



CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

SISTEMAS *Fuzzy* EVOLUTIVOS NA PREVISÃO E RECOMENDAÇÃO DE INVESTIMENTOS EM CRIPTOMOEDAS

VINÍCIUS LUIZ DO AMARAL

Orientador: Prof. Dr. Alisson Marques da Silva
Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)

BELO HORIZONTE
OUTUBRO DE 2021

VINÍCIUS LUIZ DO AMARAL

SISTEMAS *Fuzzy* EVOLUTIVOS NA PREVISÃO E RECOMENDAÇÃO DE INVESTIMENTOS EM CRIPTOMOEDAS

Dissertação apresentada ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para a obtenção do título de Mestre em Modelagem Matemática e Computacional.

Área de concentração: Modelagem Matemática e Computacional

Linha de pesquisa: Sistemas Inteligentes

Orientador: Prof. Dr. Alisson Marques da Silva
Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL
BELO HORIZONTE
OUTUBRO DE 2021

Amaral, Vinícius Luiz do
A485s Sistemas Fuzzy evolutivos na previsão e recomendação de investimentos
em criptomoedas / Vinícius Luiz do Amaral. – 2021.
104 f.

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em
Modelagem Matemática e Computacional.

Orientador: Alisson Marques da Silva.

Dissertação (mestrado) – Centro Federal de Educação Tecnológica de
Minas Gerais.

1. Sistemas difusos – Teses. 2. Finanças – Processamento de dados –
Teses. 3. Probabilidades – Teses. 4. Criptomoedas – Teses. 5. Bitcoin – Teses.
6. Blockchains (bases de dados) – Teses. I. Silva, Alisson Marques da.
II. Centro Federal de Educação Tecnológica de Minas Gerais. III. Título.

CDD 519.5



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
COORDENAÇÃO DO CURSO DE MESTRADO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

**“SISTEMAS FUZZY EVOLUTIVOS NA PREVISÃO E RECOMENDAÇÃO
DE INVESTIMENTOS EM CRIPTOMOEDAS”**

Dissertação de Mestrado apresentada por **Vinícius Luiz do Amaral**, em 26 de outubro de 2021,
ao Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG,
e aprovada pela banca examinadora constituída pelos professores:

Prof. Dr. Alisson Marques da Silva (Orientador)
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Leandro dos Santos Maciel
Universidade de São Paulo

Prof. Dr. Gray Farias Moita
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Adriano Chaves Lisboa
Centro Federal de Educação Tecnológica de Minas Gerais

Visto e permitida a impressão,

Prof^a. Dr^a. Elizabeth Fialho Wanner
Presidenta do Colegiado do Programa de Pós-Graduação em
Modelagem Matemática e Computacional

Dedico este trabalho aos meus pais pelo amor, por todos os ensinamentos e por abrirem mão do pouco que tinham para que eu tivesse acesso a uma educação melhor.

Agradecimentos

Primeiramente, sou grato a Deus por conduzir em graça todos os meus dias. Tenho fé que este momento e os caminhos que percorrerei em decorrência dele são partes importantes da minha peregrinação.

Aos meus pais, Luiz e Regina, pelos valores ensinados e por priorizarem uma educação de qualidade aos seus filhos em detrimento do próprio conforto.

À minha esposa, Karine, por todo amor, cuidado e compreensão, principalmente, nos constantes momentos de ausência. Obrigado por escolher sonhar os meus sonhos.

Aos meus irmãos, Virgínia, Vitória e Victor, pelo apoio incondicional e por me ensinarem pelo exemplo.

Ao Prof. Alisson, por me orientar com total maestria e pela disponibilidade nos horários mais diversos. Certamente o seu comprometimento com meu trabalho e ao longo das disciplinas lecionadas me inspiram, ainda mais, a seguir carreira acadêmica.

A todo o corpo docente do programa, em especial aos professores Gray Farias, Antônio Scarpelli, Marcone Jamilson, Thiago Dias, Thiago Mattos, Gustavo Menezes e Paulo Almeida (*in memoriam*), pela rica contribuição com a minha formação.

Aos professores Aluísio Eustáquio, Daniel Dalip, Maria Augusta, Rommel Carneiro e Carolina Marcelino por fazerem parte da minha trajetória e por também servirem de inspiração para seguir por este caminho.

Enfim, a todos que estiveram ao meu lado e que, de alguma forma, me apoiaram nesta caminhada.

“O que foi tornará a ser, o que foi feito se fará novamente; não há nada novo debaixo do sol.”
(Eclesiastes 1:9)

Resumo

Os Sistemas Evolutivos (*Evolving Systems* ou ES) foram propostos para atender a demanda existente no processamento *online* de dados, no qual as informações geralmente são disponibilizadas na forma de um fluxo contínuo, com características que envolvem a disponibilização em grandes quantidades, altamente dinâmicas e não-estacionárias. Adicionalmente, os Sistemas *Fuzzy* Evolutivos (*Evolving Fuzzy Systems* ou EFS) incorporam, de forma nativa, o tratamento de incertezas, o que é viabilizado pelo uso da lógica *Fuzzy* em detrimento da lógica booleana tradicional. Por sua vez, as criptomoedas são um novo tipo de ativo financeiro, que apresenta alta volatilidade se comparada a ativos tradicionais, proporcionando, portanto, um ambiente altamente dinâmico. Diante disso, neste trabalho é proposta uma nova abordagem que contempla a utilização de EFS na previsão do valor de fechamento, na previsão do sentido de variação e na recomendação de investimentos em criptomoedas, um passo adiante. Na previsão do valor de fechamento, são avaliados três EFS e a previsão do sentido de variação é realizada com base no valor previsto por eles. Para esta tarefa também é proposto um novo modelo, chamado Comitê, que consiste na combinação dos resultados apresentados pelos três EFS. Por último, propõe-se uma estratégia de recomendação de investimentos, que consiste na compra e/ou venda de criptomoedas, baseando-se na previsão do sentido de variação do seu preço. Experimentos computacionais foram realizados para avaliar as abordagens utilizando 10 intervalos (conjuntos de dados), sendo 6 para o Bitcoin e 4 para o Ether. Os resultados obtidos foram comparados entre as abordagens e, quando possível, com resultados da literatura. Os resultados evidenciam que os EFS considerados postulam como bons candidatos na solução do problema de previsão do valor e do sentido de variação do preço de criptomoedas. Os números apresentados pelos modelos estudados foram competitivos, principalmente em termos de Acurácia e Rentabilidade, superando os trabalhos da literatura. Além disso, os modelos foram capazes de proporcionar ganhos superiores à valorização das criptomoedas em grande parte dos intervalos e, até mesmo, evitar maiores prejuízos em cenários de desvalorização.

Palavras-chave: Previsão, Recomendação de Investimentos, Sistemas *Fuzzy* Evolutivos, Criptomoedas, Bitcoin, Ether.

Abstract

Evolving Systems (ES) were proposed to meet the existing demand for online data processing, in which information is made available in the form of a continuous flow, with characteristics that involve availability in massive quantities, highly dynamic and not stationary. Additionally, Evolving Fuzzy Systems (EFS) natively incorporate the treatment of uncertainties, which is made possible using Fuzzy Logic instead of traditional Boolean Logic. On the other hand, cryptocurrencies are a new type of financial asset, which has high volatility compared to traditional assets, therefore providing a highly dynamic environment. Thus, this work proposes novel approaches that contemplate EFS's use in closing value forecasting, in the direction of variation forecasting, and investments recommendation in cryptocurrencies. In predicting the closing value, three Evolving Systems are evaluated. The direction of variation is obtained based on the value forecasted by the Evolving Systems. For this task, a new model (named the Committee) is also proposed, combining the results presented by the three EFS. Finally, an investment recommendation strategy is proposed, which includes the purchase and/or sale of cryptocurrencies based on the prediction of the direction of variation. Computational experiments were conducted to evaluate the approaches using 10 intervals (data sets), 6 for Bitcoin and 4 for Ether. The results obtained compared the proposed approaches and, when possible, also with results from the literature. The computational experiments show that the proposed approaches achieved better or comparable results to those in the literature. In addition, the models were able to provide gains greater than the valuation of cryptocurrencies in most of the intervals and even avoid greater losses in devaluation scenarios. Summarizing, the results show that the proposed approaches are promising and suitable candidates for forecasting and recommending in the cryptocurrencies market.

Keywords: Forecast, Investment Recommendation, Evolving Fuzzy Systems, Cryptocurrencies, Bitcoin, Ether.

Lista de Figuras

Figura 1 – Padrões encontrados em séries temporais.	10
Figura 2 – Aprendizagem por agrupamento (<i>Bagging, Boosting e Stacking</i>).	17
Figura 3 – Bitcoins em circulação.	28
Figura 4 – Evolução do número total de transações envolvendo o Bitcoin.	29
Figura 5 – Volume diário de transações envolvendo o Bitcoin.	29
Figura 6 – Visão macro da abordagem proposta.	54
Figura 7 – Período de seleção de variáveis e intervalo de interesse.	57
Figura 8 – Representação da etapa de seleção de variáveis.	59
Figura 9 – Representação da fase de previsão do sentido de variação do preço de fechamento de criptomoedas.	61
Figura 10 – Evolução do preço de fechamento do Bitcoin nos intervalos BTC-1 e BTC-2.	66
Figura 11 – Evolução do preço de fechamento do Bitcoin nos intervalos BTC-3, BTC-4, BTC-5 e BTC-6.	66
Figura 12 – Evolução do preço de fechamento do Ether nos intervalos ETH-3, ETH-4, ETH-5 e ETH-6.	67
Figura 13 – Comparação entre o valor real e o previsto nos intervalos BTC-1 e BTC-2.	69
Figura 14 – Comparação entre o valor real e o previsto nos intervalos BTC-3, BTC-4, ETH-3 e ETH-4.	70
Figura 15 – Comparação entre o valor real e o previsto nos intervalos BTC-5, BTC-6, ETH-5 e ETH-6.	71
Figura 16 – Evolução da rentabilidade na recomendação de investimentos no intervalo BTC-1.	77
Figura 17 – Evolução da rentabilidade na recomendação de investimentos no intervalo BTC-2.	78
Figura 18 – Evolução da rentabilidade na recomendação de investimentos nos intervalos BTC-3 e BTC-4.	79
Figura 19 – Evolução da rentabilidade na recomendação de investimentos nos intervalos ETH-3 e ETH-4.	79
Figura 20 – Evolução da rentabilidade na recomendação de investimentos nos intervalos BTC-5 e BTC-6.	81
Figura 21 – Evolução da rentabilidade na recomendação de investimentos nos intervalos ETH-5 e ETH-6.	81

Lista de Tabelas

Tabela 1 – Evolução anual de importantes ativos/índices financeiros (2016 a 2020).	3
Tabela 2 – Quantitativo da volatilidade diária do preço do Bitcoin, se positiva ou negativa, em relação a intervalos de variação pré-estabelecidos (2016 a 2020).	5
Tabela 3 – Comparação acerca de algumas das principais características existentes em criptomoedas para Bitcoin (BTC), Litecoin (LTC), Ripple (XRP) e Ether (ETH).	32
Tabela 4 – Transformação de problemas de regressão em problemas de classificação.	34
Tabela 5 – Resumo dos trabalhos relacionados.	46
Tabela 6 – Conjunto de atributos e valores considerados na geração de variáveis por defasagem e médias móveis simples.	56
Tabela 7 – Resumo dos conjuntos de dados utilizados nos experimentos.	68
Tabela 8 – Desempenho na previsão do valor de fechamento nos intervalos BTC-1 e BTC-2.	68
Tabela 9 – Desempenho na previsão do valor de fechamento nos intervalos BTC-3, BTC-4, ETH-3 e ETH-4.	69
Tabela 10 – Desempenho na previsão do valor de fechamento nos intervalos BTC-5, BTC-6, ETH-5 e ETH-6.	70
Tabela 11 – Comparação entre os modelos que apresentaram o melhor e o pior desempenho na previsão do valor de fechamento.	72
Tabela 12 – Desempenho na previsão do sentido de variação do preço de fechamento nos intervalos BTC-1 e BTC-2.	72
Tabela 13 – Desempenho na previsão do sentido de variação do preço de fechamento nos intervalos BTC-3, BTC-4, ETH-3 e ETH-4.	73
Tabela 14 – Desempenho na previsão do sentido de variação do preço de fechamento nos intervalos BTC-5, BTC-6, ETH-5 e ETH-6.	74
Tabela 15 – Comparação do desempenho na previsão do sentido de variação do preço de fechamento entre os intervalos que antecedem o dia 11/03/2020 (BTC-3, BTC-4, ETH-3 e ETH-4) e os intervalos que compreendem esta data (BTC-5, BTC-6, ETH-5 e ETH-6).	75
Tabela 16 – Rentabilidade Acumulada (RA) e Retorno por Unidade de Risco (RR) na recomendação de investimentos nos intervalos BTC-1 e BTC-2.	76
Tabela 17 – Rentabilidade Acumulada (RA) e Retorno por Unidade de Risco (RR) na recomendação de investimentos nos intervalos BTC-3, BTC-4, ETH-3 e ETH-4.	78

Tabela 18 – Rentabilidade Acumulada (RA) e Retorno por Unidade de Risco (RR) na recomendação de investimentos nos intervalos BTC-5, BTC-6, ETH-5 e ETH-6.	80
Tabela 19 – Análise de evolução da rentabilidade: quando um modelo é superior aos demais.	82
Tabela 20 – Análise de evolução da rentabilidade: quando um modelo é inferior aos demais.	83
Tabela 21 – Dados numéricos do valor de fechamento ao longo dos intervalos utilizados nos experimentos.	98
Tabela 22 – Aumentos (+) ou diminuições (-) sucessivas ao longo dos intervalos utilizados nos experimentos.	98

Lista de Algoritmos

Algoritmo 1 – Generalização da Estratégia de Investimento	62
---	----

Lista de Abreviaturas e Siglas

APPP	<i>Average Profit per Position</i>
A/D	<i>Accumulation/Distribution Oscillator</i>
ANFIS	<i>Adaptive Neuro Fuzzy Inference System</i>
ANN	<i>Artificial Neural Network</i>
APARCH	<i>Asymetric Power GARCH</i>
API	<i>Application Programming Interface</i>
AR	<i>Autoregressive</i>
ARCH	<i>Autoregressive Conditional Heteroskedasticity</i>
ARMA	<i>Autoregressive Moving Average</i>
ARIMA	<i>Autoregressive Integrated Moving Average</i>
altcoins	<i>Alternative Coins</i>
BART	<i>Binary Auto Regressive Tree</i>
BDTR	<i>Boosted Decision Tree Regression</i>
BLR	<i>Bayesian Linear Regression</i>
BRNN	<i>Bayesian Regularization Neural Network</i>
BRL	<i>Brazilian Real</i>
BTC	Bitcoin
BVSP	Bolsa de Valores de São Paulo
CCI	<i>Commodity Channel Index</i>
CFS	<i>Correlation-based Feature Subset Selection</i>
CGARCH	<i>Component GARCH</i>
CNN	<i>Convolutional Neural Network</i>
CORR	<i>Correlation Analisis</i>
daaps	<i>Decentralized Applications</i>

DAX	<i>Deutscher Aktienindex</i>
DJI	<i>Dow Jones Industrial Average</i>
DNN	<i>Deep Neural Network</i>
DRN	<i>Deep Residual Network</i>
EGARCH	<i>Exponential GARCH</i>
eGNN	<i>evolving Granular Neural Network</i>
EFS	<i>Evolving Fuzzy Systems</i>
EOD	<i>End of Day</i>
ePL-I	<i>Adaptive Interval Fuzzy Modeling</i>
EL	<i>Ensemble Learning</i>
eMG	<i>Multivariable Gaussian Evolving Fuzzy Modeling System</i>
eOGS	<i>evolving Optimal Granular System</i>
ETH	<i>Ether</i>
EUR	<i>Euro</i>
FIS	<i>Fuzzy Inference System</i>
FFNN	<i>Feedforward Neural Networks</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
GARCH	<i>Generalized AutoRegressive Conditional Heteroskedasticity</i>
GARCH-M	<i>GARCH In Mean</i>
GLM	<i>Generalized Linear Model</i>
GRP	<i>Gaussian Poisson Regressions</i>
GRU	<i>Gated Recurrent Unit</i>
GTB	<i>Gradient Tree Boosting</i>
HST	<i>Decision Token</i>
IGARCH	<i>Integrated ARCH</i>

IT	Indicadores Técnicos
k-NN	<i>k-Nearest Neighbors</i>
LDA	<i>Linear Discriminant Analysis</i>
LNR	<i>Logarithmic Return</i>
LPG	<i>Largest Percentual Gain</i>
LPL	<i>Largest Percentual Loss</i>
LR	<i>Linear Regression</i>
LSTM	<i>Long Short-Term Memory</i>
LTC	Litecoin
LWI	<i>Larry William's R</i>
MA	<i>Moving Average</i>
MACD	<i>Moving Average Convergence Divergence</i>
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MDN	<i>Maximum Drawdown</i>
MSE	<i>Mean Squared Error</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
NASDAQ	<i>National Association of Securities Dealers Automated Quotations</i>
NDEI	<i>Non-Dimensional Error Index</i>
OBV	<i>On Balance Volume</i>
OSA	<i>One Step Ahead</i>
PCA	<i>Principal Component Analysis</i>
PCC	<i>Pearson's Correlation Coefficient</i>
PCR	<i>Pearson Correlation</i>
PI	<i>Permutation Importance</i>

PIB	Produto Interno Bruto
PoS	<i>Proof-of-Stake</i>
PoW	<i>Proof-of-Work</i>
PPM	<i>Percentage of Periods in Market</i>
PPP	<i>Percentage of Profitable Positions</i>
PV	<i>Parkinson's Volatility Index</i>
QDA	<i>Quadratic Discriminant Analysis</i>
RA	<i>Rentabilidade Acumulada</i>
RBFNN	<i>Radial Basis Function Neural Network</i>
RDF	<i>Random Decision Forests</i>
RF	<i>Random Forests</i>
RMSE	<i>Root Mean Square Error</i>
RNN	<i>Recurrent Neural Networks</i>
RoC	<i>Rate of Change</i>
ROI	<i>Return of Investment</i>
RR	<i>Rentabilidade por Unidade de Risco</i>
RSI	<i>Relative Strength Index</i>
RT	<i>Regression Trees</i>
SANN	<i>Stacked Artificial Neural Network</i>
SCC	<i>Spearman Correlation Coefficient</i>
SD	<i>Stochastic D%</i>
SGBM	<i>Stochastic Gradient Boosting Machine</i>
SHR	<i>SharpeRatio</i>
SK	<i>Stochastic K%</i>
SMA	<i>Simple Moving Average</i>
SOC	<i>Stochastic Oscillator</i>

SRR	<i>Shortino Ratio</i>
SVM	<i>Support Vector Machine</i>
SVR	<i>Support Vector Regression</i>
S&P 500	<i>Standard & Poor's 500</i>
TGARCH	<i>Threshold GARCH</i>
TME	<i>Triple Moving</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
USD	<i>United States Dollar</i>
UTC	<i>Universal Time Control</i>
VIF	<i>Variance Inflation</i>
VMD	<i>Variational Mode Decomposition</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>
WMA	<i>Weighted Moving Average</i>
XGB	<i>Extreme Gradient Boosting</i>
XMR	<i>Monero</i>
XRP	<i>Ripple</i>

Lista de Símbolos

c	Constante
C, C^f	Operador de agregação
d	Ordem do modelo de integração ou da parte integrada de um modelo
G	Conjunto <i>fuzzy</i>
h	Coefficientes do consequente das regras nos modelos evolutivos
r	Taxa de adaptação no modelo eGNN
i, j	Iteradores
k	Épocas de treinamento para uma Rede Neural Artificial
ln	Logaritmo natural
n	Tamanho total de uma série temporal ou conjunto de amostras
N	Número total de variáveis para uma amostra específica
m	Dias
o	Saída da camada de agregação no modelo eGNN
p	Ordem do modelo Autoregressivo ou da parte autoregressiva de um modelo
P_i	Patrimônio inicial
P_f	Patrimônio final
q	Ordem do modelo de Médias Móveis ou da parte de Médias Móveis de um modelo
R	Série de Retorno
R_i	Regra
R_f	Retorno Livre de Risco ou <i>Risk-Free Return</i>
$R\$$	Moeda brasileira (Real)
t	Instante

$US\$$	Moeda norte-americana (Dólar)
w	Peso associado a um neurônio em uma Rede Neural Artificial
x, y	Variável definida de acordo com o contexto abordado
\hat{y}	Saída estimada para um modelo de previsão
\hat{y}'	Sentido de variação previsto por um modelo evolutivo na abordagem proposta
\hat{y}''	Sentido de variação previsto pelo Comitê na abordagem proposta
α	Nível de significância no modelo eMG / ponto de corte das funções e dispersão inicial no modelo eOGS
γ	Conjunto de informações disponíveis para a realização de previsões de valores futuros em uma série temporal
δ	Pesos da camada de saída no modelo eGNN
ϵ	Série temporal do tipo ruído branco (<i>white noise</i>)
η	Taxa de crescimento dos grânulos no modelo eGNN
θ, κ	Coefficientes de um modelo ou de parte de um modelo de séries temporais
λ	Taxa de aprendizado no modelo eMG
ν	Limite para exclusão de regras por inatividade no modelo eOGS
ρ	Limiar de granularidade no modelo eGNN
σ	Desvio padrão
σ^2	Variância
v	Parte imprevisível ou ruído branco (<i>white noise</i>)
v'	Primeira parte imprevisível ou ruído branco (<i>white noise</i>)
v''	Segunda parte imprevisível ou ruído branco (<i>white noise</i>)
ϕ	Tamanho máximo dos conjuntos de variáveis na abordagem proposta
χ	Número de variáveis disponíveis após o ranqueamento na abordagem proposta
ψ	Velocidade de adaptação do grau de dispersão no modelo eOGS

ω Tamanho da janela no modelo eMG / distância mínima entre os grânulos no modelo eOGS

Σ_{init} Matriz de dispersão inicial no modelo eMG

Sumário

1 – Introdução	1
1.1 Contextualização	1
1.2 Motivação e Relevância	3
1.3 Objetivos	6
1.3.1 Objetivo Geral	6
1.3.2 Objetivos Específicos	6
1.4 Principais Contribuições	6
1.5 Publicações	7
1.6 Organização do Trabalho	8
2 – Fundamentação Teórica	9
2.1 Séries Temporais	9
2.2 Modelagem de Séries Temporais	11
2.2.1 Modelos Autorregressivos, Integrados e de Médias Móveis	12
2.2.2 Técnicas de Aprendizagem de Máquina	14
2.2.2.1 Aprendizagem por Agrupamento	15
2.2.2.2 Sistemas <i>Fuzzy</i> Evolutivos	17
2.2.2.2.1 eGNN - evolving Granular Neural Network	18
2.2.2.2.2 eMG - Multivariable Gaussian Evolving Fuzzy Modeling System	20
2.2.2.2.3 eOGS - evolving Optimal Granular System	22
2.3 Medidas de Desempenho	23
2.3.1 Regressão	23
2.3.2 Classificação	24
2.3.3 Rentabilidade e Risco	26
2.4 Blockchain e Criptomoedas	26
2.4.1 Bitcoin (BTC)	27
2.4.2 Litecoin (LTC)	29
2.4.3 Ripple (XRP)	30
2.4.4 Ether (ETH)	30
2.4.5 Análise Comparativa	31
3 – Trabalhos Relacionados	33
3.1 Abordagens	33
3.2 Conjunto de Dados	35
3.2.1 Dados de Cotação	36

3.2.2	Dados de Mercado	37
3.2.3	Dados da Rede	37
3.2.4	Dados Externos	38
3.2.5	Indicadores Técnicos	39
3.3	Seleção de Variáveis	40
3.4	Técnicas de Previsão	41
3.5	Critérios de Avaliação	43
3.6	Resumo dos Trabalhos Relacionados	44
4	– Metodologia	54
4.1	Construção do Conjunto de Dados	54
4.1.1	Extração	55
4.1.2	Pré-processamento	55
4.1.3	Informações Complementares	56
4.2	Seleção de Variáveis	56
4.2.1	Ranqueamento	57
4.2.2	Definição dos Melhores Conjuntos de Variáveis	58
4.3	Previsão e Recomendação	59
4.3.1	Previsão do Valor	60
4.3.2	Previsão do Sentido de Variação	60
4.3.3	Recomendação de Investimentos	61
5	– Experimentos Computacionais	64
5.1	Definição de Parâmetros	64
5.2	Conjuntos de Dados	65
5.3	Previsão do Valor	67
5.3.1	Intervalos BTC-1 e BTC-2	68
5.3.2	Intervalos BTC-3, BTC-4, ETH-3 e ETH-4	69
5.3.3	Intervalos BTC-5, BTC-6, ETH-5 e ETH-6	70
5.3.4	Discussão dos Resultados	71
5.4	Previsão do Sentido de Variação	72
5.4.1	Intervalos BTC-1 e BTC-2	72
5.4.2	Intervalos BTC-3, BTC-4, ETH-3 e ETH-4	73
5.4.3	Intervalos BTC-5, BTC-6, ETH-5 e ETH-6	73
5.4.4	Discussão dos Resultados	74
5.5	Recomendação de Investimentos	75
5.5.1	Intervalos BTC-1 e BTC-2	76
5.5.2	Intervalos BTC-3, BTC-4, ETH-3 e ETH-4	77
5.5.3	Intervalos BTC-5, BTC-6, ETH-5 e ETH-6	79
5.5.4	Discussão dos Resultados	80

6 – Considerações Finais	84
6.1 Propostas de Continuidade	85
Referências	86
Apêndices	95
APÊNDICE A – Detalhamento Numérico dos Intervalos	96
Anexos	99
ANEXO A – Conjuntos de Atributos Utilizados nos Trabalhos de Referência . .	100
A.1 Dados de Cotação	100
A.2 Dados de Mercado	100
A.3 Dados da Rede	101
A.4 Dados Externos	103

1 Introdução

Neste capítulo apresenta-se uma descrição inicial do problema abordado por este trabalho, juntamente com os principais desafios e motivações que norteiam essa pesquisa. Na Seção 1.1 introduz-se os temas Criptomoedas e *Blockchain*, dois tópicos principais sob os quais este trabalho está fundamentado. Em seguida, na Seção 1.2, apresenta-se as principais questões relacionadas à previsão no contexto de criptomoedas e a importância do estudo e do desenvolvimento de ferramentas que possam auxiliar na tomada de decisões sobre o investimento neste tipo de ativo. Apresentado o contexto, define-se, na Seção 1.3, os objetivos geral e específicos. Posteriormente, na Seção 1.4, destaca-se as principais contribuições, e, na Seção 1.5, são listados os trabalhos, frutos desta pesquisa. Por último, na Seção 1.6, descreve-se, de forma resumida, o conteúdo de cada um dos capítulos subsequentes.

1.1 Contextualização

No ano de 2008 foi decretada a falência do banco americano Lehman Brothers, considerada até então a maior falência da história dos Estados Unidos (AMADEO, 2020). No mesmo ano foi registrada uma grande queda do índice *Dow Jones*¹, conhecida como a queda de *subprime*, influenciada por problemas hipotecários e concessão desenfreada de crédito pelo sistema financeiro norte-americano em anos anteriores. Tudo isso contribuiu para que viesse à tona a maior crise econômica do Estados Unidos nos últimos 70 anos, gerando impacto na economia de diversos países (MAKATJANE; MOLEFE; WYK, 2018).

Em meio a este contexto, foi publicada a proposta de funcionamento da primeira moeda digital, com a promessa de provocar uma grande revolução no sistema financeiro, principalmente pela sua não suscetibilidade a intervenções estatais e entidades intermediárias (NAKAMOTO, 2008). O Bitcoin, nome dado a moeda, foi proposto por um ou mais programadores sob o pseudônimo Satoshi Nakamoto, e teve a sua primeira transação registrada em 3 de Janeiro de 2009. Trata-se do primeiro sistema de pagamentos global inteiramente descentralizado, sendo atualmente considerada a principal criptomoeda devido ao seu alto valor e volume de transações diárias (INVESTING, 2021). Desde então, com a popularização do Bitcoin, diversas outras moedas dessa natureza foram criadas, também conhecidas como *altcoins* (*Alternative Coins*, ou Moedas Alternativas), todas baseadas na mesma tecnologia, de modo que este assunto ganhou ainda mais notoriedade no mercado

¹ DJI (*Dow Jones Industrial Average*) é um índice financeiro utilizado para representar a situação do mercado de ações norte-americano, baseado na cotação das ações de 30 das principais empresas dos Estados Unidos.

financeiro mundial.

Conforme Ulrich (2014), a invenção do Bitcoin é inovadora porque, pela primeira vez, o Problema do Gasto Duplo² pode ser resolvido sem a necessidade de um intermediário central. Em uma transação financeira eletrônica convencional o Problema do Gasto Duplo é contornado a partir do envolvimento de uma terceira entidade, responsável pelo débito do valor transacionado na conta de origem e consequente crédito na conta de destino. Essa entidade garante que o mesmo dinheiro não será enviado por um único indivíduo, ao mesmo tempo, para dois destinatários (ou duas transações) diferentes. A utilização da *Blockchain*³ como ponto de partida na implementação do Bitcoin proporciona que todas as transações sejam persistidas em um registro público (por isso o nome *Blockchain*, ou cadeias de blocos de transações), distribuído a todos os usuários do sistema por meio de uma rede ponto-a-ponto. Novas transações são verificadas contra a *Blockchain* de modo a assegurar que os mesmos Bitcoins não tenham sido gastos previamente, eliminando assim o Problema do Gasto Duplo. Neste caso, a rede global ponto-a-ponto, composta por milhares de usuários, torna-se o próprio intermediário.

As criptomoedas são um tipo de ativo financeiro que pode proporcionar alta margem de lucro ao investidor, devido à alta volatilidade do seu preço. Porém, assim como é possível ganhar muito dinheiro em pouco tempo, também é possível, da mesma maneira, que se perca dinheiro muito rapidamente. Ulrich (2014) elenca alguns desafios no contexto de criptomoedas e que, inclusive, podem impactar no seu valor:

- falta legislação específica na maioria dos países, abrindo brechas para que algumas das definições basilares sobre as quais a moeda está fundamentada possam vir a ser alteradas no futuro;
- eventual possibilidade de proibição de sua utilização em determinada região, tornando ilegal o seu uso;
- alta aplicabilidade para fins criminosos, devido à possibilidade de anonimidade tanto de executores como de receptores em transações;
- violações de segurança por meio de ataques *hackers*;
- alta volatilidade, o que pode indicar a existência de bolhas especulativas.

O autor ressalta que muitos dos desafios enfrentados pelas criptomoedas também são enfrentados pelas moedas fiduciárias e outros tipos de ativos financeiros e, por isso, acredita que é possível dar prosseguimento no seu uso.

² O Problema do Gasto Duplo (*Double Spending Problem*) é um fenômeno em que uma única unidade monetária é gasta simultaneamente mais de uma vez, o que cria uma disparidade entre o registro de gastos e o valor dessa moeda disponível.

³ A *Blockchain* é uma tecnologia que funciona como um banco de dados distribuído e imutável, que pode ser utilizado para o registro de transações ou o rastreamento de ativos com maior grau de confiança, sem a dependência de um intermediário central.

1.2 Motivação e Relevância

Sabe-se que é de interesse mundial a análise do comportamento de séries temporais do mercado financeiro tais como o desempenho de índices de ações, o preço de *commodities*, índices inflacionários e o preço de criptomoedas. Por não serem reguladas/controladas por nenhuma agência oficial, a variação do preço das moedas digitais é consequência, principalmente, do aumento ou diminuição de sua oferta e demanda no mercado, o que pode ser influenciado por diversos fatores como aumento no número de carteiras digitais, transações por segundo, aceitação do ativo como meio de pagamento, consolidação do ativo como reserva de valor, dentre outros (VIANEZ; ROMAN; GOMEZ, 2020).

A Tabela 1 apresenta o valor de fechamento anual do Bitcoin nos últimos 5 anos, considerando a cotação da criptomoeda em Dólar Americano no dia 31/12 de cada ano⁴. Dados históricos e em tempo real sobre o volume de transações do Bitcoin e outros ativos financeiros podem ser obtidos em Investing (2021). É perceptível o crescente aumento de valor da criptomoeda ao longo do período observado, ficando mais acentuado no decorrer dos anos de 2017 e 2020. Para efeito de comparação, são também apresentados os valores de fechamento anuais para a cotação do Real (BRL) em Dólar Americano (USD) e Euro (EUR). Além disso, também são listadas as pontuações de fechamento de três importantes índices que medem o desempenho de ações: DAX⁵, DJI e BVSP⁶. Percebe-se, claramente, que para nenhum dos índices ou taxas de conversão avaliados houve variação semelhante à apresentada pelo Bitcoin.

Tabela 1 – Evolução anual de importantes ativos/índices financeiros (2016 a 2020).

Ano	Tipo	Moedas			Índices de Ações		
		BTC	USD	EUR	DAX	DJI	BVSP
2020	Valor	28.933,00	5,19	6,69	13.718,78	30.606,48	119.017,24
	Var. (B)	21.730,45	1,17	2,18	469,77	2.068,04	3.371,90
	Var. (%)	+301,70	+29,21	+48,35	+3,55	+7,25	+2,92
2019	Valor	7.202,55	4,02	4,51	13.249,01	58.538,44	115.645,34
	Var. (B)	3.353,33	0,14	0,05	2.690,05	5.210,98	27.758,08
	Var. (%)	+87,12	+3,56	+1,23	+25,48	+22,34	+31,58
2018	Valor	3.849,22	3,88	4,45	10.558,96	23.327,46	87.887,26
	Var. (B)	9.775,78	0,57	0,48	2.358,68	1.391,76	11.485,18
	Var. (%)	-71,75	+17,17	+12,02	-18,26	+5,63	+15,03
2017	Valor	13.625,00	3,31	3,97	12.917,64	24.719,22	76.402,08
	Var. (B)	12.626,38	0,06	0,55	1.436,58	4.956,62	16.174,79
	Var. (%)	+1.264,38	+1,78	+16,12	+12,51	+25,08	+26,86
2016	Valor	998,62	3,25	3,42	11.481,06	19.762,60	60.227,29
	Var. (B)	564,00	-0,71	-0,88	738,05	2.337,57	16.877,33
	Var. (%)	+129,77	-17,83	-20,44	+6,87	+13,42	+38,93

Fonte: dados extraídos e adaptados de Investing (2021).

⁴ A coluna Valor apresenta o valor de fechamento anual da criptomoeda e as colunas Var. (B) e Var. (%) apresentam, respectivamente, o cálculo da variação bruta e percentual do valor da criptomoeda entre o ano imediatamente anterior e o ano atual.

⁵ DAX (*Deutscher Aktienindex* ou Índice de Ações Alemão) é o índice de ações composto pelas 30 empresas de maior capitalização e liquidez listadas na Bolsa de Frankfurt (Alemanha).

⁶ BVSP (Bolsa de Valores de São Paulo) é o índice mais importante do Brasil para medir o desempenho de ações, composto por empresas de diversos ramos como: varejo, *commodities* e bens de consumo.

O ano de 2020 requer uma análise mais profunda dos dados apresentados por ser mais recente e também por trazer a maior alta do valor do Bitcoin considerando valores absolutos. Além disso, no dia 11 de Março de 2020 a Organização Mundial de Saúde (OMS) passou a reconhecer oficialmente como pandêmico o surto de Covid-19 iniciado na China no final do ano anterior. Neste momento a doença havia alcançado 114 países com 118 mil casos registrados e mais de 4 mil mortos (WHO, 2021). Para que se tenha uma dimensão do problema, o último evento sanitário adverso de proporções semelhantes havia acontecido a mais de 100 anos (PARK et al., 2021). No dia do anúncio feito pela OMS o valor unitário do Bitcoin caiu de US\$ 7.900 para US\$ 4.900, uma queda de aproximadamente 37%. Até o final de 2020, essa era considerada a maior desvalorização diária registrada na história para a criptomoeda em valores absolutos. No entanto, mesmo com a continuidade da pandemia, que atingiu a marca de 1,7 milhão de mortos em todo o mundo no final de 2020, o Bitcoin, que em Janeiro era cotado a US\$ 7.200, registrou uma alta anual de aproximadamente 300%, fechando o ano com cotação superior a US\$ 28.000.

A rápida recuperação do Bitcoin durante a pandemia frente a outros ativos ou índices tradicionais é perceptível. O índice de ações DJI registrou queda superior a 13,74% no mês de Março de 2020. Outros índices importantes relacionados ao mercado de ações como DAX e BVSP também registraram quedas acentuadas no mesmo período, apresentando 16,44% e 29,90% respectivamente. Ao final de 2020, mesmo com o avanço dos estudos sobre a doença, liberação emergencial de vacinas e o estabelecimento de protocolos de segurança, nenhum dos índices citados teve valorização expressiva se comparado ao Bitcoin. Alguns analistas apontam que a pandemia encorajou os investidores a reavaliar as perspectivas de longo prazo para o Bitcoin e outras criptomoedas, passando a enxergá-las como uma possibilidade de diversificação diante das incertezas do mercado de ações (CONLON; MCGEE, 2020).

Fazer previsões no contexto de criptomoedas é uma tarefa extremamente complexa, justamente pela alta volatilidade a qual esse tipo de ativo está suscetível. A Tabela 2 apresenta um detalhamento quantitativo diário acerca do aumento (+) e diminuição (-) do valor do Bitcoin entre 2016 e 2020. Ao analisar os dados apresentados na Tabela 2, é possível concluir que o valor diário do Bitcoin mais aumentou do que diminuiu em todos os anos retratados e que 2017 e 2020 foram os anos em que houve maior ocorrência de aumentos. Além disso, na mesma tabela, é possível visualizar como a moeda se comportou em relação ao percentual diário de variação. Percebe-se, por exemplo, que o percentual de variação diário do Bitcoin esteve mais entre os intervalos [0-1[e [1-2[do que em todos os outros intervalos somados. Percebe-se ainda que a criptomoeda apresentou alta volatilidade nos anos de 2017 e 2018, visto que nesse período foram registradas 95 (26,03%) e 72 (19,73%) variações maiores ou iguais a 5%, respectivamente. Porém, percebe-se que variações mais bruscas ocorreram menos vezes nos últimos dois anos, tendo isso sido registrado nesse período 40 (10,95%) e 38 (10,38%) vezes, respectivamente.

Tabela 2 – Quantitativo da volatilidade diária do preço do Bitcoin, se positiva ou negativa, em relação a intervalos de variação pré-estabelecidos (2016 a 2020).

Ano	Tipo	[0-1[[1-2[[2-3[[3-4[[4-5[[5-10[[10-∞[Total
2020	+	68	55	26	20	16	17	6	208
	-	70	35	18	10	10	13	2	158
	Total	138	90	44	30	26	30	8	366
2019	+	76	40	32	12	8	17	8	193
	-	71	37	20	17	12	12	3	172
	Total	147	77	52	29	20	29	11	365
2018	+	6	45	20	17	13	21	6	188
	-	51	29	23	19	10	35	10	177
	Total	117	74	3	36	23	56	16	365
2017	+	51	47	26	25	22	41	12	224
	-	40	20	23	10	6	34	8	141
	Total	91	67	49	35	28	75	20	365
2016	+	116	44	10	13	10	7	4	204
	-	103	24	20	4	2	7	2	162
	Total	219	68	30	17	12	14	6	366

Fonte: dados extraídos e adaptados de [CoinMarketCap \(2021\)](#).

