

Cryptocurrency exchange analysis through the use of artificial intelligence

Gustavo M. Penna, Ricardo M. Salgado

Abstract—The cryptocurrency exchange market is where it's possible to buy and sell cryptocurrency like *Bitcoin* and *Ethereum*, and where people try to earn money through the price difference between buying and selling, but trying to predict the future value of a currency is hard, and needs a strong mathematical analysis or even a market sentiment analysis to get close to the real value. Thus, this paper's objective is to use the *LSTM* Deep Learning model, together with techniques of data manipulation like exploratory data analysis, to predict the *Bitcoin* price and tendency, providing a reliable model that can be used to help traders make decisions on whether or not they should buy a cryptocurrency, aiming at profit.

Index Terms— Artificial intelligence, *LSTM*, Deep learning, *Bitcoin*, cryptocurrency, data analysis, statistics, stock market.

I. INTRODUÇÃO

Segundo o *Estadão*, o mercado financeiro movimentou trilhões de dólares todos os anos [10], e muitas tecnologias que o afetam também possuem grande demanda e envolvem grandes quantidades de dinheiro, é um campo que possui constantes inovações e automatizações, e a explosão do uso da inteligência artificial se inclui nisso [11].

O setor financeiro sempre busca usar tecnologias de ponta [12], e com a concretização do uso de algoritmos de inteligência artificial fora do ambiente de pesquisa, rapidamente foi descoberta sua utilização no mercado financeiro, possibilitando o desenvolvimento de estratégias de negociação mais eficientes e seguras [13].

Não apenas foi concebido o uso de algoritmos de inteligência artificial nessa área, como também foram inovados, passando a utilizar novos algoritmos e modelos como o *Deep Learning* através das *CNNs* (*Convolutional Neural Networks*) [14][3], técnicas de extração de variáveis como as *RBMs* (*Restricted Boltzmann Machines*) [15], modelos capazes de trabalhar com conjuntos de dados temporais (em que os dados são correlacionados através do tempo) como as *RNNs* (*Recurrent Neural Networks*) [3][7], conseqüentes melhorias como a *LSTM* (*Long short-term memory*) [3][2][7], e também modelos que fazem tanto o uso

dos dados estatísticos quanto de dados textuais de notícias, providas de jornais [5].

Logo, a área está em constante evolução, e aperfeiçoamentos são sempre buscados pelas empresas que atuam no ramo, investindo milhões por ano em pesquisa e desenvolvimento para buscar avanços tecnológicos e obter o maior lucro em relação aos concorrentes [16].

Devido à importância dessas inovações, as mesmas quase sempre são mantidas em sigilo empresarial, dificultando para que sejam estudadas no ambiente acadêmico e por pesquisadores independentes, e conseqüentemente dificultando sua aplicação por pessoas ou fundos que desejam investir no mercado financeiro ou analisá-lo através das tecnologias disponíveis.

Muitos trabalhos relacionados já foram realizados, usando diferentes técnicas e formatos de análise, como modelos avançados baseados nos dados estatísticos como o *ARMA* (*Autoregressive-moving average*) [17] e *ARIMA* (*Autoregressive integrated moving average*) [3], além dos próprios modelos de aprendizado de máquina.

Tais algoritmos apresentam resultados diversos, porém é constatado que a maioria possui resultados muito bons (define-se mais de 50% de acerto como lucro), como a previsão estatística usando *CNN* que possui apenas 4,89% de erro médio [3]. A previsão por texto usando *Paragraph Vector* também conseguiu ótimos resultados, quase triplicando o lucro em comparação com modelos de análises numéricas, com lucro de aproximadamente 12.119.000 yens [5].

Assim, hipóteses anteriores como a do mercado eficiente [4], suposta em 1970, que afirma que é impossível prever preços, já foram refutadas [19], pois na época não foi levado em conta a capacidade de previsão de algoritmos de inteligência artificial, visto que o poder computacional era muito menor, e o conceito de *Algorithmic Trading* (negociações feitas com o auxílio de algoritmos) acabara de dar seus primeiros passos [20].

Além dos modelos para previsão de valores, também é possível o desenvolvimento de simulações do mercado financeiro, onde é possível testar os modelos para validar métricas de desempenho [21][6].

