linear (Hastie et al., 2009).

O processo de minimização do erro redutível pode incorrer em sobreajuste da função preditora aos dados (conhecido também como *overfitting*). A função se ajusta às observações utilizadas, mas seu desempenho se deteriora quando considera bases de dados diferentes (sem capacidade de generalização). Uma solução para esse problema é efetuar a divisão da base de dados da seguinte forma: em amostras de treinamento, para a estimação dos parâmetros; amostras de validação, para a obtenção da melhor configuração do modelo – estrutura e parâmetros de controle; e amostras de teste, para aferição da capacidade preditiva dos modelos, isto é, sua capacidade de generalização. Essa abordagem é conhecida como procedimento de validação cruzada (*cross-validation*) em aprendizagem de máquinas.

#### 4.2.2 Formulação de Árvores de Decisão

Métodos baseados em árvores envolvem estratificar ou segmentar um espaço preditor<sup>13</sup> 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 (James, Witten, Hastie, & Tibshirani, 2013). Visto de outra forma, as árvores de decisão possibilitam a formação de agrupamentos entre as observações da base de dados de acordo com os valores combinados de seus atributos e, em cada um desses grupos, a mesma previsão/classificação é realizada.

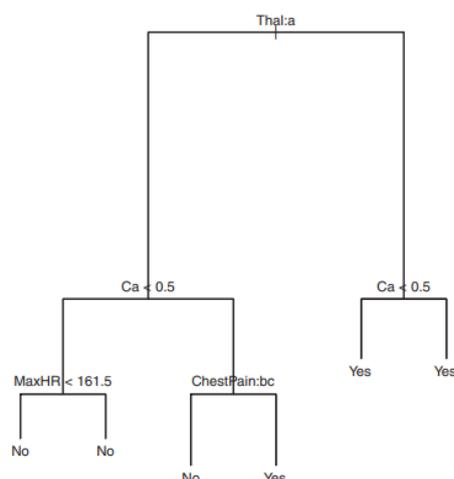
Como o conjunto de regras utilizado para segmentar esse espaço preditor é realizado em forma de árvore, essa abordagem é conhecida como árvore de decisão. As regiões finais encontradas são conhecidas como nós finais ou folhas e os pontos ao longo da árvore, em que o espaço preditor é dividido, são conhecidos como nós internos. A ideia principal é que a divisão do espaço preditor em diferentes regiões não sobrepostas possibilita a captura de relações complexas e não lineares entre as variáveis explicativas (atributos) e a variável resposta (rótulo) (James et al., 2013).

---

<sup>13</sup> Espaço preditor refere-se às divisões (partições) que podem ser criadas em uma base de dados de acordo com os valores de seus atributos. Por exemplo, em um conjunto de dados que contém o gênero de pessoas como atributo, o espaço preditor pode ser dividido em duas regiões: uma contendo apenas valores sobre homens e outra contendo apenas valores sobre mulheres, e os rótulos podem então ser estimados separadamente em cada uma dessas regiões.

A Figura 1 apresenta um exemplo simplificado de árvore de decisão aplicado em uma base de dados chamada *Heart* (Coração), como descrito em James et al. (2013). Essa base contém dados de 303 pacientes que apresentaram dor na região do peito e se submeteram a uma angiografia, que atesta a presença de doença no coração. O rótulo é binário, contendo as categorias *Yes* (Sim) e *No* (Não), indicando a presença ou não de doença no coração, respectivamente. São 13 os atributos, incluindo os utilizados pelo classificador, tais como *Thal* (teste de estresse *Thalium*), *Ca* (número de veias coloridas por fluoroscopia), *ChestPain* (tipos de dor no peito), e *MaxHR* (taxa de batimento cardíaco máxima).

Figura 1 - Exemplo de uma árvore de decisão



Fonte: James et al. (2013, p. 313).

No exemplo de árvore da Figura 1, a base de dados inicial é fragmentada binariamente em diferentes ocasiões. A primeira quebra (nó inicial da árvore) divide a base de dados entre as observações que possuem o atributo  $Thal = a$  (para a esquerda) e as que possuem  $Thal \neq a$  (para a direita). A quebras sequenciais resultaram em seis diferentes nós finais (regiões, espaços preditores). Em cada um desses conjuntos de nós, foi realizada a classificação *Yes* ou *No*, de acordo com a categoria do rótulo mais frequente dentre as observações. Esses agrupamentos se referem aos nós finais da árvore. Os pontos anteriores, em que foram realizadas as quebras, são os chamados nós internos.

As árvores de decisão podem ser categorizadas como árvores de regressão ou árvores de classificação, dependendo da variável prevista (rótulo) ser numérica ou categórica,

respectivamente (Hastie et al., 2009). Neste trabalho, está descrito o funcionamento de uma árvore de classificação.

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 essas 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 dessa região, com  $k = 1, \dots, K$ , ou, ainda, que representa a frequência relativa de cada classe em cada nó, da seguinte forma (James et al., 2013):

$$\hat{p}_{jk} = \frac{1}{N_j} \sum_{\mathbf{x}_n \in R_j} I(y_n = k) \quad (5)$$

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, como na Equação (6) (James et al., 2013).

$$E_{class} = 1 - \max_k(\hat{p}_{jk}) \quad (6)$$

Ocorre que o erro de classificação não é suficientemente sensível ao crescimento da árvore, ou seja, a medida 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 Gini<sup>14</sup> (James et al., 2013). O Índice de Gini dentro de uma região  $j$  é definido por:

---

<sup>14</sup> Não se deve confundir o índice de Gini utilizado neste trabalho – que mensura grau de pureza dentro de uma região da árvore de decisão – com o índice de mesmo nome utilizado em economia e em outras áreas da ciência para mensurar concentração de renda ou de outra medida.

$$G_j = \sum_{k=1}^K \hat{p}_{jk}(1 - \hat{p}_{jk}) \quad (7)$$

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):

$$G_j = 2\hat{p}_{jk}(1 - \hat{p}_{jk}) \quad (8)$$

Neste trabalho,  $\hat{p}$  refere-se à proporção das observações contendo a categoria “valorizar”.

Como mencionado, 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). Assim, para realizar a quebra binária do nó  $j$ , primeiramente são selecionados um atributo  $\mathbf{x}_p$  e um 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 (Alvarenga Júnior, 2018):

$$\Delta\mathbb{I}(j) = \mathbb{I}(N_j) - \hat{p}_{jL} \cdot \mathbb{I}(N_{jL}) - \hat{p}_{jR} \cdot \mathbb{I}(N_{jR}) \quad (9)$$

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  $N_j$  refere-se ao total de observações do nó  $j$ .

Esse processo – de escolha do melhor preditor e do melhor ponto de corte para a divisão das novas regiões – é, então, repetido, maximizando a redução de impureza dessas novas quebras (James et al., 2013).

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

Uma solução para tal 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 combinatória<sup>15</sup> 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 por meio da minimização da função de risco,  $\mathcal{L}(y_j, Y_j)$ , da seguinte forma (Alvarenga Júnior, 2018):

$$\Theta = \underset{\Theta}{\operatorname{argmin}} \sum_{j=1}^J \sum_{x_n \in R_j} \mathcal{L}(y_j, Y_j) \quad (10)$$

---

<sup>15</sup> Em um problema de otimização combinatória, procura-se pela solução ótima entre um conjunto finito de possibilidades (Alvarenga Júnior, 2018).

A resolução do 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, quando a região  $R_j$  é determinada, alguma aproximação deve ser realizada. Uma estratégia usual nesse 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).

Alvarenga Júnior (2018) elenca algumas vantagens que tornaram as árvores de decisão populares e difundidas em problemas de classificação: a capacidade de modelar relações complexas entre os atributos e o rótulo sem a necessidade de assumir hipóteses *a priori*; a capacidade de tratar atributos numéricos, categóricos ou ambos; e a implementação intrínseca de seleção de atributos<sup>16</sup>. Louppe (2014), por sua vez, aponta que as árvores de decisão são modelos de fácil interpretabilidade.

Por outro lado, em problemas complexos, as árvores de decisão precisam crescer de forma significativa para aprender padrões irregulares, e assim tendem a apresentar problemas de *overfitting*, em decorrência da alta variância de tal técnica (Tan, 2018). Uma pequena mudança nos dados pode resultar em uma série de quebras distintas. A razão principal dessa instabilidade é a natureza hierárquica do processo: o efeito de um erro na primeira quebra é propagado abaixo para todas as quebras seguintes (Hastie et al., 2009). Em outras palavras, lembrando que o erro de previsão de um modelo pode ser desmembrado em viés e variância, conforme a Equação (4), em problemas complexos as árvores de decisão não conseguem diminuir o viés sem um aumento consequente na variância.

### 4.3 *Random Forest*

As *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* é um modelo do tipo *ensemble*, que combina a predição de um conjunto de árvores de decisão para obter uma única resposta como saída, e tende a apresentar melhor desempenho que cada árvore do modelo em separado devido à redução da variância de suas previsões

---

<sup>16</sup> Seleção de atributos é o processo de escolha dos atributos que farão parte de um modelo. Em uma árvore de decisão, esse processo pode ser auxiliado pela criação de uma lista de importância de cada atributo, produzida pela mensuração das quebras realizadas por cada variável, em que as primeiras quebras são mais importantes.

(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 obter um modelo final de baixa variância. Assim, para cada observação de atributos  $\mathbf{x}_n$  (Hastie et al., 2009):

$$\hat{f}_{bagging}(\mathbf{x}_n) = \frac{1}{B} \sum_{b=1}^B \hat{f}_T^b(\mathbf{x}_n) \quad (11)$$

Hastie et al. (2009) explicam que a média de  $B$  variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), cada uma com variância  $\sigma^2$ , possui variância de  $\frac{\sigma^2}{n}$ . Assim, as predições realizadas pela técnica *bagging* apresentariam menor variância por se tratar de uma média de predições. 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 esse 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 do 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 das bases de dados forem apenas identicamente distribuídas (não independentes), com correlação positiva  $\rho$

entre elas, 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 da questão da correlação das criando a ferramenta *random split selection* (seleção aleatória de quebras) – uma evolução do método de randomização (aleatorização) de Kwok e Carter (1990). Nessa ferramenta, em vez de utilizar todos os atributos da base de dados como candidatos para cada quebra das árvores, apenas algumas das 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.

Breiman (2001) apresenta, com base na literatura sobre o tema, o modelo de *random forest*, cuja base de funcionamento é a utilizada 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*).

É possível 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  $\Theta_b$  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:

$$\hat{f}_{rf}(\mathbf{x}_n; \Theta) = \frac{\sum_{b=1}^B I(\hat{f}_T^b(\mathbf{x}_n; \Theta_b) = \text{"valorizar"})}{B} \quad (12)$$

Considere, como exemplo, um modelo de *random forest* composto por 300 árvores de decisão, na 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  $\hat{f}_{rf}(x_n; \Theta)$  para essa observação é 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”.

Akyildirim 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* RStudio. Os desenvolvedores do pacote se basearam no manual criado por Breiman (2002). O *software* RStudio é um programa de ampla utilização no ambiente acadêmico, de utilização gratuita e que contém diversos pacotes internos contendo modelos de previsão.

O modelo de *random forest* depende de diferentes parâmetros, que são definidos pelo usuário. A Tabela 2 exibe o nome dos parâmetros, em linha com o pacote do RStudio utilizado, em conjunto com seus valores padrão (*default*) para casos de não seleção, a faixa de possíveis valores que cada parâmetro recebe, e uma breve descrição de cada um.

O parâmetro *samplesize* refere-se à porcentagem de observações utilizadas nas amostragens da base de treinamento. É uma forma adicional que a função “*randomForest()*” permite de reduzir a variância entre as previsões das árvores de decisão. Quanto menos observações forem consideradas nas amostragens *bootstrap* das bases de treinamento, mais diferentes serão as bases resultantes e, conseqüentemente, menor a variância final das previsões das *random forest*. De acordo com o método *bootstrap* (EFRON, 1979), o processo de amostragem das bases de dados foi realizado com reposição das observações (*replace = verdadeiro*).

Tabela 2 - Parâmetros de controle da função “*randomForest()*” do pacote “*randomForest*” do *software* RStudio

Parâmetro	Valor Default	Faixa de Valores	Descrição
<i>ntree</i>	500	(1,+∞)	Número de árvores criadas.
<i>samplesize</i>	1	-0,1	Porcentagem de observações utilizadas nas amostragens das bases de treinamento.
<i>nodesize</i>	1	(1,número de amostras da base de dados)	Número mínimo de observações em um nó final.
<i>mtry</i>	raiz quadrada do número de atributos	(1,quantidade de atributos)	Número de variáveis explicativas candidatas para cada quebra dos nós.
<i>replace</i>	VERDADEIRO	(verdadeiro;falso)	Reposição ou não das observações nas amostragens da base de treinamento.
<i>cutoff</i>	-	-0,1	Valor de corte utilizado como referência na decisão de classificação de cada observação.

Fonte: elaborado pelo autor (2023).

Os demais parâmetros foram selecionados de acordo com simulações, sendo eles: o número de árvores geradas (*ntree*), a porcentagem de observações utilizada nas amostragens de cada base de treinamento (*samplesize*), o número mínimo de observações em cada nó final (*nodesize*), e o número de variáveis disponíveis no processo de quebra do nós (*mtry*).

O primeiro desses parâmetros, *ntree*, que indica o número total de árvores criadas, está diretamente relacionado com o nível de erro de classificação do modelo. O erro geral do modelo tende a diminuir com o aumento da quantidade de árvores, até atingir um ponto de estagnação. O aumento expressivo do valor desse parâmetro não influencia negativamente as previsões, mas aumenta diretamente o custo computacional associado.