A previsão do valor e do sentido de variação do preço de um ativo financeiro são importantes instrumentos para a tomada de decisão. A previsão do preço é útil, principalmente, no que se refere à gestão de carteiras e gestão de risco. Por sua vez, a previsão do sentido de variação pode ser utilizada em estratégias de *trading*, baseadas em uma análise mais técnica. Para a realização destas previsões diversas técnicas têm sido utilizadas por agentes do mercado financeiro. Dentre elas, destaca-se, principalmente, *Autoregressive Integrated Moving Average* (ARIMA), um método tradicional de Inferência Estatística, muito aplicado na previsão de valores em séries temporais do mercado financeiro. Além disso, alguns modelos de Aprendizagem de Máquina também têm sido aplicados para a resolução destes problemas, tais como: *Artificial Neural Networks* (ANN), *Support Vector Machines* (SVM) e *Ensemble Learning* (EL). As classes de modelos mencionadas têm apresentado resultados satisfatórios na realização das previsões e, por isto, têm sido utilizadas de forma prática por investidores e analistas do mercado financeiro. Entretanto, algumas das limitações apresentadas por estes modelos podem ser mencionadas, como, por exemplo, a necessidade da realização de um ajuste periódico de parâmetros e de estrutura. Além disso, essas técnicas não incorporam nativamente a habilidade de trabalhar com cenários de incertezas, o que, conforme mencionado, é uma característica iminente no contexto de criptomoedas.

Os Sistemas Evolutivos, um campo de pesquisa recente em relação aos modelos citados, foram propostos para atender a demanda existente no processamento *on-line* de dados, no qual as informações geralmente são disponibilizadas na forma de um fluxo contínuo, com características que envolvem a disponibilização em grandes quantidades, altamente dinâmicas e não estacionárias (LEITE; SKRJANC; GOMIDE, 2020). A principal característica destes modelos é que eles não precisam supor uma dinâmica inicial dos dados, além de não haver a necessidade de realizar, de forma recorrente, um ajuste de

parâmetros e estrutura. Por isto, essa classe de modelos é considerada adequada para a modelagem de sistemas dinâmicos e não-lineares. Neste caso, os parâmetros necessários para inicialização dos modelos, quando é o caso, são definidos uma única vez, e a sua estrutura evolui à medida que o sistema aprende com as novas informações identificadas no fluxo de dados. Adicionalmente, os Sistemas *Fuzzy* Evolutivos incorporam, de forma nativa, o tratamento de incertezas, visto que possuem a capacidade de tratar o conhecimento de maneira mais flexível, o que é viabilizado pelo uso da lógica *Fuzzy* em detrimento da lógica Booleana tradicional. Entende-se, portanto, como valiosa e promissora, a realização de uma avaliação do comportamento de Sistemas *Fuzzy* Evolutivos, quando aplicados aos problemas de previsão de valores e do sentido de variação do preço de criptomoedas, visto que as características inerentes à esta classe de modelos a colocam como candidata à resolução deste problema de forma assertiva e eficiente.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo geral deste trabalho é a proposição de uma nova abordagem para a realização de previsões no contexto de criptomoedas utilizando Sistemas *Fuzzy* Evolutivos.

1.3.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:

- elaborar uma revisão de literatura sobre previsão em séries financeiras relacionadas à criptomoedas, contemplando os diferentes métodos aplicados até então a este problema;
- identificar modelos evolutivos que apresentem um bom desempenho na realização de previsões de valores relacionados ao mercado de criptomoedas;
- propor novas abordagens utilizando Sistemas *Fuzzy* Evolutivos nos problemas de previsão do valor, previsão do sentido de variação e recomendação de investimentos em criptomoedas;
- avaliar o desempenho das abordagens propostas e, se possível, comparar com resultados da literatura;

1.4 Principais Contribuições

De modo geral, as principais contribuições dessa pesquisa para o meio acadêmico bem como para o mercado financeiro, são:

- a realização de um estudo profundo acerca da previsão de valores em criptomoedas, abordando pelo menos 30 trabalhos publicados sobre este assunto. Neste aspecto, algumas perguntas importantes sobre as pesquisas concluídas ou em andamento foram respondidas, tais como:
 - Quais as principais técnicas ou modelos têm sido utilizados nas previsões?
 - Quais atributos têm sido previstos?
 - Quais horizontes de previsão têm sido considerados?
 - Como os dados históricos das criptomoedas têm sido utilizados nos modelos de previsão?
 - Algum indicador externo ao contexto de criptomoedas também tem sido utilizado?
 - Como poderia ser realizada uma aplicação prática das previsões, que pudesse auxiliar o investidor de forma efetiva na tomada de decisões acerca de seus investimentos?
- a proposição de uma nova abordagem utilizando o método de seleção de características Kruskal-Wallis e Sistemas *Fuzzy* Evolutivos para a previsão do valor, previsão do sentido de variação e recomendação de investimentos em criptomoedas;
- a proposição e a implementação de uma nova abordagem que combina modelos evolutivos (Comitê) na previsão do sentido de variação do preço de criptomoedas e recomendação de investimentos.

1.5 Publicações

Nesta seção são apresentadas as publicações produzidas no decorrer desta pesquisa:

1. Amaral, V. L.; Silva, A. M.: **Sistemas *Fuzzy* Evolutivos por *Stacked Generalization* na Previsão do Sentido da Variação do Preço de Criptomoeda - Bitcoin**. In: VI Congresso Brasileiro de Sistemas *Fuzzy* (CBSF 2021). (Publicado). ([AMARAL; SILVA, 2021](#))
2. Amaral, V. L.; Silva, A. M.: ***Evolving Fuzzy Systems in Forecasting and Investments Recommendation in Cryptocurrencies***. (Trabalho em desenvolvimento, a ser submetido em periódico).

O primeiro artigo consiste na aplicação de Sistemas *Fuzzy* Evolutivos no problema de previsão do sentido de variação do preço do Bitcoin. Em contrapartida, o segundo trabalho, ainda em fase de desenvolvimento, se refere aos problemas de previsão do valor, previsão do sentido de variação e recomendação de investimentos em criptomoedas, também utilizando uma abordagem baseada em Sistemas *Fuzzy* Evolutivos.

1.6 Organização do Trabalho

Para melhor compreensão, o trabalho foi dividido em 6 capítulos. O presente capítulo introduziu o assunto, além de apresentar as motivações e os objetivos geral e específicos dessa pesquisa. A seguir, o Capítulo 2 entra em detalhes acerca de conceitos relacionados a séries temporais e modelos de previsão, além de contextualizar o leitor a respeito dos princípios de funcionamento das criptomoedas. Por sua vez, o Capítulo 3 apresenta uma revisão da literatura, onde são listados trabalhos publicados cujo objetivo é a previsão de valores no contexto de criptomoedas e a utilização dessa previsão na tomada de decisão sobre investimentos. O Capítulo 4 discorre acerca da abordagem proposta neste trabalho para previsão de valores neste mesmo contexto, além de propor a utilização das previsões para recomendação de investimentos. O Capítulo 5 apresenta os resultados obtidos nos experimentos realizados a partir da materialização da abordagem proposta no capítulo anterior. Por fim, o Capítulo 6 traz as considerações finais e os próximos passos a serem seguidos para a continuidade da pesquisa.

2 Fundamentação Teórica

Este capítulo tem por objetivo apresentar e descrever os principais conceitos abordados neste trabalho. A Seção 2.1 define o termo séries temporais a partir da demonstração das suas principais propriedades. Em seguida, na Seção 2.2, é apresentado o modelo ARIMA (*Autoregressive Integrated Moving Average*), considerado o estado da arte para o problema de previsão em séries temporais do mercado financeiro. Além disso, define-se também alguns conceitos de Aprendizagem de Máquina, campo de estudo onde se enquadram os Sistemas *Fuzzy* Evolutivos, um dos principais elementos deste trabalho. Na Seção 2.3 são apresentadas medidas de desempenho para a avaliação de modelos de séries temporais, enfatizando, principalmente, as métricas mais utilizadas para a avaliação dos modelos aplicados ao contexto de criptomoedas. Por último, a Seção 2.4 apresenta tecnicamente os temas Criptomoedas e *Blockchain*, entrando em detalhes de implementação que diferenciam algumas das criptomoedas de maior aceitação no mercado.

2.1 Séries Temporais

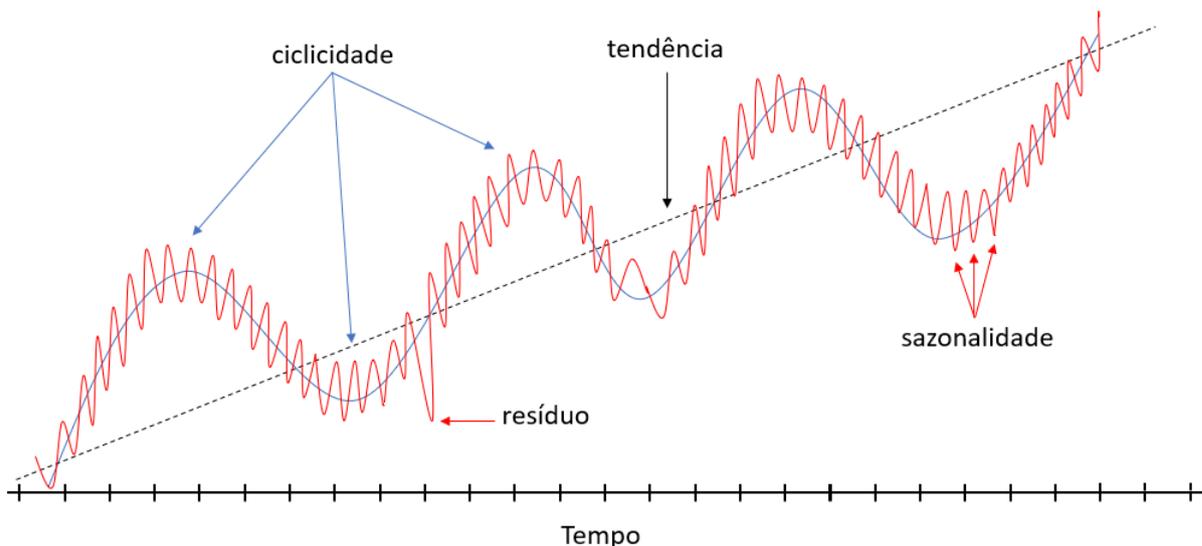
Uma série temporal consiste em um conjunto de observações geradas sequencialmente ao longo do tempo (BOX; JENKINS, 1976). O PIB (Produto Interno Bruto) brasileiro entre os anos de 1970 e 2020 é um exemplo de série temporal e pode ser descrito como uma variável aleatória y_t onde $\{y_t\}_{1970}^{2020} = \{y_{1970}, y_{1971}, \dots, y_{2020}\}$. Para que determinada série seja classificada como uma série temporal, as observações devem apresentar uma dependência serial entre si, ou seja, os dados de uma variável aleatória y , no instante t , com t variando até n , precisam, de certa maneira, conter informações que auxiliem na determinação do valor dessa variável no instante $t + 1$ (MORETTIN; TOLOI, 2006). Um dos principais objetivos da análise de séries temporais é a busca por um modelo que defina a série para previsão de valores futuros baseado em valores passados.

Existem basicamente duas formas de se modelar uma série temporal: a partir da análise no domínio do tempo ou da frequência (TSAY, 2010). A análise do domínio do tempo tem como objetivo a determinação da magnitude de cada evento nos diversos instantes da série. Na análise do domínio da frequência, o interesse está em verificar a frequência que alguns eventos ocorrem em determinado período de tempo. As séries temporais podem também apresentar comportamentos cíclicos ou variações periódicas. É conveniente que se faça a medição da mesma a partir dos dados disponíveis e inclua essa sazonalidade dentro do modelo de previsão (TSAY, 2010).

Neste sentido, segundo o modelo clássico, as séries temporais geralmente são compostas por quatro padrões: tendência, ciclicidade, sazonalidade e resíduo (HYNDMAN;

ATHANASOPOULOS, 2018). A tendência é o componente mais relevante da série, demonstrando o seu comportamento a longo prazo, causado pelo crescimento demográfico, mudança gradual nos hábitos de consumo ou qualquer outro aspecto que afete a variável de interesse a longo prazo. A ciclicidade refere-se a flutuações suaves e repetidas ao longo da componente de tendência, e que se repetem com certa periodicidade, podendo ser consequência de variações da economia como períodos de crescimento ou recessão, fenômenos climáticos, dentre outros. A sazonalidade diz respeito às flutuações em um menor intervalo de tempo (exemplo: mês, semana e dia), geralmente em função das estações do ano, feriados ou festas populares. Por último, o resíduo são as flutuações inexplicáveis, resultado de fatos fortuitos e inesperados como catástrofes naturais, atentados terroristas, decisões imprevistas de governos, etc.. A Figura 1 exemplifica uma série temporal fictícia, demonstrando cada um dos componentes mencionados.

Figura 1 – Padrões encontrados em séries temporais.



Cada série temporal é única, mas pode, em geral, ser classificada em relação a algumas de suas características: contínuas ou discretas; determinísticas ou estocásticas; univariadas ou multivariadas (BOX; JENKINS, 1976). Uma série temporal é dita contínua quando as observações que compõem a série são feitas continuamente ao longo do tempo; ou discreta, se as observações são feitas em tempos específicos, geralmente equiespaçados (horário, diário, semanal, mensal, trimestral, anual, etc.). Se os valores futuros de uma série temporal podem ser descobertos de forma exata por alguma função matemática, então esta série é denominada determinística; mas, se os valores futuros podem ser descritos apenas em termos de uma distribuição de probabilidade, então a série é chamada não-determinística ou, simplesmente, série temporal estocástica. Uma série temporal univariada consiste de observações únicas (escalares) registradas sequencialmente durante incrementos iguais de tempo; em contrapartida, em séries multivariadas as observações consistem em um conjunto de valores.

Outra característica relevante em séries temporais está relacionada ao conceito de estacionariedade. Uma série temporal é dita estacionária quando suas propriedades estatísticas, como a média e a variância, são constantes ao longo do tempo. Em contrapartida, tem-se uma série não-estacionária quando suas propriedades estatísticas variam ao longo do tempo. Quando uma série temporal é dita não-estacionária, pode-se, geralmente, transformá-la em uma série estacionária utilizando alguns métodos matemáticos (HYNDMAN; ATHANASOPOULOS, 2018). Dentre as técnicas mais utilizados neste sentido, destaca-se a Diferenciação, isto é, dada uma série y_t , cria-se uma nova série a partir da expressão $z_t = y_t - y_{t-1}$. Neste caso, a nova série conterá um ponto a menos que a série original. Embora possa-se diferenciar os dados mais que uma vez, uma única diferenciação geralmente é o suficiente para a obtenção de uma série estacionária. Embora a sazonalidade também viole a estacionariedade, isto é usualmente incorporado de forma explícita no modelo de série temporal.

2.2 Modelagem de Séries Temporais

Independentemente da classificação que se faça para uma série temporal, um número vasto de modelos pode ser considerado para a descrição de uma série particular. A construção destes modelos depende de vários fatores, tais como o comportamento do fenômeno ou o conhecimento que se tem antecipadamente acerca da natureza e do objeto da análise. Morettin e Tolo (2006) salientam que existem vários motivos para o estudo e modelagem de séries temporais: (i) investigar o mecanismo gerador; (ii) fazer previsões de valores; (iii) descrever o comportamento; e (iv) procurar periodicidades relevantes nos dados. Conforme mencionado anteriormente, o objetivo deste trabalho se concentra na aplicação de técnicas de Aprendizagem de Máquina, mais especificamente Sistemas *Fuzzy* Evolutivos, para a previsão de valores e tendências em séries temporais financeiras, associadas ao mercado de criptomoedas.

De modo geral, análise de regressão é o termo utilizado para qualificar modelos que tentam explicar os movimentos registrados em uma variável ao longo do tempo por meio da identificação da relação com movimentos registrados por outras variáveis (BROOKS, 2014). De forma mais concreta, um modelo de regressão é utilizado na tentativa de explicar os movimentos de uma variável y a partir do movimento registrado para um conjunto de variáveis x_1, x_2, \dots, x_N , formado por observações coletadas de outras séries temporais com as quais acredita-se na existência de alguma relação com a série estudada ou valores passados da própria série. Após a definição e avaliação do modelo de acordo com os critérios estabelecidos, é esperado que se possa utilizá-lo para a previsão de valores da série em questão.

Os modelos de regressão são utilizados para a previsão de valores contínuos em

séries temporais. Como exemplo, pode ser citado um modelo que realiza a previsão do valor de uma criptomoeda para o dia seguinte baseado no preço registrado para a própria criptomoeda nos 3 dias anteriores (univariado) ou baseado no valor atual de 3 diferentes criptomoedas (multivariado). Em contrapartida, outros modelos fornecem como saída a previsão se uma observação futura da série estudada se encaixará em um conjunto finito de classes, previamente conhecido. Neste caso, pode ser citado como exemplo um modelo que, baseado em um conjunto de variáveis, apresenta como saída uma previsão se o valor de uma criptomoeda no dia seguinte irá subir, permanecer o mesmo ou diminuir.

Essa seção tem por objetivo inicial descrever o modelo ARIMA (*Autoregressive Integrated Moving Average*), considerado o estado da arte para a realização de previsões em séries temporais financeiras. Em seguida, discorre-se sobre Aprendizagem de Máquina, onde detalha-se os modelos baseados em Aprendizagem por Agrupamento (*Ensemble Learning* ou EL), visto que alguns conceitos dessa classe de modelos também ajudam a compor a abordagem proposta neste trabalho para previsão e recomendação de investimentos em criptomoedas. Por último, descreve-se os Sistemas *Fuzzy* Evolutivos, tópico fundamental sobre o qual a abordagem proposta neste trabalho está fundamentada.

2.2.1 Modelos Autorregressivos, Integrados e de Médias Móveis

Conforme Frisiani (2019), geralmente pode-se pensar qualquer série temporal y_t como a soma de duas partes: o que pode e o que não pode ser previsto usando o conhecimento do passado. A Equação (1) é uma decomposição de uma série temporal sob essa perspectiva onde $E[.]$ representa o operador do valor esperado condicional, y_t é a variável de interesse na previsão, γ_{t-1} é o conjunto de informações em $t - 1$ que contém toda a informação disponível para se fazer previsões de valores futuros da série e v_t é a parte chamada imprevisível.

$$y_t = E[y_t | \gamma_{t-1}] + v_t. \quad (1)$$

Os modelos lineares são considerados os mais simples na previsão de valores em séries temporais. Esse tipo de modelo busca encontrar uma aproximação linear entre y_t e γ_{t-1} . Ao se concluir que o conjunto γ_{t-1} não contém qualquer informação que possa ser utilizada para fazer previsões em um modelo linear para y_t , a série temporal correspondente pode ser chamada de *white noise* (ruído branco). Dessa forma, denota-se essa série apenas como ϵ_t . Para isto, é necessário que esta série possua média constante e igual a zero e variância constante (MORETTIN; TOLOI, 2006).

Dentre os modelos lineares destacam-se, principalmente, o $AR(p)$ (*Autoregressive*), $MA(q)$ (*Moving Average*), $ARMA(p, d)$ (*Auto Regressive Moving Average*) e $ARIMA(p, d, q)$ (*Auto Regressive Integrated Moving Average*). No modelo $AR(p)$ ou modelo Autorregressivo

de ordem p , a variável de interesse y_t é prevista a partir de uma combinação linear de seus valores passados $\{y_{t-1}, y_{t-2}, \dots, y_{t-n}\}$. Neste caso, o termo autorregressivo indica que se trata de uma regressão da variável contra si mesma. A Equação (2) é a representação matemática do modelo $AR(p)$, onde y_t é o valor previsto para o instante t , p é a ordem do modelo, κ_i são os parâmetros ou coeficientes e v_t é o chamado ruído branco, associado ao instante t . De forma simplificada, a ideia por trás deste modelo é a determinação de uma equação linear capaz de descrever a série temporal a partir da atribuição de um coeficiente para os p valores anteriores.

$$y_t = \sum_{i=1}^p (\kappa_i y_{t-i}) + v_t + c. \quad (2)$$

O Modelo de Médias Móveis (ou *Moving Average*) de ordem q , dado por $MA(q)$, também é amplamente utilizado na previsão de séries temporais. Este modelo tenta prever o próximo valor de uma série temporal a partir do acúmulo de ruído branco, multiplicado por coeficientes previamente definidos sobre o valor médio da série. Isto faz com que o modelo tenha uma reação mais lenta a mudanças bruscas, mas também o torna mais resistente em cenários de alta volatilidade. A Equação (3) apresenta matematicamente o modelo, onde c é uma constante, $\theta_1, \theta_2, \dots, \theta_q$, são os coeficientes mencionados (a serem estimados), e v_t é o ruído branco.

$$y_t = \sum_{i=1}^q (\theta_i \epsilon_{t-i}) + v_t + c. \quad (3)$$

A combinação dos modelos $AR(p)$ e $MA(q)$ resulta no ARMA (*Autoregressive Moving Average*) (BOX; JENKINS, 1976). Denota-se $ARMA(p, q)$ um processo autorregressivo de médias móveis de ordem p e q , conforme explicitado na Equação (4). A composição dos modelos $AR(p)$ e $MA(q)$ resulta em um modelo mais flexível, haja vista que os parâmetros p e q podem assumir valores diferentes. Dessa maneira, o número de termos na parte autorregressiva do modelo é diferente do número de ruídos considerados em se tratando de médias móveis. Trata-se, portanto, de uma representação da série temporal como uma ponderação dos próprios valores e/ou erros passados da série.

$$y_t = \sum_{i=1}^p (\kappa_i y_{t-i}) + \sum_{i=1}^q (\theta_i \epsilon_{t-i}) + v_t + c. \quad (4)$$

Os modelos apresentados têm aplicação específica no contexto de séries temporais estacionárias. Porém, na modelagem de séries não-estacionárias, faz-se necessária a utilização de modelos mais complexos (TSAY, 2010). Neste sentido, denota-se por

$ARIMA(p, d, q)$ o modelo Autorregressivo Integrado e de Médias Móveis (ou *Autoregressive Integrated Moving Average*) (BOX; JENKINS, 1976). O $ARIMA(p, d, q)$ é um método amplamente utilizado na previsão de séries temporais não-estacionárias, que combina o modelo $ARMA(p, q)$ com o processo de diferenciação mencionado anteriormente, aplicado na estacionarização de uma série temporal. De uma forma geral, os modelos ARIMA postulam que as séries temporais podem ser representadas por uma sequência de choques aleatórios submetidos a três operações de filtragem (auto regressão, diferenciação e médias móveis). Os parâmetros p e q tem o mesmo significado definido para o método $ARMA(p, q)$ e o parâmetro d representa o número de vezes que se fará diferenciação de valores com o objetivo de tornar a série estacionária.

Na literatura é possível encontrar diversas variações dos modelos ARMA e ARIMA, por exemplo: ARFIMA (*Autoregressive Fractionally Integrated Moving Average*); SARIMA (*Seasonal ARIMA*); ARMAX, ARIMAX, ARFIMAX e SARIMAX (*ARMA / ARIMA / ARFIMA / SARIMA with exogenous variables*). Sabe-se que os modelos ARIMA originaram-se no trabalho de Box e Jenkins (1976) mas, no melhor do nosso conhecimento, não foi possível encontrar a origem de cada uma dessas variações. Entretanto, um detalhamento acerca desses modelos bem como alguns exemplos de aplicações podem ser encontrados em Cryer e Chan (2008) e Vagropoulos et al. (2016).

2.2.2 Técnicas de Aprendizagem de Máquina

Atualmente, o termo Aprendizagem de Máquina (*Machine Learning*, ou simplesmente ML) é utilizado para descrever o ramo da ciência focado no estudo, projeto e desenvolvimento de algoritmos que viabilizam o processo de aprendizagem por máquinas baseado em dados. A inteligência dos algoritmos de ML não é explicitamente programada, mas sim a forma como esse aprendizado se dará a medida que novos dados são apresentados ao sistema (PRAVEENA, 2017). O maior foco deste campo de pesquisa está em prover sistemas inteligentes para a resolução de problemas como o reconhecimento automático de padrões, tomada de decisões e previsão de valores em séries temporais, sendo amplamente associada aos campos da estatística, mineração de dados e inteligência artificial. Autores como Stinis (2019) categorizam os métodos de aprendizagem em três classes: supervisionada, não-supervisionada e por reforço.

O termo Aprendizagem Supervisionada (*Supervised Learning*) define os modelos de ML que utilizam variáveis de entrada previamente rotuladas com as saídas esperadas no processo de aprendizagem. Neste caso são fornecidas referências para o modelo daquilo que está certo ou errado. Um sistema capaz de prever a idade de uma pessoa a partir de uma fotografia é um exemplo passível de aplicação de modelos que utilizam aprendizagem supervisionada. Para isto, pode-se fornecer imagens de diversas pessoas rotuladas com a respectiva idade. Após o ajuste do modelo, ao apresentar ao sistema uma imagem não

utilizada no treinamento, espera-se que ele seja capaz de reconhecer a idade dessa pessoa.

Na Aprendizagem Não-Supervisionada (*Unsupervised Learning*) não existem resultados pré-definidos para o modelo utilizar como referência no processo de aprendizagem. Tomando o mesmo exemplo citado anteriormente, uma aplicação possível para esse tipo de aprendizagem seria fornecer ao sistema o mesmo conjunto de fotografias de pessoas, porém sem o respectivo rótulo. Espera-se que o sistema aprenda, sem supervisão, a fazer um agrupamento dessas imagens de acordo com alguns critérios, como, por exemplo, a idade dessas pessoas. Como não há uma referência (ou critério específico) para o modelo seguir, define-se que esse aprendizado é não-supervisionado.

Na Aprendizagem por Reforço (*Reinforcement Learning*) a proposta é que o sistema aprenda a atingir uma meta em um ambiente incerto e potencialmente complexo. Neste caso o sistema de inteligência artificial é colocado em uma situação e utiliza um método de tentativa e erro para encontrar uma solução para o problema proposto. Para que a máquina cumpra o objetivo desejado, o agente recebe um tipo de recompensa ou penalidade pelas ações que executa. O objetivo final é maximizar a recompensa total.

Ao longo dos últimos anos diversos modelos de ML foram desenvolvidos para problemas de Aprendizagem Supervisionada no contexto de classificação e regressão, onde está concentrado o foco desta pesquisa. Em um problema de regressão, tem-se por objetivo realizar a previsão de um valor futuro ou desconhecido em uma série, fornecendo como saída um valor contínuo. Em se tratando de problemas de classificação, é esperado que o sistema classifique a entrada em um conjunto finito de categorias.

2.2.2.1 Aprendizagem por Agrupamento

Aprendizagem por Agrupamento (*Ensemble Learning*, ou EL) é uma importante classe de algoritmos de *Machine Learning* que se baseia na ideia de combinar diversos modelos considerados mais simples (*weak learners*), treiná-los para uma mesma tarefa e, em sequência, produzir um modelo agrupado mais complexo (*strong learner*), que consiste basicamente na soma de suas partes (WANG et al., 2019). Dentre as diferentes formas de se combinar modelos na geração de um modelo mais efetivo, as três principais são: *Bagging*, *Boosting* e *Stacking*.

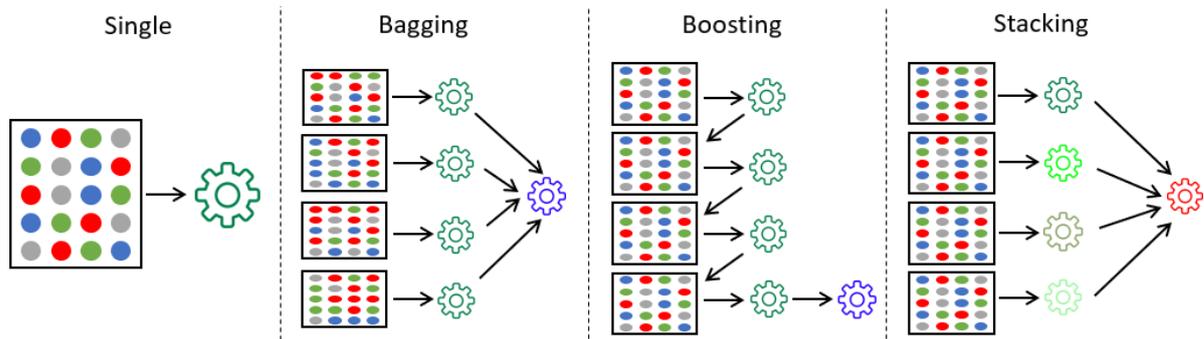
Bagging (BREIMAN, 1996) é uma combinação dos termos *Bootstrapping* e *Aggregating*. *Bootstrapping* é uma técnica estatística que consiste na criação de novos conjuntos de dados a partir da seleção aleatória de amostras do conjunto original, podendo haver repetições nos subconjuntos criados. Esses subconjuntos podem ser visualizados como uma representação independente da distribuição real dos dados. Em seguida, cada subconjunto criado no processo de *Bootstrapping* é então apresentado a uma instância de um modelo (*weak learner*), executados de forma paralela e independente. A resposta final consiste na média (*aggregating*) das respostas de cada *weak learner*. Modelos baseados em *Bagging*

Ensemble Learning geralmente são aplicados quando se faz necessária a redução da variância. Isso se deve ao fato de que uma variância alta, pode indicar que o desempenho do modelo depende dos dados apresentados, ou seja, sua capacidade de generalização é baixa. As *Random Forests*, introduzidas em Ho (1995), são um exemplo de modelo que segue essa abordagem.

Modelos baseados em *Boosting*, uma abordagem introduzida em Schapire (1990), partem do mesmo princípio mencionado para os modelos baseados em *Bagging*: os resultados fornecidos por cada *weak learner* são agregados para a obtenção de uma resposta final. Entretanto, *Boosting* é uma técnica que consiste no treinamento de múltiplos *weak learners* de forma sequencial e adaptativa (SCHAPIRE, 2003). Neste caso, cada modelo é treinado, porém, a cada iteração, é dada mais atenção às amostras para as quais foi computado maior erro nos modelos executados anteriormente, ou seja, cada novo modelo se concentra nas observações consideradas mais complexas. Ao final do treinamento, espera-se que seja obtido um modelo com baixo erro e com alto poder de generalização. *Adaboost* (FREUND; SCHAPIRE, 1999) e *Extreme Gradient Boosting* (CHEN; GUESTRIN, 2016) são exemplos de modelos construídos sob essa abordagem.

Finalmente, *Stacking* ou *Stacked Generalization* é também um tipo de aprendizagem por agrupamento, inicialmente apresentada em Wolpert (1992), onde um novo modelo é treinado a partir da combinação das previsões retornadas por dois ou mais modelos executados anteriormente. Podem existir quantas camadas forem necessárias, compostas por modelos de diferentes tipos. Conforme Fauzi e Yuniarti (2018), existem duas formas principais de realizar a agregação dos resultados apresentados pelos *weak learners*, denominadas *hard* e *soft voting*. Na votação por *hard voting*, a saída que mais se repetiu como resultado retornado pelos *weak learners* será consolidada como a resposta final do modelo. Em contrapartida, na votação por *soft voting*, pode-se atribuir um peso ao voto de cada *weak learner*, isto é, a classe prevista para cada modelo é multiplicada pelo peso atribuído ao classificador que a gerou. A classe final é derivada da classe com a maior probabilidade computada.

A Figura 2 apresenta quatro diferentes abordagens para a construção de um sistema de previsão ou classificação. Na abordagem *single*, um único modelo é utilizado para solucionar o problema. Em seguida, são apresentadas ilustrações dos modelos *Bagging* e *Boosting Ensemble Learning*, com quatro *weak learners* homogêneos (em verde) e um modelo de votação (em azul). É importante mencionar que o conjunto de amostras apresentado a cada *weak learner* na abordagem *Bagging* é diferente, mas todos são baseados no conjunto de amostras inicial. Finalmente, apresenta-se uma ilustração do modelo *Stacking Ensemble Learning*, com quatro *weak learners* heterogêneos (tons de verde) e um *strong learner* (em vermelho).

Figura 2 – Aprendizagem por agrupamento (*Bagging*, *Boosting* e *Stacking*).

2.2.2.2 Sistemas Fuzzy Evolutivos

O iminente aumento na disponibilidade de grandes quantidades de informação tem motivado o desenvolvimento de modelos baseados em algoritmos incrementais capazes de processar fluxos de dados contínuos. Nessas condições, a modelagem baseada em sistemas que possuem uma estrutura fixa pode não ser adequada e a necessidade recorrente de realizar o treinamento dos modelos normalmente não é uma opção viável (BOUCHACHIA; LUGHOFFER; SAYED-MOUCHAWEH, 2014). Conforme Leite, Skrjanc e Gomide (2020), isso acontece por alguns motivos: (i) as amostras são apresentadas ao sistema continuamente; (ii) o sistema pode não ter controle sobre a ordem de chegada dos dados; (iii) fluxos de dados não apresentam limites; e (iv) idealmente uma amostra precisa ser descartada após o processamento para evitar problemas de escalabilidade. Portanto, o desafio de processar grandes fluxos de dados variantes no tempo traz problemas únicos que requerem um esforço considerável no desenvolvimento de novas topologias e algoritmos de aprendizagem (LEMOS, 2011).

Há um campo de pesquisa emergente que aborda este problema usando metodologias que possuem mecanismos para treinar e atualizar continuamente os modelos, de maneira *online*, para que estes se adaptem à dinâmica das mudanças presentes no fluxo de dados. Os Sistemas Evolutivos (*Evolving Systems* ou EF) são caracterizados por sua capacidade de extrair conhecimento a partir de dados e adaptar sua estrutura e parâmetros em tempo real, acompanhando as mudanças do ambiente (SILVA, 2014). O aprendizado incremental permite um processamento rápido e apresenta um baixo consumo de memória, uma vez que as amostras do fluxo de dados são processadas uma única vez e, então, são descartadas (LUGHOFFER, 2013), isto é, o mecanismo de aprendizagem utiliza apenas a amostra de dados atual (TUNG; QUEK; GUAN, 2013). Dessa maneira, estes modelos são aderentes principalmente em situações onde a utilização de métodos tradicionais acarretaria em um elevado custo computacional (KASABOV; FILEV, 2006). Enquanto o aprendizado permite a aquisição contínua de novos conhecimentos modificando a estrutura e os parâmetros do sistema, este deve ser capaz de manter/armazenar o conhecimento relevante aprendido previamente (TUNG; QUEK; GUAN, 2013).

Adicionalmente, os Sistemas *Fuzzy* Evolutivos (*Evolving Fuzzy Systems* ou EFS) podem ser entendidos como uma sinergia entre Sistemas *Fuzzy* tradicionais e métodos recursivos de Aprendizagem de Máquina, e representam uma mistura de modelos *fuzzy* (geralmente lineares) locais simples, cuja combinação é não-linear e adaptativa (LEMOS, 2011). Os conjuntos *fuzzy* se diferenciam do conceito clássico de conjuntos pois existe a possibilidade de que sejam atribuídos valores intermediários (entre 0 e 1) para quantificar a percepção sobre o quão compatível são os objetos de um domínio em relação a uma classe. Neste caso, um número mais próximo de 1 indica maior relação de pertencimento e, conseqüentemente, quanto mais perto de 0 for o valor atribuído, menor o grau de relação. Valores de pertencimento (*membership values*) intermediários indicam que o elemento de um universo pode ser membro de uma classe e de outra ao mesmo tempo, com graus de pertinência distintos (KACPRZYK; PEDRYCZ, 2015). Essa forma de representar o conhecimento inerente aos Sistemas *Fuzzy* Evolutivos é favorável, principalmente, ao tratamento de problemas existentes em um contexto de incertezas.

Apresentados alguns conceitos fundamentais e destacada a relevância dos Sistemas *Fuzzy* Evolutivos, principalmente, no processamento de fluxos de dados contínuos e no tratamento de problemas em ambientes de incertezas, as próximas subseções se dedicam a caracterizar os três EFS utilizados neste trabalho, a saber: eGNN (*evolving Granular Neural Network*) (LEITE; COSTA; GOMIDE, 2010; LEITE; COSTA; GOMIDE, 2013); eMG (*Multivariable Gaussian Evolving Fuzzy Modeling System*) (LEMOS; CAMINHAS; GOMIDE, 2011); e eOGS (*evolving Optimal Granular System*) (LEITE; ŠKRJANC, 2019; LEITE et al., 2020). Estes modelos foram escolhidos por se diferenciarem em alguns aspectos relevantes, tais como: a forma de representar o conhecimento; mecanismo de criação, atualização e exclusão das regras; tipo de saída; quantidade de parâmetros; e etc.. Destaca-se, entretanto, que a abordagem proposta neste trabalho consiste na utilização de Sistemas *Fuzzy* Evolutivos de maneira geral e não de um conjunto de modelos específico.

2.2.2.2.1 eGNN - evolving Granular Neural Network

O *evolving Granular Neural Network* (eGNN), proposto em Leite, Costa e Gomide (2010) (classificação) e Leite, Costa e Gomide (2013) (regressão), utiliza neurônios de agregação *fuzzy* como unidades básicas de processamento e codifica um conjunto de regras em um sistema de inferência *fuzzy*. Para isto, o modelo trabalha na descoberta de um conhecimento granular, com alto nível de abstração, a partir de dados granulares mais precisos.

O eGNN possui quatro camadas em sua estrutura: (1) Entrada, (2) Granular, (3) Agregação e (4) Saída. Na segunda camada, o conhecimento granular de alto nível é traduzido em uma base de regras *fuzzy*, na forma

$$R_i : \text{if } (x^t \text{ is } G_i) \text{ then } \underbrace{\hat{y}_i \text{ is } \gamma_i}_{\text{linguístico}} \text{ AND } \underbrace{\hat{y}_i^t = h_{i0}^t + \sum_{j=1}^m h_{ij}^t x_j^t}_{\text{funcional}}, \quad (5)$$

onde R_i representa a i -ésima regra *fuzzy*; i é o índice das regras; x é o vetor de entrada no instante t , descrito por $[x_1^t \dots x_j^t \dots x_m^t]$; t é o instante atual; j é o índice das variáveis de entrada; G_i são os conjuntos *fuzzy* representados por funções de pertinência Trapezoidais; \hat{y} é o conseqüente da i -ésima regra, com seus componentes linguístico e funcional; e h_{i0}^t e h_{ij}^t são os parâmetros do conseqüente¹. A terceira camada da rede é composta por neurônios baseados em operadores de agregação, os quais podem variar de acordo com o problema. Estes neurônios empregam uma *T-norma* na realização de processamento sináptico e um operador de agregação, C , para unir os resultados individuais do processamento realizado nos neurônios. Portanto, a saída de um neurônio de agregação pode ser expressa por

$$o = C(\tilde{x}_1 w_1, \dots, \tilde{x}_1 w_m), \quad (6)$$

onde w_i são os pesos associados a cada neurônio da Camada 2. Finalmente, a camada de saída agrega os valores obtidos na Camada 3, também utilizando um neurônio de agregação C^f para produzir a saída singular fornecida pela rede, e pode ser definida por

$$\hat{y} = C^f(o^1 \hat{y}^1 \delta^1, \dots, o^c \hat{y}^c \delta^c), \quad (7)$$

onde δ_i são os pesos da camada de saída e \hat{y} é a saída estimada final produzida pela rede.