Previsões focadas especificamente no preço do *Bitcoin* também foram realizadas, usando *LSTM*, *RNN* e *ARIMA*, Com *RMSE* (desvio médio quadrático) de 6.87%, 5.45% e 53.74%, respectivamente [7]; *BNNs* (*Bayesian neural networks*) [8] com *RMSE* de 0.0256 e *SVR* (*Support vector regression*) [17] com *RMSE* de 0.0179. Diferente do mercado financeiro, para a previsão de valores de criptomoedas não há dados de empresas para uso, então deve-se usar dados numéricos como preços de abertura e fechamento, volume e

quantidade de operações, e dados da própria *blockchain* (sistema de rede em que funciona o *Bitcoin*) como dificuldade de mineração e taxa de *hash* [7].

Os resultados da previsão de *Bitcoin* também são muito variáveis, não dependendo apenas dos algoritmos usados, mas também do conjunto de dados usados e da época, visto que a capitalização do *Bitcoin* alterou drasticamente numa medida de poucos anos, principalmente entre 2013 e 2017 [22].

De acordo com [7], a média de acurácia da previsão do preço do *Bitcoin* usando *LSTM* é próxima de 52%, de *RNN* 50% e de *ARIMA* também 50%, utilizando dados entre 2013 e 2016.

Já de acordo com [8], o *RMSE* da regressão linear chegou a aproximadamente 9%, já a *BNN* teve aproximadamente 0.5% e o *SVR* 14%, usando dados de 2011 até 2017.

Outras abordagens já foram realizadas para tentativa de previsão do valor, como a geração de imagens de *candlesticks* (representações de dados em intervalos de tempo) para previsão visual, utilizando ações da Petrobrás de 2001 até 2016, com acurácia de 73% [2]; e o uso de indicadores macroeconômicos externos, junto dos dados de vários mercados financeiros do mundo e conseguindo acurácias de mais de 70% usando modelos como o *GARCH* (*Generalized autoregressive conditional heteroskedasticity*) [1].

Dessa forma, este estudo tem o intuito de realizar uma previsão relativamente precisa do preço do *Bitcoin* após determinado período de tempo, aplicando operações de manipulação dos dados históricos, e modelos de inteligência artificial capazes de interpretar séries temporais, sendo este um estudo quantitativo, aplicado, exploratório e experimental.

Com isso, é possível estudar o funcionamento desse mercado, e compará-lo ao mercado financeiro tradicional em relação aos efeitos dos sentimentos, possibilidade de previsão dos valores e diferenças em geral.

O conteúdo deste trabalho encontra-se organizado da seguinte forma: na Seção II são apresentados alguns conceitos sobre previsão de vazões, na Seção III é abordada a metodologia utilizada para realizar os experimentos, já na Seção IV é feita uma discussão sobre os dados utilizados para alimentar o modelo, na Seção V são mostrados os resultados obtidos e finalmente na Seção VI são feitas as considerações finais do trabalho.

II. FUNDAMENTAÇÃO TÉCNICA

Para prever o valor de uma moeda ou ação, é necessário avaliar seu comportamento passado, comparando as tendências de alta e de baixa, estabilidade, e como a variação do seu valor no último intervalo de tempo predefinido afeta o valor final.

É possível realizar medidas estatísticas para comparar médias, medianas e modas num intervalo de tempo, porém tais operações são muito superficiais comparadas com as

aplicabilidades da inteligência artificial [17], que pode ser aproveitada através de algoritmos de *Deep Learning* como o *LSTM*, que é especializado em conjuntos de dados que envolvem sequências temporais.

Além de modelos para realizar a previsão de um valor, é possível usar *Autoencoders* [2] para extrair variáveis e *PCAs* (*Principal Component Analysis*) para reduzir a dimensionalidade, otimizando ainda mais o conjunto de dados [18].

Assim, visto que o problema da previsão de valor do *Bitcoin* é temporal, foi escolhida a arquitetura *LSTM*. Essa arquitetura possui um “estado escondido” adicional, que é repassado junto de cada elemento do conjunto de dados, assim, o modelo possui uma entrada que serve como “memória” do que foi visto anteriormente, sendo ideal para problemas temporais.

III. METODOLOGIA PROPOSTA

Nesta seção, serão explicados o conjunto de dados do problema e os métodos utilizados para tratar dados, processá-los, treinar modelos e realizar previsões.

A. Conjunto de dados

O conjunto de dados usados neste trabalho é o histórico de informações dos *candlesticks* do *Bitcoin* em relação ao dólar (BTC/USDT). Também chamados de “velas” ou “*candles*”, são representações de variáveis do mercado financeiro em intervalos de tempo.