O parâmetro *samplesize*, que determina a porcentagem de observações utilizadas nas amostragens da base de treinamento, está relacionado com a possibilidade de *overfitting* do modelo resultante. Um valor maior para esse parâmetro, ou seja, a utilização de uma maior porcentagem de observações da base de treino nas amostragens, aumenta o risco de sobreajuste. Isso porque as modelagens seriam desenvolvidas utilizando uma maior parte das observações, como acontece nos modelos de árvores de decisão. Um menor valor para esse parâmetro tem a vantagem de reduzir o tempo computacional, mas pode diminuir o poder preditivo do modelo, pois incorre no risco de não haver observações suficientes na base de treinamento para um ajuste adequado.

O parâmetro *nodesize*, número mínimo de observações em cada nó final, define a extensão de cada árvore de decisão. Árvores mais extensas podem incorrer em *overfitting* e aumento do custo computacional. Árvores mais reduzidas podem perder poder preditivo, pois correm o risco de não captar as relações complexas e não lineares da estrutura da base de dados.

Por fim, o parâmetro  $mtry$ , número de variáveis disponíveis no processo das quebras dos nós das árvores de decisão, altera o grau em que as árvores estão correlacionadas no modelo de *ensemble*. A correlação entre as respostas das árvores de decisão é diretamente proporcional ao valor de  $mtry$ . Quanto maior seu valor, maior o tempo computacional consumido, pois mais variáveis podem ser selecionadas em cada quebra. Adicionalmente, a capacidade preditiva do modelo pode ser reduzida, pois as árvores geradas se tornam muito semelhantes, tornando o modelo mais parecido com uma árvore de decisão simples. Um menor valor de  $mtry$  promove uma redução da variância do modelo e pode ser útil também quando existem muitas variáveis explicativas correlacionadas. No entanto, um valor baixo para esse parâmetro pode prejudicar a capacidade preditiva do modelo porque muitas quebras são realizadas utilizando variáveis sem relação ou pouco relacionadas com a variável de interesse.

É válido lembrar que a base de dados, formada por  $N$  amostras, é simbolizada por  $D = (X, \mathbf{y})$ , em que  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]$  é a matriz do conjunto dos atributos e  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$  contém o vetor dos rótulos, e que as quebras das árvores de decisão para o nó  $j$  são realizadas utilizando um ponto de corte  $t_j$  do atributo  $\mathbf{x}_p$ . Assim, uma floresta aleatória contendo  $B$  árvores de decisão,  $T_1, T_2, \dots, T_B$ , aplicada na base de dados  $D$ , pode ser descrita pelo seguinte pseudo-código:

Para  $b = 1$  até  $B$  (parâmetro *ntree*):

- a. Realize uma amostra *bootstrap* de tamanho  $N$  do banco de dados de treinamento, com reposição (parâmetro *replace* = *verdadeiro*), selecionando aleatoriamente apenas uma porcentagem (parâmetro *sampsiz*) das  $N$  observações.
- b. Construa uma árvore de decisão  $T_b$  utilizando a amostra *bootstrap*, repetindo as seguintes etapas para cada nó da árvore, até que a quantidade mínima de observações em um nó (parâmetro *nodesize*) seja atingida:
  - i. Selecione  $m$  atributos (parâmetro *mtry*) aleatoriamente entre os  $P$  atributos;
  - ii. Realize todas as quebras binárias possíveis utilizando todos os valores de cada atributo  $\mathbf{x}_p$  selecionado;
  - iii. Escolha o par  $\{p, t_j\}$  (atributo e ponto de corte) que maximiza a redução da impureza (Gini) dos nós resultantes da quebra;
  - iv. Separe o nó em dois nós-filhos.
- c. Realize e guarde a classificação  $\hat{f}_T^b(\mathbf{x}_n)$  de cada árvore de decisão para cada observação  $y_n$  (“valorizar” ou “desvalorizar”).

Retorne:  $\hat{f}_T^b(\mathbf{x}_n) = \frac{1}{B} \sum_{b=1}^B I(\hat{f}_T^b(\mathbf{x}_n) = \text{"valorizar"})$  para cada observação de atributos  $\mathbf{x}_n$ .

## 4.4 Modelos Competidores

O poder preditivo dos modelos RF utilizados neste trabalho foi comparado com o de outros dois modelos, que são: o modelo autorregressivo integrado de médias móveis (*autoregressive integrated moving average* – ARIMA), considerado um dos modelos clássicos em econometria de séries temporais, e o modelo de regressão logística, tradicional para tarefas de classificação, utilizado como *benchmark* em diferentes áreas de aplicação.

O uso de modelos clássicos de séries temporais é vantajoso por diversos motivos. Eles são capazes de capturar a estrutura de diferentes séries temporais; podem ensejar uma clara interpretação dos coeficientes dos modelos; oferecem entendimento sobre a funcionalidade do sistema; e são facilmente executados por *softwares*, pois já são utilizados intensivamente há muitos anos (Katris, 2021). A seguir, são apresentados os modelos ARIMA e de regressão logística, respectivamente.

### 4.4.1 ARIMA

Uma série temporal pode ser expressa como um conjunto de dados/observações ordenados no tempo. Os modelos ARIMA objetivam descrever a dinâmica de uma determinada série temporal com base em seus próprios valores passados (defasados) e termos de choques aleatórios (Sun, 2021).

Seja uma série temporal  $y_t$ . Sua dinâmica pode ser descrita por meio de um processo autorregressivo (*autoregressive process*) de ordem  $p$ , ou  $AR(p)$ , de acordo com a seguinte equação (Metcalf & Cowpertwait, 2009):

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t \quad (13)$$

em que  $y_t$  é o valor da série no instante  $t$ ,  $\epsilon_t$  é um processo aleatório do tipo ruído branco, e  $\phi_1, \phi_2, \dots, \phi_p$  são os parâmetros do modelo, com  $\phi_p \neq 0$  para um processo de ordem  $p$ . O valor atual de  $y_t$ , portanto, é uma combinação linear de seus  $p$  valores mais recentes (Cryer & Chan, 2008).

Um processo ruído branco,  $\epsilon_t \sim \text{RB}(0, \sigma_\epsilon^2)$ , é definido como uma sequência de variáveis aleatórias independentes e identicamente distribuídas (Cryer & Chan, 2008). Esse processo possui média zero, variância constante  $\sigma_\epsilon^2$  e é não correlacionado no tempo (Hamilton, 1994).

Se, adicionalmente, o processo RB também segue uma distribuição normal, ou seja,  $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ , a série é chamada de ruído branco gaussiano (Metcalf & Cowpertwait, 2009).

Um processo de média móvel (*moving average*) de ordem  $q$ ,  $MA(q)$ , é uma combinação linear de termos de ruído branco, podendo ser definido como (Metcalf & Cowpertwait, 2009):

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (14)$$

em que  $\epsilon_t \sim RB(0, \sigma_\epsilon^2)$  e  $\theta_1, \theta_2, \dots, \theta_q$  são os parâmetros do modelo de médias móveis.

Uma classe de modelos é obtida quando termos AR e MA são agregados em uma única expressão. Uma série temporal  $y_t$  segue um processo autorregressivo e de médias móveis (*autoregressive moving average*) de ordem  $(p, q)$ , denotado ARMA  $(p, q)$ , se (Metcalf & Cowpertwait, 2009):

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (15)$$

A modelagem de uma série temporal por um processo ARMA requer que a série em questão seja estacionária<sup>17</sup> (Sun, 2021). Diferenciar termos de uma série pode transformar uma série não estacionária em estacionária. O operador desse processo,  $\Delta$ , é definido por (Metcalf & Cowpertwait, 2009):

$$\Delta y_t = y_t - y_{t-1} \quad (16)$$

---

<sup>17</sup> A ideia básica de estacionariedade é de que as leis de probabilidade que governam o comportamento do processo estocástico não mudem ao longo do tempo. Dessa forma, o processo está em equilíbrio estatístico. Especificamente, um processo  $\{y_t\}$ , com  $N$  observações, é considerado estritamente estacionário se a distribuição estatística conjunta de  $y_1, y_2, \dots, y_N$  é a mesma que a distribuição conjunta de  $y_{1-z}, y_{2-z}, \dots, y_{N-z}$  para toda a escolha de tempo  $t=1, 2, \dots, n$  e todas escolhas de defasagem (*lag*)  $z, z \leq t$  (Cryer & Chan, 2008). Se a série não é estritamente estacionária, mas a média e a variância são constantes ao longo do tempo e a autocovariância (covariância entre as defasagens da série) só depende da defasagem associada, então a série é dita estacionária de segunda ordem, covariância-estacionária ou fracamente estacionária. A autocorrelação entre a série no tempo  $t$ ,  $y_t$ , e a própria série defasada no tempo  $t - z$ ,  $y_{t-z}$ , só depende da defasagem  $z$ , para toda a escolha de tempo  $t$  (Metcalf & Cowpertwait, 2009).

Segundo Kırbas, Sözen, Tuncer e Kazancıoğlu (2020), uma série temporal não estacionária pode ser transformada em estacionária pela aplicação do processo de diferenciação  $d$  vezes. A série final estacionária é chamada de integrada de ordem  $d$ , ou  $I(d)$ .

O modelo geral ARIMA pode ser representado como um ARIMA  $(p, d, q)$ , em que  $p$  é o número de termos autorregressivos,  $q$  é o número de termos de médias móveis e  $d$  é o número de diferenciações necessárias para estacionariedade (Sun, 2021).

Neste trabalho, o processo de modelagem das séries temporais por meio de modelos ARIMA segue as etapas da metodologia de Box e Jenkins (1976), sendo elas: (1) identificação; (2) ajuste; e (3) análise de diagnóstico.

A fase de identificação envolve o exame da dinâmica temporal da série por meio da busca de indícios sobre a estrutura do modelo ARIMA  $(p, d, q)$  mais apropriada para aproximar a série temporal em questão. Primeiramente, se for detectada a não estacionariedade da série, o processo de diferenciação é aplicado  $d$  vezes até que a série se torne estacionária. Em seguida, como a série diferenciada é um processo estacionário, modelamos uma estrutura ARMA  $(p, q)$ , em que a preocupação consiste em selecionar os termos  $(p, q)$ . Como, na prática, o processo gerador da série é desconhecido, um conjunto de potenciais estruturas é avaliado ao variarem-se o valor dos termos autorregressivos ( $p$ ) e dos termos de médias móveis ( $q$ ) do modelo.

Neste trabalho, a estacionariedade das séries é analisada pelo teste de raiz unitária de Dickey-Fuller aumentado<sup>18</sup> (*augmented Dickey-Fuller test* – ADF) e os resultados foram validados pelo teste Phillips-Perron (1988), que faz uma correção paramétrica no teste ADF. Os testes têm como hipótese nula a presença de raiz unitária, ou seja, a não estacionariedade da série.

Com a série já diferenciada, a próxima etapa se refere à escolha da quantidade de termos autorregressivos ( $p$ ) e de médias móveis ( $q$ ) dos modelos ARIMA. Para cada conjunto de treinamento e teste, foram testadas 36 diferentes estruturas dos modelos variando os valores dos termos  $p$  e  $q$  de zero a cinco, e a escolha da estrutura mais apropriada foi realizada

---

<sup>18</sup> Considere o modelo de primeira ordem,  $AR(1)$ , assumindo que a série é estacionária e que satisfaz  $y_t = \phi y_{t-1} + \epsilon_t$ . Pode ser demonstrado que, sujeito à restrição de que  $\epsilon_t$  é independente de  $y_{t-1}, y_{t-2}, y_{t-3}, \dots$  e de que  $\sigma_\epsilon^2 > 0$ , a solução do  $AR(1)$ ,  $y_t = \phi y_{t-1} + \epsilon_t$ , será estacionária se e somente se  $|\phi| < 1$ . A exigência  $|\phi| < 1$  geralmente é chamada de condição de estacionariedade do processo  $AR(1)$ . Uma forma de não estacionariedade é obtida quando  $|\phi| = 1$ , em que a equação do modelo é satisfeita por um processo ruído branco,  $y_t = y_{t-1} + \epsilon_t$ , de outra forma,  $\Delta y_t = \epsilon_t$  (Cryer & Chan, 2008). O teste de Dickey-Fuller aumentado (Said & Dickey, 1984) utiliza três diferentes formulações para testar a presença de raiz unitária – sem constante e sem tendência, com constante e sem tendência, com constante e com tendência – e permite que a série diferenciada seja qualquer processo estacionário (Metcalfe & Cowpertwait, 2009).

utilizando como métrica de comparação o critério de informação bayesiano (*Bayesian Information Criterion – BIC*) dos diferentes modelos.

A métrica BIC acessa a qualidade dos modelos dando crédito aos modelos com erro menor enquanto aplica penalidades para modelos com muitos parâmetros, sendo representado da seguinte forma (Arunkumar et al., 2021):

$$BIC = -2 \cdot \log(L) + c \cdot \log(N) \quad (17)$$

em que  $L$  é uma função de verossimilhança do modelo,  $c$  é o número de parâmetros do modelo ( $p + q + 1$  se há intercepto) e  $N$  é o número de observações utilizadas para construir o modelo.

Para qualquer conjunto de observações,  $\{y_1, y_2, \dots, y_n\}$ , série temporal ou não, a função de verossimilhança,  $L$ , é definida como a probabilidade de densidade conjunta de se obter os dados realmente observados. Para modelos ARMA – Eq. (14), essa abordagem requer especificar uma distribuição particular para o processo de ruído branco  $\{\epsilon_t\}$ . Tipicamente é assumido que  $\epsilon_t$  é um ruído branco gaussiano,  $\epsilon_t \sim i. i. d. N(0, \sigma_\epsilon^2)$ . A função de verossimilhança ( $L$ ) de cada observação  $\epsilon_t$  é dada por (Hamilton, 1994):

$$L(\theta|\epsilon_t) = f_\theta(\epsilon_t) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma_\epsilon^2}\right) \quad (18)$$

em que  $f_\theta(\epsilon_t)$  é a função densidade de  $\epsilon_t$  e  $\theta$  é o conjunto de parâmetros que define a densidade.