O aprendizado no eGNN significa fundamentalmente a acomodação de novos dados em modelos granulares existentes (ou não), de forma recursiva. Para isto, novas regras, neurônios e conexões entre neurônios podem ser adicionadas quando necessário. Os parâmetros das funções do conseqüente das regras também são objetos de aprendizado. Isso significa que o eGNN captura novas informações do fluxo de dados e se adapta ao novo cenário, evitando redesenho e retreinamento. A estrutura neural granular pode se tornar mais genérica ou refinada, dependendo das relações entre os grânulos definidos (transparência e interpretabilidade) e dos índices de erro (precisão). O modelo é iniciado sem qualquer conhecimento *a priori*, isto é, sua estrutura evolui em tempo de execução, a medida que o conhecimento vai sendo adquirido.

¹ Independentemente da escolha de neurônios de agregação, parâmetros de rede e da natureza dos dados de entrada-saída, o termo linguístico do conseqüente fornece uma saída granular, enquanto o termo funcional apresenta uma saída singular.

Os principais parâmetros necessários para a execução do modelo são: ρ limiar de granularidade, que denota o tamanho máximo de expansão de um conjunto *fuzzy* (ou grânulo); h_r taxa de adaptação, responsável por definir a periodicidade na qual a estrutura dos grânulos será atualizada e o tempo máximo de inatividade a ser considerado na exclusão; e η taxa de crescimento dos grânulos. Para dados normalizados, ρ recebe valores entre 0 e 1, onde um valor próximo de 0 indica que os grânulos jamais irão crescer e, em contrapartida, um valor próximo de 1 indica que um único grânulo irá cobrir todo o domínio de dados, o que tornará impossível a previsão em séries não-estacionárias. Portanto, recomenda-se o uso de valores intermediários para ρ (LEITE; COSTA; GOMIDE, 2013).

Na literatura é possível encontrar aplicações do eGNN em problemas de diferentes áreas como: engenharia e medicina (LEITE; COSTA; GOMIDE, 2013); previsão de temperatura (LEITE; COSTA; GOMIDE, 2012); reconhecimento de padrões (LEITE; COSTA; GOMIDE, 2009); identificação de sistemas caóticos não-lineares (LEITE et al., 2016); e detecção de falhas em sistemas industriais (LEITE, 2019).

2.2.2.2.2 eMG - Multivariable Gaussian Evolving Fuzzy Modeling System

O *Multivariable Gaussian Evolving Fuzzy Modeling System* (eMG), proposto em Lemos, Caminhas e Gomide (2011), é um EFS funcional do tipo Takagi-Sugeno, que utiliza um algoritmo de agrupamento evolutivo Gaussiano baseado no conceito de aprendizado participativo (YAGER, 1990) para atualizar continuamente sua base de regras. No eMG, cada grupo pode ser modelado por uma função de pertinência Gaussiana multivariável caracterizada por um vetor central e uma matriz de dispersão, que representa o grau de dispersão de cada variável individualmente e a interação entre elas. As regras são definidas conforme

$$R_i : \text{if } (x^t \text{ is } G_i) \text{ then } \hat{y}_i^t = h_{i0}^t + \sum_{j=1}^m h_{ij}^t x_j^t, \quad (8)$$

onde R_i é a i -ésima regra *fuzzy*; i é o índice das regras *fuzzy* e *clusters*; x_t é o vetor de entrada no instante t , descrito por $[x_1^t \dots x_j^t \dots x_m^t]$; t é o instante atual; j é o índice das variáveis de entrada; m é o número total de variáveis de entrada; G_i são os conjuntos *fuzzy* representados pelas funções de pertinência Gaussianas multivariáveis; \hat{y}_i^t é o consequente da i -ésima regra *fuzzy* (saída da regra); e h_{i0}^t e h_{ij}^t são os parâmetros do consequente. A saída final do modelo é calculada utilizando a média ponderada da saída das regras, conforme

$$\hat{y}_t = \sum_{i=1}^{c^t} \tau_i^t y_i^t, \quad (9)$$

onde y_i^t é o consequente da i -ésima regra e τ_i^t são as funções de pertinência normalizadas.

O aprendizado participativo é um modelo de aprendizagem que assume que o processo de atualização do conhecimento adquirido sobre um sistema depende do que já foi aprendido anteriormente. Ou seja, o conhecimento adquirido sobre o sistema faz parte do processo de aprendizagem e influencia na forma como novas observações são utilizadas neste processo. A característica essencial desse modelo de aprendizagem consiste no fato de que o impacto causado por uma nova observação no aprendizado ou na revisão do conhecimento existente depende de sua compatibilidade com o que já foi aprendido (YAGER, 1990). Dessa maneira, o eMG provê um mecanismo automático para decidir se uma nova observação é incompatível com o conhecimento atual (representado por uma estrutura de grupos), isto é, se a amostra deverá ser o primeiro representante de um novo grupo ou se o novo dado pode ser caracterizado como um *outlier* (amostra discrepante), que deve ser descartado ou ter seu efeito suavizado (LEMOS, 2011).

No eMG A estrutura de grupos (número, centro e formato) é atualizada recursivamente a cada iteração e os limiares são definidos automaticamente. Este modelo implementa um mecanismo capaz de ajustar o valor do limiar de distância baseado na dimensão do espaço de entrada, o que melhora o seu desempenho mesmo em conjuntos de dados com alta dimensionalidade. Os parâmetros do consequente são atualizados usando o método recursivo dos mínimos quadrados ponderados. A cada nova amostra apresentada, a estrutura de grupos é atualizada por uma medida de compatibilidade e por um índice de alerta calculados para cada grupo. Um novo grupo é criado quando a medida de compatibilidade é menor que o limiar de compatibilidade para todos os grupos e o índice de alerta é maior que o limiar de alerta para o grupo com maior compatibilidade.

De acordo com o autor, o eMG difere da maior parte dos modelos evolutivos em alguns aspectos: (i) o fato de utilizar funções de pertinência multivariáveis Gaussianas para os conjuntos *fuzzy* ajuda a prevenir perda de informação na integração entre as variáveis; (ii) o algoritmo recursivo de clusterização é robusto mesmo na presença de *outliers*; (iii) a criação de novas regras é governada por um mecanismo que ajusta o limiar considerando a dimensão do espaço de entrada, sendo flexível mesmo para conjuntos maiores; e (iv) não há a necessidade de normalização dos dados de entrada e saída.

Na proposta do algoritmo de agrupamento utilizado no eMG é mencionado que, para sua execução, são necessários quatro parâmetros: a taxa de aprendizado λ , utilizada para atualizar os parâmetros de cada grupo, normalmente ajustada para um valor baixo como, por exemplo, $\lambda \in [10^{-1}, 10^{-5}]$; o tamanho da janela ω , usado pelo mecanismo de alerta para

monitorar a dinâmica da medida de compatibilidade, definindo o padrão anômalo utilizado para classificar cada nova observação como um *outlier* ou o primeiro membro de um novo grupo; o nível de significância α , usado para calcular os limiares de compatibilidade e alerta, cujo valor depende de ω ; e a matriz de dispersão inicial \sum_{init} , utilizada na criação de novos grupos.

Na literatura é possível encontrar aplicações do eMG em diversas áreas, como: sistemas de previsão de demanda de energia elétrica, previsão de valores da série de Mackey-Glass; identificação de sistemas não-lineares e de alta dimensionalidade (LEMOS, 2011); previsão em modelos térmicos de transformadores de potência (SOUZA et al., 2012); e detecção antecipada de falhas em sistemas industriais (INACIO; LEMOS; CAMINHAS, 2015).

2.2.2.2.3 eOGS - evolving Optimal Granular System

O *evolving Optimal Granular System* (eOGS), proposto em Leite e Škrjanc (2019) e Leite et al. (2020), assim como o eGNN, também é baseado no conceito de computação granular e apresenta uma saída granular e funcional para cada regra. Entretanto, neste modelo, as representações são dadas por funções de pertinência Gaussianas e Hiper-retangulares. A granularidade ou o tamanho dos grânulos é definido de acordo com a variabilidade dos dados, conjuntos de níveis, estimativas de erro e uma compensação entre especificidade e exatidão. A adaptação das regras consiste em aumentar ou contrair os limites dos grânulos e, simultaneamente, atualizar os parâmetros do conseqüente para comportar novos padrões identificados no fluxo de dados. O eOGS utiliza o método recursivo dos mínimos quadrados para determinar os coeficientes de suas funções. O tamanho da base de regras é controlado por um mecanismo que realiza o agrupamento de regras sobrepostas, as quais tendem a cobrir informações similares. Ainda neste sentido, regras inativas são excluídas após um número pré-determinado de iterações sem o seu respectivo acionamento.

Os principais parâmetros necessários para a utilização do modelo são: α , que denota o ponto de corte das funções Gaussianas e a dispersão inicial no momento em que os grânulos são criados, sendo o parâmetro que tende a apresentar maior influência nos resultados; ω , que está relacionado à distância mínima entre os grânulos, o que interfere na definição do agrupamento ou separação destes. Valores maiores de ω tendem a deixar o banco de regras mais compacto; ν é o limite para exclusão de regras por tempo de inatividade, para o qual deve ser atribuído um valor mais alto caso seja desejável a permanência de memória mais antiga; e ψ , usado na adaptação do grau de dispersão das funções Gaussianas, pode ser reduzido ou aumentado para definir a velocidade em que isso será realizado ao longo das iterações. Na proposta do modelo, o autor recomenda que sejam respeitados os seguintes limites para cada parâmetro: $0,01 \leq \alpha \leq 1$; $0,01 \leq \omega \leq 0,05$;

$1,5 \leq \psi \leq 2,5$; e $\nu > m$, onde m representa o número de variáveis utilizadas na execução do modelo. Adicionalmente, o autor sugere alguns valores como ponto de partida na sua implementação: $\alpha = 0.1$; $\omega = 0.01$; $\nu = 500$; $\psi = 2$.

Por se tratar de um modelo proposto recentemente, o número efetivo de aplicações em problemas reais ainda é limitado. Entretanto, nas propostas do modelo, conforme Leite e Škrjanc (2019) e Leite et al. (2020), os autores realizam experimentos na previsão de valores da série de Mackey-Glass e também em um problema da área de Engenharia relacionado a avaliação do desempenho de misturas de concreto (LEITE et al., 2020), além da previsão de temperatura, humidade e velocidade média do vento para quatro importantes cidades europeias: Paris, Frankfurt, Reykjavik e Oslo (LEITE; ŠKRJANC, 2019).

2.3 Medidas de Desempenho

Esta seção destaca as principais medidas de desempenho utilizadas na avaliação de técnicas empregadas no contexto de criptomoedas. As seções 2.3.1, 2.3.2 apresentam as métricas utilizadas para avaliação dos modelos quanto à sua capacidade de regressão e classificação, respectivamente. Por último, a Seção 2.3.1 detalha as métricas utilizadas na avaliação dos modelos considerando uma eventual aplicação em uma estratégia de recomendação de investimentos.

2.3.1 Regressão

Na aplicação de técnicas de regressão em séries temporais tem-se como objetivo primário a obtenção de um resultado que apresente a menor distância possível entre o valor previsto e o valor real. Para a realização desta análise, é possível encontrar na literatura diferentes métricas. Nesta subseção serão apresentadas as quatro métricas mais utilizadas nos trabalhos relacionados ao contexto de criptomoedas.

O *Root Mean Square Error (RMSE)* é uma medida quadrática que avalia a magnitude média do erro. Trata-se da raiz quadrada da média dos resíduos quadrados entre o valor previsto e a observação real. É uma medida de precisão dependente de escala e pode ser expressa matematicamente pela Equação (11), em que \hat{y}_t é a saída estimada, y_t é a saída desejada e n é o número total de observações.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2}. \quad (10)$$

O *Non-Dimensional Error Index (NDEI)* também é uma métrica aplicável a modelos de regressão, e representa a razão entre o RMSE e o desvio padrão. O NDEI apresenta uma medida de erro independente da dimensão dos dados, e pode ser obtido por:

$$NDEI = \frac{RMSE}{\sigma(y)}. \quad (11)$$

O *Mean Absolute Error (MAE)* também é uma medida bastante utilizada na avaliação de métodos de regressão pois quantifica a magnitude média dos erros, desconsiderando a direção. Consiste na média da diferença absoluta entre o valor previsto e a observação real. O MAE pode ser obtido por:

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t|. \quad (12)$$

Outra medida utilizada é o MAPE (*Mean Absolute Percentage Error*). O MAPE tem significado semelhante ao MAE, diferenciando-se pelo fato de que retrata o erro em termos percentuais. A Equação (13) ilustra o cálculo do MAPE.

$$MAPE = \frac{100}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{y_t}. \quad (13)$$

2.3.2 Classificação

Quando o principal interesse é qualificar o modelo acerca da sua capacidade em determinar corretamente a qual classe pertence um conjunto de dados de entrada, entende-se como necessária a aplicação de métricas diferentes das citadas para problemas de Regressão. Para defini-las é importante, de antemão, a definição de alguns termos:

- *True Positive (TP)*: quando o modelo sugere que determinada amostra pertence a uma classe de interesse e isto é verdade, ou seja, a amostra realmente pertence a esta classe;
- *False Positive (FP)*: quando o modelo sugere que determinada amostra pertence a uma classe de interesse, porém isto não é verdade;
- *True Negative (TN)*: quando o modelo sugere que determinada amostra não pertence a uma classe de interesse e isto é verdade, ou seja, a amostra não pertence à classe mencionada;

- *False Negative (FN)*: quando o modelo sugere que determinada amostra não pertence a uma classe de interesse, porém isto não é verdade.

Definidos estes conceitos, a Acurácia é uma métrica muito utilizada para avaliar modelos de classificação, e pode ser compreendida como uma medida para previsão de tendência. Dessa maneira, quando um modelo é aplicado a um problema onde deve definir se uma amostra pertence a uma de duas (ou mais) classes possíveis (A e B, por exemplo), a Acurácia apresenta a proporção de previsões (ou classificações) fornecidas corretamente, independente da classe, e é dada por:

$$\text{Acurácia} = \frac{\text{acertos}}{n} \times 100. \quad (14)$$

Mesmo sendo uma métrica amplamente utilizada, a Acurácia, quando analisada de forma isolada, pode levar a uma compreensão inadequada dos resultados. Supondo um cenário onde 90% das amostras para um dado problema são pertencentes à uma classe A e as demais pertencem à classe B, caso o o modelo avaliado tenha fornecido como resposta que 100% das amostras pertencem à classe A, mesmo considerando que a Acurácia obtida pelo modelo foi de 90%, sabe-se que ele não foi capaz de classificar corretamente qualquer amostra pertencente à classe B, definindo-se como um modelo pouco confiável. Portanto, para se combater um desbalanceamento de classes ao se analisar a Acurácia de forma isolada, algumas outras métricas precisam ser analisadas em conjunto, como por exemplo:

- Especificidade: dada por:

$$\text{Especificidade} = \frac{TN}{TN + FP}, \quad (15)$$

qualifica o modelo em relação à sua capacidade de detectar quando os elementos de um conjunto de amostras não pertencem à uma classe específica;

- Precisão: obtida por:

$$\text{Precisão} = \frac{TP}{TP + FP}, \quad (16)$$

considerando todas as respostas fornecidas para uma classe específica, consiste na avaliação do modelo em relação à quantidade de classificações realizadas corretamente;

- Revocação: calculada por:

$$\text{Revocação} = \frac{TP}{TP + FN}, \quad (17)$$

considerando todas as respostas esperadas para uma classe específica, qualifica o modelo em relação à quantidade de classificações realizadas corretamente;

- F1-Score: medida por:

$$\text{F1-Score} = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}}, \quad (18)$$

representa a média harmônica entre a Precisão e a Revocação, medidas que se preocupam com a resposta individual do modelo para uma dada classe.

2.3.3 Rentabilidade e Risco

Sabe-se da existência de métricas específicas para avaliar os modelos quando aplicados em uma estratégia de recomendação de investimentos. Dentre elas, a mais simples consiste em avaliar o Retorno de Portfólio (*Portfolio Return*), que mede a Rentabilidade Acumulada (RA), isto é, o quanto a utilização do modelo proporcionou de lucro (ou perda) ao investidor. O cálculo do percentual de RA pode ser visualizado na Equação 19, onde P_i e P_f são, respectivamente, o Patrimônio Inicial e Final do investidor.

$$RA = \frac{P_i - P_f}{P_i} \times 100. \quad (19)$$

Além da RA, também é importante que seja avaliado o risco assumido pelo investidor em uma eventual utilização de uma estratégia de recomendação de investimentos. Para isto, diversas métricas são utilizadas por agentes do mercado financeiro. Dentre elas, destaca-se o Índice de Sharpe (*Sharp Ratio*), que mede o Retorno por Unidade de Risco (RR), representado na Equação 20, onde R é a Série de Retorno²; $Avg(R)$ é a média da Série de Retorno; σ_R é o Desvio Padrão da Série de Retorno; e R_f representa o Retorno Livre de Risco (*Risk Free Ratio*)³.

$$RR = \frac{Avg(R) - R_f}{\sigma_R}. \quad (20)$$

2.4 Blockchain e Criptomoedas

No segundo semestre de 2008, foi publicada a proposta de funcionamento da primeira moeda digital, um novo tipo de ativo virtual que apresenta alto grau de inovação, com a promessa de provocar uma grande revolução no sistema financeiro (NAKAMOTO, 2008). Pouco tempo depois, em 3 de Janeiro de 2009, o Bitcoin (BTC) teve a sua primeira transação registrada, tornando-se a primeira criptomoeda da qual se tem registro. Nos anos que se seguiram outras moedas foram lançadas, também conhecidas como *altcoins* (*Alternative Coins* ou Moedas Alternativas). No início de 2021 já existiam mais de 4000 criptomoedas alternativas ao Bitcoin (COINMARKETCAP, 2021).

Todas as criptomoedas foram construídas utilizando a mesma tecnologia proposta por Nakamoto (2008), denominada *Blockchain*, que trouxe consigo grande inovação devido

² A Série de Retorno é formada pelo percentual de variação do patrimônio do investidor a cada passo da série original.

³ *Risk Free Ratio* é o retorno de um investimento sem riscos. No Brasil, a SELIC (taxa básica de juros), é geralmente utilizada para este fim.

à possibilidade de descentralização de serviços que sempre dependeram de uma entidade central, principalmente por questões de segurança. A *Blockchain* consiste literalmente em uma cadeia de blocos concatenados, compartilhados entre múltiplos usuários por meio de uma rede ponto-a-ponto. Esses blocos podem conter qualquer tipo de informação, a depender do fim para o qual a tecnologia está sendo aplicada. No caso do Bitcoin, por exemplo, os registros existentes em cada bloco nada mais são do que o histórico de todas as transações ocorridas em um intervalo de tempo (ULRICH, 2014).

É sabido que o Bitcoin foi a primeira moeda digital e possui a maior capitalização de mercado dentre todas as criptomoedas existentes (COINMARKETCAP, 2021). Acredita-se que, por este motivo, mais de 90% dos trabalhos analisados utilizam o Bitcoin na realização de experimentos. Entretanto, será visto no Capítulo 3, que a maior parte dos autores também considera pelo menos uma das seguintes criptomoedas: Litecoin (LTC), Ripple (XRP) e Ether (ETH). Ao final de 2020, juntas, as 4 criptomoedas contabilizavam mais de 70% da capitalização de mercado, considerando todas as criptomoedas existentes, compondo a lista das 10 mais relevantes (COINMARKETCAP, 2021).

Mesmo utilizando uma tecnologia em comum, os criadores de cada criptomoeda são os responsáveis por algumas definições importantes, que terão grande influência na sua adoção. Nesta subseção alguns aspectos importantes do funcionamento do Bitcoin, Litecoin, Ripple e Ether serão explicados, dando-se maior ênfase ao Bitcoin. Por último, será apresentado um comparativo das principais características de cada uma das 4 criptomoedas mencionadas.

2.4.1 Bitcoin (BTC)

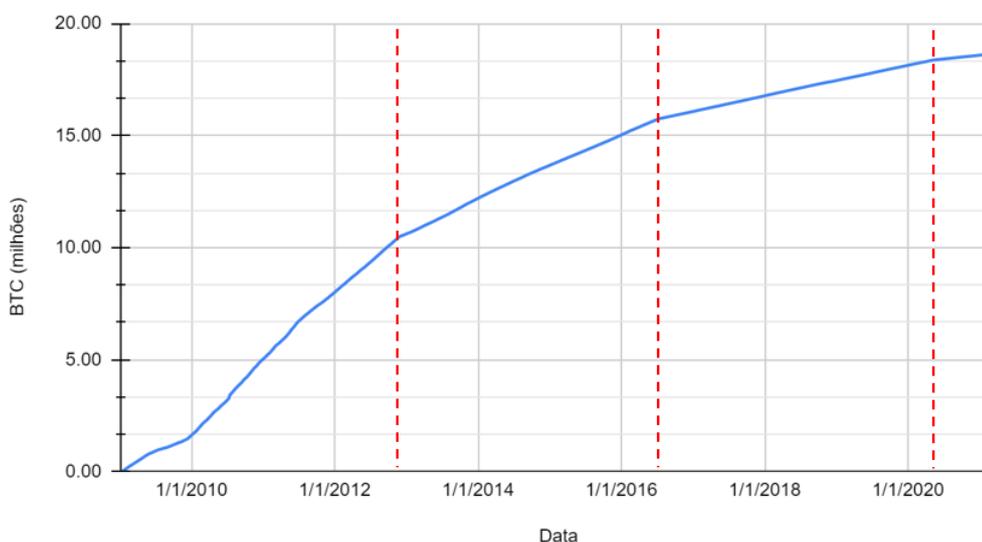
Devido à natureza pública e distribuída do Bitcoin, existe a necessidade de um consenso entre os participantes da rede, que proveem a força computacional para confirmar os registros e conciliar as transações. Em sua proposta, Nakamoto (2008) optou por um modelo baseado em mineração e prova de trabalho (*Proof-of-Work*, ou *PoW*). Neste modelo, os usuários que proveem a força computacional para o funcionamento da rede são chamados mineradores. De forma simplificada, no modelo *PoW*, novos Bitcoins são criados (ou minerados) à medida que milhares de computadores dispersos resolvem problemas matemáticos complexos que verificam as transações na *Blockchain*, sendo esta uma forma de recompensá-los por este trabalho. Este modelo é questionável visto que grande quantidade de energia é gasta todos os dias no processo de mineração. Além disso, pelo fato dos algoritmos ficarem mais complexos ao longo do tempo, a tendência é que apenas pessoas (ou empresas) que disponham de equipamentos com alto poder de processamento consigam minerar Bitcoins (O'DWYER; MALONE, 2014).

A recompensa pela mineração do Bitcoin é reduzida pela metade a cada 210.000 blocos minerados (a este processo dá-se o nome de *halving*). Inicialmente, o minerador era

recompensado com 50 Bitcoins ao processar 1 bloco. Atualmente, após o último *halving* ocorrido em 2020, a recompensa caiu para 6,25 BTC. Considerando que 1 novo bloco de Bitcoin é minerado a cada 10 minutos, sabe-se que um novo *halving* acontecerá no ano de 2024. A tendência natural é que, com a adoção da criptomoeda por mais pessoas e tornando-se difundido o seu uso como meio de pagamento, a redução da oferta irá provocar consequentemente um aumento progressivo no seu valor.

Outra característica importante se refere à uma definição prévia do número máximo de unidades da moeda que poderá existir. No caso do Bitcoin, foi estabelecido que este número é igual a 21 milhões de unidades. Isso significa que quando essa quantidade for alcançada, não há a possibilidade de a base monetária diminuir ou contrair-se. Estima-se que os mineradores colherão o último *Satoshi* (nome dado à menor fração do Bitcoin que representa 0,0000001 BTC) no ano 2140. O gráfico expresso na Figura 3 apresenta a evolução do número de Bitcoins minerados ao longo do tempo. Foram incluídos três marcadores que indicam o momento onde os *halvings* ocorreram, nos anos de 2012, 2016 e 2020, respectivamente.

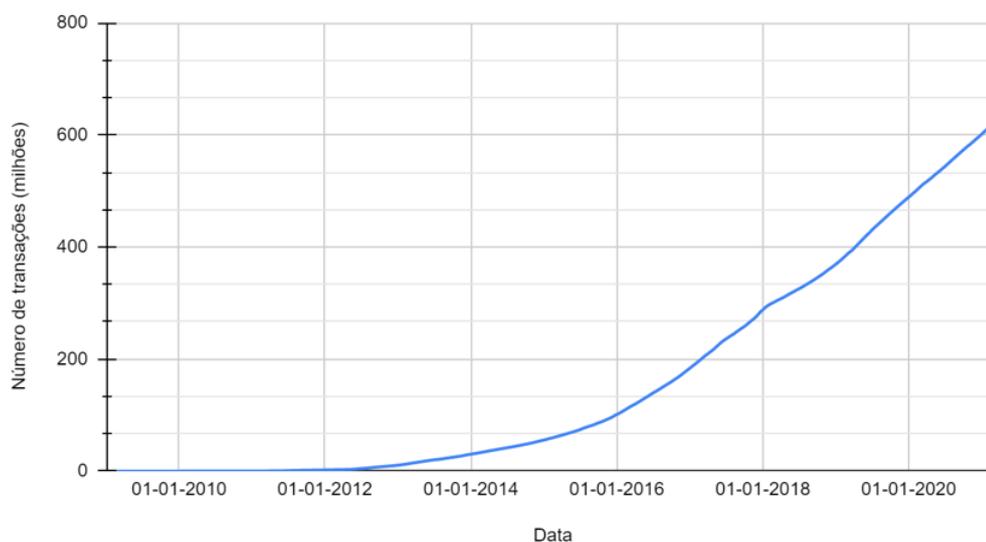
Figura 3 – Bitcoins em circulação.



Fonte: dados extraídos de [Blockchain \(2021\)](#).

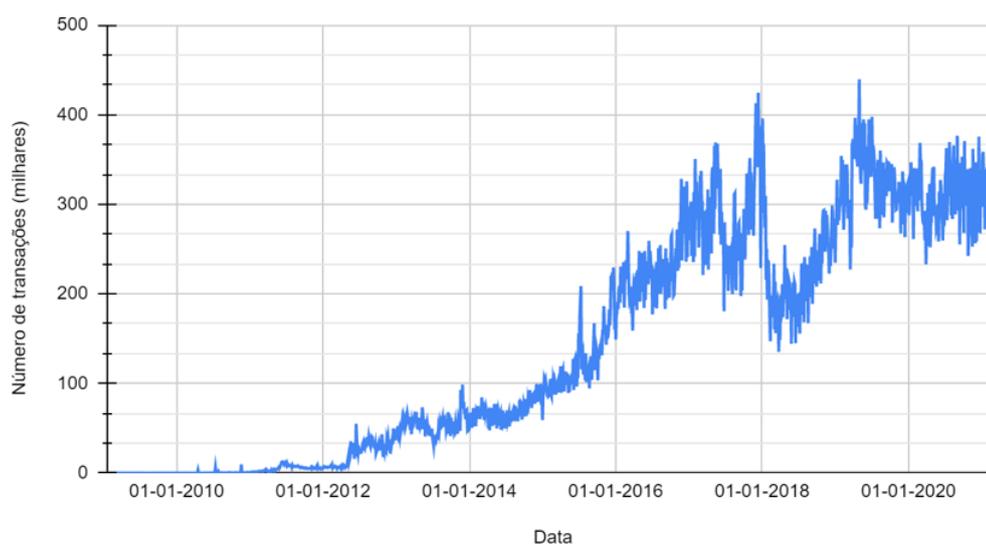
Para que se tenha uma dimensão acerca da adoção do Bitcoin pelo mercado, os gráficos ilustrados nas Figuras 4 e 5 apresentam, respectivamente, o número total de transações entre Janeiro de 2009 e Abril de 2021 e o volume de transações diárias que envolveram a criptomoeda no mesmo período. Em relação ao número total de transações, percebe-se que há um aumento progressivo a partir de 2012. No que tange ao volume diário, uma tendência de alta se dá a partir do mesmo ano, havendo uma queda expressiva em 2018, porém há uma retomada do crescimento logo na sequência.

Figura 4 – Evolução do número total de transações envolvendo o Bitcoin.



Fonte: dados extraídos de [Blockchain \(2021\)](#).

Figura 5 – Volume diário de transações envolvendo o Bitcoin.



Fonte: dados extraídos de [Blockchain \(2021\)](#).

2.4.2 Litecoin (LTC)

O Litecoin (LTC) é uma criptomoeda descentralizada baseada no Bitcoin e que compartilha, inclusive, do mesmo código-fonte. Lançada em 2011 por Charlie Lee, tem como objetivo principal se tornar uma moeda alternativa ao Bitcoin ([GIBBS; YORDCHIM, 2014](#)). Entretanto, no início de 2021, o valor do Litecoin ainda era cerca de trezentas vezes menor que o valor do Bitcoin ([COINMARKETCAP, 2021](#)).

Apesar da semelhança, existem algumas diferenças importantes ao se comparar o funcionamento do LTC com BTC. Conforme [Padmavathi e Suresh \(2019\)](#), foi previamente estabelecido, por exemplo, que o número máximo de unidades da criptomoeda que virá a

existir é de 84 milhões, uma quantidade que supera em 4 vezes o número previsto para o Bitcoin. Estima-se que este número será alcançado no ano 2142. No final de 2020 já haviam sido minerados aproximadamente 65 milhões de LTC ([COINMARKETCAP, 2021](#)). Por haver um espaço de tempo menor entre os blocos e um algoritmo de mineração que, conseqüentemente, exige menos memória, o Litecoin rapidamente conquistou adeptos, e desde então vem assegurando uma posição entre as maiores criptomoedas em capitalização de mercado ([GIBBS; YORDCHIM, 2014](#)).

2.4.3 Ripple (XRP)

Enquanto grande parte das criptomoedas foram baseadas no Bitcoin, o Ripple (XRP), por sua vez, foi desenvolvido de forma completamente apartada. Conforme [Jain et al. \(2018\)](#), não se trata unicamente de uma Moeda Alternativa (*altcoin*), mas sim de uma plataforma de pagamentos, que permite transações utilizando moedas fiduciárias (USD, EUR, BRL, etc.) ou criptomoedas (BTC, ETH, LTC, etc.). De forma simplificada, o objetivo da Ripple é substituir a forma como os bancos de todo o mundo realizam suas transações. Diversas entidades financeiras possuem projetos em andamento na plataforma, como o Banco Santander, por exemplo ([LABS, 2020](#)).

Apesar do seu desenvolvimento ter a tecnologia *Blockchain* como fundamento, o XRP é controlado por uma entidade central chamada *Ripple Labs*, uma companhia de investimento, que atua no desenvolvimento de soluções de pagamentos globais para instituições financeiras. O Bitcoin pode ser visualizado como uma moeda, pois não está associado diretamente a qualquer instituição. Em contrapartida, o XRP é, na verdade, um produto da Ripple Labs, a organização responsável por regular o seu funcionamento no mercado.

Outra característica que difere o XRP do BTC é o fato de todas as moedas terem sido previamente geradas, ou seja, não há a necessidade do trabalho de mineradores. Originalmente, a *Ripple Network* (Rede Ripple) foi criada com o número de 100 bilhões de unidades de XRP. Desse total, 20% foi entregue aos fundadores, 25% ficou mantido pela Ripple Labs e os 55% restantes foram colocados à venda ([JAIN et al., 2018](#)). Ao contrário de outras criptomoedas, em vez de existir uma vasta rede de nós, a validação das transações é realizada por alguns nós específicos (*Unique Node List*), escolhidos criteriosamente pelo comitê da Ripple Labs.

2.4.4 Ether (ETH)

Proposta pelo programador e escritor russo-canadense Vitalik Buterin ([BUTERIN, 2013](#)) e lançada oficialmente em Julho de 2015, a Ethereum foi projetada para se tornar uma grande plataforma de *software* distribuído. Diferente do Bitcoin, que nasceu para ser

utilizado como uma moeda de troca, a Ethereum foi criada como uma plataforma que permite a construção de aplicações descentralizadas (*decentralized applications* ou *dapps*) baseadas em contratos inteligentes (*smart contracts*). Conforme [Wahrstatter \(2021\)](#), por meio do uso da tecnologia *Blockchain*, a plataforma Ethereum habilita o desenvolvimento de aplicações executadas de forma totalmente distribuída que podem ser, por exemplo, soluções de armazenamento de dados, oferta de poder computacional, jogos ou aplicações para o mercado financeiro. Mais detalhes sobre *dapps* e *smart contracts* podem ser obtidos em [Antonopoulos e Wood \(2018\)](#).

O Ether, por sua vez, é o ativo digital utilizado para recompensar os desenvolvedores que utilizam a plataforma Ethereum para seus projetos. Assim como o Bitcoin, para se criar novas moedas, também é usado o processo de mineração, onde os próprios usuários verificam e validam as operações, sendo recompensados por isso (*Proof-of-Work*). Entretanto, está sendo desenvolvida uma grande atualização para a plataforma Ethereum, que recebeu o nome de Ethereum 2.0. No novo modelo proposto, a recompensa para os validadores da rede será gerada apenas através das taxas de transação e não mais da mineração de novos blocos (*Proof-of-Stake* ou *PoS*) ([SAVCHEKO M., 2019](#)). O lançamento dessa atualização foi dividido em fases, a primeira delas iniciou em 2020 e a última está prevista para o ano de 2022.

2.4.5 Análise Comparativa

Na Subseção 2.4 foram descritos aspectos relevantes que diferenciam algumas das criptomoedas com maior valor de mercado. A Tabela 3 resume as informações apresentadas e inclui outros detalhes de implementação acerca do Bitcoin, Litecoin, Ripple e Ether. A ordem das criptomoedas foi definida baseada na data da primeira transação, priorizando-se da mais antiga para a mais recente. Na tabela são detalhados alguns aspectos referentes, por exemplo, à definição de um limite acerca da quantidade máxima de unidades da moeda; existência, periodicidade e ocorrências dos *halvings*; precisão decimal; e algoritmos utilizados pela rede.

Tabela 3 – Comparação acerca de algumas das principais características existentes em criptomoedas para Bitcoin (BTC), Litecoin (LTC), Ripple (XRP) e Ether (ETH).

	Bitcoin	Litecoin	Ripple	Ether
Primeiro bloco	03/01/2009	08/10/2011	01/06/2012	30/07/2015
Tempo por bloco	10 minutos	2,5 minutos	4 segundos	12-14 segundos
Último <i>Halving</i>	11/05/2020	05/08/2019	-	-
Próximo <i>Halving</i>	13/03/2024	06/08/2023	-	-
Periodicidade <i>Halvings</i>	210k blocos	840k blocos	-	-
Recompensa inicial	50BTC	50LTC	-	-
Recompensa atual	6,25 BTC	12,5 LTC	-	-
Quantidade Inicial	0 BTC	0 LTC	100B XRP*	72M ETH
Quantidade (abril/2021)	18,7M BTC	66,8M LTC	45,4B* XRP	115,3M ETH
Quantidade Máxima	21M (total)	84M (total)	100B (total)	18M (anual)
Precisão Decimal	10^{-8}	10^{-8}	10^{-6}	10^{-18}
Menor fração	Satoshi	Litoshi	Drop	Wei
Protocolo	PoW	PoW	RPCA	PoS
Algoritmo de Hash	SHA-256	Scrypt	RPCA	Etash

Fonte: informações extraídas de [CoinMarketCap \(2021\)](#) e trabalhos mencionados no detalhamento de cada criptomoeda.

3 Trabalhos Relacionados

Diante dos objetivos definidos no Capítulo 1, um estudo sobre o que já foi feito no contexto de previsão de valores em criptomoedas fez-se necessário. De forma geral, neste capítulo serão abordadas as principais características de alguns dos trabalhos mais relevantes sobre este assunto disponíveis na literatura. Ao final, será apresentado um resumo dos 30 trabalhos utilizados como referência neste estudo, contendo as características mais relevantes de cada um deles no que tange, principalmente, às técnicas utilizadas, conjunto de dados e medidas de desempenho.

3.1 Abordagens

A maior parte dos trabalhos analisados tem por objetivo a previsão do valor de uma ou mais criptomoedas. Para isto, geralmente, as abordagens partem de uma etapa de coleta, que consiste na obtenção de dados, para o dia atual e/ou dias passados, sobre os quais acredita-se na existência de uma relação com o preço da criptomoeda estudada. Nestes trabalhos, os dados coletados podem ser informações de mercado da própria criptomoeda, outros tipos de ativos financeiros ou até mesmo informações não-financeiras. Após a coleta, os dados obtidos são utilizados como entrada para modelos, como os que foram apresentados no Capítulo 2, os quais têm por objetivo oferecer uma previsão do valor da criptomoeda que apresente o menor erro possível ao ser comparado com o valor real. Quando o objetivo de um modelo é realizar a previsão de valores discretos, de forma convencional, denomina-se este como um modelo de regressão.

É possível transformar qualquer problema de regressão em um problema de classificação (SALMAN; KECCMAN, 2012). Diante disso alguns autores trabalham com o problema em questão sob uma nova perspectiva, tendo por objetivo responder à seguinte pergunta: o valor previsto será maior, igual ou menor se comparado a um valor de referência? Para responder a esta questão duas abordagens são possíveis: converter a saída de um modelo de regressão, transformando-a em classes, ou utilizar modelos desenvolvidos especificamente para problemas de classificação. Na Tabela 4 são apresentados três exemplos sobre como transformar um problema de regressão em um problema de classificação. Nesta tabela, y_t é o valor registrado para a série no instante t e y_{t-1} é o valor assumido no instante anterior, sendo as classes 0, 1 e 2 geradas por meio de uma simples comparação entre os valores da série. Na literatura, é possível encontrar autores que trabalham com as três formas exemplificadas. Resumidamente, em relação ao objetivo, a maior parte dos trabalhos analisados se concentram na realização de previsões (1) do valor futuro ou (2) do sentido de variação do preço de uma ou mais criptomoedas.

Tabela 4 – Transformação de problemas de regressão em problemas de classificação.

Exemplo	Condição	Classe
A	$y_t > y_{t-1}$	0
	$y_t = y_{t-1}$	1
	$y_t < y_{t-1}$	2
B	$y_t \geq y_{t-1}$	0
	$y_t < y_{t-1}$	1
C	$y_t > y_{t-1}$	0
	$y_t \leq y_{t-1}$	1

O mercado de criptomoedas permanece aberto 24 horas por dia, 7 dias por semana (CATANIA; GRASSI; RAVAZZOLO, 2019). Sendo assim, o valor de uma criptomoeda pode mudar muito rapidamente. Em 24 de Março de 2021, por exemplo, o Bitcoin iniciou o dia cotado a US\$ 54.710,49, chegando a alcançar o valor de US\$ 57.262,38, porém encerrou o dia valendo US\$ 52.774,26 (COINMARKETCAP, 2021). É fato que este foi um dia atípico, mas, de qualquer forma, é do interesse de alguns autores realizar a previsão não somente do valor de fechamento mas também do maior e/ou menor valor alcançado pela criptomoeda em determinado momento, ou ainda outras medidas/indicadores associados a este contexto, conforme será detalhado mais adiante.

Independentemente do modelo aplicado, uma definição importante se refere a quantos passos a frente a saída do modelo será fornecida, definindo-se aqui o conceito de horizonte. Pode-se trabalhar, por exemplo, com a previsão do valor da criptomoeda para $1..m$ dias adiante. Autores que trabalham com este tipo de horizonte geralmente o denominam *Short Frequency Forecasts* (Previsões de Baixa Frequência). Em contrapartida, também é possível trabalhar com previsões para intervalos menores, como minutos ou horas. A este tipo de horizonte é geralmente dado o nome de *High Frequency Forecasts* (Previsões de Alta Frequência) (TSAY, 2010).