Cada *candlestick* pode ter um tempo variado entre 1 minuto até 1 mês, e armazena os dados como preço de início, preço final, volume de transações, entre outros, nesse período de tempo. Para o melhor equilíbrio entre precisão dos dados e tamanho da base de dados, foi usado o intervalo de 5 minutos para cada *candlestick*. Tal intervalo foi decidido pensando num período que não é afetado por grandes alterações nas tendências de alta e baixa, focando no curto prazo e sem precisar se preocupar nas tendências a longo prazo. As informações de cada *candlestick* desse conjunto de dados estão representadas na Tabela I.

Como objetivo da previsão, foi definido o valor de fechamento da *candle* (coluna *Close*), visto que, ao conseguir os dados básicos na abertura, pode-se prever seu valor final e decidir se a compra deve ser feita ou não. Quando não é possível obter valores como o máximo ou mínimo, usa-se o valor da *candle* anterior, visto que é a mais próxima da atual.

Tabela 1

Descrição das variáveis do conjunto de dados

Variável	Definição
<i>Open time</i>	Horário de abertura.
<i>Open</i>	Valor de abertura.

<i>High</i>	Valor máximo no intervalo.
<i>Low</i>	Valor mínimo no intervalo.
<i>Close</i>	Valor de fechamento.
<i>Close time</i>	Horário de fechamento.
<i>Quote asset volume</i>	Volume da moeda vendida (USDT).
<i>Number of trades</i>	Quantidade de trocas realizadas.
<i>Taker buy base asset volume</i>	Volume de ordens da moeda comprada (BTC).
<i>Taker buy quote asset volume</i>	Volume de ordens da moeda vendida (USDT).

Fonte: Documentação da API da Binance [9]

As variáveis da tabela 1 são as *features* (propriedades ou características mensuráveis) bases do conjunto de dados, essas serão manipuladas na análise exploratória de dados para ficarem mais adequadas para o modelo de inteligência artificial.

Os dados usados foram coletados de 31 de dezembro de 2020, até 31 de dezembro de 2021, em intervalos de 5 minutos, usando a *API (Application Programming Interface)* da corretora *Binance* [9], seu uso é gratuito e pode ser feito por qualquer pessoa com uma conta registrada. No conjunto de dados coletado há 104924 *candles* disponíveis, ou seja, medições para serem estudadas e processadas. A figura 1 representa um gráfico dessas *candles*, limitadas do dia 1 de janeiro de 2021 até 31 de janeiro de 2021, como exemplo.



Fig. 1. Candlesticks do Bitcoin no mês de janeiro de 2021

B. Etapas de processamento

O modelo proposto possui 3 etapas. A primeira etapa é a **análise exploratória de dados (EDA)**, onde será feito um estudo dos dados disponíveis e manipulações para gerar mais

informações relevantes. A segunda etapa é o **treinamento dos modelos de inteligência artificial**, em que os dados disponíveis serão separados em um conjunto de treino e outro de testes, e o conjunto de treino será alimentado ao modelo. Na terceira etapa é realizada a **previsão dos dados de testes e validação dos resultados obtidos**. A figura 2 representa as etapas do modelo, juntamente com a aquisição de dados, que já foi citada.

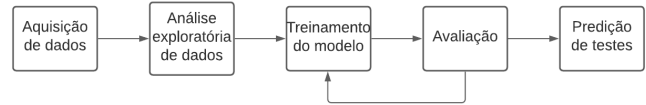


Fig. 2. Fluxograma do método proposto.

C. Análise exploratória de dados

Para aplicar a *EDA* nos dados no modelo, é necessário estudá-los, formatá-los corretamente, remover variáveis que não serão utilizadas bem como criar novas variáveis que podem ser úteis.

Devido à natureza estatística dessa base, é evidente a possibilidade de gerar novos dados com operações matemáticas simples, como a média de valores entre determinados intervalos de tempo, ou a diferença entre valores para cada entrada. Essas são úteis pois provêm mais informações para serem processadas pela IA, que se forem geradas de forma correta, aumentam sua performance.

Para o problema novas *features* adicionais foram geradas, como a *RSI (Relative Strength Index)* e *ADX (Average Directional Index)*, que são indicadores usados no mercado financeiro para análise técnica e auxílio da tomada de decisão, assim, são variáveis que podem melhorar consideravelmente a acurácia do modelo:

1) *RSI* (fórmula 1)

$$RSI = 100 - \frac{100}{(1+RS)} \quad (1)$$

$$RS = \frac{Mxa}{Mxb}$$

Onde *Mxa* é a média entre as últimas 14 altas e *Mxb* a média entre as últimas 14 baixas.