Após a estrutura (valores de  $p$  e  $q$ ) ser selecionada no processo anterior, a etapa de ajuste do modelo consiste na melhor estimação possível dos parâmetros desconhecidos dos modelos ARIMA utilizando esta estrutura. O processo de estimação dos parâmetros é realizado por meio do método de estimação por máxima verossimilhança (*Maximum Likelihood Estimation – MLE*)<sup>19</sup>.

Para modelos ARIMA Eq. (14), a função de verossimilhança,  $L$  – Eq. (17), é uma função de  $\phi$ 's,  $\theta$ 's e  $\sigma_\epsilon^2$  dadas as observações  $y_1, y_2, \dots, y_n$ . Os estimadores de máxima verossimilhança são então definidos como aqueles valores dos parâmetros para qual os dados realmente observados são mais prováveis, ou seja, os valores que maximizam a função de verossimilhança (Cryer & Chan, 2008).

---

<sup>19</sup> O processo de estimação dos parâmetros foi realizado utilizando o pacote “*forecast*” do software RStudio.

Assim,  $L(\theta|\epsilon_t)$  é a probabilidade do valor  $\epsilon_t$  dado o conjunto de parâmetros  $\theta$ . Como os erros  $\epsilon_t$  são independentes (processo ruído branco), a probabilidade conjunta das realizações amostrais  $\{\epsilon_1, \epsilon_2, \dots, \epsilon_N\}$  é (Enders, 2015):

$$L(\theta|\epsilon_1, \epsilon_2, \dots, \epsilon_N) = f_\theta(\epsilon_1) \cdot f_\theta(\epsilon_2) \cdot \dots \cdot f_\theta(\epsilon_N) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon_n^2}{2\sigma^2}\right) \quad (19)$$

Tomando o logaritmo natural da função de verossimilhança, obtém-se (Enders, 2015):

$$\ln(L) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^N \epsilon_n^2 \quad (20)$$

em que  $N$  é o número de observações utilizadas para construir o modelo.

O objetivo, então, é obter os valores de  $\theta$  e  $\sigma^2$  que maximizam  $\ln(L)$ , ou seja, os valores que fazem essa amostra mais provável. Definindo valores iniciais para  $y$ 's e  $\epsilon$ 's, por um processo de interação em  $n = 1, 2, \dots, N$ , a função log-verossimilhança do modelo ARMA( $p, q$ ) é obtida como (Hamilton, 1994):

$$\ln(L) = -\frac{N-p}{2} \ln(2\pi) - \frac{N-p}{2} \ln(\sigma^2) - \sum_{n=p+1}^N \frac{\epsilon_n^2}{2\sigma^2} \quad (21)$$

Na fase de análise de diagnóstico, o pesquisador averigua se os pressupostos dos modelos foram satisfeitos. Os resíduos do modelo devem seguir o comportamento de um ruído branco (média nula, variância constante e autocovariâncias nulas) e possuírem distribuição normal. Esses pressupostos foram analisados pelo teste Ljung-Box (Ljung & Box, 1978) de autocorrelação de resíduos e pelo teste Jarque-Bera de normalidade<sup>20</sup>.

O teste Ljung-Box é utilizado para a verificação de que o grupo das autocorrelações dos resíduos é estatisticamente não significativo. A hipótese nula é a de que as autocorrelações, utilizando  $Z$  diferentes defasagens,  $z = 1, 2, \dots, Z$ , são nulas, ou seja,  $H_0: r_1 = r_2 = \dots = r_Z =$

---

<sup>20</sup> Neste trabalho, o teste de autocorrelação dos resíduos é realizado utilizando o pacote “stats”, enquanto o teste de normalidade recorre ao pacote “tseries”. Ambos pacotes pertencem ao *software* RStudio.

0, em que  $r_z$  refere-se à correlação da série temporal com a própria série na defasagem  $z$ . A hipótese alternativa do teste é a existência de pelo menos uma autocorrelação diferente de zero,  $H_1: r_1, r_2, \dots, r_z \neq 0$  (Cryer & Chan, 2008). A estatística Ljung-Box é dada por:

$$Q = N(N + 2) \sum_{z=1}^Z r_z^2 (N - z) \quad (22)$$

em que  $Q$  segue aproximadamente uma distribuição qui-quadrado com  $Z - p - q - 1$  graus de liberdade. Ajustar um modelo incorreto tende a inflar  $Q$ . Então, um teste geral iria rejeitar o modelo ARMA( $p, q$ ) se os valores observados de  $Q$  ultrapassassem um valor crítico apropriado em uma distribuição qui-quadrado com o mesmo número de graus de liberdade (Cryer & Chan, 2008).

A verificação da normalidade dos resíduos é realizada por meio do teste Jarque-Bera. Esse teste se baseia no fato de que a distribuição normal possui assimetria (terceiro momento, distribuição em torno da média) nula e curtose (quarto momento, achatamento das caudas da distribuição) igual a três (Cryer & Chan, 2008). A estatística Jarque-Bera é definida como:

$$JB = \frac{N}{6} \left( \frac{\sum_{n=1}^N (\hat{\epsilon}_n^S)^3}{N} \right) + \frac{N}{24} \left( \frac{\sum_{n=1}^N (\hat{\epsilon}_n^S)^4}{N} - 3 \right) \quad (23)$$

em que  $\hat{\epsilon}_n^S$  são os resíduos padronizados<sup>21</sup> do modelo. Sob a hipótese nula de normalidade, a estatística do teste Jarque-Bera segue aproximadamente uma distribuição qui-quadrado com 2 (dois) graus de liberdade. O teste rejeita a premissa de normalidade se o valor calculado da estatística for maior que o valor crítico da distribuição.

As características do modelo ARIMA expostas nesta seção o tornaram tradicional no tema da previsão de séries temporais. Assim, diversos estudos utilizam os modelos ARIMA como referência na análise do poder preditivo de modelos modernos de aprendizado de máquina. Alguns exemplos incluem os trabalhos de Akyildirim et al. (2021), Ibrahim et al. (2021), Akyildirim et al. (2022) e Gradojevic et al. (2023).

---

<sup>21</sup> Resíduos padronizados são obtidos pela divisão dos valores dos resíduos pelo seu desvio padrão.

#### 4.4.2 Regressão Logística

Modelos de regressão logística<sup>22</sup> permitem a um pesquisador prever uma variável categórica partindo de variáveis explicativas que podem ser numéricas, categóricas ou uma combinação de ambas. A regressão logística trata das mesmas questões que técnicas como a função discriminante e a regressão múltipla com rótulos binários. Entretanto, a regressão logística é mais flexível do que essas outras técnicas. Diferente da função discriminante, a regressão logística não assume premissas sobre as distribuições dos atributos (não precisa ser normalmente distribuída, linearmente relacionada ou possuir a mesma variância nos mesmos grupos de categorias). Diferente da regressão linear múltipla, que também possui suposições sobre as distribuições dos atributos, a regressão logística não retorna probabilidades negativas (Tabachnick, Fidell, & Ullman, 2013).

A regressão logística modela a probabilidade condicional de um evento de interesse ou resultado acontecer e considera que o rótulo segue uma distribuição de Bernoulli<sup>23</sup> com probabilidade  $\hat{p}$  para o evento de interesse e probabilidade  $1 - \hat{p}$  para o evento preterido (Hair Jr., Black, Babin & Anderson, 2009).

Seja  $\mathbf{y}$  o rótulo de interesse, que possui  $K$  diferentes categorias, e  $X$  a matriz dos  $P$  atributos  $\{\mathbf{x}_1, \dots, \mathbf{x}_P\}$ , a saída da regressão logística pode ser descrita por (Tabachnick et al., 2013):

$$Prob(\mathbf{y} = k/X) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_P x_P}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_P x_P}} \quad (24)$$

em que  $Prob(\mathbf{y} = k/X)$  refere-se à probabilidade do rótulo  $\mathbf{y}$  pertencer à categoria  $k$ , dado o conjunto de variáveis explicativas  $X$ ,  $\beta$  denota o conjunto dos coeficientes da função logística  $\{\beta_0, \beta_1, \dots, \beta_P\}$ , e  $e$  representa o valor neperiano.

Neste trabalho, da mesma forma que os modelos de *random forest*, os modelos de regressão logística exibem como valor de saída a probabilidade da criptomoeda em questão se valorizar em determinado horizonte de previsão, dados os valores dos atributos pertencentes àquela observação, da seguinte forma:

<sup>22</sup> Neste trabalho, as previsões do modelo regressão logística são realizadas utilizando o pacote “stats” do *software* RStudio.

<sup>23</sup> Distribuição de Bernoulli é uma distribuição discreta de espaço amostral  $\{0, 1\}$  que recebe o valor 1 (um) com probabilidade de sucesso  $\hat{p}$  e valor 0 (zero) com a probabilidade de falha  $\hat{q} = 1 - \hat{p}$ .

$$Prob(\mathbf{y} = \text{"valorizar"}/X) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_P x_P}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_P x_P}} \quad (25)$$

A Equação (24) é chamada função logística ou função sigmóide (devido a seu formato em S). A função logística tem a propriedade de transformar quaisquer valores reais de entrada em valores de saída,  $Prob(\mathbf{y} = k/X)$ , nos limites entre zero e um, obtendo a probabilidade de o rótulo pertencer a determinada categoria (Borges & Neves, 2020).

A regressão logística não descreve uma relação direta entre os atributos e o rótulo – como acontece em modelos de regressão linear tradicionais –, mas sim por meio de uma transformação *logit* na base de dados. Os coeficientes da função logística são estimados utilizando o logaritmo das chances (*log-odds* ou *logit*) como rótulo (Hair Jr. et al., 2009):

$$\begin{aligned} \underbrace{Prob(\mathbf{y})}_{\text{probabilidade}} &= Prob(\mathbf{y} = k/X) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_P x_P}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_P x_P}} \\ \underbrace{\frac{Prob(\mathbf{y})}{1 - Prob(\mathbf{y})}}_{\text{odds}} &= e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_P x_P} \\ \underbrace{\ln\left(\frac{Prob(\mathbf{y})}{1 - Prob(\mathbf{y})}\right)}_{\text{log-odds (logit)}} &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_P x_P \end{aligned} \quad (26)$$

As chances (*odds*) de um evento de probabilidade  $Prob(\mathbf{y})$  acontecer são calculadas como  $Prob(\mathbf{y})/(1 - Prob(\mathbf{y}))$ .

A estimação dos parâmetros do modelo de regressão logística pode ser realizada pelo método de máxima verossimilhança (Hair Jr. et al., 2009), assim como descrito na seção 4.4.1. Com base em uma amostra de dados contendo  $N$  observações,  $n = 1, 2, \dots, N$ , o problema de otimização associado para a estimação dos parâmetros  $\beta_1, \dots, \beta_P$ ,  $p = 1, 2, \dots, P$ , consiste na minimização da seguinte função de custo (Borges & Neves, 2020):

$$\min_{\beta_0, \beta_1, \dots, \beta_P} \sum_{p=1}^P \frac{\beta_p^2}{2} + c \sum_{n=1}^N \ln(e^{-y_n(\beta_0 + \beta_1 x_1 + \dots + \beta_P x_P)} + 1) \quad (27)$$

em que  $\sum_{p=1}^P \frac{\beta_p^2}{2}$  é um termo de penalidade de regularização,  $c$  é um parâmetro de regulação,  $c \sum_{n=1}^N \ln(e^{-y_n(\beta_0 + \beta_1 x_1 + \dots + \beta_P x_P)} + 1)$  corresponde à função de log-verossimilhança, e  $\beta$  representa os coeficientes da função. Neste trabalho, o parâmetro  $c$  foi selecionado pelo método de validação cruzada utilizando as bases de dados de treinamento e de validação (Borges & Neves, 2020).

A vantagem da regressão logística é sua simplicidade e seu baixo custo computacional. Como há poucos parâmetros a serem estimados, problemas de sobreajuste ocorrem com menor frequência (Akyildirim et al., 2021). Descritos os modelos competidores, a próxima seção apresenta as métricas de desempenho utilizadas neste trabalho para a medição da qualidade de previsão dos modelos empregados.

#### 4.5 Métricas de Desempenho

A avaliação do desempenho das previsões dos modelos nos horizontes de previsão considerados é realizada pela análise conjunta 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 das classificações dos modelos e dos valores reais da variável de interesse, conforme apresentado na Figura 2.

Figura 2 - Matriz de Confusão

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

Fonte: elaborado pelo autor (2023).

Na Figura 2, a categoria “positivo” significa, neste trabalho, o aumento do preço do ativo no próximo horizonte de previsão (valorização), enquanto “negativo” refere-se a uma queda ou manutenção do preço (desvalorização). Para as saídas dos modelos de *random forest* e de regressão logística, os valores maiores ou iguais a 0,5 são classificados como “positivo” e os demais como “negativo”. Em relação às previsões do modelo ARIMA, valores acima de zero são classificados como “positivo”, enquanto os demais como “negativo”, uma vez que o modelo ARIMA prevê o retorno futuro (se positivo, valorização; se negativo, desvalorização).

As iniciais representadas na Figura 2 se referem à nomenclatura dada a cada combinação das categorias previstas com os resultados reais. Quando o modelo apresenta um resultado positivo (valorização) e a previsão se mostrou correta (o preço da criptomoeda subiu), há um verdadeiro positivo (*True Positive* – TP); de outra forma, se a previsão foi incorreta (o preço da criptomoeda caiu ou se manteve), ocorre um falso positivo (*False Positive* – FP). Quando o modelo exibe um resultado negativo (desvalorizar) e a previsão se mostrou correta (o preço da criptomoeda caiu ou se manteve), há um verdadeiro negativo (*True Negative* – TN); e se a previsão foi incorreta (o preço da criptomoeda subiu), acontece um falso negativo (*False Negative* – FN).

Qualquer modelo de previsão buscará maximizar as ocorrências corretas (TP e TN). A métrica de acurácia mensura esse cenário, sendo calculada pelo acerto total de classificação do modelo. Um modelo pode, entretanto, estar desbalanceado, apresentando taxas de acerto muito diferentes entre as categorias “positivas” e “negativas” do rótulo. A adição das medidas sensibilidade e da especificidade entre as métricas de avaliação visa capturar esse efeito. A sensibilidade é a taxa de acertos dos valores positivos (valorizações), ou seja, representa a capacidade do modelo de determinar positivos. A especificidade é a taxa de acertos entre os valores negativos (desvalorizações), representando a habilidade do modelo em evitar alarmes falsos.