Apresentado o vasto número de possibilidades, é importante mencionar que grande parte dos autores trabalham com os horizontes EOD (*End of Day*) ou OSA (*One Step Ahead*). Na previsão para o fim do dia (EOD) considera-se a execução do modelo no momento de abertura do mercado, onde o objetivo é fornecer uma previsão para o fechamento do mercado, considerado pela maior parte dos autores como 00:00 UTC e 23:59 UTC, respectivamente. Em se tratando de previsão para o dia seguinte (OSA), considera-se que o modelo seja executado até o fim do dia (instante t), realizando a previsão do valor da criptomoeda para determinado horário no dia seguinte (instante $t + 1$).

Após a definição de algumas abordagens para previsão de valores em criptomoedas, as possíveis variáveis estimadas e os horizontes geralmente considerados, pode-se mencionar uma abordagem complementar, vista como uma possível aplicação para os modelos desenvolvidos. Alguns autores propõem o uso da saída fornecida por modelos de regressão ou classificação em uma estratégia para recomendação de investimentos. Pode-se, por

exemplo, orientar ao investidor que faça a aquisição de moedas caso a previsão para o próximo passo seja de aumento de valor ou, de forma contrária, pode ser dada uma sugestão de venda, caso seja previsto que o valor da criptomoeda irá diminuir.

Diante dos objetivos listados, de forma resumida, pode-se dividir os problemas abordados pelos trabalhos analisados em três grupos: previsão do valor, previsão do sentido de variação e recomendação de investimentos. É importante mencionar que um determinado trabalho pode apresentar solução para 1, 2 ou 3 destes problemas. Além disso, quanto ao horizonte, pode-se dividi-los em *Short Frequency Forecasts* e *High Frequency Forecasts*. Por último, em relação ao que será previsto ou classificado, a maior parte dos autores trabalha com o valor, dentro do horizonte escolhido. Entretanto, é do interesse de alguns autores prever outros tipos de dados relacionados ao mercado de criptomoedas.

3.2 Conjunto de Dados

Na literatura é possível encontrar trabalhos que utilizam diversas fontes na coleta dos dados que servirão de entrada para os modelos. Independentemente da fonte de dados considerada, neste momento, faz-se necessária a definição de dois termos amplamente utilizados neste trabalho. Denomina-se atributo um tipo de informação qualquer, associado a um domínio específico. Portanto, entende-se que o menor e o maior valor alcançado pela criptomoeda em uma unidade de tempo qualquer (hora, dia, semana, mês, etc.) são dois exemplos de atributos associados ao domínio criptomoedas. Por sua vez, denomina-se variável os possíveis valores que um atributo pode assumir em um conjunto de dados. Neste caso, o valor de fechamento diário (*close*) da criptomoeda XPTO nos instantes t (períodos de 24 horas), existentes no intervalo 01/01/2021 à 01/08/2021, é um exemplo de variável, que pode ser expressa como $close_t$. Definidos estes conceitos, pode-se compreender melhor os principais conjuntos de atributos utilizados nos trabalhos consultados.

[Mallqui e Fernandes \(2019\)](#) afirmam que os dados utilizados em seus experimentos podem ser classificados como internos ou externos. Para os autores, as fontes de informações internas se referem à parte do conjunto de dados relacionada ao comportamento de diferentes parâmetros da criptomoeda ou da rede sob a qual ela opera. Por sua vez, as fontes de dados externas se referem a informações que não apresentam relação direta com o mercado da criptomoeda, nem tampouco com qualquer uma de suas características. Em outros trabalhos verifica-se a existência de termos que classificam o conjunto de dados utilizado em grupos mais específicos. [Sebastião e Godinho \(2021\)](#) consideram três grupos de atributos: *Exchange Trading Information*, *Blockchain Information* e *Deterministic Values*. O primeiro deles abrange informações de mercado da própria criptomoeda, o segundo trata de métricas da *Blockchain* e o último agrupa os atributos que não possuem qualquer relação com a criptomoeda. [Catania, Grassi e Ravazzolo \(2019\)](#) classificam em dois grupos os

atributos utilizados em seu trabalho: macroeconômicos e financeiros. Finalmente, [Jagannath et al. \(2021\)](#) e [Chen, Li e Sun \(2020\)](#) também usam formas diferentes de definir o conjunto de dados utilizado.

Devido à abrangência deste trabalho, onde buscou-se alcançar o maior número possível dentre o vasto número de propostas disponíveis, na apresentação dos atributos mais utilizados na literatura, optou-se por dividi-los em quatro categorias: Cotação, Mercado, Rede e Externos. Nas subseções a seguir será feito um resumo dos principais elementos que compõem cada categoria de atributos. Em seguida, a Subseção 3.2.5 apresenta o conceito de Indicadores Técnicos, artifício amplamente utilizado por diversos autores. No Anexo A é feito um detalhamento de todos os atributos considerados nos trabalhos de maior relevância no assunto.

3.2.1 Dados de Cotação

No estudo do comportamento de séries temporais envolvendo ativos financeiros é comum que seja realizada uma divisão da série em pequenos intervalos, comumente denominados *candles*. Um *candle* possui quatro componentes, sendo todos eles medidas do preço de determinado ativo durante o intervalo compreendido, em momentos distintos, a saber: o preço de abertura, fechamento, o maior e o menor valor alcançado ([LUCARELLI; BORROTTI, 2019](#)). A duração de um *candle* é variável e pode ser, por exemplo, 1 minuto, 5 minutos, 1 hora, 1 dia, 1 semana, etc.. Sendo assim, uma série temporal composta por 1 ano de observações do valor de uma criptomoeda pode possuir 52 *candles* semanais ou 365 *candles* diários (considerando um ano não bissexto).

Diversos autores trabalham com a previsão do valor da criptomoeda para o próximo dia e, para isso, necessitam de informações de dias passados, utilizando portanto a definição de *candle* diário. Neste contexto, para a formação de um *candle* em um dia qualquer, considera-se o valor da criptomoeda, por exemplo, as 00:00 UTC (abertura), as 23:59 UTC (fechamento) e também o maior e o menor valor alcançado neste intervalo. Outros autores utilizam 12:00 UTC como referência ([ATSALAKIS et al., 2019](#)).

Assim como outros tipos de ativos, uma criptomoeda não possui um valor fixo de cotação. O preço praticado, portanto, varia entre cada *Exchange*. A plataforma *CoinMarket-Cap*¹ fornece uma API (*Application Programming Interface*) gratuita na qual possível obter dados de *candles* diários do Bitcoin e outras criptomoedas. As informações disponibilizadas nessa plataforma consistem na média do preço praticado entre algumas das principais *Exchanges* (corretoras) existentes. Também é possível obter informações acerca dos preços praticados diretamente por alguma corretora específica, como em *Bitfinex*² e *Binance*³.

¹ <https://www.coinmarketcap.com>

² <https://www.bitfinex.com>

³ <https://www.binance.com>

3.2.2 Dados de Mercado

Além dos dados de cotação, diversas outras informações podem ser extraídas na observação do mercado de criptomoedas. No que tange à quantidade de moedas em circulação e ao seu valor de mercado podemos citar, por exemplo, as Capitalizações de Mercado e Realizada. A Capitalização de Mercado (*Market Capitalization*) consiste na multiplicação do preço de referência da criptomoeda (em *USD*) por uma estimativa aproximada do número de unidades atualmente em circulação (SEBASTIÃO; GODINHO, 2021). A Capitalização Realizada (*Realized Capitalization*) é um outro indicador de mercado, semelhante ao *Market Capitalization*, com a diferença de que considera o preço de cada moeda no momento em que foi movimentada pela última vez na *Blockchain* (CHRISTOFOROU; EMIRIS; FLORAKIS, 2020).

Em relação às transações realizadas, pode-se considerar também o quantitativo absoluto de transações diárias (*Number of Transactions*) e o Volume Transacionado (*Volume of Trades*), calculados em Bitcoin ou *USD*. Outras medidas de mercado relevantes são o volume de transações por minuto (*Trades per Minute*), a taxa de crescimento do volume de transações (*Volume of Trades Growth Rate*), a quantidade de unidades da criptomoeda existente (*Total Units*), o número de unidades da criptomoeda enviado em todas as transações realizadas no dia (*Sent*) e o tempo médio necessário para que uma transação seja confirmada pela rede (*Transaction Median Time*).

É possível ainda extrair informações relevantes dos endereços (ou carteiras) envolvidos nas transações (MUDASSIR et al., 2020). Uma métrica relevante representa a proporção de Bitcoins concentrados nas 100 carteiras que acumulam a maior quantidade de moedas em relação ao número total de moedas existente (*Top 100 to Total*). Outras medidas importantes se referem ao número de endereços distintos envolvidos em transações diariamente (*Sent From Adresses*) e o número de transações diárias que não envolveram as 100 carteiras com maior rentabilidade (*Non Popular Transactions*).

3.2.3 Dados da Rede

Conforme explicado na Seção 2.4, as criptomoedas implementam em seu protocolo algumas condições que tornam o uso da rede atraente tanto para os mineradores, que são os responsáveis pela garantia do funcionamento da rede, como dos usuários, que realizam as movimentações. Sobre o mecanismo de recompensa pelo uso de poder computacional implementado nas moedas baseadas em Prova de Trabalho (*Proof-of-Work*), o valor médio pago aos mineradores pelo processamento de uma transação (*Median Fees*) e a proporção das taxas enviadas em transações como recompensa em comparação ao volume total transacionado (*Fee-to-Reward Ratio*) podem ser consideradas medidas relevantes (JI; KIM; IM, 2019).

Em relação aos blocos, componentes fundamentais da *Blockchain* de uma criptomoeda, pode-se extrair informações como o tamanho e a quantidade de blocos minerados (*Block size* e *Block Count*, respectivamente). Conforme Jang e Lee (2017), o número total de blocos criados e já incluídos na *Blockchain* e o número médio de transações registrados por bloco estão diretamente relacionados ao volume de suprimento e demanda da criptomoeda pelo mercado e por isso também são medidas que podem contribuir bastante no desempenho dos modelos.

Muitos autores também utilizam variáveis relacionadas à dificuldade de mineração, que é proporcional ao aumento do número de mineradores (JAGANNATH et al., 2021). A capacidade computacional diária total da rede (*Hash Rate*) e a dificuldade média diária para mineração da criptomoeda (*Mining Difficulty*) são medidas que podem ser extraídas nesse sentido. A dificuldade é computada pela rede depois que um número especificado de blocos foi criado. Por definição, sabe-se que o tempo médio necessário para o processamento de um bloco é de geralmente 10 minutos para o Bitcoin e 2,5 minutos para o Litecoin, entretanto esse número pode variar.

Alguns dados de Mercado e da Rede podem ser extraídos diretamente de APIs disponibilizadas por corretoras. Existem também APIs públicas que disponibilizam essas informações como a oferecida pela plataforma *Coin Metrics*⁴. Nessa API são agregados dados de mais de 20 corretoras, disponibilizando gratuitamente mais de 42 métricas de dados históricos, para cerca de 50 criptomoedas. Além disso, há ainda uma versão paga, que possibilita o consumo dessas e outras métricas em tempo real.

3.2.4 Dados Externos

Em alguns trabalhos é possível encontrar autores investigando a relação de medidas externas ao contexto de criptomoedas que podem apresentar influência na variação do seu preço. Para isto, alguns autores consideram a pontuação de índices de ações de relevância mundial como NASDAQ⁵ e S&P 500⁶. Outros autores acrescentam ainda a evolução do preço de *commodities*, como ouro e petróleo, por exemplo. A pergunta que esses trabalhos buscam responder geralmente é: existe relação entre uma oscilação positiva (ou negativa) de determinado índice de ações ou valor de *commodity*, em instantes passados, com uma oscilação positiva (ou negativa) no valor de uma criptomoeda? (MALLQUI; FERNANDES, 2019).

No que tange ao desenvolvimento e evolução da plataforma, o código-fonte responsável pela implementação das criptomoedas geralmente é armazenado em um repositório

⁴ <https://community-api.coinmetrics.io>

⁵ NASDAQ (*National Association of Securities Dealers Automated Quotations*) é um mercado de ações norte-americano composto principalmente por empresas do ramo de tecnologia.

⁶ S&P 500 (*Standard Poor's 500*) é índice composto por quinhentos ativos cotados na bolsa de Nova York, qualificados devido ao seu tamanho de mercado, liquidez e representação de grupo industrial.

público chamado *Github*⁷. Neste site é possível encontrar o repositório de código do *Bitcoin* e diversas *altcoins*, como *Ether* e *Litecoin*. [Fahmi et al. \(2018\)](#) entende que um maior número de usuários interessados no código de uma criptomoeda pode apresentar uma relação positiva com o aumento do seu preço. Para que se tenha uma dimensão, atualmente o Bitcoin já possui mais de 400 *Pull Requests* finalizadas e mais de 28.000 *Forks*. Uma *Pull Request* é uma solicitação de mudança aberta por qualquer pessoa que possua uma conta na plataforma, podendo ser para correção de um problema existente, melhoria ou evolução do código. *Forks* indicam quantas vezes o código foi clonado, para qualquer fim como, por exemplo, a criação de uma nova moeda baseada em uma outra de origem.

Outro fator externo relevante são os chamados indicadores sociais. De acordo com [Chen, Li e Sun \(2020\)](#), trata-se de importantes medidas acerca da atenção de investidores e cobertura da mídia sobre o assunto, o que tende a influenciar de modo geral no sentimento do mercado especulativo. É possível que se faça uma investigação, por exemplo, da relação entre o volume de postagens consideradas positivas, negativas e neutras no *Twitter*⁸ com a variação do preço de criptomoedas ([JAIN et al., 2018](#)). [Kristoufek \(2013\)](#) afirma ter encontrado uma correspondência positiva entre o aumento do valor do Bitcoin e o crescimento do número de pesquisas por termos relacionado à criptomoeda em plataformas de busca como *Google*⁹ e *Wikipedia*¹⁰.

3.2.5 Indicadores Técnicos

Os atributos definidos até este momento consistem basicamente em valores brutos, extraídos de diversos contextos, sobre os quais os autores acreditam haver alguma relação com o valor ou o sentido de variação do preço de criptomoedas. Além de utilizar dados brutos como entrada para os modelos é comum encontrar autores que extraem novas informações destes dados por meio de Indicadores Técnicos. Segundo [Mudassir et al. \(2020\)](#), os Indicadores Técnicos são capazes de demonstrar propriedades que não são facilmente perceptíveis em dados brutos. Eles mostram, por exemplo, como o valor da criptomoeda está relacionado ao desvio padrão do número de transações em um período de 30 dias em vez de utilizar somente dados brutos sobre este atributo de forma individual (por *candle*).

A Média Móvel Simples (*Simple Moving Averages* ou *SMA*) é um exemplo de Indicador Técnico. Ao aplicar este indicador pode-se, por exemplo, calcular a média aritmética dos valores de fechamento (*close*) registrados nos últimos 3 dias ($close_t$, $close_{t-1}$ e $close_{t-2}$) e utilizar o resultado como entrada de um modelo que realiza a previsão do valor de fechamento para o próximo dia ($close_{t+1}$) ([ATSALAKIS et al., 2019](#)). Além disso, é possível

⁷ <https://github.com>

⁸ <https://www.twitter.com>

⁹ <https://www.google.com>

¹⁰ <https://www.wikipedia.com>

utilizar o valor obtido na definição das *labels* para as amostras na resolução do problema de classificação (subida/descida), ou seja, para definir a classe à qual uma amostra pertence, em vez de comparar $close_t$ com $close_{t-1}$ é comparado se $close_t$ é superior ou inferior a média móvel deste atributo para os últimos m dias (CHRISTOFOROU; EMIRIS; FLORAKIS, 2020).

Alguns trabalhos exploram com mais profundidade o uso de Indicadores Técnicos em uma etapa de preparação do conjunto de dados. Após a obtenção dos valores brutos, é realizado um pré-processamento com a inclusão de valores computados para os Indicadores Técnicos sobre os dados brutos de interesse. Para isto, nos trabalhos analisados, além de SMA, são também considerados os indicadores *Weighted Moving Average (WMA)*, *Relative Strength Index (RSI)*, *Accumulation/Distribution Oscillator (A/D)*, *Commodity Channel Index (CCI)*, *Larry William's R (LWI)*, *Momentum*, *Moving Average Convergence Divergence (MACD)*, *Stochastic D% (SD)* e *Stochastic K% (SK)*, *Stochastic Oscillator (SOC)* e *On Balance Volume (OBV)*. É sugerida consulta complementar aos trabalhos de Atsalakis et al. (2019), Borges e Neves (2020), Pabuccu, Ongan e Ongan (2020), Alonso-Monsalve et al. (2020), onde são detalhados com clareza o que cada indicador técnico mencionado representa e como podem ser calculados para um dado atributo.

3.3 Seleção de Variáveis

Após a extração e preparação do conjunto de dados de interesse, é comum os autores se depararem com um grande conjunto de variáveis. O Anexo A apresenta uma lista composta por mais de 80 atributos candidatos, que podem ser utilizados como entrada para os modelos. Sabe-se, porém, da existência de outros atributos não utilizados nos trabalhos analisados. Além disso, a aplicação de Indicadores Técnicos sobre todo o conjunto de dados resultaria em uma expansão considerável do número de variáveis disponíveis. Mudassir et al. (2020), por exemplo, considera apenas 16 atributos em seu trabalho. Entretanto, após a aplicação de alguns Indicadores Técnicos para diferentes *lags* (passos), o autor chega a um número superior a 700 variáveis.

Existe ainda a possibilidade de que sejam extraídas outras informações do mercado e da rede de uma criptomoeda, e também de que sejam investigadas a relação do seu preço com outros fatores externos não mencionados. O site oficial da plataforma *Blockchain*¹¹, por exemplo, apresenta uma série de métricas específicas do Bitcoin, onde nem todas elas foram consideradas nos trabalhos analisados. Diante deste cenário, técnicas de seleção de variáveis (seleção de características ou *feature selection*) são importantes para que sejam apresentadas aos modelos apenas as variáveis mais relevantes.

Waikato Environment for Knowledge Analysis (WEKA) é um *software* amplamente

¹¹ <https://www.blockchain.com>

utilizado neste sentido pois implementa diversos métodos de seleção de características (MALLQUI; FERNANDES, 2019). Para utilizá-lo, após a importação do conjunto de dados completo, é necessário escolher o método de Avaliação de Atributos (*Attribute Evaluator*) e o Método de Busca (*Search Method*). O *Attribute Evaluator* é a técnica pela qual cada atributo do conjunto de dados será avaliado no contexto da variável de saída definida, enquanto o *Search Method* representa o método utilizado na navegação entre as diferentes combinações de atributos para que se chegue a uma lista reduzida de variáveis. A lista dos *Attribute Evaluators* e *Search Methods* disponíveis na WEKA e detalhes de implementação podem ser obtidos no site oficial da ferramenta¹².

Boruta é outro método de seleção de variáveis, não implementado na WEKA, proposto pelos cientistas poloneses Miron Kurasa e Witold Rudnicki, baseado na técnica de *Machine Learning* denominada *Random Forests* (RF) (KURSA; RUDNICKI, 2010). Devido à natureza de aprendizagem por agrupamento (*Ensemble Learning*) presente nas *Random Forests*, neste método são considerados os votos de múltiplos classificadores. Na avaliação da importância de uma variável específica, compara-se a relevância de uma variável aleatória, adicionada ao conjunto de dados, com o valor definido por cada classificador para a variável em questão. Como resultado final são consideradas relevantes apenas as variáveis do conjunto original cuja importância seja maior do que a importância definida para as variáveis aleatórias (MCNALLY; ROCHE; CATON, 2018).

Existem ainda autores que apresentam métodos próprios de seleção de variáveis. Mudassir et al. (2020), por exemplo, realiza repetidas iterações sobre o conjunto de dados de treinamento. A cada iteração, o autor remove uma das variáveis presentes no conjunto de dados e verifica o desempenho do modelo. As variáveis que, ao serem removidas, gerarem maior degradação nos resultados são consideradas as mais relevantes. Sabe-se ainda da existência de outros métodos de seleção de variáveis, não contemplados nos trabalhos analisados sobre criptomoedas, mas com amplo uso em pesquisas de outras áreas. Como exemplo, podem ser citados: *Fisher Linear Discriminant* (DUDA; STORK; HART, 2000) e *Kruskal-Wallis* (KRUSKAL W. H., 1952).

3.4 Técnicas de Previsão

Apesar do tema criptomoedas ser um assunto relativamente recente, ganhando maior visibilidade a partir de 2014 quando houve um aumento expressivo no valor do Bitcoin (COINMARKETCAP, 2021), o comportamento de séries temporais financeiras é um problema que vem sendo estudado a mais tempo. Diversos métodos estatísticos e econométricos tradicionais foram aplicados na previsão de séries temporais associadas ao mercado financeiro como, por exemplo, previsão do preço de ações (GALDI; PEREIRA,

¹² <https://www.cs.waikato.ac.nz/ml/weka/index.html>

2007), índices (LIM; SEK, 2013), taxa de conversão de moedas (PARIKAKIS; MERIKA, 2009) e preço de *commodities* (YAZIZA et al., 2013). Após a introdução do conceito de criptomoedas, alguns autores também aplicaram técnicas semelhantes a problemas de previsão em séries associadas a este novo mercado.

Dentre os métodos tradicionais utilizados, é possível citar o modelo Auto Regressivo Integrado e de Médias Móveis (*Autoregressive Integrated Moving Average* ou ARIMA) (BOX; JENKINS, 1976) como técnica de previsão, onde a variável de interesse é estimada a partir de seus próprios valores defasados (AYGUN; KABAKCI, 2021). Em termos de previsão de volatilidade (*log return*), o modelo Auto Regressivo de Heteroscedasticidade Condicional (*Autoregressive Conditional Heteroskedasticity* ou ARCH) (ENGLE, 1982) já foi utilizado, juntamente com algumas das diversas variações existentes, tais como *Generalized ARCH* (GARCH) (BOLLERSLEV, 1986), *Exponential GARCH* (EGARCH) (NELSON, 1991), *Threshold GARCH* (TGARCH) (GLOSTEN; JAGANNATHAN; RUNKLE, 1993), *GARCH in Mean* (GARCH-M) (BOLLERSLEV; ENGLE; WOOLDRIDGE, 1988), *Asymmetric Power ARCH* (APARCH) (DING; GRANGER; ENGLE, 1993), *Component GARCH* (CGARCH) (LEE; ENGLE, 1993). A motivação dos autores em utilizar o modelo ARCH e suas variantes reside no fato de que séries temporais financeiras geralmente exibem agrupamento de volatilidade variante com o tempo, ou seja, existem períodos de instabilidade intercalados com períodos de relativa estabilidade (ANGELINI; EMILI, 2018).

Além das técnicas tradicionais mencionadas, a maior parte dos trabalhos encontrados na literatura sobre previsão e classificação no contexto de criptomoedas utiliza métodos de *Machine Learning* para atingir o seu objetivo. No que tange aos sistemas de Aprendizagem por Agrupamento (*Ensemble Learning*) já foram utilizados: *Random Forests* (RF) (HO, 1995); *Regression Trees* (RT) (MORGAN; SONQUIST, 1963); e *Extreme Gradient Boosting* (XGB) (CHEN; GUESTRIN, 2016).

Algumas técnicas de *Machine Learning*, as quais se assemelham a métodos pré-existentes nos campos da matemática e estatística, também já foram aplicadas. Regressão Linear (*Linear Regression* ou LR), por exemplo, é uma técnica aplicada em problemas de previsão, utilizada para descrever a dependência linear da variável de saída com uma ou mais variáveis preditoras, a partir do desenho de uma equação linear que melhor represente uma aproximação dos dados (FAHMI et al., 2018). Por sua vez, Regressão Logística (*Logistic Regression* ou LR) é uma técnica de classificação utilizada para estimação da probabilidade de ocorrência de um evento binário, que utiliza uma função logística sigmoideal para gerar um valor de probabilidade, o qual determina o grau de associação entre determinada entrada e uma de duas classes possíveis (AKYILDIRIM; CORBET, 2020). Análise Discriminante Linear (*Linear Discriminant Analysis* ou LDA) é outra técnica neste mesmo sentido que, além de ser utilizada como método de seleção de variáveis, pode também ser aplicada a problemas de classificação (CHEN; LI; SUN, 2020).

As Redes Neurais Artificiais representam uma importante classe dentre os modelos de *Machine Learning*, amplamente aplicada a este contexto. Nestes trabalhos são utilizadas redes de diferentes arquiteturas, com número de neurônios (*neurons*) e camadas escondidas (*hidden layers*) variável. Além disso, o método de treinamento da rede e a direção dos sinais também diverge entre as abordagens propostas. No que tange às arquiteturas de Redes Neurais mais aplicadas a este problema, podem ser citadas duas que predominam: *Feedforward Neural Networks (FFNN)* e *Recurrent Neural Networks (RNN)*. Outros tipos de Redes Neurais consideradas nos trabalhos analisados, porém em menor escala, são: *Bayesian Regularization Neural Network (BRNN)*, *Radial Basis Function Neural Network (RBFNN)*, *Deep Residual Network (DRN)* e *Deep Neural Network (DNN)*. Os Sistemas Neuro-Fuzzy também foram aplicados por diversos autores na previsão de outras séries temporais do mercado financeiro, como pode ser visto, por exemplo, em [Billah, Waheed e Hanifa \(2015\)](#), [Lincy \(2016\)](#) e [Rajab e Sharma \(2019\)](#). [Atsalakis et al. \(2019\)](#) propõe o uso de um Sistema Neuro-Fuzzy de Inferência Adaptativa (*Adaptive Neuro Fuzzy Inference System* ou ANFIS), apresentado inicialmente em [Jang \(1993\)](#), na previsão do valor e do sentido de variação do preço de criptomoedas.

Recentemente, os Sistemas Evolutivos, um campo de pesquisa emergente, tem sido explorado em tarefas de previsão no contexto de criptomoedas. Exemplos de aplicação de Sistemas Evolutivos neste contexto podem ser encontrados em [Garcia, Leite e Skrzjanc \(2019\)](#), [Maciel, Ballini e Gomide \(2020\)](#). Neste mesmo sentido, essa pesquisa pretende contribuir com a aplicação de três Sistemas Fuzzy Evolutivos pouco explorados no contexto em questão.

3.5 Critérios de Avaliação

Conforme detalhado no Capítulo 2, existem métricas específicas que ajudam a mensurar o desempenho de modelos aplicados a problemas de Regressão e Classificação. Em se tratando de trabalhos cujo objetivo é a previsão do valor, a métrica mais utilizada é o *Root Mean Squared Error (RMSE)*, considerada pelo menos na metade dos trabalhos analisados. De forma complementar, muitos autores utilizam *Mean Absolute Error (MAE)*, *Mean Absolute Percentage Error (MAPE)*, *Mean Squared Error (MSE)* e *Non-Dimensional Error Index (NDEI)*. Em contrapartida, autores que trabalham com a previsão do sentido de variação do preço da criptomoeda, na grande maioria dos trabalhos, utilizam Acurácia como medida de desempenho. Outra métrica muito utilizada pelos autores são Revocação (*Recall*), Precisão e *F1-Score*.

Para os trabalhos que utilizam os modelos propostos como estratégia de investimentos existem ainda outras medidas de desempenho. [Borges e Neves \(2020\)](#) elenca algumas das métricas que podem ajudar a mensurar a qualidade dos modelos desenvolvidos para

este fim: *Return of Investment (ROI)*, *Maximum Drawdown (MDN)*, *Sharpe Ratio (SHR)*, *Shortino Ratio (SRR)*, *Percentage of Periods in Market (PPM)*, *Percentage of Profitable Positions (PPP)*, *Average Profit per Position (APPP)*, *Largest Percentual Gain (LPG)* e *Largest Percentual Loss (LPL)*.

3.6 Resumo dos Trabalhos Relacionados

Trabalhos que tratam do problema de previsão de valores em séries temporais relacionadas à criptomoedas possuem certa semelhança na sua estrutura, visto que algumas perguntas comuns geralmente precisam ser respondidas em relação à abordagem proposta pelos autores. Apesar de ser um tema bastante explorado, existem algumas questões para as quais ainda não há um consenso entre os autores, principalmente no que se refere à escolha do conjunto de variáveis/atributos de entrada bem como em relação à melhor técnica para a realização das previsões. Geralmente, em trabalhos que consideram múltiplos intervalos, os modelos que apresentam o melhor resultado se diferem entre os intervalos considerados (ou, pelo menos, as configurações destes), o que demonstra o iminente desafio na busca de um único modelo que apresente bom desempenho de maneira geral. Sabe-se, entretanto, que o modelo ARIMA é amplamente utilizado de forma prática por agentes do mercado financeiro. Porém, os modelos de Aprendizagem de Máquina também vêm apresentado bons resultados em pesquisas desta área e se mostram como uma boa alternativa aos métodos tradicionais.

Diante disso, as seções anteriores trouxeram um resumo sobre as características mais recorrentes entre os trabalhos analisados. De forma complementar, achou-se necessária a compilação dessas características, considerando os principais trabalhos analisados, com o entendimento de que esse resumo poderá servir como material de apoio na continuidade desta pesquisa ou em novos trabalhos sobre o assunto. Portanto, a Tabela 5 é o produto desta análise, e contém algumas das informações mais relevantes acerca dos 30 trabalhos considerados. Abaixo, descreve-se o tema de cada coluna presente nessa tabela:

- **Proposta:** elenca as técnicas (modelos de *Machine Learning*, Métodos de Inferência Estatística, etc.) utilizadas na abordagem;
- **Entrada (x_t):** lista as variáveis de entrada consideradas pelos autores na realização dos experimentos. As variáveis são classificadas conforme definido na Seção 3.2. Na coluna da esquerda são apresentadas variáveis relacionadas à Dados de Cotação (C), Dados Externos (E), Dados de Mercado (M) e Dados da Rede (R). Na segunda coluna são listados os Indicadores Técnicos (IT) aplicados sobre algumas ou todas as variáveis mencionadas na coluna anterior;
- **Saída (\hat{y}_t):** lista as variáveis de saída consideradas na realização dos experimentos, ou seja, aquilo que está sendo previsto;

- **Seleção:** método de seleção de variáveis (caso o autor tenha utilizado este recurso);
- **Horizonte:** horizontes de previsão considerados nos experimentos (minuto, fim do dia (EOD), semana, mês, etc.);
- **Moedas:** moedas consideradas nos experimentos;
- **Intervalos:** períodos para os quais os dados foram extraídos e, quando disponíveis, a quantidade de amostras e percentual de dados utilizados para treinamento (T) e validação (V);
- **Métricas:** medidas de desempenho consideradas na avaliação dos modelos propostos.

Tabela 5 – Resumo dos trabalhos relacionados.

Proposta	Entrada (x_t)	Indicadores Técnicos	Saída (\hat{y}_t)	Seleção	Horizonte	Moedas	Intervalos	Métricas
	Valores Brutos (C/E/M/R)							
		C: Daily Returns, High, Low						
		M: Active Addresses, Adjusted On-Chain Volume, Market Cap., Median Transaction Value, Number of Transactions, On-Chain Transaction Volume, Payment Count, Volume of Trades						
		R: Block Count, Block Size, Generated Coins, Median Fees, Mining Difficulty, Transaction Fees						
		E: Daily Dummies						
RF		RR	C: Daily Returns	-	1 dia	BTC	15/08/2015 a 03/03/2019 1305 amostras	Acurácia RMSE MAE MAPE Theil's U ²
SVM		PV				ETH LTC	75% T / 25% V	
SVR								
Sebastião e Godinho (2021)								
LSTM				PCC	1 dia	BTC	2016 a 2020 1400 amostras (apx.) 10/50/90%T / 90/50/10%V	MAE
Jagannath et al. (2021)								
		M: Active Addresses, Existing Address, Trans. Rate, Unspent Trans., Volume of Trades						
		R: Average Transactions per Block, Balances of Addresses, Block Count, Block Height, Block Size, Exchange Balance, Exchange Deposits, Exchange Inflow/Outflow, Exchange Withdrawals, Hash Rate, Miners Balance, Miners Inflow Volume, Miners Outflow Volume, Miners to Exchanges, Mining Difficulty, Total Miners Revenue						
			C: Close					
Derbentsev et al. (2021)								
RF							01/01/2015 a 31/12/2019 1826 amostras	MAPE RMSE
SGBM		SMA	C: Close	-	1 dia	BTC ETH XRP	07/08/2015 a 31/12/2019 1608 amostras 80%T / 20%V	
			C: Close					
Jana, Ghosh e Das (2021)								
SVR			C: Close	-	1 dia	BTC	10/01/2013 a 23/02/2019 1959 amostras 80%T / 20%V	DA MSE NSD NSE

Proposta	Valores Brutos (C/E/M/R)	Entrada (x_t)	Indicadores Técnicos	Saída (\hat{y}_t)	Seleção	Horizonte	Moedas	Intervalos	Métricas
GARCH MS-GARCH	C: Close		LNR	C: Daily Returns	-	1 dia	BTC ETH LTC XRP	17/08/2017 a 17/09/2020 1127 amostras 32% V 13/12/2017 a 17/09/2020 1009 amostras 36% V 04/05/2018 a 17/09/2020 867 amostras 42% V	CT SF
ANN SVM SANIN SVM	C: Close		SMA RSI WMA SD σ^2 TME RoC		RDF VIF PCR	EOD 7 dias 30 dias 90 dias	BTC	19/08/2013 a 19/07/2016 1066 amostras 80% T / 20% V 01/04/2013 a 01/04/2017 1461 amostras 80% T / 20% V 01/04/2013 a 31/12/2019 2465 amostras 80% T / 20% V	Especificidade Precisão Acurácia Recall F1-Score AUC ROC MAE MAPE RMSE
CNN CNN-LSTM MLP RBFNN	C: Close, High, Low		A/D CCI LWI Momentum MACD RSI SMA SD SK WMA			1 min	BTC ETH LTC DASH XMR XRP	01/07/2018 a 30/06/2019 (1m) 525600 amostras 85% T / 15% V	Acurácia

Trucios e Taylor (2020)

Mudassir et al. (2020)

Alonso-Monsalve et al. (2020)

M: Active Addresses, Average Transaction Value, Median Transaction Value, Number of Transactions Sent, Sent from Addresses, Top 100 / Total
R: Average Transaction Fee, Block Size, Block Time, Fee-to-Reward Ratio, Hash Rate, Median Transaction Fee, Mining Difficulty, Mining Profitability

Proposta	Valores Brutos (C/E/M/R)	Entrada (x_t)	Indicadores Técnicos	Saída (\hat{y}_t)	Seleção	Horizonte	Moedas	Intervalos	Métricas
LR LDA RF XGB QDA SVM LSTM	M: Confirmed Transactions per Day, Market Cap., Number of Transactions, Trades per Minute R: Block Size, Block Time, Hash Rate, Mempool Size, Mempool Transaction Count, Mining Difficulty, Transaction Fees USD E: Baidu Media, Closing Gold Spot Price, Google Trend Search Index	Chen, Lie Sun (2020)	-	C: Close	Boruta	5 min EOD	BTC	02/02/2017 a 01/02/2019 (EOD) 730 amostras 80% T / 20% V 17/07/2017 a 17/01/2018 (5m) 50000 amostras 75% T / 25% V	Precisão Acurácia Recall F1-Score
LR MLP RF SVM	C: Close Direction, High Direction, Low Direction, Price Direction	Pabuccu, Ongan e Organ (2020)	A/D LWI MACD Momentum RSI SD SK SMA WMA	C: Close	-	EOD	BTC	2008 a 2019 95%T / 5%V	Acurácia Precisão Recall F Statistics MAE RAE RMSE
VMD k-NN SVR FFNN GLM Cubist	C: High, Low, Open	Silva et al. (2020)	-	C: Close	PCA BoxCox CORR	EOD 2 dias 3 dias	BTC	18/07/2010 a 21/02/2016 2045 amostras 70% T / 30% V	RRMSE sMAPE APE
CNN LSTM GRU	C: Close, High, Low, Open M: Volume of Trades	Zoumppekas, Houstis e Vavalis (2020)	WMA	C: Close	-	5 min	ETH	08/09/2015 a 28/05/2018	MDA RMSE nRMSE MAE nMAE

Proposta	Valores Brutos (C/E/M/R)	Entrada (x_t)	Indicadores Técnicos	Saída (\hat{y}_t)	Seleção	Horizonte	Moedas	Intervalos	Métricas
ePL-I	C: High, Low		-	C: High, Low	-	EOD	BTC ETH	01/01/2016 a 31/12/2019 1461 amostras 75%T / 25%V	DA NSD R^C R^E
<p><i>Maciel, Ballini e Gomide (2020)</i></p>									
<p><i>Ji, Kim e Im (2019)</i></p>									
CNN		M: Active Addresses, Confirmed Transactions per Day, Existing Address, Market Cap., Market Price, Non Popular Transactions, Number of Transactions, On-Chain Transaction Volume, Total Transactions Output, Unspent Transactions, Volume of Trades							Acurácia
CNN-LSTM									Especificidade
DNN		R: Average Transactions per Block, Block Size, Blockchain Size, Hash Rate, Median Fees, Median Fees (%)							F1-Score
DRN				C: Close	SCC	1 dia	BTC	29/11/2011 a 31/12/2018 2590 amostras 80%T / 20%V	Precisão Recall
LSTM									
RNN									
<p><i>Mempool Transaction Count, Miners Revenue, Mining Difficulty, No Longer than 100 Transactions, Total Units, Transaction Fees BTC, Transaction Fees USD, Transaction Median Time, Transactions per second</i></p>									
<p><i>Atsalakis et al. (2019)</i></p>									
ANFIS	C: Price		SMA	C: Price	-	MOD	BTC ETH LTC XRP	13/09/2011 a 12/10/2017 2201 amostras 97%T / 3%V	Acurácia MAE MSE RMSE
<p><i>Derbentsev et al. (2019)</i></p>									
BART	C: Close		LNR	C: Close	-	5 dias 10 dias 14 dias 21 dias	BTC ETH XRP	01/01/2017 a 01/03/2019 790 amostras 80%T / 20%V	RMSE
<p><i>Desev, Kabaivanov e Desev (2019)</i></p>									

Proposta	Valores Brutos (C/E/M/R)	Entrada (x_t)	Indicadores Técnicos	Saída (\hat{y}_t)	Seleção	Horizonte	Moedas	Intervalos	Métricas
ARIMA	C: Price		-	C: Current Price	-	1 hora	ETH HST NEO	01/01/2018 a 31/12/2018 8590 observações 10%T / 90%V	MAE MAPE RMSE
MLP RNN K-Means	C: Close, High, Low, Open, Price Direction M: Number of Transactions, Volume of Trades R: Hash Rate, Median Fees, Mining Difficulty, Transaction Fees USD E: Closing Crude Price, Closing DAX, Closing Gold Spot Price, Closing Nasdaq, Closing S&P 500		SMA	C: Close, High, Low	Corr Relief InfoGain PCA CFS	1 dia	BTC	19/08/2013 a 19/07/2016 1066 amostras 80% T / 20% V 01/04/2013 a 01/04/2017 1322 amostras 75% T / 25% V	Sensibilidade Especificidade Acurácia RMSE MAE MAPE
FFNN	C: Close E: CBOE Volatility Index		SMA	C: Close	-	EOD	BTC	22/04/2011 a 10/01/2013 11/01/2013 a 30/09/2014 01/10/2014 a 15/06/2016 16/06/2016 a 02/03/2018 1781 amostras 85%T / 15%V	Acurácia MSPE
LSTM AR	C: Close M: Volume of Trades		-	C: Close	-	EOD	BTC	01/01/2018 a 28/07/2018 208 observações 66%T / 34%V	MAE MAPE MSE RMSE
Fuzzy Transform Fourier Transform FIS	C: Open		-	C: Open	-	EOD	BTC ETH LTC	08/07/2015 a 24/12/2018 100%V	Acurácia

Mallqui e Fernandes (2019)

Adcock e Gradojevic (2019)

Chih-Hung et al. (2019)

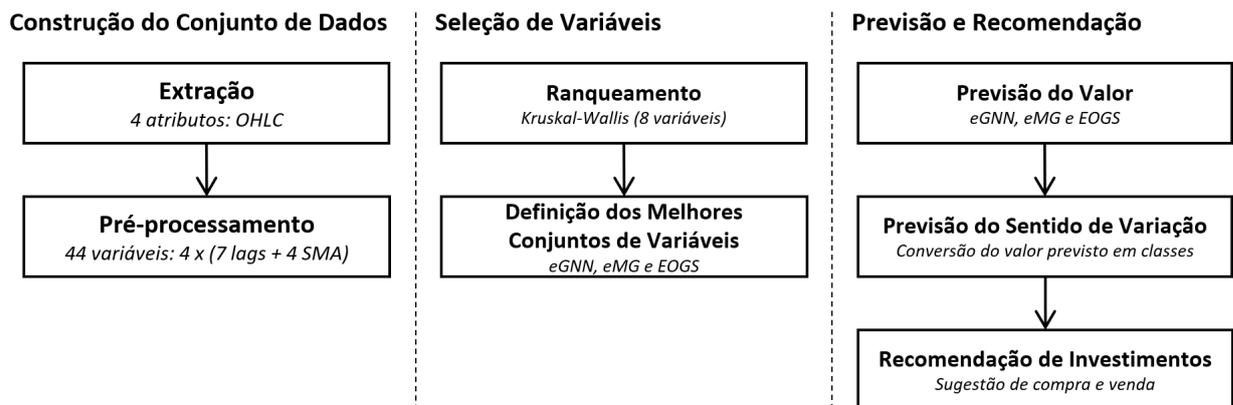
Poljedica, Petrovic e Milosevic (2019)

Angelini e Ermili (2018)

4 Metodologia

De maneira geral, a abordagem proposta pode ser dividida em três etapas: Construção do Conjunto de Dados; Seleção de Variáveis; e Previsão e Recomendação. Na primeira etapa é realizada a Extração e o Pré-processamento do conjunto de dados, o qual é utilizado como entrada para a etapa subsequente. Em seguida, na segunda etapa, os dados extraídos e pré-processados são submetidos a um procedimento de Seleção de Variáveis, onde é realizado um Ranqueamento das variáveis por ordem de relevância, bem como a Definição do Melhor Conjunto de Variáveis para cada modelo evolutivo. Finalmente, após essa definição, a etapa de Previsão e Recomendação poderá ser iniciada¹, sendo composta por três fases: (I) Previsão do Valor; (II) Previsão do Sentido de Variação; e (III) Recomendação de Investimentos. A Figura 6 apresenta uma visão expandida do funcionamento de cada etapa mencionada, com suas respectivas fases, as quais serão detalhadas nas seções subsequentes.