2) *ADX* (fórmula 2)

$$ADX = MMA(DX, 14) \quad (2)$$

$$DX = \left(\frac{|DP-DN|}{|DP+DN|} \right) \times 100$$

Onde *MMA* é a média móvel aritmética, *DP* é o índice direcional positivo e *DN* é o índice direcional negativo. O *ADX* também foi calculado usando os últimos 14 valores.

Além disso, foi formatado o *timestamp* (carimbo de tempo) de abertura para uma data no formato *YYYY-MM-DD H:M:S*, que é usada como índice para o modelo, para melhor visualização. Por fim, foi calculado o preço médio da *candle* (fórmula 3), e criada uma variável booleana “Deve comprar” que considera se a compra na abertura deveria ter sido efetuada, levando em consideração se o preço de fechamento é maior que o preço de abertura.

$$Média = \frac{Preço\ mais\ alto + Preço\ mais\ baixo}{2} \quad (3)$$

Ao final, obteve-se o conjunto de dados demonstrado na Tabela II.

Tabela II

Conjunto de dados após análise exploratória

Data (A-M-D H:M)	Horário de abertura (Timesta mp)	Valor de abertura	Valor mais alto	Valor mais baixo
2020-12-31 21:00	1609459 200000	28923.63	29017.5	28913.12
2020-12-31 21:05	1609459 500000	28975.65	28979.53	28846.28
...
2021-12-31 20:55	1640994 900000	46285.79	46285.8	46132.04
2021-12-31 21:00	1640995 200000	46216.93	46391.49	46208.37

Valor de fechame nto	Volume	Horário de fechame nto	Volume da moeda vendida	Quantida de trocas realizada s
28975.65	182.89	1609459 499999	5300787	5614
28858.94	214.57	1609459 799999	6201532	4928
...
46216.93	145.02	1640995 199999	6703859	3870
46321.34	185.68	1640995 499999	8598314	4551

Volume de ordens da moeda comprada	Volume de ordens da moeda vendida	RSI	ADX	Deve comp rar	Valor médio
80.03	2319247	1.0	1.0	1	28965.31
113.76	3287213	1.0	1.0	0	28912.90

...
77.52	3584387	42.59	30.93	0	46208.92
106.55	4933083	48.94	29.25	1	46299.93

As propriedades desses dados estão mais bem representadas na análise descritiva da tabela III, exibindo a média, mínima, máxima e variação dos principais valores.

Tabela III

Análise descritiva de alguns dados do conjunto

	Valor mais alto	Valor mais baixo	Valor de fechame nto	Valor de abertura
Média	47442.07	47271.40	47357.35	47357.19
Mínima	28846.46	28130.00	28752.80	28752.80
Máxima	69000.00	68681.00	68734.26	68734.26
Variação	9822.40	9820.39	9821.77	9821.90

Após a geração dessas novas *features* artificiais, é necessário formatar os dados para inserção no modelo. Para isso, primeiro é realizada a separação entre dados de treino e teste, nesse problema foi feita a separação de 67% dos dados para treino (Contemplando do dia 31 de dezembro de 2020 até 2 de setembro de 2021) e 33% restantes para testes (de 2 de setembro de 2021 até 31 de dezembro de 2021), dessa forma garante-se que o modelo irá ser avaliado com base em dados não vistos anteriormente.

Vale ressaltar que por ser um problema temporal, não é possível usar técnicas de validação cruzada, visto que os dados devem ser treinados e testados em ordem, dada a natureza dos modelos *LSTM*. Os dados também são separados em dois conjuntos, sendo um as variáveis usadas para treino, e o outro a variável que deve ser prevista, no caso, a coluna *Close*, que representa o valor de fechamento da *candle*, essa está representada na figura 3.

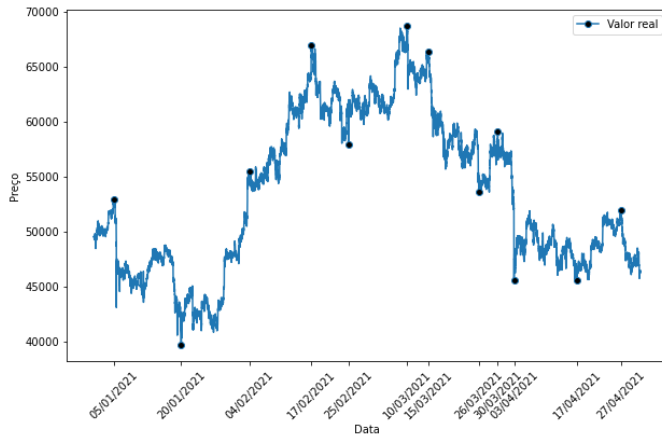


Fig. 3. Gráfico dos valores que tentarão ser previstos (valor de fechamento X data).