Resumidamente, a forma de cálculo dessas métricas é a seguinte:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (28)$$

$$\text{Sensibilidade} = \frac{TP}{TP + FN} \quad (29)$$

$$\text{Especificidade} = \frac{TN}{TN + FP} \quad (30)$$

em que:  $TP$  indica o número de classificações positivas corretas,  $TN$  é o número de classificações negativas corretas,  $FP$  representa o número de classificações positivas incorretas, e  $FN$  é o número de classificações negativas incorretas.

Como mencionado, nas classificações dos modelos *random forest* e regressão logística foi utilizado o valor de corte de 0,5 para transformar as respostas numéricas (probabilidade de valorização) na classificação final do atributo (“valorizar” ou “desvalorizar”). Entretanto, dependendo do objetivo do trabalho, qualquer outro valor entre zero e um poderia ser utilizado como ponto de corte, o que geraria infinitas matrizes de confusão. Por esse motivo, foi adicionada a métrica AUC (Hanley & McNeil, 1982) neste trabalho, por possuir a vantagem de abranger todos os valores de corte em seu cálculo. Essa métrica apresenta valores entre 0,5 e 1, em que o valor de 0,5 revela que as predições não são melhores que uma previsão aleatória, enquanto o valor de 1 denota que as predições 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, representada a seguir:

$$AUC = \int_0^1 TPR(FPR) dFPR \quad (31)$$

## 5 RESULTADOS

Este capítulo apresenta os resultados empíricos da pesquisa e está dividido em três seções. A primeira trata da análise descritiva dos dados utilizados e da definição dos atributos (*inputs*) dos modelos. A segunda aborda o processo de modelagem (seleção dos parâmetros de controle e estimação) e apresenta os resultados das previsões dos modelos de *random forest*. A última seção traz uma comparação geral dos resultados das modelagens RF com os dos competidores regressão logística e ARIMA.

Todas as etapas de processamento de dados, modelagens e aferição dos resultados foram executadas na linguagem de programação R, utilizando a interface do software RStudio<sup>24</sup>. O *hardware* utilizado para execução dos experimentos computacionais possui processador Intel i7 7500U de 7ª geração, memória RAM de 8GB e armazenamento do tipo SSD de 256 GB.

### 5.1 Análise Descritiva dos Dados

Os dados do Bitcoin foram coletados por intermédio do *website* da corretora Binance<sup>25</sup>, que oferece gratuitamente uma ampla variedade de conjuntos de dados do mercado de criptomoedas em diferentes frequências. 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

---

<sup>24</sup> Versão 4.2.0, disponível em: <https://rstudio.com/products/rstudio/download/>.

<sup>25</sup> Fonte: <https://data.binance.vision/>.

$t + h$ , 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 18 diferentes log-retornos<sup>26</sup> passados (defasados), calculados a partir dos preços de fechamento (*Close*), máximos (*High*) e mínimos (*Low*) das séries. 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.

A intenção é encontrar um conjunto de retornos que represente, além das variações pontuais (preços de fechamento) da criptomoeda, suas variações máximas e mínimas ocorridas nos preços passados do Bitcoin em comparação ao seu valor atual. As variações máximas e mínimas passadas contém uma informação adicional usualmente utilizada por investidores do mercado financeiro, relacionada possivelmente a uma sobre-reação recente dos investidores, que não é capturada quando da utilização apenas dos retornos pontuais.

Para tal, 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,  $P_t = Close_t$ . Em relação aos preços defasados na fórmula de cálculo,  $P_{t-x}$ , foram considerados três subgrupos de valores.

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

Os seguintes cinco atributos abordam as variações mínimas passadas (maiores quedas no preço em alguns intervalos de tempo) e 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

---

<sup>26</sup> Log-retornos, diferenças de logaritmos de preços, são amplamente utilizados em pesquisas sobre séries temporais.

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 tratam das variações máximas de preço (maiores subidas) e 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,

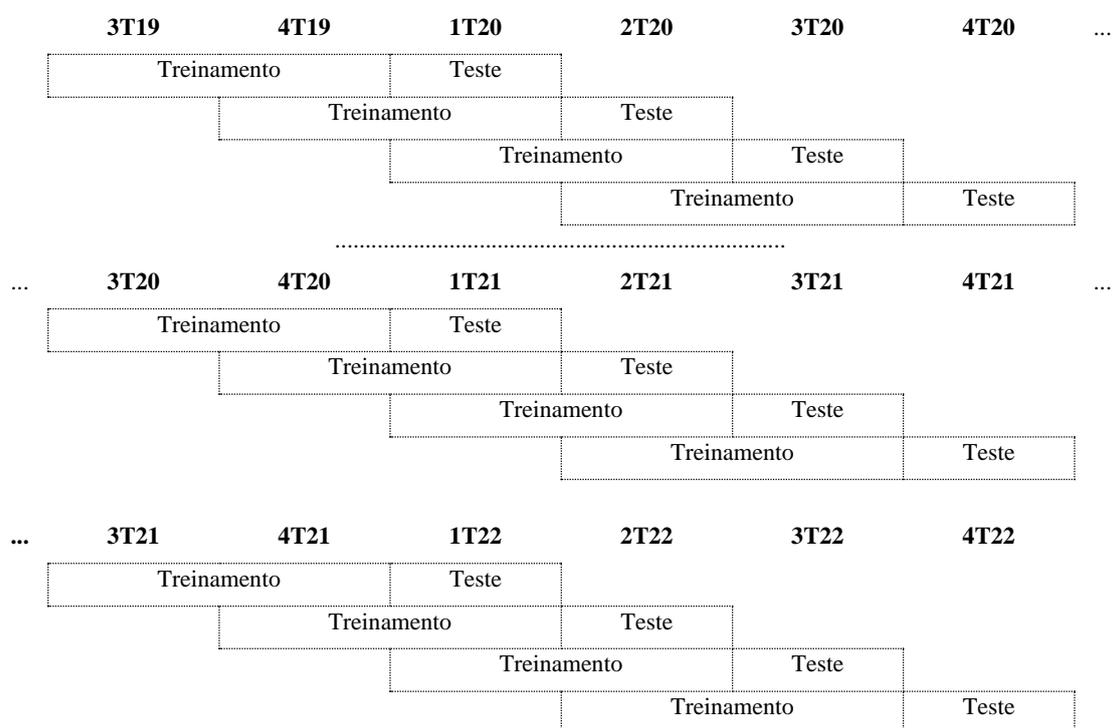
$$\left\{ \begin{array}{l} \min_{0 \leq x < 1} Low_{t-x}; \min_{1 \leq x < 5} Low_{t-x}; \min_{5 \leq x < 15} Low_{t-x}; \\ \min_{15 \leq x < 60} Low_{t-x}; \min_{60 \leq x < 360} Low_{t-x}; \min_{360 \leq x < 1440} Low_{t-x} \end{array} \right\}.$$

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.

As observações referentes ao ano de 2019 foram utilizadas apenas para a escolha da melhor configuração dos parâmetros de controle do modelo de RF. Para a modelagem e a aferição da capacidade preditiva dos modelos de previsão, foram criados 12 diferentes conjuntos de amostra teste, referentes aos trimestres dos anos de 2020, 2021 e 2022. Os conjuntos de treinamento referentes a cada conjunto de teste abrangem sempre o semestre anterior a cada amostra teste, conforme demonstra a Figura 3.

Em cada rodada de treinamento, houve o balanceamento na quantidade de rótulos “valorizar” e “desvalorizar” da base de treinamento, a fim de se evitar modelos desbalanceados (com previsões exageradas de apenas um rótulo). Todas as observações contendo o rótulo menos frequente foram utilizadas. Em mesma quantidade que as observações do rótulo menos frequente, foram selecionadas aleatoriamente observações contendo o rótulo mais frequente.

Figura 3 - Divisão da base de dados nos 12 distintos conjuntos de treinamento/teste para aferição da capacidade preditiva dos modelos de previsão

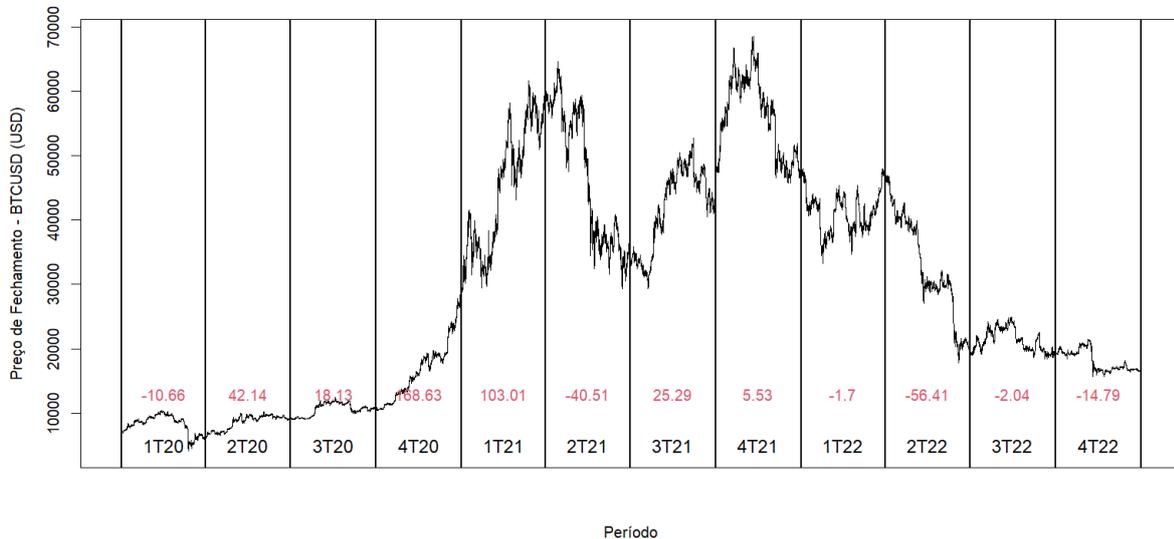


Fonte: elaborada pelo autor (2023).

A Figura 4 ilustra o comportamento do preço do Bitcoin no período de janeiro de 2020 a dezembro de 2022, período utilizado nas amostras teste dos modelos. As linhas horizontais indicam a divisão dos períodos por trimestres de amostras teste. É costume na literatura do tema de previsão de ativos financeiros a utilização de apenas um ou dois períodos de amostra teste. Este trabalho utilizou uma diversidade maior nos períodos de amostras teste com a finalidade de trazer maior robustez aos resultados.

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, 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. As variações percentuais do Bitcoin dentro de cada trimestre, indicados pelos valores dentro da Figura 4, reforçam a alta volatilidade da moeda digital.

Figura 4 - Evolução do preço de fechamento diário do Bitcoin no período de janeiro de 2020 a 2022



Nota: as linhas horizontais indicam a divisão dos períodos pelos 12 trimestres de amostras teste. Os valores dentro do gráfico exibem a variação percentual do preço do Bitcoin naquele período.

Fonte: dados da pesquisa (2023).

A Tabela 3 exibe as principais estatísticas descritivas dos retornos intradiários do Bitcoin, segmentados de acordo com a caracterização dos atributos a serem considerados nesse trabalho.

Os valores da Tabela 3 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.

Tabela 3 - Estatísticas descritivas dos retornos intradiários do Bitcoin no período de janeiro de 2020 e dezembro de 2022

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%

Nota: os retornos foram segmentados de acordo com a periodicidade dos atributos a serem considerados nos modelos preditivos.

Fonte: dados da pesquisa (2023).

A análise dos valores de desvio padrão e de medidas de posição, como máximo, mínimo e quartis, indicam alta volatilidade do preço da criptomoeda (ver Tabela 3). 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 1,58%. 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 três). Por outro lado, segundo valores da Tabela 3, a curtose decresce com o aumento do horizonte de cálculo dos retornos.

A seção seguinte apresenta o processo de modelagem e os respectivos resultados da aplicação dos modelos *random forest* na tarefa de previsão da direção do preço intradiário do Bitcoin.

## 5.2 Modelagem e Resultados – *Random Forest*

O primeiro passo do processo de modelagem das RF consistiu na escolha da melhor configuração dos parâmetros de controle do modelo RF. Essa 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: *samplesize* (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 dos nós das árvores. Nas modelagens, todos os atributos das bases de dados foram utilizados e foram realizadas previsões para todos os horizontes de previsão.

O parâmetro *nntree*, o número de árvores produzidas, foi mantido em 200, pois não houve efeito significativo no nível de erro da amostra de validação com a adição de mais árvores de decisão. Em conformidade com o método *bootstrap* (Efron, 1979), o processo de amostragem das bases de dados foi realizado com reposição das observações (parâmetro *replace* = *verdadeiro*). Finalmente, o valor de corte utilizado como referência na decisão de

classificação de cada observação (parâmetro *cutoff*) foi de 0,5 (valor padrão). A qualidade das configurações de parâmetros de controle foi mensurada de acordo com a métrica AUC.

Os resultados encontrados nas simulações decorrem de menores valores para o parâmetro *sampsize*, maiores valores para *nodesize* e menores valores para *mtry*. Em outras palavras, os modelos com maior poder preditivo utilizam de árvores mais curtas, o que protege os modelos contra *overfitting*, menos observações nas amostragens de cada base de treinamento e um menor número de variáveis candidatas no processo das quebras dos nós das árvores, características que geram árvores mais distintas e reduzem a variância dos modelos. A Tabela 4 apresenta os resultados, ordenados pela média da métrica AUC, para a amostra de validação considerada, nos diferentes horizontes de previsão.

Como foi selecionada apenas uma configuração de parâmetros de controle para uso nas previsões dos modelos *random forest* nos diferentes horizontes de previsão considerados, o melhor conjunto foi selecionado com base na melhor média de AUC encontrada utilizando todos os horizontes de previsão.