Figura 6 – Visão macro da abordagem proposta.



4.1 Construção do Conjunto de Dados

Nesta seção, detalha-se o procedimento para a construção do conjunto de dados a ser utilizado como entrada para a execução dos modelos evolutivos.

¹ Considera-se que a solução será executada diariamente no momento de abertura do mercado. O momento de abertura e fechamento do mercado neste caso equivale a 00:00 e 23:59 UTC, respectivamente. Portanto, a previsão será fornecida com cerca de 24 horas de antecedência.

4.1.1 Extração

Sabe-se que o mercado de criptomoedas permanece aberto 24 horas por dia, 7 dias por semana. A plataforma CoinMarketCap² disponibiliza uma API gratuita que agrega dados diários de mercado para as corretoras consideradas mais relevantes, agrupando esses dados em intervalos de 24 horas, considerando o fuso horário UTC. Por meio dela, é possível obter dados históricos diários de mais de 1000 criptomoedas, como valor de abertura e fechamento do mercado, maior e menor valores alcançados, volume transacionado, capitalização de mercado, dentre outros. Portanto, os dados históricos necessários para a execução da abordagem proposta serão extraídos dessa API.

Os atributos considerados, para os quais os dados serão efetivamente extraídos, consistem nos valores diários de abertura (*Open*) e fechamento (*Close*), além do maior (*High*) e menor (*Low*) valor alcançados pela criptomoeda no dia. Na continuação do texto, esse conjunto de atributos será referenciado como OHLC (*Open, High, Low e Close*). Os dados extraídos são então normalizados entre 0 e 1, uma condição necessária para o funcionamento de alguns dos modelos evolutivos utilizados.

4.1.2 Pré-processamento

Após a Extração, é necessário que os dados históricos diários obtidos para os 4 atributos mencionados sejam preparados, o que ocorre na fase de Pré-processamento. Nesta fase parte-se da premissa que, na previsão do valor da criptomoeda para o final do dia ($close_t$), serão utilizados dados brutos de até 7 defasagens para cada atributo, totalizando 28 (7 x 4) variáveis³. Adicionalmente, novas variáveis são criadas a partir do cálculo de Média Móvel Simples (SMA_q), de diferentes ordens, para cada atributo OHLC mencionado, o que é dado por:

$$SMA_q = \frac{1}{q - i + 1} \sum_{i=1}^q x_{t-i}, \quad (21)$$

em que t é o instante inicial ($t = 0$), x_{t-i} é o valor de um dado atributo para o instante $t - i$ e q é a ordem da média móvel. Para cada atributo foram considerados dados de até 4 semanas anteriores ao instante atual⁵. Considerando 4 médias móveis para cada um dos 4 atributos, chega-se ao número de 16 variáveis de médias móveis. Conclui-se, portanto,

² <https://www.coinmarketcap.com>

³ Entende-se que dados diários para a semana anterior ao instante atual guardem informações que tendem a influenciar de forma mais brusca no preço da criptomoeda um passo adiante.

⁴ Para o atributo *Open*, tem-se $i = 0$.

⁵ Dessa maneira, os dados dos últimos 28 dias também estão sendo utilizados, de forma a apresentar ao modelo uma tendência de mais longo prazo.

que cada observação é composta por 44 ($28 + 16$) variáveis de entrada, todas baseadas nos atributos *OHLC*. A Tabela 6 apresenta, para cada atributo, os valores considerados na formação das variáveis no que tange às defasagens ($t - i$) e ordens (q) de médias móveis.

Tabela 6 – Conjunto de atributos e valores considerados na geração de variáveis por defasagem e médias móveis simples.

Atributo	Defasagens	Médias Móveis
<i>Open</i>	$i = \{0, 1, 2, 3, 4, 5, 6\}$	$q = \{6, 13, 20, 27\}$
<i>High</i>	$i = \{1, 2, 3, 4, 5, 6, 7\}$	$q = \{7, 14, 21, 28\}$
<i>Low</i>	$i = \{1, 2, 3, 4, 5, 6, 7\}$	$q = \{7, 14, 21, 28\}$
<i>Close</i>	$i = \{1, 2, 3, 4, 5, 6, 7\}$	$q = \{7, 14, 21, 28\}$

4.1.3 Informações Complementares

Neste ponto, faz-se necessária uma distinção entre o que será referenciado, na continuidade do texto, como Período de Seleção de Variáveis e Intervalo de Interesse. O Período de Seleção de Variáveis é composto pelas 360 observações anteriores ao instante t_0 , isto é $[t_{-360}, t_{-1}]$, onde t_0 representa o primeiro instante para o qual pretende-se realizar previsões⁶. Em contrapartida, Intervalo de Interesse é o nome atribuído ao período para o qual deseja-se efetivamente realizar previsões, isto é, $[t_0, t_{n-1}]$, onde n representa o número total de observações existentes no Intervalo de Interesse.

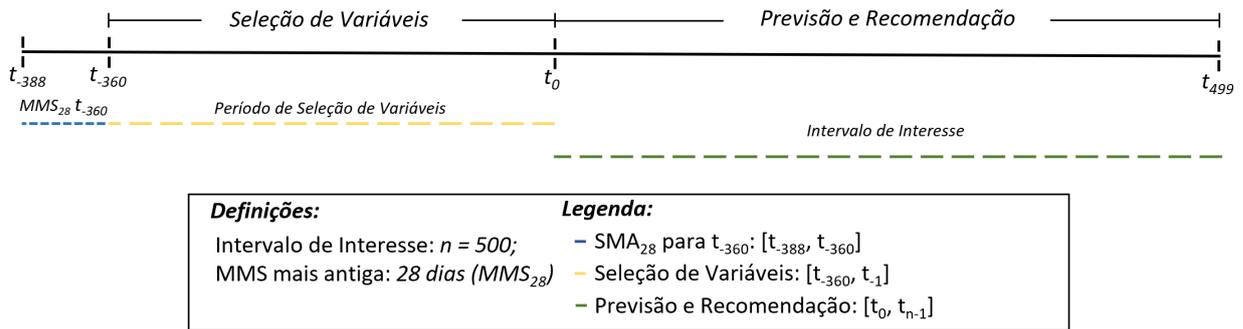
Para melhor compreensão, a Figura 7 apresenta uma ilustração que retrata com clareza a distribuição dos dados extraídos e pré-processados na etapa de Construção do Conjunto de Dados. Tomando-se como exemplo o Intervalo de Interesse $[t_0, t_{499}]$, composto por 500 observações ($n = 500$). Diante do cenário exemplificado, o processo de Seleção de Variáveis se dará no intervalo $[t_{-360}, t_{-1}]$. Para isso, será necessário extrair dados históricos da criptomoeda no período $[t_{-388}, t_{-1}]$. Na etapa de pré-processamento, os dados referentes ao intervalo $[t_{-388}, t_{-361}]$ servirão exclusivamente para o cálculo das Médias Móveis Simples (SMA) que serão calculadas para os últimos 7, 14, 21 e 28 dias. Portanto, ao final da etapa de pré-processamento dos dados, tem-se observações compostas por 44 variáveis para os últimos 360 dias, isto é, $[t_{-360}, t_{-1}]$.

4.2 Seleção de Variáveis

Conforme apresentado, a abordagem proposta incorpora modelos que utilizam dados passados para a realização de previsões. Originalmente, os Sistemas *Fuzzy* Evolutivos não necessitam de uma etapa de treinamento visto que sua estrutura e parâmetros são ajustados em tempo real. À medida que novos dados são apresentados ao modelo, o

⁶ Utilizar um período maior que um ano poderia ser prejudicial, fazendo uma seleção de variáveis sobre dados muito defasados, principalmente considerando a alta volatilidade da criptomoeda.

Figura 7 – Período de seleção de variáveis e intervalo de interesse.



sistema é inteligente o suficiente para adaptar sua estrutura e parâmetros a partir do cálculo do erro entre o valor previsto e o valor realizado. Entretanto, diante do universo considerado, composto por 44 variáveis, entende-se como necessária a implementação de um mecanismo que reduza o número de variáveis apresentadas aos modelos evolutivos. A seleção de um conjunto reduzido de variáveis é realizada, principalmente, para evitar um problema conhecido como Maldição da Dimensionalidade (*Curse of Dimensionality*), um termo introduzido em Bellman (1966) utilizado para descrever o problema causado pelo aumento exponencial no volume de dados associado à inclusão de dimensões extras ao espaço Euclidiano. Na prática este problema implica que, para um dado conjunto de amostras, existe um número máximo de variáveis/características a partir do qual o desempenho do modelo possivelmente começará a degradar. Portanto, nesta seção, são apresentadas as duas fases que compõem a etapa de Seleção de Variáveis: Ranqueamento e Definição dos Melhores Conjuntos de Variáveis.

4.2.1 Ranqueamento

Na fase de Ranqueamento, é realizada uma ordenação do conjunto de variáveis quanto à sua relevância. Para a determinação da relevância de cada variável, optou-se por utilizar o método não-paramétrico de *Kruskal-Wallis* (KRUSKAL W. H., 1952). Dessa forma, o método é executado uma única vez sobre as 44 variáveis pré-processadas (V_1, V_2, \dots, V_{44}), para as 360 observações que compõem o Período de Seleção de Variáveis. Ao final desta etapa, o conjunto de variáveis estará ordenado, de forma decrescente, priorizando as variáveis mais relevantes. Após a ordenação do conjunto de variáveis, as 8 variáveis consideradas mais relevantes pelo método de *Kruskal-Wallis* são selecionadas para a continuidade do procedimento de seleção. Portanto, a partir deste momento a solução proposta passa a considerar apenas as 8 variáveis mais relevantes, descartando as demais.

4.2.2 Definição dos Melhores Conjuntos de Variáveis

Sabe-se que cada modelo pode se comportar melhor com um conjunto específico de variáveis, de dimensões distintas. Dessa forma, na segunda e última fase do procedimento, as 8 variáveis consideradas mais relevantes, selecionadas na fase anterior, são combinadas com o objetivo de descobrir o melhor conjunto de variáveis para cada modelo evolutivo. Neste passo, os modelos validam individualmente todas as formações possíveis de conjuntos, compostos por 1..4 variáveis. O cálculo do número máximo de combinações é dado por:

$$\sum_{i=1}^{\phi} \frac{\chi!}{i!(\chi-i)!}, \quad (22)$$

onde ϕ representa o tamanho máximo dos conjuntos e χ é o número de elementos disponíveis. Neste caso, como, por definição, o tamanho máximo dos conjuntos de variáveis é igual 4 ($\phi = 4$) e o número de variáveis disponíveis após o Ranqueamento é igual a 8 ($\chi = 8$), para cada modelo evolutivo serão testadas 375 possíveis formações de conjuntos de variáveis⁷. Este passo é realizado sob o mesmo conjunto de dados considerado na primeira fase, isto é, o Período de Seleção de Variáveis. No entanto, desta vez, utiliza-se apenas os dados computados para as 8 variáveis selecionadas no Ranqueamento. Após as execuções de cada modelo evolutivo, com cada formação de conjunto de variáveis possível, o conjunto que proporcionou ao modelo o menor erro de previsão (RMSE) será considerado na etapa de Previsão e Recomendação.

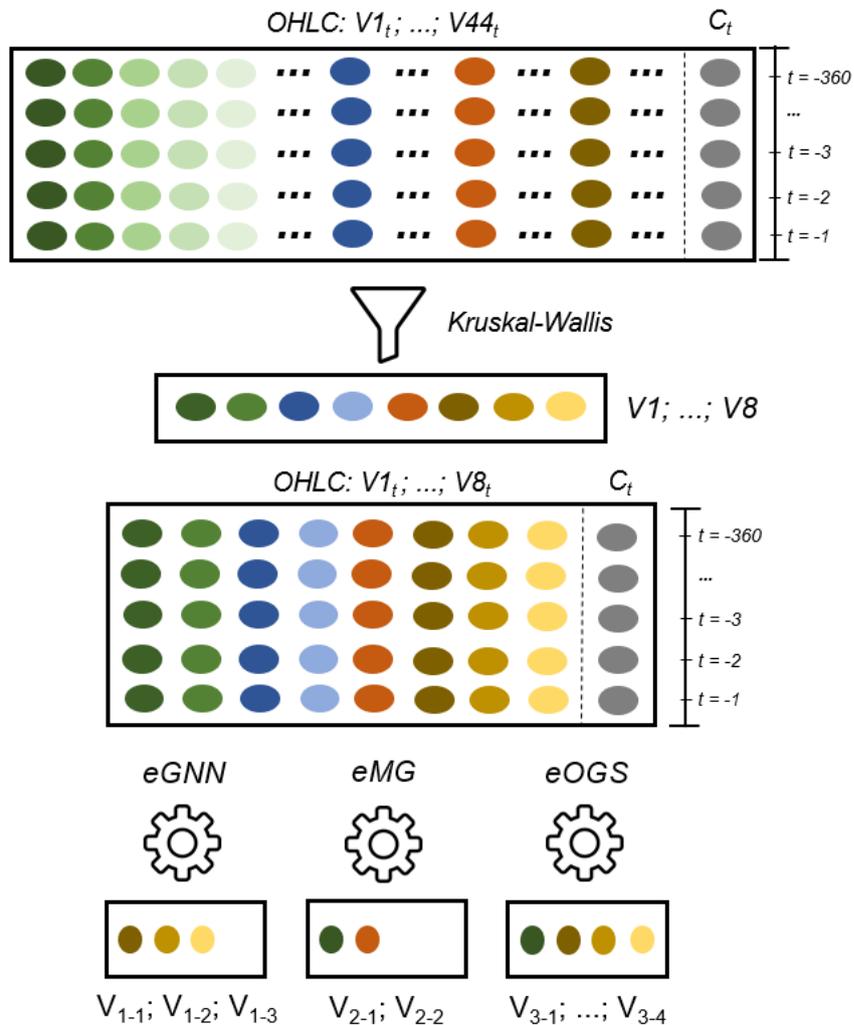
Para melhor compreensão, a Figura 8 traz uma ilustração do procedimento de Seleção de Variáveis na sua totalidade. Nesta figura, é possível visualizar graficamente algumas das informações descritas acerca deste procedimento, tais como:

- os 4 atributos OHLC, representados pelas cores Verde, Azul, Laranja e Marrom. Para cada atributo, supõe-se a existência de 11 tonalidades de cores diferentes, que representam as 7 defasagens calculadas para o referido atributo no instante t , além das 4 Médias Móveis Simples (SMA), totalizando 44 (11 x 4) variáveis;
- o atributo $Close_t$, em cinza, representa o valor esperado para cada observação, isto é, o valor de fechamento da criptomoeda para o dia atual;
- após o Ranqueamento, realizado pelo método de *Kruskal-Wallis*, as 8 variáveis mais relevantes são selecionadas e as demais são descartadas;
- na fase de Definição dos Melhores Conjuntos de Variáveis, são formados grupos compostos por até 4 das 8 variáveis restantes, para os mesmos dados correspondentes utilizados na primeira fase, o que resultará em 375 execuções por modelo,

⁷ Percebe-se que o número de execuções necessárias cresce exponencialmente, a depender dos valores atribuídos aos parâmetros χ e ϕ , o que motivou a escolha de $\chi = 8$ e $\phi = 4$.

uma com cada formação. Para cada modelo, o conjunto de variáveis responsável pela execução que apresentar o menor erro é selecionado para a etapa de Previsão e Recomendação.

Figura 8 – Representação da etapa de seleção de variáveis.



4.3 Previsão e Recomendação

Após as etapas de preparação do sistema, com os três grupos de variáveis definidos (um para cada Sistema *Fuzzy* Evolutivo), a solução projetada estará pronta para uso. Nesta seção, detalha-se a etapa de Previsão e Recomendação, que é a aplicação efetiva do sistema ao Intervalo de Interesse, executada de forma iterativa para cada observação, a partir do detalhamento de cada fase que a compõe, a saber: Previsão do valor, Previsão do Sentido de Variação e Recomendação de Investimentos.

4.3.1 Previsão do Valor

Na fase de Previsão do Valor, cada um dos três Sistemas *Fuzzy* Evolutivos será executado individualmente, utilizando como entrada dados para o respectivo conjunto de variáveis selecionado na fase anterior. Dessa forma, caso tenha sido definido no procedimento de Seleção de Variáveis que o melhor conjunto de variáveis para o modelo eOGS equivale a $open_t, close_{t-1}, close_{t-2}, close_{t-3}$, por exemplo, este modelo receberá como entrada, ao longo do Intervalo de Interesse, os valores computados para as variáveis mencionadas, e assim será capaz de realizar a previsão do preço de fechamento da criptomoeda para o dia atual.

4.3.2 Previsão do Sentido de Variação

Na previsão do sentido de variação do preço da criptomoeda, são utilizadas as saídas dos modelos evolutivos. Porém, os valores fornecidos pelos modelos precisam ser interpretados, visto que os modelos considerados foram criados com o propósito de regressão, ou seja, eles sempre fornecerão uma saída discreta. Deste modo, faz-se necessária uma conversão da saída retornada por cada modelo para um valor binário, que indicará o sentido de variação previsto, isto é, (0) redução ou (1) aumento de valor. A conversão é realizada conforme:

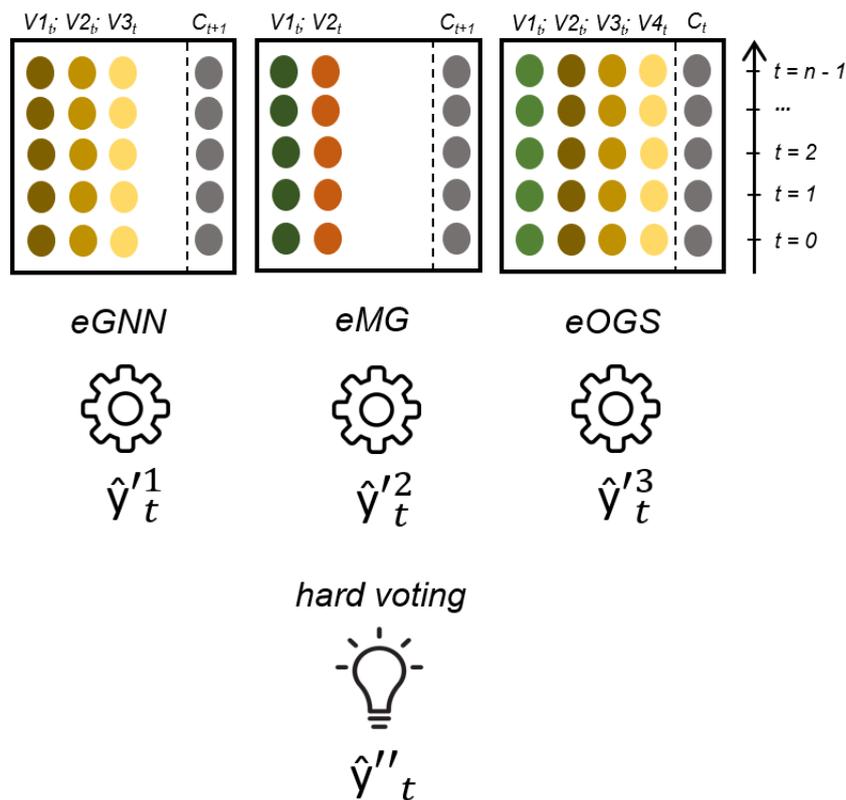
$$\hat{y}_t'' = \begin{cases} 0, & \text{se } \hat{close}_t < close_{t-1}, \\ 1, & \text{senão.} \end{cases} \quad (23)$$

Neste caso, \hat{close}_t representa o valor de fechamento previsto pelo modelo para o dia atual; $close_{t-1}$ é o valor de fechamento registrado para o dia anterior; e \hat{y}_t'' é o sentido de variação previsto.

Além da resposta, \hat{y}_t'' , fornecida por cada Sistema *Fuzzy* Evolutivo, adicionalmente, é proposta a formação de um Comitê, o que remete ao conceito de *Stacked Generalization* (SG), uma técnica de combinação de modelos proposta em sistemas baseados em Aprendizagem por Agrupamento (*Ensemble Learning* ou EL). Na aplicação dessa técnica, os três Sistemas *Fuzzy* Evolutivos utilizados na fase de Previsão do Valor podem ser considerados como *weak learners*, fazendo-se necessária a definição de um mecanismo de votação que realize uma agregação da resposta fornecida por cada um deles. Na proposta apresentada, optou-se por utilizar uma abordagem *hard voting*. Como foi utilizado um número ímpar de modelos evolutivos, registra-se vitória por maioria simples. Neste caso, a classe que recebeu a maior quantidade de votos será apresentada como o resultado final fornecido pelo Comitê para a observação em questão.

A Figura 9 ilustra o funcionamento proposto para a fase de Previsão do Sentido de Variação. Conforme pode ser visualizado e, novamente, para efeito de exemplificação, caso a etapa de Seleção de Variáveis, executada sob as últimas 360 observações, retorne que o conjunto $open_t, close_{t-1}, close_{t-2}, close_{t-3}$ representa as variáveis mais relevantes para o modelo eOGS, as próximas iterações irão fornecer para este modelo os valores registrados para as quatro variáveis mencionadas. Conforme ilustrado, é possível que os demais modelos recebam como entrada os valores computados para variáveis diferentes, de acordo com o melhor conjunto definido para o respectivo modelo na etapa de Seleção de Variáveis. Como exemplo, é retratado na figura que os modelos eGNN e eMG utilizam na etapa de Previsão e Recomendação três e duas variáveis, respectivamente. Após o processamento realizado pelos *weak learners* e a conversão da saída, realizada conforme a Equação 23, tem-se três sentidos de variação previstos: $\{\hat{y}_t^1, \hat{y}_t^2, \hat{y}_t^3\}$. Os três valores calculados são então submetidos ao *strong learner* que classifica em \hat{y}_t'' , por maioria simples, se o comportamento previsto do valor da criptomoeda para o fim do dia, em relação ao dia anterior, será de redução (0) ou aumento (1) de valor.

Figura 9 – Representação da fase de previsão do sentido de variação do preço de fechamento de criptomoedas.



4.3.3 Recomendação de Investimentos

Entende-se que a previsão do sentido de variação do preço de uma criptomoeda pode servir de auxílio ao investidor na tomada de decisão acerca de seus investimentos.

Neste sentido, Ji, Kim e Im (2019) propõem em seu trabalho uma estratégia que consiste em orientar ao investidor que utilize todos os seus fundos na compra de criptomoedas, caso seja previsto que o valor no dia seguinte será superior ou igual ao registrado para o dia atual; de modo contrário, caso a previsão retornada pelo modelo seja de queda, o investidor deverá vender todas as criptomoedas que possui a fim de evitar prejuízos com a provável redução de valor. De forma geral, entende-se que a estratégia proposta é agressiva, visto que todos os fundos seriam deslocados de acordo com a previsão fornecida, o que poderia proporcionar ganhos consideráveis em caso de acerto mas, em contrapartida, poderia gerar prejuízo elevado caso o modelo forneça uma previsão incorreta.

Diante disso, no Algoritmo 1 é proposta uma generalização da abordagem de Ji, Kim e Im (2019), onde o valor aplicado em cada transação de compra ou venda de moedas seria apenas uma proporção do patrimônio do investidor. Isto é importante haja vista que existem diferentes perfis de investimento, isto é, existem investidores mais agressivos que se importam menos com uma perda de curto prazo quando há possibilidade de ganhos elevados; mas também existem investidores que optam por ganhos mais baixos, onde há menor risco de perda de patrimônio.

Algoritmo 1: Generalização da Estratégia de Investimento

Entrada: *dinheiro*; *criptomoeda*; *limiar*;
Saída: *dinheiro*; *criptomoeda*;

```

1 início
2   se  $previsao_t \geq real_{t-1}$  então
3     se  $dinheiro > 0$  então
4        $saida \leftarrow dinheiro \times limiar$ ;
5        $dinheiro \leftarrow dinheiro - saida$ ;
6        $criptomoeda \leftarrow criptomoeda + (saida/real_t)$ ;
7     fim
8   senão
9     se  $criptomoeda > 0$  então
10       $saida \leftarrow criptomoeda \times limiar$ ;
11       $criptomoeda \leftarrow criptomoeda - saida$ ;
12       $dinheiro \leftarrow dinheiro + (saida \times real_t)$ ;
13    fim
14  fim
15 fim
```

No algoritmo apresentado os parâmetros *dinheiro* e *criptomoeda* indicam, respectivamente, o patrimônio atual do investidor em moeda fiduciária e na criptomoeda para a qual a estratégia será aplicada. Além disso, o parâmetro *limiar* representa a proporção do patrimônio a ser considerada em uma transação. Neste caso, por exemplo, se o valor atribuído for igual a 0,1, apenas 10% do patrimônio do investidor em *dinheiro* será utilizado na compra de criptomoedas e, de modo semelhante, apenas 10% do patrimônio em

criptomoeda será considerado em uma situação de venda. Assim como em [Ji, Kim e Im \(2019\)](#), os eventuais custos de transação foram desconsiderados por motivo de simplicidade e também pela possibilidade de negociação direta com a rede, sem intermediários (corretoras).

5 Experimentos Computacionais

Neste capítulo são detalhados os experimentos realizados bem como os resultados alcançados. Acerca do código necessário para a realização dos experimentos, as implementações dos três Sistemas *Fuzzy* Evolutivos (eGNN, eMG e eOGS), escritas na linguagem Matlab, foram gentilmente disponibilizadas pelos seus respectivos autores. Para a fase de Ranqueamento, a implementação nativa do método de Kruskal-Wallis, existente no Matlab, também foi utilizada. Além disso, foi implementada uma aplicação, na linguagem C#, que extrai dados históricos de criptomoedas, e materializa a solução apresentada no Capítulo 4 para dados históricos, a partir da informação do Intervalo de Interesse e da criptomoeda desejada.

O detalhamento dos experimentos é realizado neste capítulo da seguinte forma: na Seção 5.1 são apresentados os valores atribuídos aos parâmetros necessários para a execução de cada Sistema *Fuzzy* Evolutivo; em seguida, na Seção 5.2, são descritos os 10 conjuntos de dados utilizados na realização dos experimentos; posteriormente, na Seção 5.3, discorre-se acerca dos resultados obtidos pelos Sistemas *Fuzzy* Evolutivos na tarefa de Previsão do Valor; na Seção 5.4, são descritos os resultados alcançados em termos de Previsão do Sentido de Variação; e, finalmente, na Seção 5.5, são apresentados os resultados referentes à Recomendação de Investimentos.

5.1 Definição de Parâmetros

Os Sistemas *Fuzzy* Evolutivos utilizados requerem a definição prévia de alguns parâmetros intrínsecos ao respectivo modelo, conforme explicado individualmente de forma detalhada na Seção 2.2.2.2. Acerca dos parâmetros necessários, os valores atribuídos foram definidos de acordo com os limites recomendados pelos autores. Sendo assim, os valores considerados são:

- eGNN: $\rho = 0,4$; $h_r = 105$; $\eta = 2,0$;
- eMG: $\alpha = 0,01$; $\lambda = 0,05$; $\omega = 350$; $\sum_{init} = 0,1$;
- eOGS: $\alpha = 0,25$; $\omega = 0,01$; $\nu = 100$; $\psi = 2,0$

Sabe-se da possibilidade de realizar um ajuste fino de parâmetros, o que poderia melhorar o desempenho individual dos modelos. Entretanto, neste trabalho foi estabelecido o objetivo de que seja utilizado um conjunto de parâmetros comum a todos os intervalos, o que possivelmente tornaria a solução mais genérica e aplicável a diferentes intervalos. Portanto, os mesmos valores foram utilizados na execução realizada para cada conjunto de dados.

5.2 Conjuntos de Dados

Na realização dos experimentos foram considerados 10 conjuntos de dados, para duas criptomoedas, sendo 6 para o Bitcoin (BTC) e 4 para o Ether (ETH), as duas criptomoedas mais relevantes e com a maior capitalização de mercado ([COINMARKETCAP, 2021](#)). Dentre os conjuntos utilizados, dois deles (ambos relacionados ao Bitcoin) estão presentes em trabalhos da literatura e, por isso, foram utilizados para fins de comparação. Os demais conjuntos, 6 para o Bitcoin e 4 para o Ether, são mais recentes, e abrangem diferentes intervalos ao longo dos anos de 2019 e 2020.

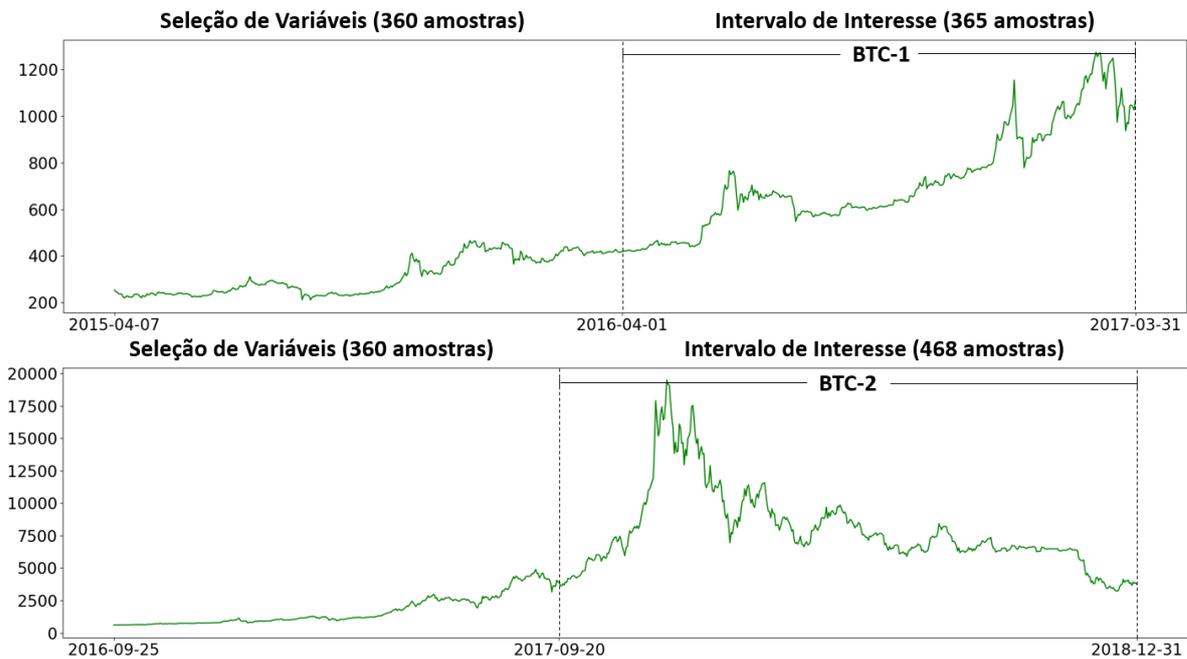
Em seu trabalho, [Mallqui e Fernandes \(2019\)](#) realizaram experimentos no período 01/04/2016 à 31/03/2017, avaliando os modelos utilizados no que se refere à sua capacidade de prever o valor do Bitcoin para o dia seguinte, bem como na previsão do sentido de variação do preço para a mesma criptomoeda. Para estas tarefas, os autores propõem a utilização de 5 modelos, a saber: *Multilayer Perceptron* (MLP); *Support Vector Machine* (SVM); e três combinações de modelos, as quais dá-se o nome de *Ensemble A*, *B* e *C*. Para avaliação dos resultados, os autores realizam uma comparação com dois tipos de ANN, *Recurrent Neural Network* (RNN) e *Long Short-Term Memory* (LSTM), além do modelo ARIMA.

Por sua vez, [Ji, Kim e Im \(2019\)](#) realizaram experimentos no intervalo 20/09/2017 à 31/12/2018, no que se refere à previsão do sentido de variação do preço do Bitcoin e também quanto ao uso dessa previsão em uma estratégia de recomendação de investimentos. Em sua proposta, o autor realiza uma comparação entre vários métodos de Deep Learning, tais como: Deep Neural Network (DNN), Long Short-Term Neural Network (LSTM), Convolutional Neural Network (CNN) e Deep Residual Neural Network. Na avaliação dos resultados, os autores utilizam como linha de base os métodos *Support Vector Machine* (SVM) e *Random Walk*. A Figura 10 apresenta a evolução do preço de fechamento do Bitcoin nos dois intervalos utilizados nos trabalhos de [Mallqui e Fernandes \(2019\)](#) e [Ji, Kim e Im \(2019\)](#), os quais, a partir deste momento, serão referenciados no texto como BTC-1 e BTC-2, respectivamente.

Além dos conjuntos BTC-1 e BTC-2, quatro novos conjuntos foram selecionados para a realização de experimentos com o Bitcoin. O principal critério considerado é que fossem selecionados dois períodos anteriores à 11/03/2020, data em que a OMS (Organização Mundial de Saúde) reconheceu como pandêmico o surto de Covid-19 iniciado na China, no final do ano anterior. Sabe-se que nessa data foi registrada a maior queda de preço da criptomoeda, em valores absolutos, até aquele momento, quando houve uma redução de aproximadamente 37% no seu valor ([COINMARKETCAP, 2021](#)). Portanto, para este cenário, foram selecionados os intervalos 24/01/2019 à 22/07/2019 e 24/01/2019 à 23/01/2020.

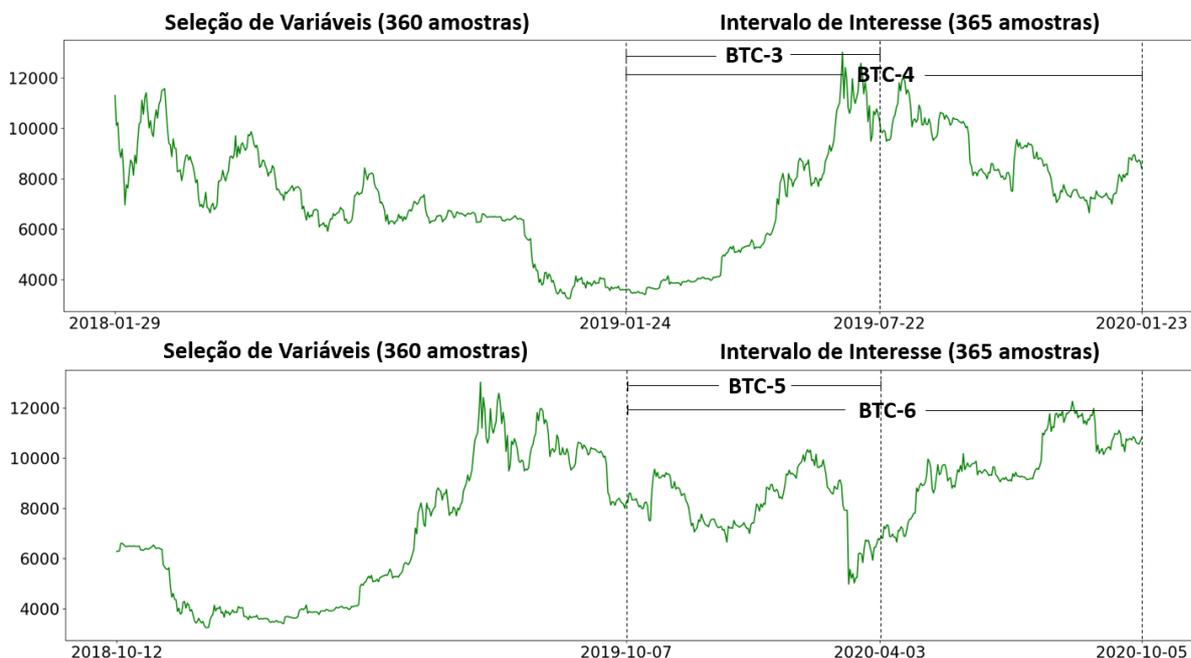
De forma complementar, foram selecionados dois novos períodos, os quais, por sua

Figura 10 – Evolução do preço de fechamento do Bitcoin nos intervalos BTC-1 e BTC-2.