Em sequência, é feita a normalização dos dados, usando as fórmulas de *MinMax* (fórmula 4) e *Standard Score* (fórmula 5), formatando os valores para o intervalo entre 0 e 1 para diminuir as discrepâncias entre valores, de forma que valores muito altos ou muito baixos ainda serão contabilizados corretamente, e colocá-los no intervalo usado pela função de ativação *ReLU*.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (4)$$

$$x' = \frac{x - \text{media}(x)}{\text{std}(x)} \quad (5)$$

Onde X é um valor, dentro do conjunto de valores, à ser normalizado.

D. Treinamento dos modelos

O modelo desse problema foi instanciado com uma camada de entrada de *LSTM* com 13 neurônios, uma camada oculta linear de 128 neurônios, e uma camada de saída linear com 1 classe de retorno (pois é um problema de regressão). Como função de ativação entre camadas, foi usada a *ReLU* (*Rectified Linear Unit*), para capturar melhor a interação entre as variáveis.

Após o instanciamento, foi feita a parametrização e inserção de demais filtros desejados, definindo o número de épocas para treino como 1000, a taxa de aprendizado para 0.001, e os critérios para cálculo de perda e otimização como o *Adam* e *L1* (*Least absolute deviations*).

Com o modelo e os dados prontos, foi realizada a iteração pela quantidade de épocas, em cada época são passados os tensores de treino pela rede neural, feita a otimização dos pesos, e então calculado o erro com base nos tensores de teste.

Antes de cada iteração é guardado o dicionário de estados do modelo, que contém os pesos e equações dos neurônios, dessa forma, na próxima iteração, pode-se recarregar o estado anterior caso o treino atual cause uma queda na porcentagem de acerto, garantindo que os treinos irão apenas causar um aumento na acurácia, e não diminuição.

Ao final do treino, também é guardado esse dicionário de estados do modelo, possibilitando carregar o modelo já treinado para prever dados futuros ou para treiná-lo novamente.

A figura 4 demonstra o gráfico dos valores previstos no treino comparados com os valores reais, pode-se notar que nas mudanças pequenas o modelo foi capaz de prever com precisão, mas não teve uma performance tão boa nas mudanças bruscas de valor.

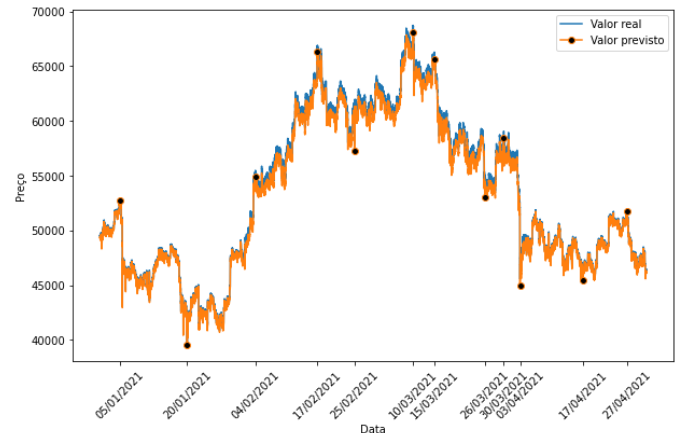


Fig. 4. Gráfico de previsões do treino (valor de fechamento X data).

E. Previsão

Com o modelo treinado, basta executá-lo novamente, mas dessa vez em dados novos ao invés dos dados de treino, operação chamada de previsão. Dessa forma, é possível verificar a acurácia do modelo como se estivesse prevendo dados do mundo real, visto que os dados de teste são desconhecidos pelo modelo no treino.

Ao final da previsão, são aplicadas diversas métricas de acurácia para calcular os resultados do modelo, são usadas as métricas *RMSE* (desvio médio quadrático), *MAPE* (porcentagem média absoluta) e *MAE* (erro médio absoluto), essas métricas foram escolhidas pois são as mais indicadas para calcular o erro em problemas de regressão.

Os dados usados para a previsão foram do dia 31 de dezembro de 2021, onde acaba o conjunto de teste do treino, até o dia 19 de janeiro de 2022, com um total de 5473 candles para previsão. Todos dados também passaram pelo processo de *EDA*, pois necessitam estar na mesma estrutura dos dados de treino para serem corretamente processados pelo modelo.