A configuração de parâmetros de controle com maior média AUC (ver Tabela 4) derivou de valores para *sampsize* de 0,25, *nodesize* de 100 e *mtry* de 2. Esse conjunto de valores foi selecionado para ser utilizado nos modelos de RF para a aferição de capacidade preditiva nos 12 conjuntos de treinamento e teste considerados.

Tabela 4 - Resultados das simulações dos modelos RF com diferentes configurações de parâmetros de controle

Parâmetros de Controle			AUC nos Horizontes de Previsão						
<i>Sampize</i>	<i>Nodesize</i>	<i>Mtry</i>	1 min.	5 min.	15 min.	1 hora	6 horas	1 dia	Média
0,25	100	2	0,5434	0,5611	0,5666	0,5698	0,5595	0,5217	0,5537
0,25	75	2	0,5420	0,5603	0,5665	0,5692	0,5592	0,5241	0,5535
0,50	100	2	0,5447	0,5602	0,5631	0,5670	0,5572	0,5211	0,5522
0,75	100	2	0,5436	0,5599	0,5640	0,5642	0,5569	0,5234	0,5520
0,50	75	2	0,5436	0,5598	0,5634	0,5674	0,5571	0,5202	0,5519
0,25	50	2	0,5422	0,5596	0,5630	0,5663	0,5574	0,5222	0,5518
0,75	75	2	0,5430	0,5598	0,5620	0,5655	0,5542	0,5205	0,5508
0,50	50	2	0,5424	0,5595	0,5622	0,5638	0,5540	0,5207	0,5504
0,25	25	2	0,5403	0,5573	0,5623	0,5642	0,5548	0,5231	0,5503
0,25	75	4	0,5455	0,5614	0,5636	0,5631	0,5513	0,5152	0,5500
0,75	50	2	0,5430	0,5579	0,5606	0,5635	0,5524	0,5224	0,5500
0,25	100	4	0,5438	0,5611	0,5643	0,5643	0,5502	0,5161	0,5500
0,25	100	6	0,5447	0,5614	0,5638	0,5593	0,5483	0,5175	0,5492
0,25	50	4	0,5429	0,5576	0,5607	0,5608	0,5500	0,5203	0,5487
0,50	100	4	0,5439	0,5602	0,5598	0,5584	0,5491	0,5182	0,5483
0,50	75	4	0,5440	0,5590	0,5596	0,5581	0,5486	0,5194	0,5481
0,50	25	2	0,5408	0,5563	0,5573	0,5623	0,5524	0,5191	0,5480
0,25	75	6	0,5437	0,5600	0,5608	0,5575	0,5487	0,5171	0,5480
0,50	50	4	0,5430	0,5567	0,5577	0,5571	0,5479	0,5204	0,5471
0,75	25	2	0,5406	0,5536	0,5576	0,5600	0,5502	0,5204	0,5471
0,75	100	4	0,5439	0,5590	0,5584	0,5548	0,5456	0,5207	0,5471
0,25	50	6	0,5420	0,5577	0,5610	0,5569	0,5474	0,5175	0,5471
0,50	100	6	0,5439	0,5599	0,5578	0,5545	0,5451	0,5180	0,5465
0,75	75	4	0,5437	0,5564	0,5577	0,5556	0,5466	0,5191	0,5465
0,25	25	4	0,5411	0,5531	0,5573	0,5588	0,5488	0,5195	0,5464
0,50	75	6	0,5437	0,5583	0,5572	0,5527	0,5459	0,5190	0,5461
0,75	100	6	0,5432	0,5581	0,5555	0,5509	0,5436	0,5205	0,5453
0,25	25	6	0,5409	0,5556	0,5559	0,5546	0,5442	0,5199	0,5452
0,75	75	6	0,5429	0,5571	0,5549	0,5501	0,5436	0,5212	0,5450
0,75	50	4	0,5411	0,5563	0,5545	0,5545	0,5450	0,5172	0,5448
0,50	25	4	0,5398	0,5530	0,5557	0,5545	0,5456	0,5187	0,5446
0,50	50	6	0,5429	0,5557	0,5543	0,5517	0,5428	0,5199	0,5445
0,75	25	4	0,5391	0,5527	0,5546	0,5536	0,5454	0,5202	0,5443
0,50	25	6	0,5410	0,5536	0,5545	0,5483	0,5433	0,5216	0,5437
0,75	50	6	0,5418	0,5549	0,5529	0,5480	0,5414	0,5223	0,5436
0,75	25	6	0,5396	0,5513	0,5511	0,5483	0,5407	0,5216	0,5421

Nota: Para cada horizonte de predição, são apresentados os resultados da métrica AUC para a amostra de validação.

Fonte: dados da pesquisa (2023).

A função de modelagem RF utilizada, do pacote “*randomForest()*” do *software* RStudio, oferece automaticamente uma informação sobre a importância de cada atributo, ou

seja, a relevância das entradas dos modelos RF. A importância de cada atributo é medida pelo aumento médio da acurácia gerado nas diferentes quebras das árvores de decisão em cada floresta aleatória. A Tabela 5 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 5 - Importância média de cada atributo - RF

Atributo	Importância Média do Atributo nos 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	<b>6,11</b>	<b>9,53</b>	9,32	7,67	6,41
Max. 6 horas	1,26	3,96	8,53	<b>13,44</b>	16,76	12,91
Max. 1 dia	1,24	3,52	7,41	13,44	<b>20,21</b>	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.	<b>4,84</b>	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	<b>19,41</b>

Nota: a importância média de cada atributo foi 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 para cada horizonte de previsão.

Fonte: dados da pesquisa (2023).

Os resultados da Tabela 5 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 6 ilustra o poder preditivo dos modelos de RF nos diferentes horizontes de previsão para a antecipação da direção futura do preço do Bitcoin nas diversas amostras teste consideradas. A capacidade preditiva do modelo é medida pelo conjunto das métricas: acurácia, sensibilidade, especificidade e AUC.

Os resultados exibidos na Tabela 6 apontam que os modelos RF possuem poder preditivo superior a 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 6, 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 6 – 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 refere-se ao horizonte de quinze minutos (0,536), enquanto o pior resultado está associado ao horizonte mais longo de previsão, de um dia (0,513) - horizontes mais longos apresentam maior volatilidade e maior dificuldade na tarefa de previsão. Em relação aos conjuntos trimestrais de teste (Painel A da Tabela 6), 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.

Tabela 6 - 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 nos anos de 2020 a 2022

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

Nota: 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.

Fonte: dados da pesquisa (2023).

A métrica sensibilidade, exibida no Painel B da Tabela 6, 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 dificuldade maior de 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 6), 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 e 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 6), 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 superiores 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 especificidade foram superiores aos de sensibilidade em todos os horizontes de previsão quando considerado o prazo completo de teste (últimas colunas dos Painéis B e C da Tabela 6).

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.

Os resultados encontrados de AUC, presentes no Painel D da Tabela 6, indicam que as predições são melhores 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) e 0,520 (um dia), considerando o período total de teste (última coluna do Painel D da Tabela 6). De mesma forma que nos resultados da métrica acurácia (Painel A da Tabela 6), 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

6) 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),  $H_1$ , em que é verificado se o poder preditivo dos modelos, medido pela acurácia, é estatisticamente superior a 0,5, como hipótese nula. A hipótese alternativa atesta acurácia maior que 0,5.

A Tabela 7 exhibe 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 foi rejeitada, ou seja, a acurácia naquela situação é estatisticamente superior a 0,5.

Tabela 7 - Valores-p (*p-values*) dos testes de proporção para a acurácia dos modelos RF, realizados para cada conjunto de teste nos diferentes horizontes de previsão

Horizonte de Previsão	Período do Conjunto Teste												Total	
	1T20	2T20	3T20	4T20	1T21	2T21	3T21	4T21	1T22	2T22	3T22	4T22		
1 min.	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>
5 min.	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>
15 min.	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>
1 hora	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,002</b>	0,212	<b>0,005</b>	<b>0,008</b>	<b>0,010</b>	<b>0,001</b>	<b>0,001</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>
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	<b>0,001</b>	
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	

Nota: 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.

Fonte: dados da pesquisa (2023).

Os resultados da Tabela 7, 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.

Neste trabalho, os modelos de RF oferecem, inicialmente, como saída, a probabilidade de ocorrer uma valorização do ativo no horizonte de previsão considerado. Em seguida, esse valor de probabilidade é comparado com o limiar de 0,5 para classificar a saída de acordo com o rótulo: “valorizar” ou “desvalorizar”. Em outras palavras, o valor das saídas dos modelos não é utilizado na classificação, mas apenas o fato de ser maior ou menor e igual a 0,5.

Para reforçar a hipótese de previsibilidade dos modelos RF, essa informação (probabilidade de valorização) pode ser aplicada em uma análise adicional, em que se busca verificar a existência de relação entre diferentes faixas de valores de saída dos modelos RF e a categoria da variável resposta utilizada na pesquisa, a direção do preço de fechamento da criptomoeda. Para esta análise, os valores de probabilidade de saída foram categorizados em 20 diferentes faixas de previsão, organizadas a cada 5% de valor (“0%-5%”, “5%-10%”, ..., “95%-100%”), e foi computada a direção do preço da criptomoeda, medida pelo percentual de valorizações da criptomoeda em cada uma dessas faixas de probabilidade. A Tabela 8 apresenta os resultados dessa análise, em que são exibidas as porcentagens de valorização em cada faixa de probabilidade para os horizontes de previsão considerados. Estão ilustrados apenas os resultados nas situações com mais de 30 ocorrências.

Os resultados da Tabela 8 revelam uma relação direta entre os valores de saídas dos modelos RF e a direção do preço futuro do Bitcoin. Nota-se uma porcentagem maior de valorizações do ativo em faixas de saídas do modelo de maior valor, padrão mais evidente para os horizontes de previsão mais curtos (um, cinco e quinze minutos). Especificamente, os resultados para o horizonte de quinze minutos, em que foram encontrados os melhores resultados de poder preditivo (ver Tabela 6), se comportaram como o esperado (ver Tabela 8), pois todas as faixas de maior probabilidade apresentaram maiores porcentagens de valorizações do Bitcoin. Tomando o mesmo horizonte de quinze minutos como exemplo, na faixa “90%-95%” de saída do modelo, ou seja, nas ocasiões em que o modelo apresentou como resposta valores entre 0,90 e 0,95 (indicando alta probabilidade de valorização do Bitcoin), houve valorização em 73,2% das situações. Por outro lado, na faixa de resposta “5%-10%” (indicando alta probabilidade de desvalorização da moeda), houve valorização em apenas 29,8% das situações.

Tabela 8 - Porcentagem de valorização do Bitcoin por faixas de previsão das saídas dos modelos RF nos diferentes horizontes de previsão

Faixa de Previsão	Horizonte de Previsão					
	1 min.	5 min.	15 min.	1 hora	6 horas	1 dia
0%-5%	-	-	-	-	42,30%	66,60%
5%-10%	-	-	29,80%	41,00%	51,10%	60,70%
10%-15%	33,30%	36,00%	34,00%	37,70%	46,30%	53,80%
15%-20%	38,00%	37,10%	38,40%	41,70%	46,40%	50,70%
20%-25%	39,90%	39,30%	39,50%	43,20%	45,90%	50,60%
25%-30%	41,30%	41,40%	41,40%	44,30%	46,40%	50,30%
30%-35%	43,10%	43,40%	43,60%	45,40%	48,20%	50,60%
35%-40%	44,80%	45,40%	45,80%	46,90%	49,00%	51,00%
40%-45%	46,90%	47,40%	47,60%	48,20%	49,90%	50,80%
45%-50%	49,20%	49,30%	49,50%	50,20%	50,80%	51,20%
50%-55%	50,90%	51,20%	51,40%	51,70%	51,90%	52,10%
55%-60%	52,60%	52,80%	53,10%	53,50%	53,10%	53,20%
60%-65%	55,10%	54,60%	54,60%	54,80%	54,40%	54,80%
65%-70%	56,90%	56,70%	57,00%	56,20%	55,20%	55,10%
70%-75%	59,20%	59,50%	59,10%	57,50%	56,10%	54,80%
75%-80%	62,10%	62,40%	61,40%	58,50%	57,60%	54,30%
80%-85%	64,20%	65,10%	65,20%	60,50%	59,00%	54,70%
85%-90%	65,70%	67,60%	68,00%	61,00%	57,90%	54,70%
90%-95%	-	-	73,20%	65,00%	63,10%	54,90%
95%-100%	-	-	-	77,50%	85,70%	62,10%

Nota: as faixas de previsão foram criadas a partir da categorização dos valores de saída dos modelos RF. Estão ilustrados apenas os resultados com mais de 30 ocorrências.

Fonte: dados da pesquisa (2023).

A análise relativa à direção da criptomoeda segmentada por faixas de probabilidade abre caminho para a busca de situações nas quais o modelo apresenta maior poder preditivo, principalmente nas faixas de previsão mais extremas, nas quais os valores de saída estão próximo de zero ou um, como nos casos do exemplo presente no parágrafo anterior.

Esta seção tratou da questão da modelagem e dos resultados da aplicação do modelo *random forest* nas bases de teste da pesquisa. Foi descrito o processo de escolha dos parâmetros de controle do modelo, foram apresentados os valores de importância de cada atributo e debatido o poder preditivo das RF, analisado de diferentes formas e contendo a discussão sobre a primeira hipótese de pesquisa do trabalho ( $H1$ ).

O processo de escolha dos parâmetros de controle do modelo apontou que as melhores RF decorreram da utilização de menos observações nas amostragens de cada base de treinamento, árvores mais curtas e um menor número de variáveis candidatas no processo das quebras do nós.