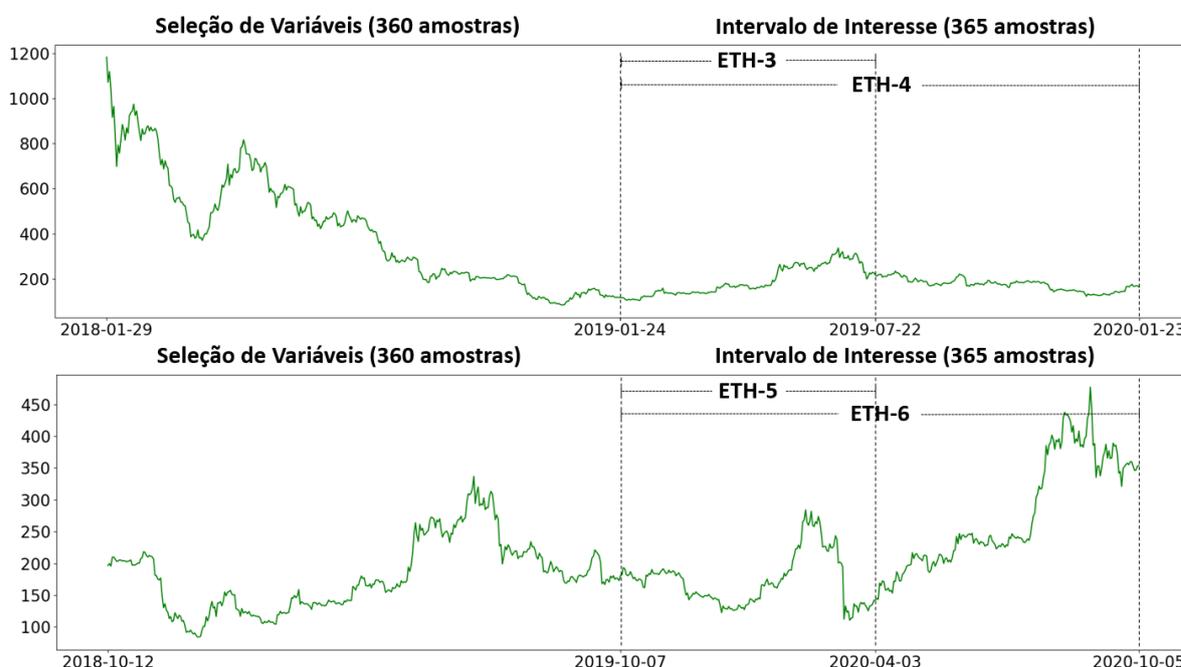
vez, compreendem o dia 11/03/2020. Dessa forma, pretende-se verificar se a presença de uma data, onde reconhecidamente houve uma variação brusca no preço da criptomoeda (*outlier*), tende a influenciar de forma significativa no desempenho final dos modelos. Neste cenário, foram selecionados os intervalos 07/10/2019 à 03/04/2020 e 07/10/2019 à 05/10/2020. A Figura 11 apresenta a evolução do preço de fechamento do Bitcoin nos quatro intervalos mencionados, os quais, a partir deste momento, são referenciados no texto como BTC-3, BTC-4, BTC-5 e BTC-6, respectivamente.

Figura 11 – Evolução do preço de fechamento do Bitcoin nos intervalos BTC-3, BTC-4, BTC-5 e BTC-6.



Diante da popularização das criptomoedas a partir da criação do Bitcoin no ano de 2008 e da posterior disponibilização de um vasto número de criptomoedas no mercado, pretende-se verificar o desempenho da abordagem proposta em alguma *altcoin* de relevância. Nos últimos anos, o Ether (ETH) tem se apresentado como a segunda criptomoeda com maior capitalização de mercado, posicionando-se muito a frente das demais neste quesito (COINMARKETCAP, 2021). Portanto, quatro intervalos também foram selecionados para esta criptomoeda, os quais são idênticos aos quatro últimos intervalos detalhados para o Bitcoin. Cria-se, portanto, no que se refere a datas, uma correspondência: BTC-3/ETH-3, BTC-4/ETH-4, BTC-5/ETH-5 e BTC-6/ETH-6. A Figura 12 apresenta a evolução do preço de fechamento do Ether nos quatro intervalos mencionados.

Figura 12 – Evolução do preço de fechamento do Ether nos intervalos ETH-3, ETH-4, ETH-5 e ETH-6.



A Tabela 7 é um resumo dos 10 intervalos utilizados na realização dos experimentos. Nessa tabela é possível visualizar o número de amostras, além da data de início e fim, tanto para o Período de Seleção de Variáveis quanto para o respectivo Intervalo de Interesse. De forma complementar, no Apêndice A, é apresentado um detalhamento numérico acerca da evolução do preço de fechamento da criptomoeda referente a cada intervalo.

5.3 Previsão do Valor

Nesta seção apresenta-se o desempenho dos modelos na Previsão do Valor. Dentre as métricas detalhadas na Seção 2.3.1, são considerados o RMSE (*Root Mean Squared Error*) e o NDEI (*Non-Destructive Error Index*). Sabe-se que, em ambos os casos, o melhor desempenho é obtido pelo modelo que apresenta os menores valores calculados.

Tabela 7 – Resumo dos conjuntos de dados utilizados nos experimentos.

Nome	Seleção de Variáveis		Intervalo de Interesse	
	Período	Amostras	Período	Amostras
BTC-1	07/04/2015 à 31/03/2016	360	01/04/2016 à 31/03/2017	365
BTC-2	25/09/2016 à 19/09/2017	360	20/09/2017 à 31/12/2018	468
BTC-3	29/01/2018 à 23/01/2019	360	24/01/2019 à 22/07/2019	180
BTC-4	29/01/2018 à 23/01/2019	360	24/01/2019 à 23/01/2020	365
BTC-5	12/10/2018 à 06/10/2019	360	07/10/2019 à 03/04/2020	180
BTC-6	12/10/2018 à 06/10/2019	360	07/10/2019 à 05/10/2020	365
ETH-3	29/01/2018 à 23/01/2019	360	24/01/2019 à 22/07/2019	180
ETH-4	29/01/2018 à 23/01/2019	360	24/01/2019 à 23/01/2020	365
ETH-5	12/10/2018 à 06/10/2019	360	07/10/2019 à 03/04/2020	180
ETH-6	12/10/2018 à 06/10/2019	360	07/10/2019 à 05/10/2020	365

5.3.1 Intervalos BTC-1 e BTC-2

Conforme descrito na apresentação dos intervalos objetos de estudo deste trabalho, sabe-se que BTC-1 e BTC-2 são intervalos também estudados por outros autores. Entretanto, para o intervalo BTC-1, apesar dos autores apresentarem o RMSE calculado para os modelos utilizados, os dados não foram normalizados na realização dos experimentos, o que inviabiliza uma comparação neste sentido. Em relação ao intervalo BTC-2, os autores apresentam resultados apenas no que tange à previsão do sentido de variação do preço do Bitcoin e à utilização dessas previsões em uma estratégia de recomendação de investimentos, também não sendo possível uma comparação dos resultados neste aspecto.

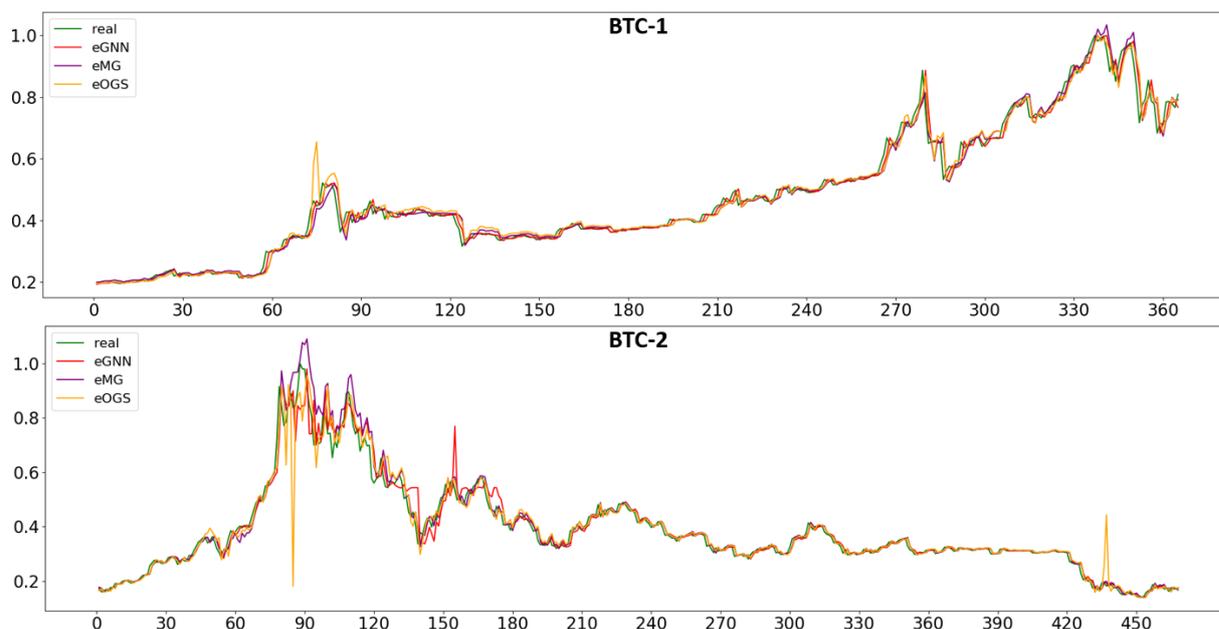
A Tabela 8 apresenta os valores calculados para os três Sistemas *Fuzzy* Evolutivos (eGNN, eMG e eOGS), com destaque (em negrito) para os melhores resultados por intervalo/métrica. Ao analisar os resultados é possível perceber que o modelo eGNN obteve melhor desempenho nos dois intervalos, apresentando RMSE e NDEI inferiores aos demais modelos. Destaca-se, entretanto, que, em se tratando de RMSE, no intervalo BTC-2, o eMG obteve um resultado comparável ao alcançado pelo eGNN (diferença de apenas 0,0001). Dentre os três modelos, o eOGS é aquele que apresentou o pior desempenho para as duas medidas calculadas. Este fato é visualmente constatável na Figura 13, a qual ilustra o desempenho de cada modelo (em vermelho, roxo e laranja), nos intervalos BTC-1 e BTC-2, comparando os resultados previstos com o preço de fechamento real do Bitcoin (em verde).

Tabela 8 – Desempenho na previsão do valor de fechamento nos intervalos BTC-1 e BTC-2.

Intervalo	eGNN		eMG		eOGS	
	RMSE	NDEI	RMSE	NDEI	RMSE	NDEI
BTC-1	0,0170	0,1188	0,0192	0,1335	0,0213	0,1482
BTC-2	0,0291	0,2499	0,0292	0,2514	0,0350	0,3005

Na Figura 13 é possível perceber que, no intervalo BTC-1, os modelos evoluíram de forma relativamente equilibrada, com apenas um pequeno ponto de desequilíbrio para o eOGS, próximo ao instante 70. Entretanto, em BTC-2, um intervalo mais longo, percebe-se que há uma instabilidade maior no resultado apresentado pelos modelos, de forma mais evidente entre os instantes 85 e 160, principalmente para o eOGS, o qual registra um novo momento de instabilidade próximo ao instante 440. Mesmo diante disso destaca-se que, na

Figura 13 – Comparação entre o valor real e o previsto nos intervalos BTC-1 e BTC-2.



maioria absoluta das amostras, as linhas que representam o valor real e os valores previstos encontram-se visualmente próximas, retratando de forma positiva a capacidade preditiva dos modelos.

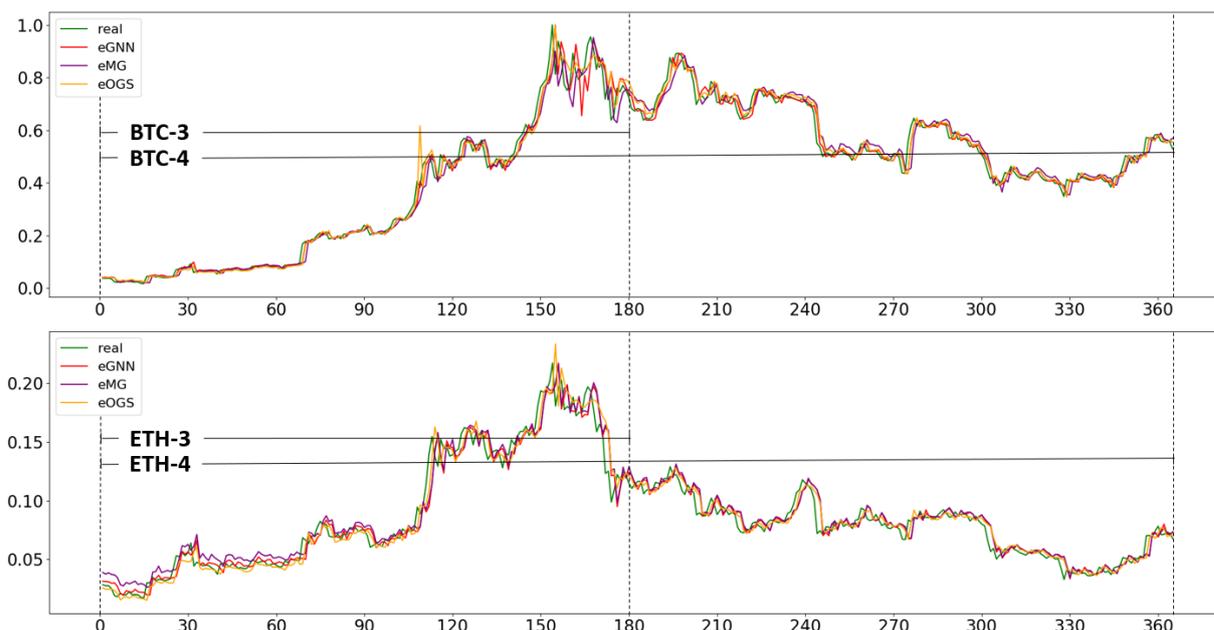
5.3.2 Intervalos BTC-3, BTC-4, ETH-3 e ETH-4

A Tabela 9 apresenta o resultado dos modelos evolutivos nos intervalos BTC-3, BTC-4, ETH-3 e ETH-4. Destaca-se que o eOGS foi aquele que obteve o melhor desempenho em três dos quatro períodos analisados. Além disso, a diferença entre o resultado obtido por eGNN e eOGS no intervalo BTC-4 (primeiro e segundo colocados, respectivamente), foi de apenas 0,0002 (RMSE) e 0,0011 (NDEI), o que confirma o excelente desempenho do eOGS, inclusive neste período. Também é importante ressaltar que, dentre os três modelos, o eMG foi o que apresentou o pior desempenho nos quatro intervalos, para as duas métricas analisadas. Entretanto, em se tratando de RMSE, apenas no intervalo BTC-4 a diferença para o segundo colocado foi superior a 0,0005, o que demonstra certo equilíbrio entre os modelos neste aspecto, fato confirmado visualmente na Figura 14.

Tabela 9 – Desempenho na previsão do valor de fechamento nos intervalos BTC-3, BTC-4, ETH-3 e ETH-4.

Intervalo	eGNN		eMG		eOGS	
	RMSE	NDEI	RMSE	NDEI	RMSE	NDEI
BTC-3	0,0244	0,1488	0,0249	0,1518	0,0240	0,1461
BTC-4	0,0251	0,1404	0,0283	0,1585	0,0253	0,1415
ETH-3	0,0065	0,2107	0,0070	0,2225	0,0062	0,1950
ETH-4	0,0069	0,2294	0,0072	0,2394	0,0062	0,2067

Figura 14 – Comparação entre o valor real e o previsto nos intervalos BTC-3, BTC-4, ETH-3 e ETH-4.



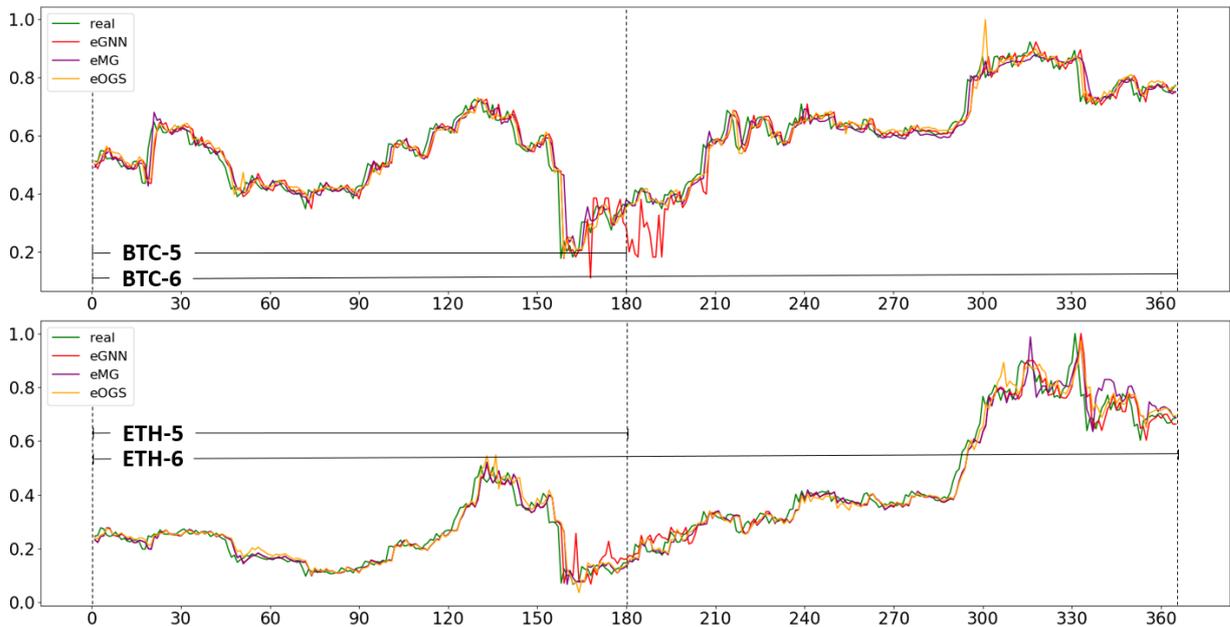
5.3.3 Intervalos BTC-5, BTC-6, ETH-5 e ETH-6

Finalmente, a Tabela 10 apresenta o desempenho dos modelos evolutivos no que se refere à previsão do valor de fechamento nos intervalos BTC-5, BTC-6, ETH-5 e ETH-6. Destaca-se que o modelo eMG obteve o melhor desempenho em três dos quatro intervalos analisados (BTC-6, ETH-5 e ETH-6). Em relação ao intervalo BTC-5, o melhor desempenho foi registrado pelo eOGS, porém, novamente o eMG apresentou um bom desempenho visto que a diferença para o eOGS, em termos de RMSE, foi de apenas 0,0003. Percebe-se ainda que o eGNN foi o modelo que apresentou o pior desempenho em três dos quatro intervalos analisados (BTC-5, BTC-6 e ETH-6). Alguns momentos de instabilidade deste modelo em relação aos demais tornam-se nítidos ao analisar a Figura 15, principalmente entre os instantes 160 e 180 no intervalo BTC-5 e entre os instantes 160 e 195 no intervalo BTC-6.

Tabela 10 – Desempenho na previsão do valor de fechamento nos intervalos BTC-5, BTC-6, ETH-5 e ETH-6.

Intervalo	eGNN		eMG		eOGS	
	RMSE	NDEI	RMSE	NDEI	RMSE	NDEI
BTC-5	0,0270	0,3869	0,0235	0,3366	0,0228	0,3275
BTC-6	0,0358	0,3258	0,0260	0,2369	0,0263	0,2390
ETH-5	0,0300	0,3276	0,0297	0,3249	0,0397	0,4335
ETH-6	0,0289	0,1876	0,0282	0,1833	0,0284	0,1848

Figura 15 – Comparação entre o valor real e o previsto nos intervalos BTC-5, BTC-6, ETH-5 e ETH-6.



5.3.4 Discussão dos Resultados

Ao realizar uma análise conjunta do desempenho dos três Sistemas *Fuzzy* Evolutivos empregados na previsão do valor de fechamento para os 10 intervalos estudados, percebe-se que o eOGS é o modelo que apresentou o melhor desempenho no maior número de intervalos (4), seguido por eGNN (3) e eMG (3). Entretanto, nota-se que três dos quatro intervalos onde o eOGS se destacou em relação aos demais são períodos com o menor número de amostras. O fato do eOGS ter registrado o melhor desempenho nos intervalos BTC-3 e BTC-5 mas não ter conseguido manter essa superioridade nos intervalos BTC-4 e BTC-6, respectivamente, é relevante, por se tratar de intervalos complementares. Entretanto, nestes intervalos em que o eOGS foi superado por outros modelos, a diferença para o modelo que o ultrapassou, nas duas métricas analisadas, foi relativamente pequena: em BTC-4 foram registradas diferenças de apenas 0,0002 (RMSE) e 0,0011 (NDEI) para o eGNN; e em BTC-6 as diferenças foram de 0,003 (RMSE) e 0,0021 (NDEI) para o eMG.

De modo inverso, o modelo que apresentou o pior desempenho na previsão do valor de fechamento no maior número de intervalos foi o eMG (4), seguido por eGNN (3) e eOGS (3). Destaca-se que, para o eMG, em relação aos quatro intervalos onde este fato foi registrado, todos se referem a períodos que antecedem o dia 11/03/2020 (BTC-3, BTC-4, ETH-3 e ETH-4). Neste mesmo sentido, os três intervalos onde o eGNN foi o modelo que registrou o pior desempenho, compreendem o dia 11/03/2020 (BTC-5, BTC-6 e ETH-6). Também é relevante mencionar que, em três dos quatro intervalos considerados possíveis pontos de maior instabilidade (BTC-5, BTC-6 e ETH-5), foram registradas as

maiores diferenças entre os modelos que apresentaram o melhor e o pior desempenho em termos de NDEI, medida independente de escala. Este fato também é registrado para o RMSE em dois destes intervalos. Na Tabela 11, são apresentadas essas diferenças para as métricas RMSE (Dif. RMSE) e NDEI (Dif. NDEI), com destaque em vermelho para os pontos mencionados.

Tabela 11 – Comparação entre os modelos que apresentaram o melhor e o pior desempenho na previsão do valor de fechamento.

Intervalo	Melhor Modelo	Pior Modelo	Dif. RMSE	Dif. NDEI
BTC-1	eGNN	eOGS	0,0043	0,0294
BTC-2	eGNN	eOGS	0,0059	0,0506
BTC-3	eOGS	eMG	0,0009	0,0057
BTC-4	eGNN	eMG	0,0032	0,0181
ETH-3	eOGS	eMG	0,0008	0,0275
ETH-4	eOGS	eMG	0,0010	0,0327
BTC-5	eOGS	eGNN	0,0042	0,0594
BTC-6	eMG	eGNN	0,0098	0,0889
ETH-5	eMG	eOGS	0,0100	0,1086
ETH-6	eMG	eGNN	0,0007	0,0043

5.4 Previsão do Sentido de Variação

Nesta seção é apresentado o desempenho dos modelos na tarefa de Previsão do Sentido de Variação. Neste aspecto, para avaliação dos resultados, foram utilizadas as métricas Acurácia (%) e *F1-Score*.

5.4.1 Intervalos BTC-1 e BTC-2

A Tabela 12 apresenta o desempenho dos modelos em termos de Acurácia (Ac. %) e *F1-Score* para os intervalos BTC-1 e BTC-2. Além disso, também são apresentados nessa tabela os resultados de outros trabalhos que utilizaram estes intervalos em seus experimentos¹. Para o intervalo BTC-1 o trabalho de Mallqui e Fernandes (2019) foi considerado como *Benchmark (Bench.)* e, para o intervalo BTC-2, foi considerado o trabalho de Ji, Kim e Im (2019). Destaca-se que foram considerados como *Benchmark* os melhores modelos, para ambos os trabalhos.

Tabela 12 – Desempenho na previsão do sentido de variação do preço de fechamento nos intervalos BTC-1 e BTC-2.

Intervalo	eGNN		eMG		eOGS		Comitê		<i>Bench.</i> Ac. (%)
	Ac. (%)	F1-Score	Ac. (%)	F1-Score	Ac. (%)	F1-Score	Ac. (%)	F1-Score	
BTC-1	60,27	0,6985	59,18	0,6850	56,99	0,6681	61,10	0,7171	59,45
BTC-2	54,70	0,6045	52,99	0,6541	55,56	0,6414	55,77	0,6601	53,06

Diante dos resultados, percebe-se que o Comitê, formado pelos três modelos evolutivos, apresentou desempenho superior, para as duas métricas analisadas, nos dois

¹ A Acurácia é a única métrica comum utilizada nos dois trabalhos considerados como *Benchmark*.

intervalos. Além disso, destaca-se que o Comitê superou, inclusive, os resultados apresentados na literatura. No intervalo BTC-1, para a métrica Acurácia, o Comitê foi seguido por eGNN, eMG, *Benchmark* e eOGS. Em se tratando da métrica *F1-Score*, o Comitê também foi superior aos demais, seguido por eGNN, eOGS e eMG. Para o intervalo BTC-2, a ordem de desempenho dos modelos em termos de Acurácia foi: Comitê, eOGS, eGNN, *Benchmark* e eMG. Destaca-se ainda que, mesmo o eMG tendo apresentado o pior desempenho neste intervalo no que se refere a Acurácia, este foi, ao mesmo tempo, o modelo que obteve o segundo melhor resultado em termos de *F1-Score*, ficando atrás apenas do Comitê. Isto indica que houve um melhor balanceamento entre as respostas fornecidas por este modelo, para as diferentes classes.

5.4.2 Intervalos BTC-3, BTC-4, ETH-3 e ETH-4

A Tabela 13 apresenta os resultados dos modelos no que se refere à previsão do sentido de variação do preço das criptomoedas nos intervalos BTC-3, BTC-4, ETH-3 e ETH-4. Percebe-se que Comitê e eGNN obtiveram os melhores resultados em termos de Acurácia. Neste caso, o Comitê apresentou o melhor desempenho nos intervalos compostos pelo menor número de amostras (BTC-3 e ETH-3) e o eGNN, conseqüentemente, desempenhou melhor nos maiores intervalos (BTC-4 e ETH-4). Entretanto, destaca-se que, em termos de *F1-Score*, o Comitê obteve resultado superior ao eGNN em todos os quatro intervalos analisados. Ainda neste aspecto (*F1-Score*), dentre todos os modelos, o Comitê apresentou o melhor desempenho em três dos quatro intervalos (BTC-3, BTC-4 e ETH-3). Portanto, de forma geral, para estes quatro intervalos, entende-se que o Comitê apresentou desempenho superior aos modelos evolutivos, quando executados de forma individual.

Tabela 13 – Desempenho na previsão do sentido de variação do preço de fechamento nos intervalos BTC-3, BTC-4, ETH-3 e ETH-4.

Intervalo	eGNN		eMG		eOGS		Comitê	
	Ac. (%)	F1-Score	Ac. (%)	F1-Score	Ac. (%)	F1-Score	Ac. (%)	F1-Score
BTC-3	56,67	0,6777	56,11	0,6220	53,33	0,6000	60,00	0,6923
BTC-4	54,25	0,5474	52,05	0,5958	53,15	0,6242	53,15	0,6242
ETH-3	57,78	0,6122	53,33	0,6250	53,33	0,4000	58,33	0,6269
ETH-4	56,16	0,6000	53,15	0,6069	53,42	0,4654	55,62	0,6049

5.4.3 Intervalos BTC-5, BTC-6, ETH-5 e ETH-6

Finalmente, a Tabela 14 apresenta os resultados para os intervalos BTC-5, BTC-6, ETH-5 e ETH-6. Destaca-se que o Comitê registrou o melhor desempenho, de forma isolada, em termos de Acurácia, nos dois intervalos referentes ao Bitcoin (BTC-5 e BTC-6). Ainda sobre esta métrica, para os intervalos referentes ao Ether, o desempenho do Comitê foi equivalente ao apresentado pelo eOGS em ETH-5, e o eMG foi superior aos demais modelos no intervalo ETH-6. Em se tratando da métrica *F1-Score*, tem-se o seguinte cenário: o eOGS apresentou melhor desempenho, de forma isolada, em dois intervalos (BTC-6 e

ETH-6); o desempenho do eOGS e do Comitê foram equivalentes em um intervalo (ETH-5); e o Comitê, por sua vez, foi o melhor, de forma isolada, em um intervalo (BTC-5).

Tabela 14 – Desempenho na previsão do sentido de variação do preço de fechamento nos intervalos BTC-5, BTC-6, ETH-5 e ETH-6.

Intervalo	eGNN		eMG		eOGS		Comitê	
	Ac. (%)	F1-Score	Ac. (%)	F1-Score	Ac. (%)	F1-Score	Ac. (%)	F1-Score
BTC-5	57,22	0,5650	54,44	0,5119	53,89	0,5829	58,33	0,5856
BTC-6	57,53	0,5584	54,52	0,4813	53,97	0,5942	58,36	0,5801
ETH-5	52,78	0,5405	51,67	0,5348	55,00	0,5668	55,00	0,5668
ETH-6	52,60	0,5312	54,79	0,5623	53,70	0,5908	54,25	0,5662

5.4.4 Discussão dos Resultados

De modo geral, para o problema de previsão do sentido de variação do preço de fechamento das criptomoedas, em termos de Acurácia, o Comitê apresentou desempenho superior aos demais modelos em cinco dos seis intervalos referentes ao Bitcoin (BTC-1, BTC-2, BTC-3, BTC-5 e BTC-6). Além disso, para os intervalos relacionados ao Ether, o Comitê foi superior aos demais modelos em ETH-3 e apresentou desempenho equivalente ao eOGS em ETH-5. Portanto, ao todo, o Comitê registrou os melhores números alcançados em sete dos dez intervalos. Além disso, destaca-se que o Comitê e o eGNN foram os únicos modelos que superaram, ao mesmo tempo, os resultados apresentados pelos modelos *benchmark* nos dois intervalos onde essa comparação foi possível (BTC-1 e BTC-2). Em relação ao *F1-Score*, o Comitê registrou resultado superior aos demais modelos no mesmo número de intervalos (7), o que também caracteriza, de forma positiva, sua qualidade em balancear os acertos entre as diferentes classes.

Ao realizar uma análise no sentido inverso, percebe-se que eMG e eOGS são os modelos que registraram o pior desempenho na métrica Acurácia para o maior número de intervalos, fato que foi registrado cinco vezes, para cada modelo, em nove intervalos (no intervalo ETH-3 os dois modelos apresentaram o mesmo resultado). Neste caso, a única exceção é o intervalo ETH-6, no qual o eGNN teve desempenho inferior aos demais modelos. Em relação à métrica *F1-Score*, os modelos apresentaram maior equilíbrio, tendo este fato sido registrado para o eOGS quatro vezes, e, para eGNN e eMG, isto ocorreu em três intervalos. Destaca-se, ainda, que, para as duas métricas, o Comitê não foi apontado como o modelo que apresentou o pior desempenho em qualquer dos intervalos.

Entende-se, ainda, como relevante uma comparação do desempenho dos modelos nos intervalos que antecedem o dia 11/03/2020 com os resultados apresentados pelos modelos nos intervalos que compreendem esta data, para que seja verificado se a presença de um *outlier* expressivo poderia afetar significativamente o desempenho dos modelos. Portanto, a Tabela 15 consiste em um resumo desta análise, comparando intervalos para criptomoedas e tamanhos equivalentes, isto é: BTC-3/BTC-5, BTC-4/BTC-6, ETH-3/ETH-5 e ETH-4/ETH-6.

Tabela 15 – Comparação do desempenho na previsão do sentido de variação do preço de fechamento entre os intervalos que antecedem o dia 11/03/2020 (BTC-3, BTC-4, ETH-3 e ETH-4) e os intervalos que compreendem esta data (BTC-5, BTC-6, ETH-5 e ETH-6).

Modelo	Acurácia	F1-Score
eGNN	Nos dois intervalos referentes à criptomoeda Ether, o desempenho do modelo foi negativamente afetado. Na comparação, para estes intervalos foram registradas perdas de 5,00% (ETH-5) e 3,56% (ETH-6).	O desempenho do modelo foi consideravelmente impactado nos intervalos BTC-5, ETH-5 e ETH-6, registrando diferenças de 0,1127; 0,0717; e 0,0688, respectivamente.
eMG	Não houve variação superior a 3%.	O desempenho do modelo foi bruscamente afetado nos quatro intervalos que compreendem o dia 11/03/2020, registrando perdas de 0,1101 (BTC-5); 0,1145 (BTC-6); 0,0902 (ETH-5); e 0,0446 (ETH-6).
eOGS	Não houve variação superior a 3% para qualquer dos intervalos. Entretanto, mesmo tendo sido registradas apenas pequenas diferenças nas quatro comparações realizadas, o desempenho do modelo foi superior nos quatro intervalos que compreendem o dia 11/03/2020, um fato até então inesperado.	Também de modo diferente do que se esperava, o modelo apresentou desempenho consideravelmente superior nos dois intervalos para a criptomoeda Ether, que compreendem o dia 11/03/2020, com diferenças positivas de 0,1668 (ETH-5) e 0,1254 (ETH-6). Nos intervalos relacionados ao Bitcoin, a diferença teve menor significância, ficando baixo de 0,0300.
Comitê	O desempenho foi negativamente afetado em três dos quatro intervalos, com a maior diferença tendo sido registrada em ETH-5, quando houve redução de 3,33% para esta métrica. O principal motivo disso ter ocorrido se refere ao fato de que, em todos os intervalos que compreendem o dia 11/03/2020, pelo menos dois modelos apresentaram desempenho inferior ao registrado para os períodos correspondentes, que não compreendem esta data.	Em três dos quatro intervalos o desempenho do modelo foi bruscamente impactado, com perdas de 0,1067 (BTC-4); 0,0601 (ETH-5); e 0,0387 (ETH-6). Em contrapartida, na comparação BTC-5/ETH-5, o modelo apresentou melhor desempenho em BTC-5, com diferença positiva de 0,0441.

Acerca da comparação realizada, destaca-se que, em geral, os modelos eGNN, eMG e Comitê apresentaram desempenho inferior nos intervalos que compreendem o dia 11/03/2020. A principal métrica afetada foi o *F1-Score*, impactando, conseqüentemente, de forma negativa, na capacidade de balanceamento das respostas fornecidas. Neste caso, o Comitê foi o maior prejudicado. Uma justificativa para isto reside no fato de que, em todos os intervalos, pelo menos dois modelos, de forma alternada, apresentaram desempenho inferior nos períodos que compreendem a data mencionada. Em contrapartida, o modelo eOGS é considerado um ponto fora da curva, visto que apresentou pouca diferença na avaliação das duas métricas.

5.5 Recomendação de Investimentos

Nesta seção é apresentada uma análise de Rentabilidade a partir da aplicação de estratégias de Recomendação de Investimentos aos intervalos estudados. Considerando o aspecto genérico da abordagem demonstrado no Algoritmo 1, apresentado na Subseção

4.3.3, sua utilização é proposta a partir da atribuição de dois valores distintos ao parâmetro *limiar*, isto é, na primeira estratégia (EST-1) assume-se $limiar = 0,1$ e, na segunda estratégia (EST-2), considera-se $limiar = 1,0$. Nas simulações ainda é considerado que o investidor não possuirá, inicialmente, qualquer quantia em criptomoedas, portanto tem-se $criptomoeda = 0$. Em contrapartida, este mesmo investidor possuirá US\$ 10.000,00 em caixa, que poderão ser utilizados para a compra de criptomoedas, de acordo com cada estratégia, portanto $dinheiro = 10000$.

5.5.1 Intervalos BTC-1 e BTC-2

A Tabela 16 apresenta a Rentabilidade Acumulada (RA) e o Retorno por Unidade de Risco (RR)² computados para os três modelos evolutivos e para o Comitê ao final dos intervalos BTC-1 e BTC-2. Além disso também são apresentados os resultados do *Benchmark*. Destaca-se em azul os resultados positivos, em vermelho os resultados negativos e em negrito ambos os extremos.

Tabela 16 – Rentabilidade Acumulada (RA) e Retorno por Unidade de Risco (RR) na recomendação de investimentos nos intervalos BTC-1 e BTC-2.

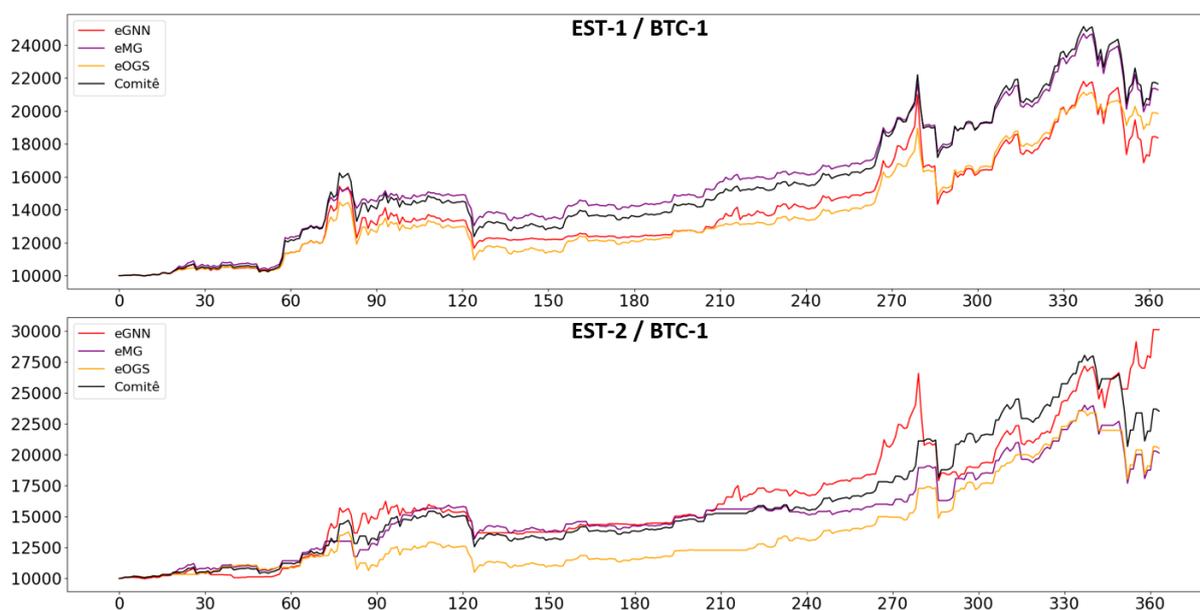
Intervalo	Var. (%)	EST	Tipo	eGNN (%)	eMG (%)	eOGS (%)	Comitê (%)	Benchmark (%)
BTC-1	156,43	1	RA	+87,84	+117,36	+102,06	+121,06	-
			RR	0,12	0,16	0,13	0,14	-
		2	RA	+210,43	+107,63	+111,44	+142,50	-
			RR	0,19	0,16	0,12	0,15	-
BTC-2	-4,17	1	RA	+51,89	-12,85	+10,32	+14,91	-
			RR	0,04	0,01	0,02	0,03	-
		2	RA	+57,99	+24,78	+74,47	+57,38	+8,77
			RR	0,04	0,03	0,05	0,04	-

Para o intervalo BTC-1, quando o Bitcoin valorizou 156,43%, o eGNN apresentou o melhor resultado em termos de RA, com ganhos de US\$ 21.042,92 (+210,43%), utilizando a estratégia EST-2. Além disso, nesta mesma estratégia, o eGNN também foi o modelo que apresentou o melhor RR, registrando 0,19. Em contrapartida, para a estratégia EST-1, o eGNN foi o modelo que apresentou o pior resultado, para as duas métricas analisadas. Nessa estratégia, em termos de RA, o Comitê apresentou o melhor resultado mas, em contrapartida, o eMG foi o modelo que apresentou o melhor RR neste intervalo/estratégia. Para o intervalo BTC-2, apresenta-se também a RA para o modelo que obteve o melhor resultado no trabalho de Ji, Kim e Im (2019), considerando a única estratégia aplicada pelos autores (EST-2). Percebe-se que o eOGS obteve o melhor resultado, proporcionando um ganho de US\$ 7.446,59 (+74,47%) na estratégia EST-2, neste período em que o Bitcoin sofreu desvalorização de 4,17%. Além disso, o eOGS também foi o modelo que apresentou o melhor RR, registrando 0,05. Destaca-se ainda que todos os modelos considerados nesta abordagem se mostraram superiores ao *benchmark*, em termos de RA, para o intervalo BTC-2.