IV. RESULTADOS

Com os dados obtidos na etapa de previsão, foi possível verificar os resultados do modelo num ambiente simulando o mundo real, com dados nunca vistos antes tanto no treino quanto no teste.

A figura 5 demonstra o gráfico dos valores previstos comparados com os valores reais, é evidente a precisão obtida, não sempre prevendo o preço exato, mas conseguindo prever as tendências de alta e baixa.

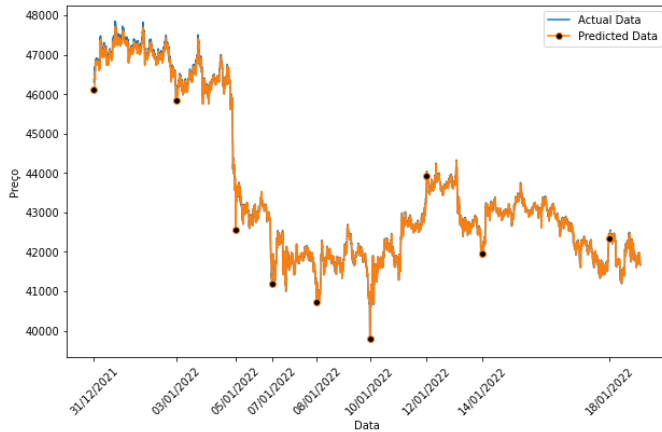


Fig. 5. Gráfico de previsões da classificação (valor de fechamento X data).

Como é possível observar, o modelo foi capaz de acompanhar as diversas tendências desse período. Ao final do ano de 2021, houve uma grande queda no preço do *Bitcoin*, devido às preocupações com as novas variantes do *Covid-19* e com os bloqueios de viagens. Assim, vários fundos que usam criptomoedas para combater a inflação decidiram vendê-las para continuarem com lucro mesmo com baixa na economia, e com o aumento na oferta da moeda veio a baixa do preço [24].

Aliado ao surgimento de novas *altcoins* (moedas alternativas) e com o aumento da inflação, principalmente na economia norte-americana, criou-se um período de alta volatilidade (alta frequência de oscilação) do valor das criptomoedas, causando as quedas e altas que podem ser visualizadas no gráfico [25].

Como exemplo, pode-se citar a queda de 05/01/2022 de 1.05%, causada principalmente pela reação às decisões econômicas do governo norte-americano, e a queda de 10/01/2022 de 1.04%, acompanhando a baixa do mercado de ações *Nasdaq* e a inflação [26].

Na figura 6, é possível ver com mais detalhes o comportamento do modelo na queda do dia 05/01/2022, que apesar de não conseguir prever os valores com exatidão, identifica corretamente as tendências. Essa identificação pode ser explicada através da queda acentuada às 16:40h, que primeiramente não havia sido identificada, mas após o uso dessa informação na próxima iteração, o modelo notou que deveria continuar em queda, e só parou após a estabilização às 17:15h. O mesmo ocorreu na queda das 19:30h, onde o modelo previu que haveria uma estabilização, mas depois de notar a queda da *candle* anterior, seguiu a tendência de queda.

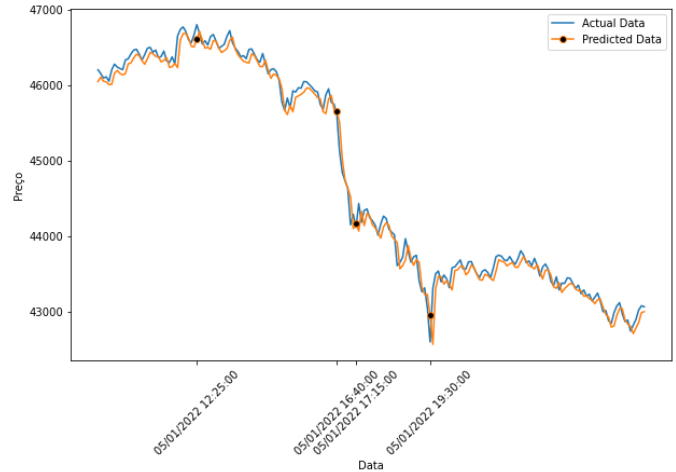


Fig. 6. Gráfico de previsões da classificação, focado na queda do dia 05/01/2022 (valor de fechamento X data).