O poder preditivo dos modelos RF foi inicialmente estimado para cada trimestre de teste e para o período total de testes, e mensurado pelas métricas: acurácia, sensibilidade, especificidade e AUC. Nesse contexto, a primeira hipótese de pesquisa do trabalho, *H1*, foi avaliada. Resumidamente, os resultados mostraram que: (1) os modelos RF foram hábeis em prever a direção do preço de criptomoedas com acurácia maior que um passeio aleatório em quase todos os horizontes de previsão (com exceção do horizonte de um dia), rejeitando a hipótese de eficiência de mercado de Fama (1970) em sua forma fraca; (2) o poder preditivo dos modelos RF apresentou melhores resultados para os horizontes de previsão intermediários, de cinco minutos, quinze minutos e uma hora, com destaque positivo para o horizonte de quinze minutos e negativo para o horizonte de um dia; (3) os modelos RF exibiram maior habilidade em identificar desvalorizações da criptomoeda, apresentando valores maiores de especificidade que os de sensibilidade; e (4) a análise dos resultados por faixas de saída dos modelos salientaram uma relação direta entre os valores de saída dos modelos RF e a direção do preço de fechamento da criptomoeda.

São escassos os trabalhos na literatura que tratam especificamente da previsão de direção de preço do Bitcoin no horizonte intradiário utilizando modelos de *random forest*. Foram encontrados apenas sete trabalhos com resultados de acurácia para horizontes de previsão menores que um dia. Ainda, quatro desses trabalhos apresentam resultados com evidência de sobreajuste nos modelos (Alonso-Monsalve et al., 2020; Ibrahim et al., 2021; Gradojevic et al.; Orte et al., 2023), com valores de acurácia de até 100%. A Tabela 9 apresenta um comparativo dos valores de acurácia deste trabalho com os valores encontrados nas pesquisas da literatura em que não houve sobre-ajuste, considerando diferentes horizontes de previsão.

Tabela 9 - Comparativo dos valores de acurácia entre os trabalhos presentes na literatura sobre previsão de direção do preço do Bitcoin no horizonte intradiário utilizando modelos *Random Forest*

Referência	Período de Teste	Horizonte de Previsão				
		1 min.	5 min.	15 min.	30 min.	1 hora
Este trabalho	jan20-dez22	0,525	0,532	0,536	-	0,534
Akyildirim et al. (2022)	ago/20	-	0,53	0,54	0,56	0,54
Jaquart et al. (2021)	set19-nov19	0,512	0,527	0,535	-	0,556
Akyildirim et al. (2021)	jan18-jun18	-	-	0,53	0,55	0,52

Fonte: dados da pesquisa (2023).

Previamente à análise dos resultados demonstrados na Tabela 9, deve-se tomar o cuidado necessário na comparação dos diferentes valores, pois cada pesquisa se utilizou de bases de dados e períodos de teste distintos, bem como possivelmente diferentes séries históricas sobre o Bitcoin (a criptomoeda é negociada em diferentes corretoras com diferenças de preços).

Os valores da Tabela 9 apontam para a semelhança dos resultados encontrados nesta pesquisa com os presentes na literatura do tema, principalmente em relação ao trabalho de Akyildirim et al. (2022). Tomando o horizonte de previsão de cinco minutos como exemplo, a acurácia para o período completo de teste encontrada neste trabalho foi de 0,532, contra 0,53 em Akyildirim et al. (2022) e 0,527 em Jaquart et al. (2021). Neste trabalho, o poder preditivo das RF foi testado em uma base de teste notadamente mais ampla que nos estudos da literatura (ver coluna “Período de Teste” da Tabela 9).

O comparativo de poder preditivo entre as pesquisas do tema reforçou o fato de que neste trabalho foram desenvolvidos modelos parcimoniosos, ou seja, menos complexos, mas com poder preditivo compatível com os da literatura. No trabalho de Jaquart et al. (2021), por exemplo, foi utilizado um conjunto amplo de 169 atributos, incluindo indicadores técnicos, de cadeia de blocos (*blockchain*), de sentimento e econômicos, utilizados como entrada nos modelos RF. Akyildirim et al. (2022), de outra forma, utilizaram uma gama de 40 atributos, incluindo defasagens passadas e diversos indicadores técnicos. De uma forma mais concisa, o trabalho apresentado aqui utilizou-se apenas de 18 atributos como entradas para os modelos RF, referentes aos retornos passados de um minuto, cinco minutos, quinze minutos, uma hora, seis horas e um dia, utilizando como preço inicial em seu cálculo valores de fechamento, máximos e mínimos do Bitcoin nesses diferentes horizontes. Ainda assim, os resultados da Tabela 9 demonstram que este trabalho obteve resultados similares ou superiores aos encontrados na literatura.

A execução das modelagens RF deste trabalho consumiu um tempo total de 165,3 minutos, ou seja, aproximadamente duas horas e quarenta minutos. No Apêndice A estão apresentados os valores de tempo dispendidos nas modelagens utilizando os 12 diferentes horizontes de previsão, realizadas em todas as bases de treinamento e teste. O tempo médio dispendido nas modelagens foi de 6,12 minutos, com desvio padrão de 0,62 minuto. Houve uma tendência de queda no tempo gasto com o aumento do horizonte de previsão.

### 5.3 Comparativo dos Resultados dos Modelos

Esta seção contém a análise comparativa do poder preditivo dos modelos *random forest*, regressão logística e ARIMA, em suas aplicações no período completo de teste, que compreende os 12 trimestres de treinamento-teste.

O processo de elaboração, aplicação e aferição dos resultados relativos ao modelo de regressão logística seguiu a estrutura das modelagens RF. Foram tratados os seguintes temas: análise da importância de cada atributo, poder preditivo dos modelos em cada base de teste, segmentação das saídas dos modelos por faixa de previsão, e tempo de processamento das modelagens. Os resultados completos estão expostos no Apêndice B.

O processo de modelagem das séries temporais por meio de modelos ARIMA seguiu os estágios da metodologia de Box e Jenkins (1976), contendo as etapas de identificação, ajuste e análise de diagnóstico. Na fase de identificação, foi verificada a estacionariedade das séries e realizada a busca da melhor estrutura do modelo em cada período de teste. Na etapa de ajuste, os parâmetros da função de previsão foram estimados. Na etapa final, de análise de diagnóstico, as séries dos resíduos dos modelos foram testadas para autocorrelação e normalidade. O poder preditivo dos modelos foi mensurado em seguida. Não há a análise do poder preditivo segmentado por faixas de previsão, pois os modelos ARIMA possuem outra forma de previsão dos rótulos. Os resultados detalhados estão exibidos no Apêndice C.

A Tabela 10 apresenta conjuntamente o poder preditivo dos diferentes modelos, mensurado pelas métricas: acurácia, sensibilidade, especificidade e AUC. Os valores de cada métrica foram calculados considerando o período completo de teste, de janeiro de 2020 a dezembro de 2020.

A leitura da Tabela 10 demonstra a superioridade de poder preditivo do modelo de *random forest*, seguido pela regressão logística, com resultados próximos, e pelo ARIMA, com resultados aparentemente inferiores.

Tabela 10 - Comparativo do poder preditivo dos modelos RF, regressão logística e ARIMA

Horizonte de Previsão	Resultados do Modelo		
	<i>Random Forest</i>	Regressão Logística	ARIMA
<b>Painel A: Acurácia</b>			
1 min.	<b>0,525</b>	0,516	0,505
5 min.	<b>0,532</b>	0,528	0,515
15 min.	<b>0,536</b>	0,531	0,509
1 hora	<b>0,534</b>	0,528	0,511
6 horas	<b>0,526</b>	0,521	0,517
1 dia	0,513	<b>0,523</b>	0,520
<b>Painel B: Sensibilidade</b>			
1 min.	0,501	<b>0,532</b>	0,460
5 min.	0,513	<b>0,532</b>	0,507
15 min.	0,524	<b>0,526</b>	0,527
1 hora	0,523	<b>0,525</b>	0,464
6 horas	<b>0,496</b>	0,472	0,480
1 dia	0,492	0,421	<b>0,583</b>
<b>Painel C: Especificidade</b>			
1 min.	<b>0,549</b>	0,501	0,548
5 min.	<b>0,550</b>	0,524	0,523
15 min.	<b>0,547</b>	0,536	0,49
1 hora	<b>0,546</b>	0,531	0,561
6 horas	0,558	<b>0,572</b>	0,555
1 dia	0,537	<b>0,635</b>	0,454
<b>Painel D: AUC</b>			
1 min.	<b>0,536</b>	0,525	0,504
5 min.	<b>0,546</b>	0,539	0,515
15 min.	<b>0,551</b>	0,542	0,509
1 hora	<b>0,548</b>	0,540	0,510
6 horas	<b>0,538</b>	0,534	0,517
1 dia	0,520	<b>0,537</b>	0,496

Nota: o poder preditivo dos modelos foi testado por meio do conjunto de métricas: acurácia, sensibilidade, especificidade e AUC, 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 em cada horizonte de previsão. Para os modelos *random forest* e regressão logística, valores de saída dos modelos maiores que 0,5 (valor de corte) são classificados como uma valorização. Para o ARIMA, o valor de corte é zero.

Fonte: dados da pesquisa (2023).

Com exceção do horizonte de um dia, o modelo *random forest* apresentou valores superiores de acurácia e de AUC (Painel A e Painel D da Tabela 10) 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 10), 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 referem-se ao horizonte de previsão de quinze minutos. Aparentemente, o horizonte de um minuto é curto para realizar previsões de direção de preços, enquanto o horizonte de um dia descola do contexto intradiário do trabalho.

A segunda hipótese de pesquisa,  $H_2$ , avalia a existência de superioridade estatística do poder preditivo dos modelos *random forest* quando comparados aos modelos de regressão logística e ARIMA. Para a execução dessa tarefa, foram realizados testes de hipótese paramétricos de proporções par a par (Morettin & Bussab, 2017) com os resultados de cada modelo, utilizando a métrica acurácia – calculada no período completo de teste – como medida de comparação. A hipótese nula de cada teste se refere à igualdade dos valores de acurácia, enquanto a hipótese alternativa testa a superioridade do valor mais alto de acurácia entre os modelos. A finalidade dessa análise é encontrar a ordenação de poder preditivo entre os modelos, ou seja, definir quais modelos obtiveram os melhores resultados em cada horizonte de previsão. A Tabela 11 exibe os resultados da análise sobre a segunda hipótese de pesquisa,  $H_2$ .

A Tabela 11 atesta para a superioridade estatística da acurácia do modelo RF nos horizontes de previsão de um minuto, cinco minutos e quinze minutos, seguido pelo modelo de regressão logística. Quanto ao horizonte de previsão de uma hora, os valores de acurácia dos modelos RF e regressão logística não foram considerados estatisticamente diferentes, porém foram maiores que os do modelo ARIMA (deve-se lembrar que este modelo não foi desenvolvido para a tarefa específica de classificação). Com relação aos horizontes de seis horas e de um dia, os valores de acurácia dos três modelos foram considerados estatisticamente iguais.

Tabela 11 - Ordenação do poder preditivo entre os modelos RF, regressão logística e ARIMA por horizonte de previsão

Horizonte de Previsão	Ordenação de Poder Preditivo
1 minuto	RF > Reg.Log. > ARIMA
5 minutos	RF > Reg.Log. > ARIMA
15 minutos	RF > Reg.Log. > ARIMA
1 hora	RF = Reg.Log. > ARIMA
6 horas	RF = Reg.Log. = ARIMA
1 dia	RF = Reg.Log. = ARIMA

Nota: foi considerado o período completo de teste (janeiro de 2020 a dezembro de 2020) e a acurácia como métrica de desempenho. Em cada horizonte de previsão, foram realizados testes de proporção par a par entre os valores de acurácia de cada modelo.

Fonte: dados da pesquisa (2023).

A Tabela 12 apresenta um comparativo das porcentagens de valorização do Bitcoin por faixa de previsão dos modelos RF e regressão logística (o modelo ARIMA não foi considerado nesta análise), utilizando o período completo de teste e o horizonte de previsão de 15 minutos, que foi o de maior acurácia para os modelos RF e regressão logística.

Tabela 12 - Comparativo da porcentagem de valorização do Bitcoin por faixas de previsão dos modelos RF e regressão logística no horizonte de previsão de 15 minutos

Faixa de Previsão	% de Valorizações por Modelo	
	<i>Random Forest</i>	Regressão Logística
0%-5%	-	-
5%-10%	29,8%	33.3%
10%-15%	34,0%	40.1%
15%-20%	38,4%	42.5%
20%-25%	39,5%	42.4%
25%-30%	41,4%	41.8%
30%-35%	43,6%	45.0%
35%-40%	45,8%	44.7%
40%-45%	47,6%	45.4%
45%-50%	49,5%	47.6%
50%-55%	51,4%	53.0%
55%-60%	53,1%	55.1%
60%-65%	54,6%	56.5%
65%-70%	57,0%	57.8%
70%-75%	59,1%	57.6%
75%-80%	61,4%	62.3%
80%-85%	65,2%	65.5%
85%-90%	68,0%	68.3%
90%-95%	73,2%	73.6%
95%-100%	-	-

Nota: as faixas de previsão foram criadas a partir da categorização dos valores de saída dos modelos.

Fonte: dados da pesquisa (2023).

Ambos modelos, RF e regressão logística, foram hábeis em discriminar a variável predita por meio das diferentes faixas de previsão de saída de seus modelos (ver Tabela 12), pois apresentaram valores menores de valorização do Bitcoin nas faixas mais baixas (próximo de zero) e maiores nas faixas mais altas (próximo de um). Houve superioridade do modelo RF, em que cada faixa superior de saída do modelo apresenta uma maior porcentagem de valorizações do Bitcoin. Os resultados relativos ao modelo de regressão logística apresentaram algumas inconsistências, como as quedas nas porcentagens de valorização na quinta, sexta, oitava e na décima quinta faixa de previsão em relação à faixa anterior.

A habilidade do modelo RF em discriminar a direção do preço da criptomoeda em diferentes faixas de saída pode ser explorada futuramente em pesquisas sobre o tema, pois gera a oportunidade de se encontrarem ocasiões com poder preditivo mais alto que a média.