² Em todos os experimentos realizados, por questões de simplicidade, considera-se $R_f = 0$ no cálculo do Retorno por Unidade de Risco (RR).

As figuras 16 e 17 apresentam a evolução da Rentabilidade ao longo dos intervalos BTC-1 e BTC-2, respectivamente. No intervalo BTC-1, destacaram-se de forma positiva eMG e Comitê, para a estratégia EST-1. Acerca da estratégia EST-2, o eGNN alternou em diversos momentos com o Comitê, sendo aqueles que apresentaram Rentabilidade superior aos demais no maior número de amostras ao longo deste intervalo/estratégia. Também é importante mencionar que, tanto para EST-1 como para EST-2, o modelo eOGS apresentou Rentabilidade inferior aos demais na maior parte do intervalo BTC-1. Por sua vez, no intervalo BTC-2, percebe-se que o modelo eGNN foi soberano ao longo da maior parte do intervalo para a estratégia EST-1 e, para a estratégia EST-2, os modelos eGNN e eOGS alternaram entre si.

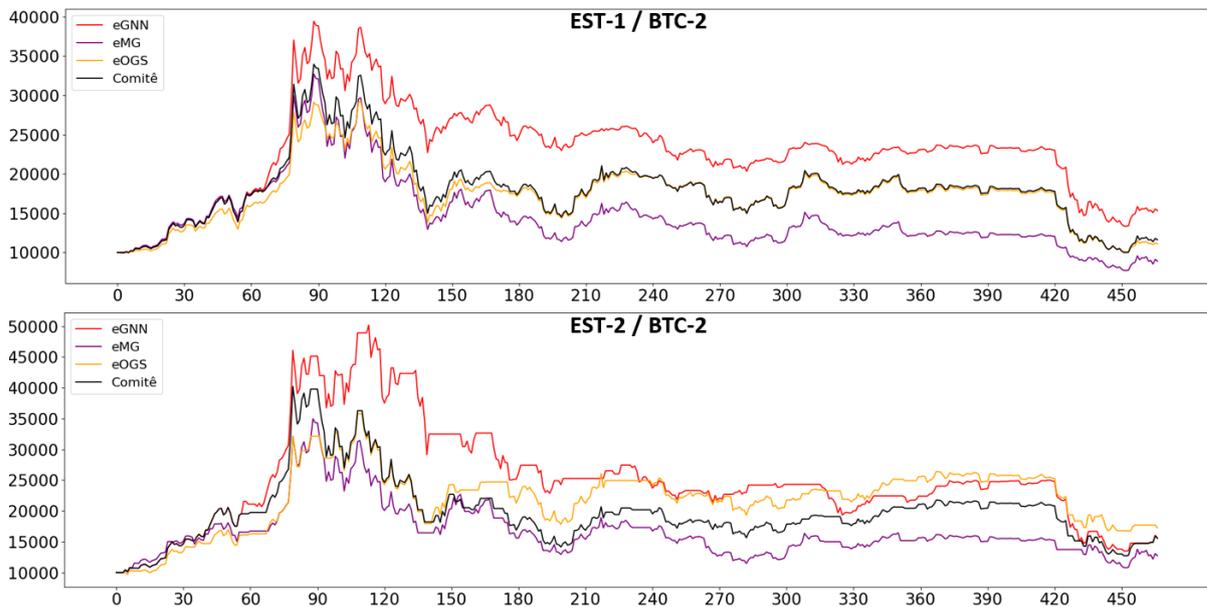
Figura 16 – Evolução da rentabilidade na recomendação de investimentos no intervalo BTC-1.



5.5.2 Intervalos BTC-3, BTC-4, ETH-3 e ETH-4

A Tabela 17 apresenta a Rentabilidade Acumulada (RA) e o Retorno por Unidade de Risco (RR) alcançados pelos modelos nos intervalos BTC-3, BTC-4, ETH-3 e ETH-4. Destaca-se que o modelo eGNN alcançou a maior RA, por meio da estratégia EST-2, em três dos quatro intervalos. Além disso, no intervalo BTC-4, este modelo também proporcionou maior RA mesmo com a estratégia EST-1. Também é importante ressaltar que em todos os intervalos/estratégias em que o eGNN se mostrou superior aos demais modelos em termos de RA, este modelo também apresentou melhor resultado no que se refere à métrica RR. Em se tratando do Comitê, este modelo alcançou o melhor resultado no intervalo BTC-3, para as duas estratégias. Porém, em termos de RR, o Comitê se mostrou superior aos demais em apenas um destes casos. Por último, destaca-se que o eOGS apresentou desempenho inferior aos demais em sete dos oito experimentos realizados

Figura 17 – Evolução da rentabilidade na recomendação de investimentos no intervalo BTC-2.



para a métrica RA e, para a métrica RR, este modelo foi inferior aos demais em todos os intervalos/estratégias. Entretanto, destaca-se que apenas em um deles o modelo não alcançou RA positiva (ETH-4/EST-2).

Tabela 17 – Rentabilidade Acumulada (RA) e Retorno por Unidade de Risco (RR) na recomendação de investimentos nos intervalos BTC-3, BTC-4, ETH-3 e ETH-4.

Intervalo	Var. (%)	EST	Tipo	eGNN (%)	eMG (%)	eOGS (%)	Comitê (%)
BTC-3	187,24	1	RA	+126,79	+112,61	+99,91	+137,54
			RR	0,18	0,22	0,16	0,19
		2	RA	+162,42	+115,81	+125,73	+221,34
			RR	0,17	0,19	0,15	0,21
BTC-4	133,46	1	RA	+142,13	+88,90	+68,86	+112,17
			RR	0,13	0,09	0,07	0,10
		2	RA	+152,93	+65,91	+55,65	+122,68
			RR	0,10	0,07	0,06	0,09
ETH-3	85,38	1	RA	+82,89	+88,89	+9,76	+77,43
			RR	0,15	0,13	0,04	0,13
		2	RA	+101,17	+82,37	+13,77	+96,08
			RR	0,14	0,11	0,04	0,13
ETH-4	38,83	1	RA	+67,92	+70,75	+2,42	+62,38
			RR	0,07	0,07	0,01	0,07
		2	RA	+101,02	+86,54	-0,45	+93,80
			RR	0,08	0,07	0,01	0,07

Na análise da Tabela 17 foi relatado que o eGNN superou os demais modelos em boa parte dos intervalos, tanto para RA quanto para RR. Entretanto, a Figura 18, que apresenta a evolução da Rentabilidade ao longo de BTC-3 e BTC-4, torna perceptível que o eGNN se distanciou dos demais modelos, sobretudo, a partir da segunda metade do intervalo BTC-4 (EST-1 e EST2), próximo ao instante 270, para as duas estratégias. No que se refere ao Ether, que tem a evolução da Rentabilidade retrata para os modelos na Figura 19, o desempenho do Comitê, eGNN e eMG, neste aspecto, esteve sempre muito próximo ao longo dos dois intervalos, com destaque negativo para o eOGS, que se mostrou inferior

aos demais na maior parte das amostras.

Figura 18 – Evolução da rentabilidade na recomendação de investimentos nos intervalos BTC-3 e BTC-4.

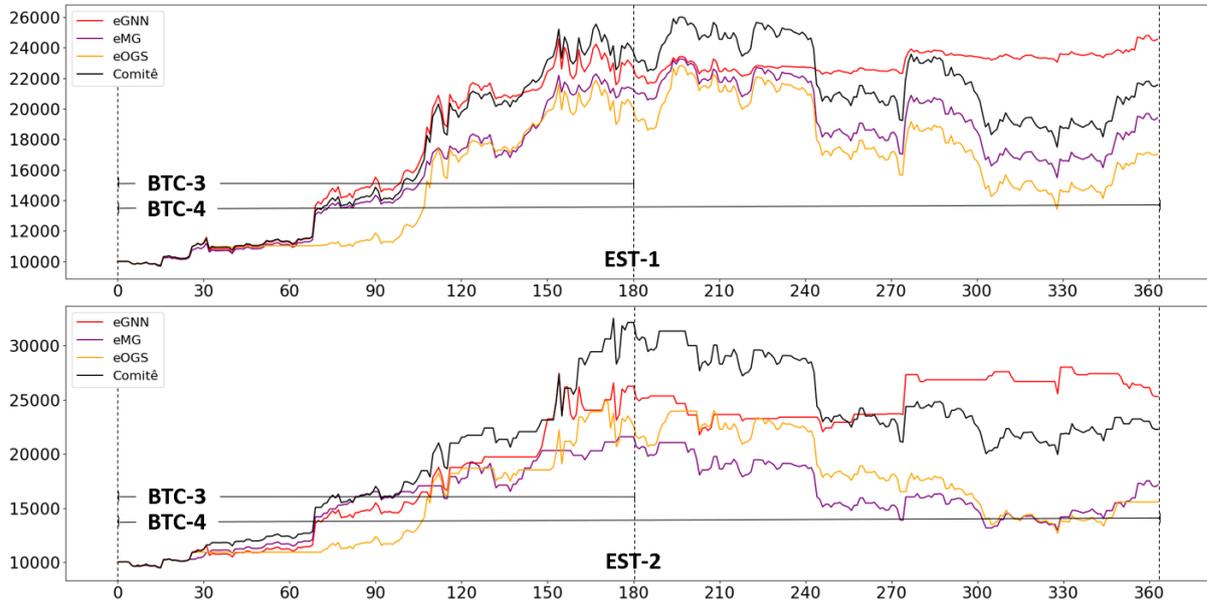
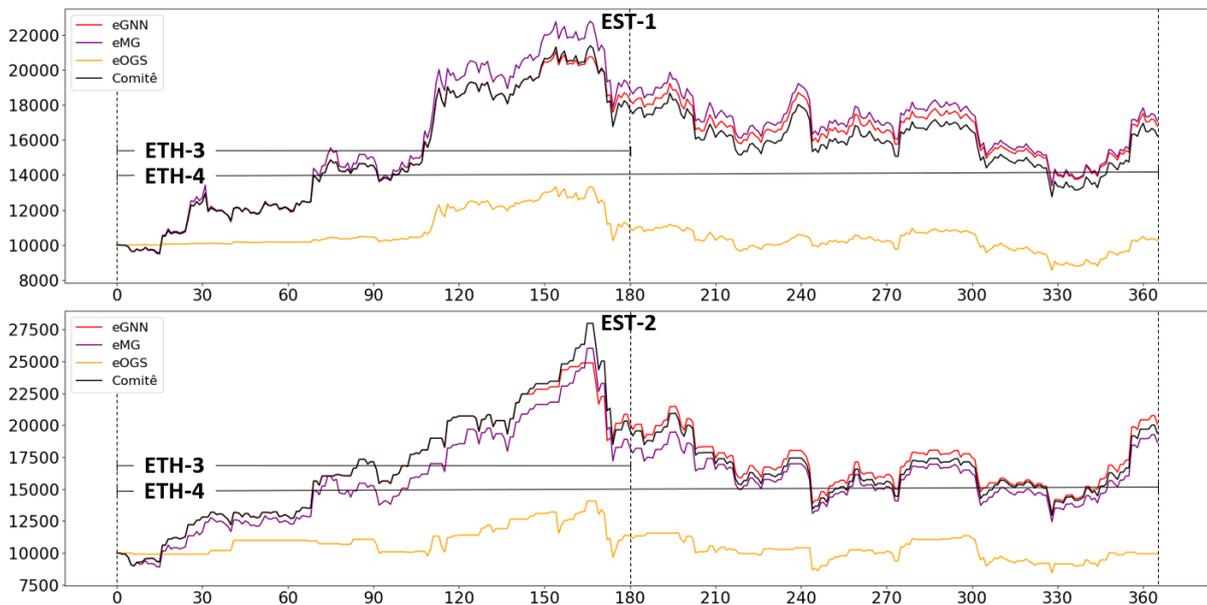


Figura 19 – Evolução da rentabilidade na recomendação de investimentos nos intervalos ETH-3 e ETH-4.



5.5.3 Intervalos BTC-5, BTC-6, ETH-5 e ETH-6

Finalmente, a Tabela 18 apresenta a Rentabilidade alcançada nos intervalos BTC-5, BTC-6, ETH-5 e ETH-6. Destaca-se que o modelo eOGS, desta vez, foi o melhor em três ocasiões no que se refere a RA, o que ocorreu duas vezes com a estratégia EST-2 e uma vez com a estratégia EST-1. Nos intervalos BTC-5 e ETH-5, quando houve forte

desvalorização das duas criptomoedas, todos os modelos geraram prejuízo em pelo menos um intervalo/estratégia. Entretanto, quando o eOGS foi aplicado utilizando a estratégia EST-2, para os dois intervalos, o modelo foi capaz de registrar Rentabilidade positiva, mesmo neste cenário adverso, também apresentando RR equivalente ou superior aos demais.

Tabela 18 – Rentabilidade Acumulada (RA) e Retorno por Unidade de Risco (RR) na recomendação de investimentos nos intervalos BTC-5, BTC-6, ETH-5 e ETH-6.

Intervalo	Var. (%)	EST	Tipo	eGNN (%)	eMG (%)	eOGS (%)	Comitê (%)
BTC-5	-18,34	1	RA	-2,70	-1,64	-6,27	-3,02
			RR	0,01	0,01	0,00	0,01
		2	RA	87,14	3,25	40,78	61,33
			RR	0,13	0,02	0,08	0,10
BTC-6	30,90	1	RA	20,70	13,16	26,34	22,87
			RR	0,04	0,03	0,04	0,04
		2	RA	207,88	25,15	74,89	155,45
			RR	0,13	0,04	0,07	0,11
ETH-5	-21,58	1	RA	-6,75	-2,36	-17,93	-6,39
			RR	0,00	0,01	-0,01	0,00
		2	RA	-21,20	-26,94	1,59	-23,84
			RR	0,00	-0,01	0,03	-0,01
ETH-6	95,35	1	RA	56,64	50,94	56,19	57,84
			RR	0,06	0,06	0,06	0,06
		2	RA	36,92	43,57	80,01	56,10
			RR	0,04	0,05	0,06	0,05

A Figura 20 apresenta a evolução da Rentabilidade alcançada ao longo dos intervalos BTC-5 e BTC-6. Destaca-se que, a partir do uso da estratégia EST-1, os modelos apresentaram desempenho visualmente próximos na maior parte das amostras. Entretanto, ao final do intervalo BTC-6, o modelo eOGS se destacou positivamente em relação aos demais, ficando a cargo do eMG o destaque negativo. Em contrapartida, na estratégia EST-2, os modelos apresentaram maior diferença entre si em um número consideravelmente maior de amostras, sobretudo a partir do instante 150, com evolução progressiva dessa diferença ao longo dos intervalos. Deste modo, ao final do intervalo BTC-6, a Rentabilidade alcançada pelo eGNN foi muito superior aos demais.

Por fim, para os intervalos ETH-5 e ETH-6, a evolução da Rentabilidade retratada na Figura 21 demonstra que os modelos apresentaram resultados semelhantes. Destaca-se, entretanto, que, para a estratégia EST-1, a partir do instante 160, o desempenho do eOGS se apresenta significativamente inferior aos demais modelos. Entretanto, ao final do intervalo ETH-6, a partir do instante 300, percebe-se que o eOGS melhora consideravelmente o seu desempenho, se equilibrando em relação aos demais modelos. De modo contrário, para a EST-2, a partir do instante 120, em ETH-5 e ETH-6, o eOGS apresenta desempenho superior aos demais, conseguindo manter o bom desempenho até o fim destes intervalos.

5.5.4 Discussão dos Resultados

Na apresentação dos resultados referentes à Rentabilidade proporcionada por cada modelo, os dados foram apresentados de forma tabular, considerando cada intervalo na sua completude, isto é, a Rentabilidade dos modelos foi avaliada somente ao final de

Figura 20 – Evolução da rentabilidade na recomendação de investimentos nos intervalos BTC-5 e BTC-6.

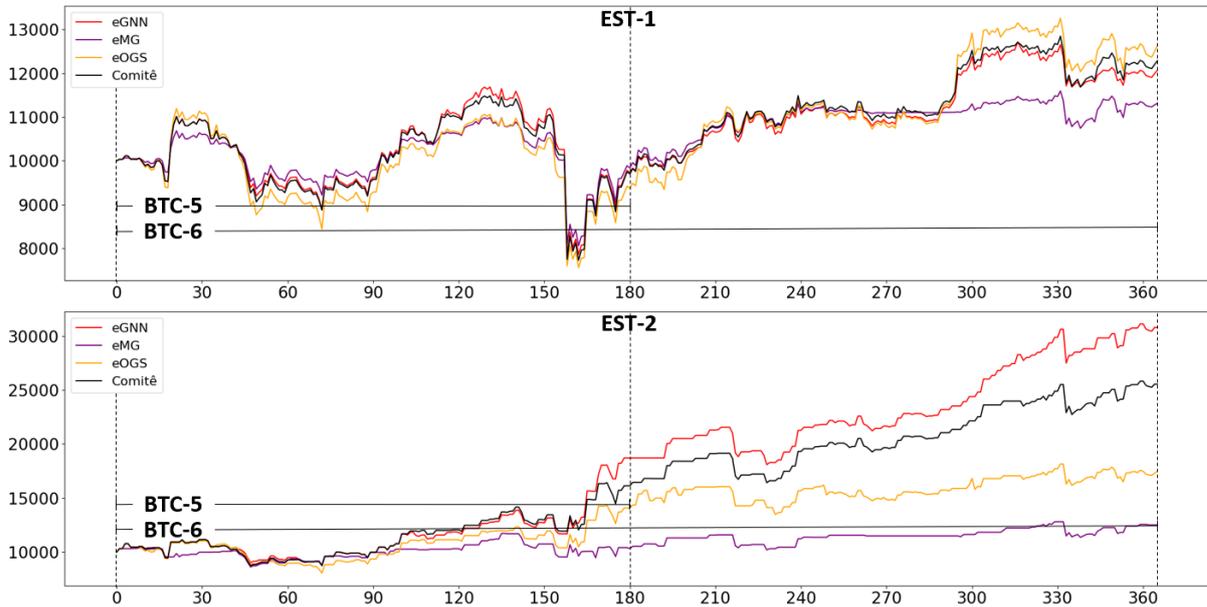
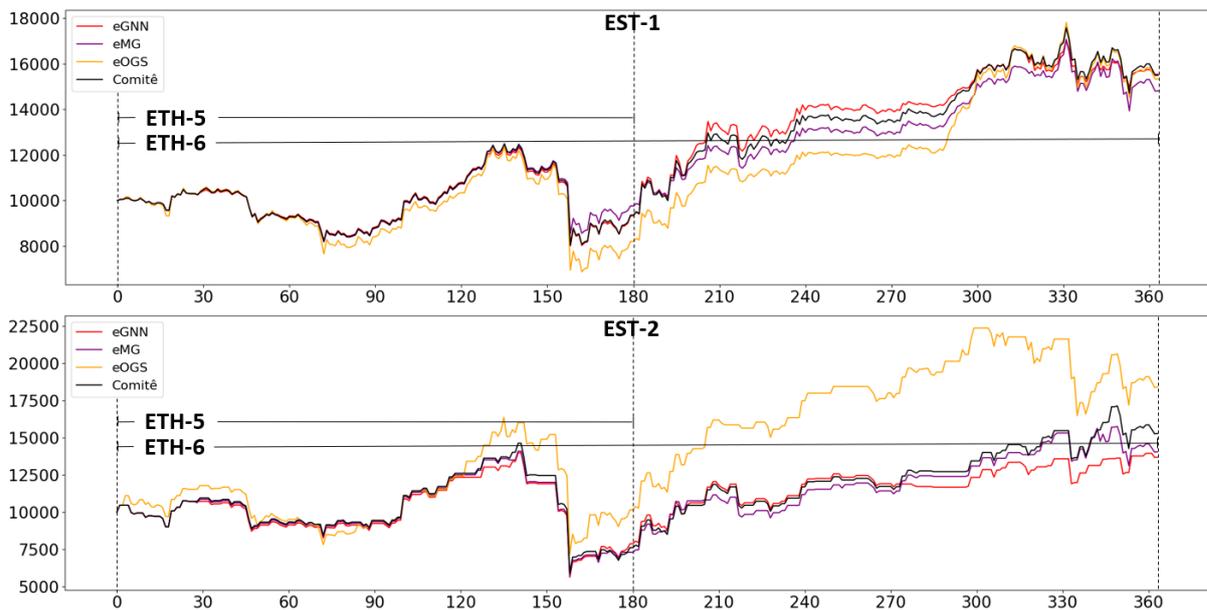


Figura 21 – Evolução da rentabilidade na recomendação de investimentos nos intervalos ETH-5 e ETH-6.



cada intervalo. Esta análise faz mais sentido em um contexto onde o investidor iniciará o uso da estratégia escolhida no primeiro dia do período e deixará de utilizá-la somente no último dia do respectivo intervalo. Entretanto, entende-se que o investidor poderia desistir da abordagem em algum momento intermediário, deslocando seus recursos para outra criptomoeda ou até mesmo outro tipo de ativo de forma antecipada. Dessa forma, é importante a realização de uma análise mais granular, que considere a possibilidade de que o investidor abandone a estratégia escolhida a qualquer momento. As figuras 16, 17, 18, 19, 20 e 21 ilustraram o desempenho dos modelos em termos de Rentabilidade,

trazendo uma noção visual acerca da evolução de desempenho neste aspecto ao longo dos intervalos. Neste momento, pretende-se aprofundar nos dados visualizados nestas figuras, com números que indiquem as possibilidades de ganho de forma mais explícita.

A Tabela 19 apresenta, para cada intervalo, o número de amostras (#) onde cada modelo alcançou desempenho superior aos demais. Entende-se que, quanto maior for este número para um dado modelo, mais chances o investidor teria de abandonar a estratégia de recomendação de investimentos em um momento onde o modelo escolhido apresenta Rentabilidade superior ao computado para os demais. Dessa maneira, estão destacados em negrito os modelos que apresentaram os melhores e piores resultados. Além disso, é apresentada a proporção (%) destas ocorrências em relação ao número total de amostras (#). Neste caso, para o cálculo do número total de amostras foram desconsideradas aquelas onde todos os modelos obtiveram resultados idênticos. Entretanto, quando dois modelos apresentaram o mesmo resultado, a amostra foi contabilizada para ambos. Ainda sobre esta tabela, é importante ressaltar que os intervalos BTC-3, BTC-5, ETH-3 e ETH-5 foram desconsiderados por estarem compreendidos nos intervalos BTC-4, BTC-6, ETH-4 e ETH-6, respectivamente.

Tabela 19 – Análise de evolução da rentabilidade: quando um modelo é superior aos demais.

Intervalo	A.	EST	eGNN		eMG		eOGS		Comitê	
			#	%	#	%	#	%	#	%
BTC-1	365	1	0	0,00	268	73,42	14	3,84	94	25,75
		2	171	46,85	99	27,12	32	8,77	73	20,00
BTC-2	468	1	417	89,10	50	10,68	3	0,64	6	1,28
		2	279	59,62	27	5,77	164	35,04	28	5,98
BTC-4	359	1	213	59,33	6	1,67	23	6,41	145	40,39
		2	123	34,26	23	6,41	5	1,39	218	60,72
BTC-6	363	1	64	17,63	142	39,12	110	30,30	53	14,60
		2	254	69,97	16	4,41	14	3,86	114	31,40
ETH-4	365	1	16	4,38	334	91,51	15	4,11	16	4,38
		2	319	87,40	2	0,55	15	4,11	156	42,74
ETH-6	363	1	221	60,88	55	15,15	46	12,67	65	17,91
		2	0	0,00	0	0,00	324	89,26	39	10,74

Para a análise da evolução da Rentabilidade, foram contabilizados 12 experimentos (seis intervalos, com duas estratégias cada). Nos dados apresentados é possível perceber que o modelo eGNN foi superior em um maior número de amostras que os demais em 7 dos 12 experimentos realizados, seguido pelo modelo eMG, que foi superior em 3 situações. O Comitê e o eOGS empataram, sendo superiores aos demais neste aspecto em apenas um intervalo. Percebe-se que o eMG conseguiu superar os demais apenas na aplicação da estratégia EST-1, em situações onde, geralmente, o modelo eGNN apresentou fraco desempenho. Outro fato relevante é que o modelo eGNN, detentor do melhor resultado na maioria dos casos, não conseguiu ser superior aos demais em qualquer momento para os intervalos BTC-1/EST-1 e ETH-6/EST-2. No total, para a estratégia EST-1, o modelo eGNN foi superior em 931 de 2283 amostras (40,78%), seguido por eMG (855 ou 37,45%), Comitê (379 ou 16,60%) e eOGS (211 ou 9,24%). Para a estratégia EST-2, o modelo eGNN também foi superior aos demais, o que foi registrado ao todo em 1146 amostras (50,20%), seguido

por Comitê (628 ou 27,51%), eOGS (554 ou 24,27%) e eMG (167 ou 7,31%).

Após a análise da probabilidade de o investidor abandonar a estratégia em um momento onde o modelo escolhido apresenta Rentabilidade superior aos demais, entende-se como necessário, neste mesmo sentido, analisar a probabilidade de que isso aconteça em um momento onde o modelo escolhido apresenta um resultado inferior. Portanto, a Tabela 20 apresenta, para cada intervalo, o número de amostras (#) e a proporção em relação ao total (%) onde cada modelo obteve desempenho inferior aos demais. Desta vez, os menores números obtidos indicam resultados melhores, visto que é desejável que determinado modelo se enquadre nesta situação no menor número de amostras possível.

Tabela 20 – Análise de evolução da rentabilidade: quando um modelo é inferior aos demais.

Intervalo	A.	EST	eGNN		eMG		eOGS		Comitê	
			#	%	#	%	#	%	#	%
BTC-1	365	1	108	29,59	0	0,00	257	70,41	0	0,00
		2	52	14,25	42	11,51	271	74,25	0	0,00
BTC-2	468	1	0	0,00	375	80,13	93	19,87	0	0,00
		2	0	0,00	385	82,26	80	17,09	3	0,64
BTC-4	359	1	6	1,67	57	15,88	302	84,12	6	1,67
		2	31	8,64	202	56,27	141	39,28	15	4,18
BTC-6	363	1	44	12,12	138	38,02	185	50,96	6	1,65
		2	0	0,00	310	85,40	53	14,60	0	0,00
ETH-4	365	1	8	2,19	14	3,84	350	95,89	8	2,19
		2	7	1,92	15	4,11	350	95,89	7	1,92
ETH-6	363	1	35	9,64	87	23,97	253	69,70	13	3,58
		2	238	65,56	110	30,30	17	4,68	47	12,95

Nesta nova análise proposta, o Comitê foi superior aos demais modelos em 10 dos 12 experimentos realizados (individual ou empate). Os dados apresentados confirmam ainda o bom desempenho do modelo eGNN também neste aspecto, visto que obteve os melhores resultados em 6 dos 12 experimentos. Entretanto, no intervalo ETH-6, este modelo apresentou Rentabilidade inferior aos demais em 65,56% das amostras. No total, para a estratégia EST-1, o Comitê foi o modelo que apresentou a pior Rentabilidade em apenas 33 de 2283 amostras (1,45%), seguido por eGNN (201 ou 8,80%), eMG (671 ou 29,39%) e eOGS (1440 ou 63,07%). Para a estratégia EST-2, esse fato também foi registrado para o Comitê em um menor número de amostras (72 ou 3,15%), seguido por eGNN (328 ou 14,37%), eOGS (912 ou 39,95%) e eMG (1064 ou 46,61%).

6 Considerações Finais

Neste trabalho, inicialmente, foi apresentada uma revisão de literatura sobre previsão de valores em séries temporais relacionadas à criptomoedas. A partir do estudo realizado, foi proposta uma nova abordagem que contempla a utilização de Sistemas *Fuzzy* Evolutivos na previsão do valor de fechamento, na previsão do sentido de variação e na recomendação de investimentos em criptomoedas. Na previsão do valor de fechamento foram avaliados três sistemas evolutivos e a previsão do sentido de variação foi realizada com base no valor previsto por estes sistemas. Para esta tarefa, também foi proposto um novo modelo, chamado de Comitê, que consiste na combinação dos resultados apresentados pelos três modelos evolutivos. Por último, foi proposta uma estratégia de recomendação de investimentos, que consiste na compra e/ou venda de criptomoedas, baseando-se na previsão do sentido de variação.

Experimentos computacionais para avaliar a abordagem proposta foram realizados com o Bitcoin (BTC) e com o Ether (ETH), as duas criptomoedas consideradas mais relevantes até o final de 2020, principalmente por apresentarem os maiores números no que tange ao volume transacionado diariamente e à capitalização de mercado. Os experimentos foram conduzidos, inicialmente, em 2 intervalos estudados em outros trabalhos publicados relacionados ao Bitcoin, o que possibilitou uma comparação com modelos aplicados por outros autores. Além disso, também foram realizados experimentos em 4 intervalos mais recentes, nos anos de 2019 e 2020, para o Bitcoin e para o Ether. Portanto, ao todo, foram utilizados 10 conjuntos de dados, 6 para o Bitcoin e 4 para o Ether.

A partir dos experimentos realizados, é possível concluir que os Sistemas *Fuzzy* Evolutivos postulam como bons candidatos na solução do problema de previsão do valor futuro e do sentido de variação do preço de criptomoedas. Os números apresentados pelos modelos estudados foram competitivos, principalmente em termos de Acurácia e Rentabilidade, quando comparados a outros trabalhos publicados. Destaca-se ainda que a proposta do Comitê é promissora, visto que promove, de forma nativa, uma espécie de compensação dos momentos em que um modelo apresenta desempenho inferior aos demais. Isto pôde ser confirmado, principalmente, na previsão do sentido de variação, onde o Comitê registrou 7 dos melhores resultados alcançados, sendo superior, inclusive, aos trabalhos utilizados como *benchmark*. Além disso, em termos de Rentabilidade, os modelos foram capazes de proporcionar ganhos superiores à valorização das criptomoedas em boa parte dos intervalos, e, até mesmo, evitar maiores prejuízos em cenários de desvalorização. Neste caso, percebe-se que a Rentabilidade contabilizada para o Comitê, geralmente, não o posiciona como o melhor nem tampouco como o pior modelo neste aspecto, sendo mais recomendado, aparentemente, para investidores de perfil moderado.

6.1 Propostas de Continuidade

Como proposta de continuidade para este trabalho, sugere-se:

- realizar testes com outros Sistemas *Fuzzy* Evolutivos pois acredita-se que, ao buscar por modelos que apresentem desempenho similar ou superior ao eGNN, o Comitê tende a apresentar resultados ainda melhores, com maiores garantias de sucesso na utilização da solução como auxílio na tomada de decisões acerca do investimento em criptomoedas;
- avaliar o desempenho do Comitê, a partir da inclusão de novos modelos evolutivos em sua composição;
- aprimorar a etapa de Seleção de Variáveis da abordagem proposta. Atualmente, as 8 variáveis mais relevantes são selecionadas na fase de Ranqueamento e, além disso, são formados grupos com até 4 elementos na busca dos melhores conjuntos de variáveis para cada modelo evolutivo. Acredita-se que a parametrização destes valores (8 e 4) poderia sensibilizar o modelo de forma a apresentar melhores resultados;
- avaliar os resultados a partir da execução da etapa de Seleção de Variáveis de forma periódica ou utilizar modelos que implementam nativamente mecanismos para seleção adaptativa de entradas, como, por exemplo, o trabalho de [Silva et al. \(2015\)](#);
- verificar o desempenho a partir da utilização de outros atributos e indicadores técnicos na construção do conjunto de dados históricos utilizado como entrada para os modelos evolutivos;
- aplicar a abordagem proposta em outros conjuntos de dados para as criptomoedas Bitcoin e Ether. Além disso, neste mesmo sentido, também é recomendada a aplicação da abordagem proposta em conjuntos de dados relacionados a outras criptomoedas de relevância, como Litecoin (LTC) e Ripple (XRP), para que a capacidade de generalização da solução seja avaliada de forma ainda mais profunda;
- realizar um refinamento da abordagem de recomendação de investimentos, passando a considerar, por exemplo, limiares de compra e/ou venda variáveis de acordo com a quantidade de modelos que forneceram a mesma resposta para uma observação específica. Desta forma, a estratégia poderia ser mais agressiva em momentos onde todos os modelos retornam a mesma previsão ou menos agressiva em momentos onde as respostas são divergentes;
- explorar a saída linguística fornecida pelo Sistemas *Fuzzy* Evolutivos que oferecem esse tipo de resposta, utilizando-a, por exemplo, na definição do parâmetro *limiar* da estratégia de recomendação de investimentos.

Referências

- ADCOCK, R.; GRADOJEVIC, N. Non-fundamental, non-parametric bitcoin forecasting. **Physica A: Statistical Mechanics and its Applications**, v. 531, 2019. Citado 2 vezes nas páginas 52 e 103.
- AGGARWAL, D.; CHANDRASEKARAN, S.; ANNAMALAI, B. A complete empirical ensemble mode decomposition and support vector machine-based approach to predict bitcoin prices. **Journal of Behavioral and Experimental Finance**, v. 27, 2020. Citado na página 47.
- AKYILDIRIM, E.; CORBET, S. Forecasting mid-price movement of bitcoin futures using machine learning. **Operations Research**, 2020. Citado 2 vezes nas páginas 42 e 50.
- ALONSO-MONSALVE, S. et al. Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. **Expert Systems with Applications**, v. 149, 2020. Citado 2 vezes nas páginas 40 e 48.
- AMADEO, K. **thebalance.com**. 2020. <https://www.thebalance.com/lehman-brothers-collapse-causes-impact-4842338>. Acessado em: 06/01/2020. Citado na página 1.
- AMARAL, V. L.; SILVA, A. M. Sistemas fuzzy evolutivos na classificação de valores futuros do bitcoin. VI Congresso Brasileiro de Sistemas Fuzzy (CBSF 2021). Submetido. 2021. Citado na página 7.
- ANGELINI, G.; EMILI, S. Forecasting cryptocurrencies: A comparison of garch models. **Econometric Modeling: International Financial Markets - Volatility Financial Crises eJournal**, v. 15, 2018. Citado 2 vezes nas páginas 42 e 52.
- ANTONOPOULOS, A.; WOOD, G. **Mastering Ethereum: Building Smart Contracts and Dapps**. Sebastopol, California, USA: OReilly, 2018. ISBN 978-1491971949. Citado na página 31.
- ATSALAKIS, G. et al. Bitcoin price forecasting with neuro-fuzzy techniques. **European Journal of Operational Research**, v. 276, n. 2, p. 770–780, 2019. Citado 5 vezes nas páginas 36, 39, 40, 43 e 51.
- AYGUN, B.; KABAKCI, E. Comparison of statistical and Machine Learning algorithms for forecasting daily Bitcoin returns. **European Journal of Science and Technology**, v. 9, n. 21, p. 44 –454, 2021. Citado 2 vezes nas páginas 42 e 47.
- BELLMAN, R. Dynamic programming. **Science**, v. 153, p. 34–37, 1966. Citado na página 57.
- BILLAH, M.; WAHEED, S.; HANIFA, A. Predicting closing stock price using artificial neural network and adaptive neuro fuzzy inference system (anfis): The case of the dhaka stock exchange. **International Journal of Computer Applications**, v. 129, 2015. Citado na página 43.
- BLOCKCHAIN. **blockchain.com**. 2021. <https://www.blockchain.com/>. Acessado em: 02/04/2021. Citado 2 vezes nas páginas 28 e 29.

BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**, v. 31, n. 3, p. 307–327, 1986. Citado na página 42.

BOLLERSLEV, T.; ENGLE, R. F.; WOOLDRIDGE, J. M. A capital asset pricing model with time-varying covariances. **Journal of Political Economy**, v. 96, n. 1, 1988. Citado na página 42.

BORGES, T.; NEVES, R. Ensemble of machine learning algorithms for cryptocurrency investment with different data resampling methods. **Applied Soft Computing Journal**, v. 90, 2020. Citado 3 vezes nas páginas 40, 43 e 47.

BOUCHACHIA, A.; LUGHOFER, E.; SAYED-MOUCHAWEH, M. Special issue: Evolving soft computing techniques and applications. **Applied Soft Computing**, v. 14, p. 141–143, 2014. Citado na página 17.

BOX, G.; JENKINS, G. **Time Series Analysis: Forecasting and Control**. Hoboken, Nova Jersey, EUA: Wiley, 1976. Citado 5 vezes nas páginas 9, 10, 13, 14 e 42.

BREIMAN, L. Bagging predictors. **Machine Learning**, v. 24, p. 123—140, 1996. Citado na página 15.

BROOKS, C. **Introductory Econometrics for Finance**. Cambridge, United Kingdom: Cambridge University Press, 2014. ISBN 978-1107661455. Citado na página 11.

BUTERIN, V. **A Next Generation Smart Contract Decentralized Application Platform**. [S.l.]: ethereum.org, 2013. https://blockchainlab.com/pdf/Ethereum_white_paper_a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf. Acessado em: 31/03/2021. Citado na página 30.

CATANIA, L.; GRASSI, S.; RAVAZZOLO, F. Forecasting cryptocurrencies under model and parameter instability. **International Journal of Forecasting**, v. 35, p. 485–501, 2019. Citado 2 vezes nas páginas 34 e 35.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: **XXII International Conference on Knowledge Discovery and Data Mining (SIGKDD 2016)**. San Francisco, USA: ACM, 2016. p. 785–794. Citado 2 vezes nas páginas 16 e 42.

CHEN, Z.; LI, C.; SUN, W. Bitcoin price prediction using machine learning: An approach to sample dimension engineering. **Journal of Computational and Applied Mathematics**, v. 365, 2020. Citado 8 vezes nas páginas 36, 39, 42, 49, 100, 101, 102 e 104.

CHIH-HUNG, W. et al. A new forecasting framework for bitcoin price with lstm. In: **IEEE. 2019 IEEE International Conference on Data Mining Workshops (ICDMW)**. Hong Kong, 2019. p. 168–175. Citado na página 52.

CHRISTOFOROU, E.; EMIRIS, I.; FLORAKIS, A. Neural networks for cryptocurrency evaluation and price fluctuation forecasting. In: SPRINGER. **Mathematical Research for Blockchain Economy**. Cham, Switzerland, 2020. v. 1, p. 133–149. Citado 5 vezes nas páginas 37, 40, 47, 100 e 104.

COINMARKETCAP. 2021. <https://coinmarketcap.com>. Acessado em: 10/01/2021. Citado 10 vezes nas páginas 5, 26, 27, 29, 30, 32, 34, 41, 65 e 67.