O mesmo pode ser observado na figura 7, em que as tendências de alta das 12:05h e 13:45h foram identificadas após a primeira *candle* de alta, e só estabilizaram após a próxima *candle* fechar com valor sem crescimento acentuado.

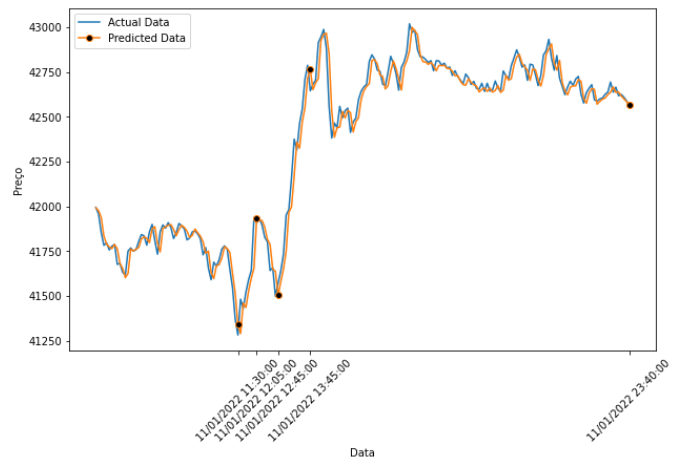


Fig. 7. Gráfico de previsões da classificação, focado na queda do dia 05/01/2022 (valor de fechamento X data).

Para uso real do modelo, é necessário principalmente validar se o valor previsto é maior ou menor que o valor inicial, pois isso é o que define se uma compra resultaria em perda ou ganho. Assim, é realizada essa validação e calculada uma taxa de acerto, que foi definida como “Acerto para valor maior ou menor”, explícito na tabela IV.

Na tabela IV, também são exibidos os erros calculados usando as diversas métricas citadas anteriormente.

Tabela IV
Resultado da validação do modelo

Métrica	Valor
Acerto para valor maior ou menor	78.82223%
RMSE	0.01299
MAPE	0.01086
MAE	6.88625

Assim, pode-se notar que o modelo possui uma boa performance, capaz de prever com certa acurácia os futuros valores de fechamento das velas, além de, em aproximadamente 78% das vezes, conseguir prever se o valor de fechamento será maior ou menor do que o de abertura.

Como a métrica MAE representa valores absolutos, e o alvo da previsão se dá em dólares, nota-se que o modelo obteve, em média, um desvio de 6.88 dólares do valor final. Apesar de parecer um valor alto, calculou-se a média de diferença entre os valores de abertura e fechamento, que resultou em 53.83 dólares para o período de testes. Assim, o desvio do modelo foi de aproximadamente 13% da diferença total de valor, o que pode ser considerado baixo.

V. CONCLUSÃO

Este trabalho apresentou uma proposta de modelo de inteligência artificial capaz de prever o valor de criptomoedas, focando na propriedade temporal dos elementos do mercado financeiro. Assim, contribuindo para que mais pessoas entendam o funcionamento do mercado de criptomoedas, incentivando seu uso, expandindo as contribuições tecnológicas, e popularizando-o.

Observou-se que o modelo proposto usando *LSTM* é mais eficiente que outros modelos como *ARIMA* e *CNN*, com seu melhor resultado na previsão da alta ou baixa dos preços, e a maior quantidade de dados, assim como um melhor ajuste na parametrização, foram capazes de melhorar consideravelmente a performance do modelo, comparado com trabalhos como [3] e [7].

Com os resultados obtidos neste estudo, percebe-se que é possível criar um *software* para auxílio na tomada de decisões, que também pode ser utilizado no escopo de diversas criptomoedas, além do *Bitcoin*, mas para melhor confiança deve ser acompanhada de outros modelos tomadores de decisão como definição de riscos e definição dos preços de compra e venda [23].

Através desse projeto, foi possível concluir que, com o uso de algoritmos de inteligência artificial, é possível prever valores com certa precisão, sem depender de sistemas caros e potentes ou análises estatísticas complexas, de forma que qualquer pessoa pode conseguir prever as tendências de preço.

Para continuidade deste trabalho, pode-se considerar a criação de *ensembles* de vários modelos diferentes, uma

análise mais profunda das *features* e da análise exploratória de dados, além de hiper-parametrização do modelo.