## 6 CONCLUSÃO

Neste trabalho, foi abordado o tema da 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*. A direção do preço de fechamento do Bitcoin, o rótulo da pesquisa, foi prevista nos horizontes de um minuto, cinco minutos, quinze minutos, uma hora, seis horas e um dia. Os atributos utilizados nos modelos se referem aos retornos das séries, utilizando como base de cálculo os preços passados de fechamento, de máximo e de mínimo da criptomoeda.

O poder preditivo dos modelos *random forest*, regressão logística e ARIMA foi avaliado em um conjunto amplo de 12 bases de treinamento/teste, em que as bases de teste são não sobrepostas e se referem aos trimestres dos anos de 2020, 2021 e 2022. O poder preditivo dos modelos foi avaliado pela seguinte coleção de métricas de desempenho: acurácia, que mensura o acerto total de classificação dos modelos; sensibilidade, relativa ao acerto entre as valorizações da criptomoeda; especificidade, que diz respeito ao acerto entre as desvalorizações da criptomoeda; e AUC, que é uma métrica mais completa que engloba as três anteriores.

Foram encontradas evidências de que modelos de *random forest* possuem capacidade de previsão da direção do preço do Bitcoin na frequência intradiária, apresentando acurácia superior a de um passeio aleatório ou de um classificador ingênuo, o que aponta para violações na hipótese de forma fraca de eficiência de Fama (1970), principalmente em relação aos horizontes de previsão mais curtos, de até uma hora. A acurácia dos modelos RF apresentou-se estatisticamente superior quando comparada com a dos 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.

O conjunto amplo de períodos de teste ofereceu maior robustez aos resultados em relação aos trabalhos na literatura, pois permitiram que os resultados dos modelos fossem mensurados para períodos mais longos e que apresentam dinâmicas distintas.

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 sobre o tema. Outras criptomoedas podem ser incluídas como objeto de pesquisa. O mercado de criptomoedas cresceu vertiginosamente nos últimos anos e a pesquisa com outros ativos digitais pode trazer maior robustez aos resultados. Diferentes modelos de previsão também podem ser explorados. O potencial econômico das previsões dos modelos também pode ser investigado. Os valores de saída dos modelos RF podem servir de

auxílio para a criação de diferentes estratégias de investimento e também como ferramenta de apoio aos investidores do mercado de criptomoedas. A análise mais detalhada da relação entre as faixas de probabilidade de saída dos modelos RF com a direção dos preços de criptomoedas pode servir como caminho inicial para essa tarefa.

## REFERÊNCIAS

- Akyildirim, E., Çepni, O., Corbet, S., & Uddin, G. S. (2022). Forecasting Mid-Price Movement of Bitcoin Futures Using Machine Learning. *Annals of Operations Research*, 1-32. doi: 10.1007/s10479-021-04205-x
- Akyildirim, E., Goncu, A., & Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research*, 297(1), 3-36. doi: 10.1007/s10479-020-03575-y
- 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(1), 1-15. doi: 10.1016/j.eswa.2020.113250
- Alvarenga Júnior, W. J. (2018). *Métodos de otimização hiperparamétrica: um estudo comparativo utilizando árvores de decisão e florestas aleatórias na classificação binária* (Dissertação de Mestrado). Escola de Engenharia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais. Recuperado de <https://repositorio.ufmg.br/handle/1843/BUBD-AX2NLF>
- Arunkumar, K. E., Kalaga, D. V., Kumar, C. M. S., Chilkoor, G., Kawaji, M., & Brenza, T. M. (2021). Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). *Applied soft computing*, 103, 1-26. doi: 10.1016/j.asoc.2021.107161
- Basher, S. A., & Sadorsky, P. (2022). Forecasting Bitcoin price direction with random forest: How important are interest rates, inflation, and market volatility? *Machine Learning with Applications*, 9. doi: 10.1016/j.mlwa.2022.100355
- Borges, T. A., & Neves, R. F. (2020). Ensemble of machine learning algorithms for cryptocurrency investment with different data resampling methods. *Applied Soft Computing*, 90, 1-20. doi: 10.1016/j.asoc.2020.106187
- Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. San Francisco, Calif: Holden-Day.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140. doi: 10.1007/BF00058655
- Breiman, L. (2001). Random forest. *Machine learning*, 45(1), 5-32. doi: 10.1023/A:1010933404324
- Breiman, L. (2002). Manual On Setting Up, Using, And Understanding Random Forest V3.1. Recuperado de [https://www.stat.berkeley.edu/~breiman/Using\\_random\\_forest\\_V3.1.pdf](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*. New York: Chapman and Hall/CRC.

- Cryer, J. D., & Chan, K. S. (2008). *Time series analysis: with applications in R* (Vol. 2). New York: Springer.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economics Statistics*, 13(3), 134-144. doi: <https://doi.org/10.1198/073500102753410444>
- 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.
- Dyhrberg, A. H. (2016). Hedging capabilities of bitcoin. Is it the virtual gold? *Finance Research Letters*, 16, 139-144. doi: 10.1016/j.frl.2015.10.025
- Efron, B. (1979). Computers and the theory of statistics: thinking the unthinkable. *SIAM review*, 21(4), 460-480. doi: 10.1137/1021092
- Enders, W. (2015). *Applied Econometric Time Series*. 4th ed. Hoboken: John Wiley & Sons.
- 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. doi: 10.1016/j.eswa.2005.06.024
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417. doi: 10.2307/2325486
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861-874. doi: 10.1016/j.patrec.2005.10.010
- 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, 1-8. doi: 10.1016/j.frl.2021.102280
- Gil, C. (2022). Intraday Trading of Precious Metals Futures Using Algorithmic Systems. *Chaos, Solitons & Fractals*, 154. doi: 10.1016/j.chaos.2021.111676
- 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. doi: 10.1016/j.ijforecast.2021.08.001
- Gyamerah, S. A. (2019). Are Bitcoins price predictable? Evidence from machine learning techniques using technical indicators. *arXiv*. doi: 10.48550/arXiv.1909.01268
- Hair Jr., J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate Data Analysis*. Prentice Hall.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
- 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. doi: 10.1148/radiology.143.1.7063747

- 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.
- Huang, J. Z., Huang, W., & Ni, J. (2019). Predicting bitcoin returns using high-dimensional technical indicators. *The Journal of Finance and Data Science*, 5(3), 140-155. doi: 10.1016/j.jfds.2018.10.001
- Ibrahim, A., Kashef, R., & Corrigan, L. (2021). Predicting market movement direction for bitcoin: A comparison of time series modeling methods. *Computers & Electrical Engineering*, 89. doi: 10.1016/j.compeleceng.2020.106905
- 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., & Martin, C. (2020). *Machine Learning for Bitcoin Pricing-A Structured Literature Review*. In 15th International Conference on Wirtschaftsinformatik (pp. 174-188).
- 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. doi: 10.1016/j.jfds.2021.03.001
- Katris, C. (2021). A time series-based statistical approach for outbreak spread forecasting: Application of COVID-19 in Greece. *Expert Systems with Applications*, 166, 1-9. doi: 10.1016/j.eswa.2020.114077
- Kırbaş, İ., Sözen, A., Tuncer, A. D., & Kazancıoğlu, F. Ş. (2020). Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. *Chaos, Solitons & Fractals*, 138, 1-7. doi: 10.1016/j.chaos.2020.110015
- Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forest: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689-702. doi: 10.1016/j.ejor.2016.10.031
- Kwok, S. W., & Carter, C. (1990). Multiple decision trees. In *Machine intelligence and pattern recognition* (Vol. 9, pp. 327-335). North-Holland: Elsevier Science Publishers. doi: 10.1016/B978-0-444-88650-7.50030-5
- Lahmiri, S., & Bekiros, S. (2020). Intelligent forecasting with machine learning trading systems in chaotic intraday Bitcoin market. *Chaos, Solitons & Fractals*, 133, 1-7. doi: 10.1016/j.chaos.2020.109641
- 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. doi: 10.1016/S0169-2070(99)00048-5
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303. doi: 10.1093/biomet/65.2.297
- Louppe, G. (2014). *Understanding random forest: From theory to practice* (Tese de Doutorado). Faculty of Applied Sciences, University of Liège, Liège. Recuperado de <https://arxiv.org/pdf/1407.7502.pdf>

- Maciel, L., & Ballini, R. (2019). On the predictability of high and low prices: The case of Bitcoin. *Brazilian Review of Finance*, 17(3), 66-84. doi: 10.12660/rbfin.v17n1.2019.77578
- Mallqui, D. C., & Fernandes, R. A. (2019). Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques. *Applied Soft Computing*, 75, 596-606. doi: 10.1016/j.asoc.2018.11.038
- Metcalfe, A. V., & Cowpertwait, P. S. (2009). *Introductory time series with R* (pp. 259-259). New York: Springer-Verlag New York.
- Morettin, P. A., & Bussab, W. O. (2017). *Estatística Básica*. São Paulo: Saraiva.
- Morgan, J. N., & Messenger, R. C. (1973). *THAID, a sequential analysis program for the analysis of nominal scale dependent variables*. Institute for Social Research, University of Michigan, Ann Arbor, MI.
- Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American statistical association*, 58(302), 415-434.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review*, 1-9.
- Nelson, D. M., Pereira, A. C., & Oliveira, R. A. (2017). *Stock market's price movement prediction with LSTM neural networks*. In 2017 International Joint Conference on Neural Networks (IJCNN) (pp. 1419-1426). doi: 10.1109/IJCNN.2017.7966019
- 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. doi: 10.1016/j.ribaf.2022.101829
- Pradeepkumar, D., & Ravi, V. (2016). *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).
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106. doi: 10.1007/BF00116251
- Quinlan, J. R. (1993). C4. 5: Programming for machine learning. *Morgan Kaufmann*, 38(48).
- Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3), 599-607. doi: 10.1093/biomet/71.3.599
- Sun, J. (2021). Forecasting COVID-19 pandemic in Alberta, Canada using modified ARIMA models. *Computer Methods and Programs in Biomedicine Update*, 1. doi: 10.1016/j.cmpbup.2021.100029
- Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2013). *Using multivariate statistics* (Vol. 6, pp. 497-516). Boston: Pearson.

## APÊNDICES

**APÊNDICE A** – Resultados adicionais *Random Forest*

**APÊNDICE B** – Resultados regressão logística

**APÊNDICE C** – Resultados ARIMA

## APÊNDICE A – Resultados adicionais *Random Forest*

Tempo de processamento (em minutos) para aplicação dos modelos de RF para previsão da direção futura do preço do Bitcoin nos diferentes horizontes de previsão e nas diferentes amostras de treinamento consideradas

Horizonte de Previsão	Período do Conjunto de Treinamento												Média
	1T20	2T20	3T20	4T20	1T21	2T21	3T21	4T21	1T22	2T22	3T22	4T22	
1 min.	7,00	7,23	6,92	7,03	7,04	6,83	6,85	7,12	7,40	7,04	6,85	6,85	7,01
5 min.	6,40	6,39	6,28	6,32	6,30	6,30	6,65	6,63	6,78	6,61	6,51	6,69	6,49
15 min.	6,33	6,37	6,17	6,16	6,13	6,12	6,21	6,27	6,44	6,34	6,28	6,27	6,26
1 hora	6,10	6,07	5,94	5,96	5,79	5,58	5,88	5,84	6,03	6,03	5,90	5,96	5,92
6 horas	5,68	5,82	5,63	5,56	5,02	5,03	5,86	5,95	5,96	5,92	5,73	5,93	5,67
1 dia	5,26	5,54	5,46	4,81	4,27	4,28	6,16	6,02	5,76	6,06	5,59	5,29	5,37

Nota: O termo “1T20” refere-se ao primeiro trimestre do ano 2020.

Fonte: dados da pesquisa (2023).

## APÊNDICE B – Resultados Regressão Logística

### B1 - Valor médio de cada parâmetro do modelo de regressão logística (Eq. 30)

Atributo	Horizontes de Previsão					
	1 min.	5 min.	15 min.	1 hora	6 horas	1 dia
Fechamento 1 min.	14,25	-30,71	-33,28	-9,76	8,79	15,55
Fechamento 5 min.	0,79	-16,63	-7,17	0,68	10,89	19,47
Fechamento 15 min.	6,23	11,99	1,35	5,65	9,95	20,37
Fechamento 1 hora	1,43	4,40	3,75	5,26	13,67	14,78
Fechamento 6 horas	0,22	0,56	1,46	2,72	5,33	9,24
Fechamento 1 dia	0,26	0,17	0,33	-0,20	-2,83	-0,22
Max. 1 min.	-37,99	21,75	21,86	4,47	-8,09	-15,38
Max. 5 min.	-0,34	7,46	3,32	-3,94	-9,64	-13,95
Max. 15 min.	-10,32	-12,83	-2,94	-9,46	-10,99	-15,95
Max. 1 hora	-5,53	-14,42	-15,31	-10,61	-8,27	-12,33
Max. 6 horas	0,01	-1,06	-3,29	-5,86	-10,84	-10,49
Max. 1 dia	-0,05	0,19	-0,01	-0,59	0,01	-4,65
Min. 1 min.	-38,02	17,7	11,31	7,67	-1,42	0,25
Min. 5 min.	-1,21	2,49	4,82	-4,99	-4,25	-2,59
Min. 15 min.	-12,01	-12,35	-2,88	-4,65	-4,77	-7,81
Min. 1 hora	-3,38	-7,68	-6,61	-2,54	-6,24	-4,91
Min. 6 horas	-0,10	-1,08	-2,90	-5,52	-8,31	-5,81
Min. 1 dia	-0,15	0,00	-0,29	-0,63	0,65	-2,63

Nota: Os valores são calculados utilizando os parâmetros das funções encontradas nas modelagens com as 12 diferentes bases de treinamento.

Fonte: dados da pesquisa (2023).