CONLON, T.; MCGEE, R. Safe haven or risky hazard? bitcoin during the covid-19 bear market. **Finance Research Letters**, v. 35, 2020. Citado na página 4.

CRYER, J. D.; CHAN, K. **Time Series Analysis With Applications in R**. New York, USA: Springer-Verlag, 2008. ISBN 978-0-387-75958-6. Citado na página 14.

DERBENTSEV, V. et al. Comparative performance of machine learning ensemble algorithms for forecasting cryptocurrency prices. **International Journal of Engineering, Transactions A: Basics**, v. 34, 2021. Citado 2 vezes nas páginas 46 e 100.

DERBENTSEV, V. et al. Forecasting cryptocurrency prices time series using machine learning approach. In: **VIII International Conference on Monitoring, Modeling Management of Emergent Economy (M3E2 2019)**. Web Conference: EDP Sciences, 2019. v. 65, p. 320–334. Citado na página 51.

DESEV, K.; KABAIVANOV, S.; DESEVN, D. Forecasting cryptocurrency markets through the use of time series models. **Business and Economic Horizons (BEH)**, v. 15, p. 242–253, 2019. Citado na página 51.

DING, Z.; GRANGER, C. W.; ENGLE, R. F. A long memory property of stock market returns and a new model. **Journal of Empirical Finance**, v. 1, n. 1, p. 83–106, 1993. Citado na página 42.

DUDA, R. O.; STORK, D. G.; HART, P. E. **Pattern Classification**. Hoboken, Nova Jersey, EUA: Wiley-Interscience, 2000. ISBN 978-0471056690. Citado na página 41.

ENGLE, R. F. Autoregressive Conditional Heteroskedasticity with estimates of the variance of United Kingdom inflation. **Econometrica**, v. 50, n. 4, p. 987–1007, 1982. Citado na página 42.

FAHMI, A. et al. Regression based analysis for bitcoin price prediction. **International Journal of Engineering Technology**, v. 7, 2018. Citado 3 vezes nas páginas 39, 42 e 53.

FAUZI, M. A.; YUNIARTI, A. Ensemble method for indonesian twitter hate speech detection. **Indonesian Journal of Electrical Engineering and Computer Science**, v. 11, n. 1, p. 294–299, 2018. Citado na página 16.

FREUND, Y.; SCHAPIRE, R. E. A short introduction to boosting. **Journal of Japanese Society for Artificial Intelligence**, v. 14, n. 5, p. 771–780, 1999. Citado na página 16.

FRISIANI, N. **Deep Learning Algorithms for Forecasting Cryptocurrency Time Series**. Dissertação (Mestrado) — The University of Manchester, Manchester, England, 4 2019. Citado na página 12.

GALDI, F. C.; PEREIRA, L. M. . Value at Risk (VaR) using volatility forecasting models: EWMA, GARCH and Stochastic Volatility. **Brazilian Business Review**, v. 4, 2007. Citado na página 42.

GARCIA, C.; LEITE, D.; SKRJANC, I. Incremental missing-data imputation for evolving fuzzy granular prediction. **IEEE Transactions on Fuzzy Systems**, IEEE, 2019. Citado na página 43.

- GIBBS, T.; YORDCHIM, S. Thai perception on Litecoin value. **International Journal of Social, Education, Economics and Management Engineering**, v. 8, 2014. Citado 2 vezes nas páginas 29 e 30.
- GLOSTEN, L.; JAGANNATHAN, R.; RUNKLE, D. On the relation between the expected value and the volatility of the nominal excess return on stocks. **Journal of Finance**, v. 48, n. 5, p. 1779–1801, 1993. Citado na página 42.
- HO, T. K. Random decision forests. In: **III International Conference on Document Analysis and Recognition (ICDAR'95)**. Washington, USA: IEEE Computer Society, 1995. Citado 2 vezes nas páginas 16 e 42.
- HYNDMAN, R.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. [S.l.]: OTexts, 2018. Citado 2 vezes nas páginas 10 e 11.
- INACIO, M.; LEMOS, A.; CAMINHAS, W. Fault diagnosis with Evolving Fuzzy Classifier based on Clustering Algorithm and Drift Detection. **Mathematical Problems in Engineering**, v. 2015, n. 3, p. 1–14, 2015. Citado na página 22.
- INVESTING. **Investing.com**. 2021. <https://br.investing.com>. Acessado em: 05/01/2021. Citado 2 vezes nas páginas 1 e 3.
- JAGANNATH, N. et al. A Self-Adaptive Deep Learning-Based Algorithm for predictive analysis of Bitcoin price. **IEEE Access**, v. 9, p. 34054–34066, 2021. Citado 6 vezes nas páginas 36, 38, 46, 101, 102 e 103.
- JAIN, A. et al. Forecasting price of cryptocurrencies using tweets sentiment analysis. In: **2018 Eleventh International Conference on Contemporary Computing (IC3 2018)**. Noida, India: IEEE, 2018. p. 1–7. Citado 2 vezes nas páginas 30 e 39.
- JANA, R.; GHOSH, I.; DAS, D. A differential Evolution-Based Regression Framework for forecasting Bitcoin price. **Annals of Operations Research**, p. 1–26, 2021. Citado na página 46.
- JANG, H.; LEE, J. An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. **IEEE Access**, v. 6, p. 5427–5437, 2017. Citado na página 38.
- JANG, J. Anfis adaptive-network-based fuzzy inference system. **IEEE Transactions on Systems Man and Cybernetics**, v. 23, p. 665–685, 1993. Citado na página 43.
- JI, S.; KIM, J.; IM, H. A comparative study of bitcoin price prediction using deep learning. **Mathematics**, v. 7, 2019. Citado 11 vezes nas páginas 37, 51, 62, 63, 65, 72, 76, 96, 100, 101 e 102.
- KACPRZYK, J.; PEDRYCZ, W. **Springer Handbook of Computational Intelligence**. Berlin, Germany: Springer, 2015. v. 2. Citado na página 18.
- KASABOV, N.; FILEV, D. Evolving intelligent systems: Methods, learning applications. In: **2006 International Symposium on Evolving Fuzzy Systems**. Ambleside, United Kingdom: Institute of Electrical and Electronics Engineers (IEEE), 2006. p. 8–18. Citado na página 17.

KRISTOUFEK, L. Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. **Scientific Reports**, v. 3, n. 3415, 2013. Citado na página 39.

KRUSKAL W. H., W. W. Use of ranks in one-criterion variance analysis. **Journal of the American Statistical Association**, v. 47, n. 260, 1952. Citado 2 vezes nas páginas 41 e 57.

KURSA, M.; RUDNICKI, W. Feature selection with the boruta package. **Journal of Statistical Software**, v. 36, 2010. Citado na página 41.

LABS, R. **Santander and Ripple Bring Faster, Cheaper Cross-Border Payments to Millions of Retail and Commercial Customers**. 2020. <https://ripple.com/customer-case-study/santander>. Acessado em: 31/03/2021. Citado na página 30.

LAHMIRI, S.; BEKIROS, S. Intelligent forecasting with machine learning trading systems in chaotic intraday bitcoin market. **Chaos, Solitons and Fractals**, v. 133, 2020. Citado na página 50.

LEE, G. J.; ENGLE, R. F. A permanent and transitory component model of stock return volatility. **Social Science Research Network (SSRN)**, 1993. Citado na página 42.

LEITE, D. Comparison of genetic and incremental learning methods for neural network-based electrical machine fault detection. In: LUGHOFER, E.; SAYED-MOUCHAWEH, M. (Ed.). **Predictive Maintenance in Dynamic Systems**. Cham, Switzerland: Springer, 2019. Citado na página 20.

LEITE, D. et al. Optimal rule-based granular systems from data streams. **IEEE Transactions on Fuzzy Systems**, v. 28, p. 583–296, 2020. Citado 3 vezes nas páginas 18, 22 e 23.

LEITE, D.; COSTA, P.; GOMIDE, F. Evolving granular classification neural networks. In: **IEEE International Joint Conference on Neural Networks (IJCNN'09)**. Atlanta, USA: Institute of Electrical and Electronics Engineers (IEEE), 2009. p. 1736–1743. Citado na página 20.

LEITE, D.; COSTA, P.; GOMIDE, F. Evolving granular neural network for semi-supervised data stream classification. In: **The 2010 International Joint Conference on Neural Networks (IJCNN 2010)**. Barcelona, Spain: Institute of Electrical and Electronics Engineers (IEEE), 2010. p. 1–8. Citado na página 18.

LEITE, D.; COSTA, P.; GOMIDE, F. Evolving granular neural network for fuzzy time series forecasting. In: **The 2012 International Joint Conference on Neural Networks (IJCNN 2012)**. Brisbane, Australia: Institute of Electrical and Electronics Engineers (IEEE), 2012. p. 1–8. Citado na página 20.

LEITE, D.; COSTA, P.; GOMIDE, F. Evolving granular neural networks from fuzzy data streams. **Neural Networks**, v. 38, p. 1–16, 2013. Citado 2 vezes nas páginas 18 e 20.

LEITE, D. et al. Fuzzy granular neural network for incremental modeling of nonlinear chaotic systems. In: **2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2016)**. Vancouver, Canada: Institute of Electrical and Electronics Engineers (IEEE), 2016. p. 64–71. Citado na página 20.

LEITE, D.; ŠKRJANC, I. Ensemble of evolving optimal granular experts, owa aggregation, and time series prediction. **Information Sciences**, v. 504, p. 95–112, 2019. Citado 3 vezes nas páginas 18, 22 e 23.

- LEITE, L.; SKRJANC, I.; GOMIDE, F. An overview on evolving systems and learning from stream data. **Evolving Systems**, v. 11, p. 181–198, 2020. Citado 2 vezes nas páginas 5 e 17.
- LEMOS, A. **Modelagem Nebulosa Evolutiva: Novas Topologias e Algoritmos de Aprendizagem**. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, R. Dr. Roberto Frias 464, Porto, Portugal, 2 2011. Citado 4 vezes nas páginas 17, 18, 21 e 22.
- LEMOS, A.; CAMINHAS, W.; GOMIDE, F. Multivariable gaussian evolving fuzzy modeling system. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 19, n. 1, p. 91–104, 2011. Citado 2 vezes nas páginas 18 e 20.
- LIM, C.; SEK, S. Comparing the performances of garch-type models in capturing the stock market volatility in malaysia. **Procedia Economics and Finance**, v. 5, p. 478–487, 2013. ISSN 2212-5671. International Conference On Applied Economics (ICOAE) 2013. Citado na página 42.
- LINCY, R. A multiple fuzzy inference systems framework for daily stock trading with application to nasdaq stock exchange. **Expert Systems with Applications**, v. 44, p. 13–21, 2016. ISSN 0957-4174. Citado na página 43.
- LUCARELLI, G.; BORROTTI, M. A deep reinforcement learning approach for automated cryptocurrency trading. In: **Proceedings of the 2019 Artificial Intelligence Applications and Innovations (AIAI 2019)**. Crete, Greece: Springer, Cham, 2019. p. 1–6. Citado na página 36.
- LUGHOFER, E. On-line assurance of interpretability criteria in evolving fuzzy systems – achievements, new concepts and open issues. **Information Sciences**, v. 251, p. 22–46, 2013. Citado na página 17.
- MACIEL, L. Cryptocurrencies value-at-risk and expected shortfall: Do regime-switching volatility models improve forecasting? **International Journal of Finance and Economics**, 2020. Citado na página 50.
- MACIEL, L.; BALLINI, R.; GOMIDE, R. Adaptive interval fuzzy modeling from stream data and application in cryptocurrencies forecasting. In: **Proceedings of the Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) 2020**. Redmond, USA: NAFIPS, 2020. v. 1, p. 1–12. Citado 2 vezes nas páginas 43 e 51.
- MAKATJANE, K.; MOLEFE, E.; WYK, R. The analysis of the 2008 us financial crisis: An intervention approach. **Journal of Economics and Behavioral Studies**, Taylor & Francis, v. 10, n. 1, p. 59–68, 2018. Citado na página 1.
- MALLQUI, D.; FERNANDES, R. Predicting the direction, maximum, minimum and closing prices of daily bitcoin exchange rate using machine learning techniques. **Applied Soft Computing Journal**, v. 75, p. 596–606, 2019. Citado 8 vezes nas páginas 35, 38, 41, 52, 65, 72, 103 e 104.
- MCNALLY, S.; ROCHE, J.; CATON, S. Predicting the price of bitcoin using machine learning. In: **26th Euromicro International Conference on Parallel, Distributed and Network-based Processing**. Cambridge, United Kingdom: Institute of Electrical and Electronics Engineers (IEEE), 2018. v. 1, p. 339–343. Citado 2 vezes nas páginas 41 e 53.

- MORETTIN, P.; TOLOI, C. **Análise de Séries Temporais**. [S.l.]: Blucher, 2006. Citado 3 vezes nas páginas [9](#), [11](#) e [12](#).
- MORGAN, J. N.; SONQUIST, J. A. Problems in the analysis of survey data, and a proposal. **Journal of the American Statistical Association**, v. 58, n. 302, p. 415–434, 1963. Citado na página [42](#).
- MUDASSIR, M. et al. Time-series forecasting of bitcoin prices using high-dimensional features: a machine learning approach. **Neural Computing and Applications**, 2020. Citado 8 vezes nas páginas [37](#), [39](#), [40](#), [41](#), [48](#), [101](#), [102](#) e [103](#).
- NAKAMOTO, S. Bitcoin: A Peer-to-Peer Electronic Cash System. Acessado:02/03/2021. 2008. Citado 3 vezes nas páginas [1](#), [26](#) e [27](#).
- NELSON, D. B. Conditional heteroskedasticity in asset returns: A new approach. **Econometrica**, v. 59, n. 2, p. 347–370, 1991. Citado na página [42](#).
- O'DWYER, K.; MALONE, D. Bitcoin mining and its energy footprint. In: **China-Ireland International Conference on Information and Communications Technologies**. Limerick, Ireland: Institute of Electrical and Electronics Engineers (IEEE), 2014. v. 2, p. 844–849. Citado na página [27](#).
- PABUCCU, H.; ONGAN, S.; ONGAN, A. Forecasting the movements of bitcoin prices: an application of machine learning algorithms. **Quantitative Finance and Economics**, v. 4, 2020. Citado 2 vezes nas páginas [40](#) e [49](#).
- PADMAVATHI, M.; SURESH, R. Secure p2p intelligent network transaction using litecoin. **Mobile Netw Appl**, v. 24, p. 318–326, 2019. Citado na página [29](#).
- PARIKAKIS, G. S.; MERIKA, A. Evaluating volatility dynamics and the forecasting ability of markov switching models. **Journal of Forecasting**, 2009. Citado na página [42](#).
- PARK, Y. et al. Fighting the war against covid-19 via cell-based regenerative medicine: Lessons learned from 1918 spanish flu and other previous pandemics. **Stem Cell Reviews and Reports volume**, v. 17, p. 9–32, 2021. Citado na página [4](#).
- POLEDICA, A.; PETROVIC, B.; MILOSEVIC, P. Forecasting cryptocurrency time series using fuzzy transform , fourier transform and fuzzy inference system. In: **Proceedings of the 11th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2019)**. Czech Republic: Atlantis Press, 2019. p. 634–640. Citado na página [52](#).
- PRAVEENA, M. A literature review on supervised machine learning algorithms and boosting process. **International Journal of Computer Applications**, v. 169, p. 32–35, 2017. Citado na página [14](#).
- RAJAB, S.; SHARMA, V. An interpretable neuro-fuzzy approach to stock price forecasting. **Soft Comput**, v. 23, p. 921–936, 2019. Citado na página [43](#).
- SALMAN, R.; KECMAN, V. Regression as classification. In: **2012 Proceedings of IEEE Southeastcon**. Orlando, Florida, USA: Institute of Electrical and Electronics Engineers (IEEE), 2012. Citado na página [33](#).

SAVCHEKO M., T. V. A. O. An approach to transaction delegation in self-protected decentralized data platforms. In: **19th International Scientific and Practical conference: Information Technologies and Security (ITS 2019)**. Kyiv, Ukraine: CEUR-WS, 2019. Citado na página 31.

SCHAPIRE, R. E. The strength of weak learnability. **Machine Learning**, v. 5, n. 2, p. 197–227, 1990. Citado na página 16.

SCHAPIRE, R. E. The boosting approach to machine learning: An overview. In: DENISON, D. D. et al. (Ed.). **Nonlinear Estimation and Classification. Lecture Notes in Statistics**. New York, NY, USA: Springer, 2003. Citado na página 16.

SEBASTIÃO, H.; GODINHO, P. Forecasting and trading cryptocurrencies with machine learning under changing market conditions. **Financial Innovation**, v. 7, n. 3, 2021. Citado 7 vezes nas páginas 35, 37, 46, 100, 101, 103 e 104.

SILVA, A. **Sistemas Neuro-Fuzzy Evolutivos: Novos Algoritmos de Aprendizado e Aplicações**. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, Avenida Antônio Carlos, 6627 - Escola de Engenharia - Bloco I sala 2303 CEP 31270-901, 4 2014. Citado na página 17.

SILVA, A. M. et al. Adaptive input selection and evolving neural fuzzy networks modeling. **International Journal of Computational Intelligence Systems**, v. 8, n. 1, p. 3–14, 2015. Citado na página 85.

SILVA, R. et al. Multi-step ahead bitcoin price forecasting based on vmd and ensemble learning methods. In: **2020 International Joint Conference on Neural Networks (IJCNN)**. Glasgow, United Kingdom: Institute of Electrical and Electronics Engineers (IEEE), 2020. v. 1, p. 1–8. Citado na página 49.

SOUZA, L. et al. Thermal modeling of power transformers using evolving fuzzy systems. **Engineering Applications of Artificial Intelligence**, v. 25, n. 5, p. 980–988, 2012. Citado na página 22.

STINIS, P. Enforcing constraints for time series prediction in supervised, unsupervised and reinforcement learning. Acessado: 04/03/2021. 2019. Citado na página 14.

TRUCiOS, C.; TAYLOR, J. Forecasting value-at-risk and expected shortfall of cryptocurrencies using combinations based on jump-robust and regime-switching models. **SSRN Electronic Journal**, 2020. Citado na página 48.

TSAY, R. **Analysis of Financial Time Series**. New Jersey, USA: John Wiley & Sons, 2010. ISBN 978-0470414354. Citado 3 vezes nas páginas 9, 13 e 34.

TUNG, S. W.; QUEK, C.; GUAN, C. et2fis: An evolving type-2 neural fuzzy inference system. **Information Sciences**, v. 220, p. 124–148, 2013. Citado na página 17.

ULRICH, F. **Bitcoin: A moeda da era digital**. São Paulo, Brazil: LVM Editora, 2014. ISBN 978-8581190761. Citado 2 vezes nas páginas 2 e 27.

VAGROPOULOS, S. I. et al. Comparison of sarimax, sarima, modified sarima and ann-based models for short-term pv generation forecasting. In: **2016 IEEE International Energy Conference (ENERGYCON)**. Ku Leuven, Belgium: Institute of Electrical and Electronics Engineers (IEEE), 2016. p. 1–6. Citado na página 14.

- VIANEZ, J.; ROMAN, C.; GOMEZ, R. Economic policy uncertainty and bitcoin. is bitcoin a safe-haven asset? **European Journal of Management and Business Economics**, v. 29, n. 3, p. 347–363, 2020. Citado na página 3.
- WAHRSTATTER, A. Stablecoin billionaires - a descriptive analysis of the ethereum-based stablecoin ecosystem. **Social Science Research Network (SSRN)**, 2021. Citado na página 31.
- WANG, Y. et al. Stacking-based ensemble learning of decision trees for interpretable prostate cancer detection. **Applied Soft Computing**, v. 77, p. 188–204, 2019. Citado na página 15.
- WHO. **covid19.who.int**. 2021. <https://covid19.who.int>. Acessado em: 10/01/2021. Citado na página 4.
- WOLPERT, D. H. Stacked generalization. **Neural Networks**, v. 5, n. 2, p. 241–259, 1992. Citado na página 16.
- YAGER, R. A model of participatory learning. **IEEE Transactions on Systems Man and Cybernetics**, v. 20, n. 5, p. 1229–1234, 1990. Citado 2 vezes nas páginas 20 e 21.
- YAZIZA, S. et al. The performance of hybrid arima-garch modeling in forecasting gold price. In: **20th International Congress on Modelling and Simulation (MODSIM 2013)**. Adelaide, Australia: Australian Mathematical Sciences Institute, 2013. Citado na página 42.
- ZOUMPEKAS, T.; HOUSTIS, E.; VAVALIS, M. Eth analysis and predictions utilizing deep learning. **Expert Systems with Applications**, v. 162, 2020. Citado na página 49.

Apêndices

APÊNDICE A – Detalhamento Numérico dos Intervalos

A Tabela 21 apresenta dados numéricos acerca do valor de fechamento das criptomoedas para cada período utilizado na realização dos experimentos, considerando apenas os Intervalos de Interesse. Nessa tabela são apresentados os valores inicial e final, isto é, o valor de fechamento no primeiro e no último dia de cada intervalo. Também são apresentados o maior e o menor valor de fechamento alcançados pela criptomoeda ao longo do intervalo, além da média, mediana e desvio padrão. Em relação aos 10 intervalos representados, pode-se destacar:

- a maior valorização, considerando o valor inicial e final, em termos percentuais e valores absolutos, ocorreu no intervalo BTC-3, quando foi registrado um aumento de 187,24% no valor do Bitcoin, equivalente a US\$ 6.742,24. Um fato interessante é que isso ocorreu em um dos intervalos com o menor número de amostras (180), indicando que houve um aumento expressivo no valor da criptomoeda em um curto espaço de tempo, o que fica evidente na Figura 11;
- a maior desvalorização, considerando o valor inicial e final, em termos percentuais, ocorreu no intervalo ETH-5, quando o Ether sofreu uma desvalorização de 21,58%, equivalente a US\$ 39,10. Considerando valores absolutos, a maior desvalorização ocorreu no intervalo BTC-5, onde o Bitcoin perdeu US\$ 1.512,23 em seu valor, uma desvalorização de 18,34%;
- no intervalo BTC-2, proposto em [Ji, Kim e Im \(2019\)](#), considerando o valor inicial e o maior valor alcançado, o Bitcoin chegou a registrar um aumento de aproximadamente 400%. Entretanto, ao final do período, o resultado foi uma desvalorização de 4,17%. O desvio padrão calculado reforça a alta dispersão nos dados da série;
- o Bitcoin chegou a sofrer uma desvalorização de 39,72% nos intervalos BTC-5 e BTC-6, quando foi registrada uma queda de US\$ 4.970,79 em seu valor, períodos estes que contemplam o dia 11/03/2020. Entretanto, ao final destes intervalos, a variação final foi de -18,34% (US\$ -1.512,23) e +30,90% (US\$ +2.547,72), respectivamente, evidenciando uma recuperação progressiva da criptomoeda. Algo semelhante ocorreu com Ether nos intervalos correspondentes ETH-5 e ETH-6, nos quais chegou a ser registrada uma desvalorização de 38,95% no preço de fechamento da criptomoeda. Entretanto, ao final destes intervalos, foram registradas variações de -21,58% (US\$ -39,10) e +95,35% (US\$ +172,77), respectivamente.

Uma visão diferente dos intervalos é apresentada na Tabela 22. Nela é possível

visualizar informações acerca do número de variações positivas ou negativas sucessivas em cada um dos 10 Intervalos de Interesse. Algumas das informações que podem ser extraídas dessa tabela são:

- no intervalo BTC-6, o sentido de variação do preço de fechamento do Bitcoin não se repetiu em 36,44% (133/365) das amostras, isto é, 65 variações positivas foram seguidas por variações negativas e 68 variações negativas foram seguidas por variações positivas;
- no intervalo ETH-3, o sentido de variação do preço de fechamento do Ether se repetiu por até uma vez em 65,56% (118/180) das amostras, isto é, nestas amostras aconteceram no máximo dois aumentos ou duas reduções sucessivas no valor de fechamento da criptomoeda;
- ao somar todas as amostras existentes nos 10 intervalos estudados, o sentido de variação do preço da respectiva criptomoeda não se repetiu em 29,46% (1206/4093) das amostras;
- dentre os intervalos estudados, no intervalo BTC-2 foi registrada a maior ocorrência de repetição no sentido de variação do preço de uma criptomoeda. Neste intervalo foi registrado o aumento do preço de fechamento do Bitcoin uma vez por 10 dias consecutivos. Considerando que este intervalo é composto por 468 amostras, isto representa 2,14% de todo o intervalo;

Tabela 21 – Dados numéricos do valor de fechamento ao longo dos intervalos utilizados nos experimentos.

Nome	Período	Tamanho	Início	Fim	Menor	Maior	Média	Mediana	Desvio
BTC-1	01/04/2016 à 31/03/2017	365	417,96	1.071,79	417,96	1.274,99	722,78	663,25	663,25
BTC-2	20/09/2017 a 31/12/2018	468	3.905,50	3.742,70	3.236,76	19.497,40	7.867,32	6.961,79	3.109,59
BTC-3	24/01/2019 à 22/07/2019	180	3.600,87	10.343,11	3.399,47	13.016,23	6.460,47	5.309,17	2.783,13
BTC-4	24/01/2019 à 23/01/2020	365	3.600,87	8.406,52	3.399,47	13.016,23	7.673,91	8.047,53	2.468,30
BTC-5	07/10/2019 à 03/04/2020	180	8.245,62	6.733,39	4.970,79	10.326,05	8.116,29	8.205,76	1.182,87
BTC-6	07/10/2019 à 05/10/2020	365	8245,62	10793,34	4.970,79	12.554,40	8.933,96	9.132,49	1.516,29
ETH-3	24/01/2019 à 22/07/2019	180	117,36	217,56	104,53	336,75	191,76	167,73	63,97
ETH-4	24/01/2019 à 23/01/2020	360	117,36	162,93	104,53	336,75	182,88	174,22	49,44
ETH-5	07/10/2019 à 03/04/2020	180	181,19	142,09	110,61	284,22	173,15	170,24	40,06
ETH-6	07/10/2019 à 05/10/2020	365	181,19	353,96	110,61	477,05	226,73	202,95	85,18

Tabela 22 – Aumentos (+) ou diminuições (-) sucessivas ao longo dos intervalos utilizados nos experimentos.

Intervalo	Total	1	2	3	4	5	6	7	8	9	10
BTC-1	+	222	31	23	19	6	4	-	-	-	-
	-	143	58	21	7	3	-	-	-	-	-
BTC-2	+	249	58	28	14	9	2	1	1	-	1
	-	219	61	33	15	6	3	-	-	-	-
BTC-3	+	107	21	7	8	2	2	-	1	-	-
	-	73	25	15	6	-	-	-	-	-	-
BTC-4	+	190	50	16	13	3	2	1	1	-	-
	-	175	43	29	14	4	1	-	-	-	-
BTC-5	+	85	32	8	5	3	-	-	-	-	-
	-	95	27	14	5	2	-	1	-	-	-
BTC-6	+	186	65	19	11	8	-	-	1	-	-
	-	179	68	21	9	5	-	1	-	-	-
ETH-3	+	94	31	13	4	5	-	-	-	-	-
	-	86	33	14	7	1	-	-	-	-	-
ETH-4	+	179	60	24	10	7	-	-	1	-	-
	-	186	53	29	16	3	-	-	-	-	-
ETH-5	+	89	28	11	7	2	-	-	-	-	-
	-	91	24	16	7	1	-	-	-	-	-
ETH-6	+	193	57	22	13	6	2	1	-	-	-
	-	172	60	24	14	3	-	-	-	-	-

Anexos

ANEXO A – Conjuntos de Atributos Utilizados nos Trabalhos de Referência

A.1 Dados de Cotação

- **Close:** valor diário de fechamento do mercado (*USD*).
- **High:** maior valor alcançado no dia (*USD*).
- **Low:** menor valor alcançado no dia (*USD*).
- **Open:** valor diário de abertura do mercado (*USD*).
- **Price:** valor da criptomoeda no momento atual (*USD*). Usado principalmente em previsões de alta frequência (quando o modelo foi projetado para oferecer como saída previsões a cada minuto ou hora, por exemplo).

A.2 Dados de Mercado

- **Market Capitalization:** consiste na multiplicação do preço de referência da criptomoeda (em *USD*) por uma estimativa aproximada do número de unidades da criptomoeda atualmente em circulação (SEBASTIÃO; GODINHO, 2021).
- **Market Price:** o preço médio da criptomoeda considerando as *Exchanges* mais relevantes, ou seja, aquelas que administram a maior parte das unidades existentes da criptomoeda (JI; KIM; IM, 2019).
- **Realized Capitalization:** semelhante ao *Market Capitalization*, exceto por considerar o preço de cada moeda no momento em que foi movimentada pela última vez na *Blockchain* (CHRISTOFOROU; EMIRIS; FLORAKIS, 2020).
- **Volume of Trades:** o valor total estimado de transações diárias nas *Exchanges* mais relevantes (SEBASTIÃO; GODINHO, 2021).
- **On-Chain Transaction Volume:** o valor total estimado (em *USD* ou *BTC*) das transações diárias registradas na *Blockchain* (CHRISTOFOROU; EMIRIS; FLORAKIS, 2020).
- **Adjusted on-chain Volume:** *On-chain Transaction Volume* após o filtro de transações econômicas pouco relevantes (SEBASTIÃO; GODINHO, 2021).
- **Trades per Minute:** o número de unidades da criptomoeda transacionadas por minuto nas *Exchanges* mais relevantes (CHEN; LI; SUN, 2020).
- **Volume of Trades Growth Rate:** índice percentual de variação do volume de transações, comparando o instante anterior ao atual (DERBENTSEV et al., 2021).

- **Number of Transactions:** o número de transações registradas para a criptomoeda (JI; KIM; IM, 2019).
- **Non Popular Transactions:** o número de transações que não envolvem as 100 carteiras com maior capital (JI; KIM; IM, 2019).
- **Confirmed Transactions per Day:** o número de transações confirmadas para a criptomoeda (CHEN; LI; SUN, 2020).
- **Transaction Rate:** o número de transações realizadas por segundo (JAGANNATH et al., 2021).
- **Average Transaction Value:** o valor médio das transações (em BTC) (MUDASSIR et al., 2020).
- **Median Transaction Value:** a média do valor das transações (em BTC) (MUDASSIR et al., 2020).
- **Top 100 to Total:** a proporção de criptomoedas armazenadas nas 100 contas de maior valor em proporção ao número de unidades da criptomoeda existente (MUDASSIR et al., 2020).
- **Total Transactions Output:** o valor total de todas as saídas de transações por dia (JI; KIM; IM, 2019).
- **Sent:** o número de unidades da criptomoeda enviada diariamente (MUDASSIR et al., 2020).
- **Sent From Addresses:** endereços distintos de onde as transações foram enviadas (MUDASSIR et al., 2020).
- **Active Addresses:** endereços distintos envolvidos em transações, seja no envio ou recebimento de criptomoedas (MUDASSIR et al., 2020).
- **Existing Address:** o número total de endereços (ou carteiras) existentes (JI; KIM; IM, 2019).
- **Payment Count:** número de destinatários das transações, que pode ser maior que um, dada a possibilidade de pagamento em lote (SEBASTIÃO; GODINHO, 2021).
- **Unspent Transactions:** o número de transações não gastas (MUDASSIR et al., 2020).

A.3 Dados da Rede

- **Transaction Median Time:** o tempo médio para uma transação ser aceita em um bloco minerado (JI; KIM; IM, 2019).
- **Mempool Transaction Count:** o número de transações aguardando confirmação (JI; KIM; IM, 2019).
- **Mempool Size:** o tamanho agregado das transações aguardando confirmação (JI; KIM; IM, 2019).
- **Mempool Growth:** a taxa de crescimento do *pool* de memória (*mempool*) por

segundo (JI; KIM; IM, 2019).

- **Transactions per second**: o número de transações adicionadas ao *mempool* por segundo.
- **Average Transactions per Block**: o número médio de transações por bloco minerado (JI; KIM; IM, 2019).
- **No Longer than 100 Transactions**: o número total de transações por dia, excluindo as cadeias maiores que 100. (JI; KIM; IM, 2019).
- **Total Units**: o número total de unidades da criptomoeda já minerada (JI; KIM; IM, 2019).
- **Transaction Fees BTC**: o valor total (BTC) de todas as taxas de transação pagas aos mineradores (JI; KIM; IM, 2019).
- **Transaction Fees USD**: o valor total (USD) de todas as taxas de transação pagas aos mineradores (JI; KIM; IM, 2019).
- **Miners Revenue**: o valor total das recompensas do bloco *Coinbase* e taxas de transação pagas aos mineradores. (JI; KIM; IM, 2019).
- **Median Fees**: valor médio (em USD) das taxas pagas aos mineradores, obtido ao dividir o somatório das taxas pagas pelo número de transações (JI; KIM; IM, 2019).
- **Median Fees (%)**: proporção de *Median Fees* pelo volume de transações (JI; KIM; IM, 2019).
- **Hash Rate**: o número estimado de Tera Hashes por Segundo (Trilhões de Hashes por Segundo) que a rede está realizando (CHEN; LI; SUN, 2020).
- **Block Count**: o número total de blocos minerados na rede. (JAGANNATH et al., 2021).
- **Block Height**: o número total de blocos já criados e incluídos no *Blockchain* primário (JAGANNATH et al., 2021).
- **Block Size**: o tamanho médio (em *Megabytes*) dos blocos minerados (JI; KIM; IM, 2019).
- **Blockchain Size**: o tamanho total de todos os blocos minerados. (JI; KIM; IM, 2019).
- **Mining Difficulty**: o número de *hashes* necessários para minerar um bloco (JAGANNATH et al., 2021).
- **Mining Profitability**: a lucratividade (em USD/dia) por 1 *Terahash* por Segundo (MUDASSIR et al., 2020).
- **Fee-to-Reward Ratio**: a proporção da taxa enviada em uma transação como recompensa para validação da transação por outros usuários (MUDASSIR et al., 2020).
- **Median Transaction Fee**: a mediana das taxas de transação (em BTC) (MUDASSIR et al., 2020).
- **Average Transaction Fee**: cada transação pode ter uma taxa de transação associada determinada pelo remetente. A taxa de transação é recebida pelos mineradores que verificam a transação. As transações com taxas mais altas incentivam os mineradores

a priorizá-las frente às demais (MUDASSIR et al., 2020).

- **Block Time:** o tempo necessário para processar 1 bloco (MUDASSIR et al., 2020).
- **Generated Coins:** o número de novas moedas geradas (SEBASTIÃO; GODINHO, 2021).
- **Miners Balance:** o número total realizado em endereços de mineradores (JAGANNATH et al., 2021).
- **Total Miners Revenue:** recompensas paga aos mineradores (*fees*) somada às novas moedas geradas (JAGANNATH et al., 2021).
- **Miners Inflow Volume:** o número total de moedas transferidas para endereços de mineradores (JAGANNATH et al., 2021).
- **Miners Outflow Volume:** o número total de moedas transferidas de endereços de mineradores (JAGANNATH et al., 2021).
- **Miners to Exchanges:** o número total de moedas transferidas de endereços de mineradores para carteiras de *Exchanges* (JAGANNATH et al., 2021).
- **Exchange Inflow/Outflow:** o número total transferido de carteiras de *Exchanges* para carteiras fora de *Exchanges* (JAGANNATH et al., 2021).
- **Exchange Balance:** o número total de moedas retidas em carteiras de *Exchanges* conhecidas (JAGANNATH et al., 2021).
- **Exchange Withdrawals:** o número de retiradas de carteiras de *Exchanges* conhecidas (JAGANNATH et al., 2021).
- **Exchange Deposits:** o número de depósitos para carteiras de *Exchanges* conhecidas (JAGANNATH et al., 2021).
- **Balances of Addresses:** o número total de moedas mantidas por um único endereço (JAGANNATH et al., 2021).

A.4 Dados Externos

- **Closing Crude Price:** são contratos padronizados e negociados em bolsa que são negociados na *New York Mercantile Exchange* (NYMEX). Cada contrato representa 1000 barris de óleo, ou 42.000 galões. (MALLQUI; FERNANDES, 2019).
- **Closing Gold Spot Price:** contrato de compra de ouro fino, sob forma de lingote, fundido por empresa refinadora e custodiado em instituição depositária, ambas credenciadas (MALLQUI; FERNANDES, 2019).
- **CBOE Volatility Index:** mensura a volatilidade das opções sobre ações do S&P 500. Seu cálculo é feito pela CBOE (*Chicago Board Options Exchange*), com base na média dos preços das opções do S&P 500 e representa a expectativa da volatilidade implícita dessas opções nos próximos 30 dias. (ADCOCK; GRADOJEVIC, 2019).
- **Closing S&P 500 Price:** abreviação de *Standard Poor's 500*, ou simplesmente S&P. Trata-se de um índice composto por 500 ativos cotados nas bolsas de NYSE

ou NASDAQ, qualificados devido ao seu tamanho de mercado, sua liquidez e sua representação de grupo industrial (MALLQUI; FERNANDES, 2019).

- **Closing Nasdaq Price:** contempla as 100 maiores instituições não financeiras listadas na Nasdaq, segunda maior bolsa de valores do mundo. A grande diferença do Nasdaq 100 e da Bolsa que o nomeia, é o maior foco em empresas de tecnologia (MALLQUI; FERNANDES, 2019).
- **Closing DAX Price:** DAX é uma relação das 30 companhias abertas de melhor performance financeira da Alemanha, com base no sistema Xetra da Bolsa de Valores de Frankfurt (MALLQUI; FERNANDES, 2019).
- **Daily Dummies:** consiste em 7 indicadores, um para cada dia da semana. Dessa forma, se a amostra analisada atualmente se refere à uma segunda-feira, haverá um indicador 1 para *dummy-monday* e 0 para os demais dias da semana, quando acredita-se que o dia da semana interfere na cotação das criptomoedas (SEBASTIÃO; GODINHO, 2021).
- **Google Trend Search Volume Index:** o volume normalizado de pesquisas pelo nome da criptomoeda por dia (CHEN; LI; SUN, 2020).
- **Baidu Media Search Volume:** o volume ponderado para cobertura da mídia pelo nome da criptomoeda (CHEN; LI; SUN, 2020).
- **Closed Issues:** número de problemas relatados considerados concluídos (CHRISTOFOROU; EMIRIS; FLORAKIS, 2020).
- **Total Issues:** número de problemas relatados (CHRISTOFOROU; EMIRIS; FLORAKIS, 2020).
- **Commit Count:** *commits* realizados no repositório da criptomoeda no *Github* nas últimas 4 semanas (CHRISTOFOROU; EMIRIS; FLORAKIS, 2020).
- **Forks:** número de vezes que o código da criptomoeda foi copiado, o que pode ser feito, por exemplo, para dar origem a uma nova criptomoeda baseada em alguma outra de origem (CHRISTOFOROU; EMIRIS; FLORAKIS, 2020).
- **Pull Request Contributors:** número de diferentes usuários que solicitaram alteração no código da criptomoeda (CHRISTOFOROU; EMIRIS; FLORAKIS, 2020).
- **Pull Requests Merged:** número de alterações realizadas no código da criptomoeda (CHRISTOFOROU; EMIRIS; FLORAKIS, 2020).
- **Number of Stars:** auxilia na medição da popularidade da criptomoeda (CHRISTOFOROU; EMIRIS; FLORAKIS, 2020).
- **Number of Subscribers:** número de usuários da plataforma que desejam receber notificações quando o repositório de código da criptomoeda for alterado (CHRISTOFOROU; EMIRIS; FLORAKIS, 2020).