**B2 - 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 nos anos de 2020 a 2022**

Horizonte de Previsão	Período do Conjunto Teste												
	1T20	2T20	3T20	4T20	1T21	2T21	3T21	4T21	1T22	2T22	3T22	4T22	Total
<b>Painel A: Acurácia</b>													
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.	<b>0,537</b>	<b>0,541</b>	0,534	0,527	0,533	<b>0,534</b>	<b>0,529</b>	0,516	0,523	0,517	0,518	0,523	0,528
15 min.	0,533	0,540	0,542	<b>0,539</b>	<b>0,538</b>	0,529	0,525	0,525	0,529	0,515	0,527	0,527	<b>0,531</b>
1 hora	0,526	0,540	0,537	0,524	0,523	0,522	0,521	0,522	0,529	0,526	<b>0,542</b>	0,520	0,528
6 horas	0,518	0,530	0,523	0,514	0,520	0,522	0,511	<b>0,533</b>	0,514	<b>0,532</b>	0,519	0,514	0,521
1 dia	0,467	0,533	<b>0,620</b>	0,490	0,534	0,502	0,500	0,504	<b>0,543</b>	0,506	0,530	<b>0,545</b>	0,523
<b>Painel B: Sensibilidade</b>													
1 min.	0,503	<b>0,541</b>	0,435	0,476	0,533	0,516	<b>0,557</b>	<b>0,55</b>	<b>0,508</b>	<b>0,550</b>	0,541	0,671	<b>0,532</b>
5 min.	<b>0,519</b>	0,514	0,49	<b>0,502</b>	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	<b>0,595</b>	0,590	0,403	0,393	0,505	0,505	<b>0,590</b>	<b>0,783</b>	0,525
6 horas	0,433	0,480	0,454	0,425	0,502	<b>0,66</b>	0,323	0,361	0,476	0,517	0,475	0,583	0,472
1 dia	0,366	0,359	<b>0,667</b>	0,433	0,509	0,649	0,224	0,243	0,453	0,400	0,358	0,375	0,421
<b>Painel C: Especificidade</b>													
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	<b>0,528</b>	0,612	0,540	0,557	0,569	0,520	0,287	0,536
1 hora	<b>0,627</b>	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	<b>0,601</b>	<b>0,640</b>	0,541	0,387	0,707	0,703	0,555	0,546	0,563	0,451	0,572
1 dia	0,576	<b>0,756</b>	0,554	0,599	<b>0,569</b>	0,377	<b>0,815</b>	<b>0,78</b>	<b>0,636</b>	<b>0,585</b>	<b>0,684</b>	<b>0,695</b>	<b>0,635</b>
<b>Painel D: AUC</b>													
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.	<b>0,549</b>	0,558	0,550	0,537	0,546	<b>0,548</b>	<b>0,540</b>	0,522	0,531	0,523	0,522	0,538	0,539
15 min.	0,542	0,553	0,554	<b>0,552</b>	0,553	0,544	0,534	0,536	0,539	0,520	0,533	0,548	<b>0,542</b>
1 hora	0,539	0,549	0,547	0,542	0,531	0,527	0,534	0,526	0,540	0,534	<b>0,555</b>	<b>0,551</b>	0,540
6 horas	0,531	0,539	0,535	0,551	0,539	0,538	0,523	<b>0,545</b>	0,526	<b>0,543</b>	0,523	0,513	0,534
1 dia	0,486	<b>0,569</b>	<b>0,621</b>	0,528	<b>0,557</b>	0,523	0,531	0,526	<b>0,563</b>	0,487	0,520	0,534	0,537

Nota: 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.

Fonte: dados da pesquisa (2023).

**B3 - Porcentagem de valorização do Bitcoin por faixas de previsão das saídas dos modelos de regressão logística nos diferentes horizontes de previsão**

Faixa de Previsão	Horizonte de Previsão					
	1 min.	5 min.	15 min.	1 hora	6 horas	1 dia
0%-5%	-	-	-	-	-	-
5%-10%	-	35.0%	33.3%	46.7%	7.0%	53.5%
10%-15%	44.2%	39.6%	40.1%	37.6%	12.4%	46.1%
15%-20%	46.4%	40.9%	42.5%	49.9%	43.3%	51.7%
20%-25%	48.3%	43.4%	42.4%	49.3%	43.8%	57.2%
25%-30%	44.5%	43.9%	41.8%	45.4%	40.7%	55.3%
30%-35%	44.9%	43.9%	45.0%	45.0%	50.8%	52.3%
35%-40%	44.7%	44.4%	44.7%	46.9%	47.9%	56.1%
40%-45%	45.5%	44.4%	45.4%	46.8%	48.9%	51.8%
45%-50%	48.3%	47.6%	47.6%	48.2%	49.5%	49.2%
50%-55%	51.3%	52.4%	53.0%	53.3%	52.9%	55.4%
55%-60%	54.0%	55.7%	55.1%	54.1%	55.7%	55.5%
60%-65%	55.8%	56.4%	56.5%	55.7%	56.6%	58.1%
65%-70%	56.9%	57.4%	57.8%	56.0%	57.1%	58.6%
70%-75%	54.5%	58.9%	57.6%	60.4%	62.3%	56.7%
75%-80%	62.0%	58.8%	62.3%	63.6%	67.4%	65.4%
80%-85%	59.3%	63.7%	65.5%	66.6%	73.6%	68.1%
85%-90%	50.7%	60.6%	68.3%	70.2%	73.6%	77.7%
90%-95%	53.3%	70.8%	73.6%	77.3%	84.8%	52.4%
95%-100%	45.5%	67.3%	65.6%	55.7%	81.7%	66.9%

Nota: As faixas de previsão foram criadas a partir da categorização das saídas dos modelos. Estão ilustrados apenas os resultados nas situações com mais de 30 ocorrências.

Fonte: dados da pesquisa (2023).

**B4** - Tempo de processamento (em minutos) para aplicação dos modelos de regressão logística para previsão da direção futura do preço do Bitcoin nos diferentes horizontes de previsão e nas diferentes amostras de treinamento consideradas

Horizonte de Previsão	Período do Conjunto de Treinamento												Média
	1T20	2T20	3T20	4T20	1T21	2T21	3T21	4T21	1T22	2T22	3T22	4T22	
1 min.	2,7	2,4	1,9	2,0	2,6	2,4	2,2	2,1	3,5	2,9	0,5	0,8	2,2
5 min.	2,8	2,5	1,7	2,3	1,5	1,8	1,5	1,0	2,4	1,4	1,7	1,1	1,8
15 min.	3,1	1,9	1,1	1,4	1,4	2,1	31,3	2,2	2,2	2,2	1,4	2,0	4,4
1 hora	0,5	2,1	1,6	2,1	1,7	2,5	2,4	1,4	0,8	1,4	2,4	2,5	1,8
6 horas	1,4	2,3	2,1	2,3	1,3	2,9	1,8	1,4	1,4	1,0	1,4	2,5	1,8
1 dia	2,0	0,9	0,9	1,8	1,1	0,7	1,7	1,5	1,7	1,4	1,3	1,3	1,4

Nota: O termo “1T20” refere-se ao primeiro trimestre do ano 2020.

Fonte: dados da pesquisa (2023).

## APÊNDICE C – Resultados ARIMA

**C1** - Estrutura ARMA( $p,q$ ) ótima encontrada em cada conjunto de treinamento para cada horizonte de previsão

Horizonte de Previsão	Período do Conjunto de Treinamento											
	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)

Nota: 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.

Fonte: dados da pesquisa (2023).

**C2** - *p-values* dos testes Ljung-Box de autocorrelação dos resíduos dos modelos ARIMA nos diferentes períodos de conjunto de treinamento

Horizonte de Previsão	Período do Conjunto de Treinamento												
	1T20	2T20	3T20	4T20	1T21	2T21	3T21	4T21	1T22	2T22	3T22	4T22	Média
1 min.	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>
5 min.	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>
15 min.	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,013</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,005</b>	0,404	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>
1 hora	<b>0,004</b>	<b>0,000</b>	0,901	0,108	<b>0,000</b>	<b>0,004</b>	<b>0,004</b>						
6 horas	0,837	0,341	<b>0,000</b>	<b>0,012</b>	<b>0,001</b>	0,191	0,05	<b>0,001</b>	0,333	0,377	0,081	<b>0,019</b>	0,837
1 dia	0,748	0,786	0,849	0,897	0,37	0,136	0,804	0,287	0,944	0,504	0,284	0,576	0,748

Nota: Valores menores que 0,05, demarcados em negrito, indicam existência de autocorrelação na série dos resíduos.

Fonte: dados da pesquisa (2023).

**C3** - Tempo de processamento (em segundos) para aplicação dos modelos ARIMA para previsão da direção futura do preço do Bitcoin nos diferentes horizontes de previsão e nas diferentes amostras de treinamento consideradas

Horizonte de Previsão	Período do Conjunto de Treinamento												
	1T20	2T20	3T20	4T20	1T21	2T21	3T21	4T21	1T22	2T22	3T22	4T22	Média
1 min.	10,7	40	41,8	41,7	33,9	21,9	22,7	21,4	45,9	42,6	18,0	7,8	29,0
5 min.	5,8	4,1	6,8	5,1	5,9	2,8	1,4	6,9	4,4	3,4	10,8	4,2	5,1
15 min.	2,4	2,0	4,9	1,0	1,9	0,8	1,0	0,9	0,4	0,4	2,8	1,6	1,7
1 hora	2,4	4,1	0,3	0,2	0,5	0,8	0,3	0,2	3,6	0,1	0,1	0,2	1,1
6 horas	0,3	0,2	0,1	0,7	1,6	0,3	0,2	0,1	0,0	0,0	0,0	0,0	0,3
1 dia	0,2	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Nota: O termo “1T20” refere-se ao primeiro trimestre do ano 2020.

Fonte: dados da pesquisa (2023).

**C4 - 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**

Horizonte de Previsão	Período do Conjunto Teste												
	1T20	2T20	3T20	4T20	1T21	2T21	3T21	4T21	1T22	2T22	3T22	4T22	Total
<b>Painel A: Acurácia</b>													
1 min.	0,515	0,525	0,501	0,503	0,486	0,513	0,505	0,501	0,513	0,500	0,492	0,503	0,505
5 min.	0,531	0,536	0,523	0,528	0,458	0,516	0,519	0,503	0,514	<b>0,523</b>	0,503	0,518	0,515
15 min.	0,497	0,514	0,526	0,544	0,472	0,519	0,529	0,519	0,493	0,507	0,496	0,489	0,509
1 hora	0,474	0,522	0,549	0,552	0,479	0,501	<b>0,532</b>	<b>0,54</b>	<b>0,535</b>	0,464	0,495	0,493	0,511
6 horas	0,518	<b>0,564</b>	0,520	0,544	0,508	<b>0,526</b>	0,518	0,501	0,507	0,488	0,501	0,507	0,517
1 dia	<b>0,584</b>	0,371	<b>0,571</b>	<b>0,600</b>	<b>0,523</b>	0,494	0,478	0,516	0,517	0,467	<b>0,549</b>	<b>0,571</b>	<b>0,52</b>
<b>Painel B: Sensibilidade</b>													
1 min.	0,441	0,496	0,527	0,535	0,488	0,597	0,505	<b>0,487</b>	0,533	0,503	<b>0,406</b>	0,011	0,460
5 min.	0,533	0,48	0,522	0,565	0,470	0,787	0,516	0,483	0,516	0,527	0,393	<b>0,294</b>	0,507
15 min.	0,502	0,514	0,519	0,637	0,500	0,744	0,545	0,486	0,844	<b>0,591</b>	0,309	0,149	0,527
1 hora	0,477	0,521	0,551	0,726	0,516	0,707	0,497	0,474	0,533	0,482	0,002	0,060	0,464
6 horas	0,505	<b>0,538</b>	0,536	0,486	0,578	0,823	0,532	0,376	0,906	0,469	0,000	0,011	0,480
1 dia	<b>0,568</b>	0,135	<b>1,000</b>	<b>0,831</b>	<b>0,979</b>	<b>1,000</b>	<b>0,867</b>	0,133	<b>0,978</b>	0,250	0,000	0,026	<b>0,583</b>
<b>Painel C: Especificidade</b>													
1 min.	0,586	0,554	0,475	0,471	<b>0,483</b>	<b>0,43</b>	0,504	0,515	0,493	0,496	0,579	<b>0,988</b>	0,548
5 min.	0,529	0,592	0,524	0,489	0,447	0,252	0,522	0,523	0,513	0,52	0,616	0,743	0,523
15 min.	0,492	0,514	0,533	0,444	0,444	0,298	0,512	0,552	0,141	0,424	0,687	0,830	0,490
1 hora	0,470	0,524	<b>0,547</b>	0,349	0,438	0,299	<b>0,569</b>	0,603	<b>0,537</b>	0,446	0,993	0,922	<b>0,561</b>
6 horas	0,532	0,590	0,503	<b>0,622</b>	0,433	0,247	0,503	0,624	0,101	0,505	<b>1,000</b>	0,979	0,555
1 dia	<b>0,600</b>	<b>0,703</b>	0,025	0,161	0,000	0,000	0,089	<b>0,891</b>	0,045	<b>0,64</b>	<b>1,000</b>	0,981	0,454
<b>Painel D: AUC</b>													
1 min.	0,514	0,525	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	0,536	0,523	0,527	0,458	0,519	0,519	0,503	0,514	<b>0,524</b>	<b>0,504</b>	<b>0,519</b>	0,515
15 min.	0,497	0,514	0,526	0,541	0,472	0,521	0,529	0,519	0,492	0,507	0,498	0,489	0,509
1 hora	0,473	0,522	<b>0,549</b>	0,538	0,477	0,503	<b>0,533</b>	<b>0,539</b>	<b>0,535</b>	0,464	0,497	0,491	0,510
6 horas	0,519	<b>0,564</b>	0,519	<b>0,554</b>	<b>0,505</b>	<b>0,535</b>	0,517	0,500	0,504	0,487	0,500	0,495	<b>0,517</b>
1 dia	<b>0,584</b>	0,419	0,513	0,496	0,489	0,500	0,478	0,512	0,512	0,445	0,500	0,503	0,496

Nota: 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.

Fonte: dados da pesquisa (2023).