REFERÊNCIAS

- [1] MACIEL, Leandro S.; BALLINI, Rosângela. Neural networks applied to stock market forecasting: An empirical analysis. **Learning and Nonlinear Models**, vol. 8, no. 1, p. 3–22, 2010. <https://doi.org/10.21528/lnlm-vol8-no1-art1>.
- [2] DIAS PAIVA, Felipe; MAGDA DA SILVA ROMA, Carolina. Métodos de deep learning aplicados a candlestick como estratégia de investimento. 2014.
- [3] SELVIN, Sreelekshmy; VINAYAKUMAR, R.; GOPALAKRISHNAN, E. A.; MENON, Vijay Krishna; SOMAN, K. P. Stock price prediction using LSTM, RNN and CNN-sliding window model. **ICACCI 2017**, vol. 2017-Jan, p. 1643–1647, 2017. <https://doi.org/10.1109/ICACCI.2017.8126078>.
- [4] FAMA, Eugene F. Session Topic: Stock market price behavior session chairman: Burton G. Malkiel efficient capital markets: A review of theory and empirical work. **The Journal of Finance**, vol. 25, no. 2, p. 383–417, 1970.
- [5] AKITA, Ryo. Deep learning for stock prediction using numerical and textual information. **2016 IEEE/ACIS 15th ICIS**, p. 1–6, 2016.
- [6] WANG, Jianliang; FANG, Linshan; ZHUANG, Xiang. Study and application of stock robot kaburobo based on artificial intelligence. **IJCAI**, p. 260–262, 2009. <https://doi.org/10.1109/IJCAI.2009.47>.
- [7] MCNALLY, Sean; ROCHE, Jason; CATON, Simon. Predicting the price of Bitcoin using machine learning. **Proceedings - 26th, PDP 2018**, p. 339–343, 2018. <https://doi.org/10.1109/PDP2018.2018.00060>.
- [8] JANG, Huisu; LEE, Jaewook. An empirical study on modeling and prediction of Bitcoin prices with Bayesian Neural Networks based on Blockchain information. **IEEE Access**, vol. 6, p. 5427–5437, 2017. <https://doi.org/10.1109/ACCESS.2017.2779181>.
- [9] *BINANCE*, Official documentation for the Binance APIs and streams. Available at: <https://github.com/binance/binance-spot-api-docs/blob/master/rest-api.md>. Access on: 01 Apr. 2021.
- [10] ANDRADE, Jenne. B3 bate recorde e movimentada R\$ 26 bilhões por dia em 2020. *eInvestidor.estadao*, 2021. Available at: <https://investidor.estadao.com.br/investimentos/b3-recorde-26-bilhoes-dia>. Access on: 23-mar-2022.
- [11] Conheça 5 tendências tecnológicas que impulsionam o mercado financeiro, *economia.ig*, 2021. Available at: <https://economia.ig.com.br/1bilhao/2021-10-28/conheca-5-tendencias-tecnologicas-que-impulsionam-o-mercado-financeiro.html>. Access on: 24-mar-2022.
- [12] DINIZ, Bruno. **O fenômeno fintech: tudo sobre o movimento que está transformando o mercado financeiro no Brasil e no mundo**. Alta Books, 2020.
- [13] Piotroski, Joseph D. "Value investing: The use of historical financial statement information to separate

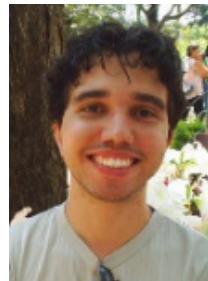
winner from losers." *Journal of Accounting Research* (2000): 1-41.

- [14] HOSEINZADE, E.; HARATIZADEH, S. Cnnpred: Cnn-based stock market prediction using a diverse set of variables. *Expert Systems with Applications*, Elsevier, v. 129, p. 273-285, 2019.
- [15] Liang, Qiubin, et al. "Restricted Boltzmann machine based stock market trend prediction." *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017.
- [16] SOARES, Rebeca. CFOs brasileiros estão mais propensos a investir em tecnologia. *eInvestidor.estadao*, 2021. Available at: <https://einvestidor.estadao.com.br/investimentos/cfos-investem-em-tecnologia>. Access on: 24-mar-2022.
- [17] MARWALA, Lufuno Ronald. Forecasting the Stock Market Index Using Artificial Intelligence Techniques. 2014. 67-70 f. 2014. Available at: <https://core.ac.uk/download/pdf/39667613.pdf>.
- [18] Waqar, M., Dawood, H., Guo, P., Shahnawaz, M. B., & Ghazanfar, M. A. (2017). *Prediction of Stock Market by Principal Component Analysis*. *2017 13th International Conference on Computational Intelligence and Security (CIS)*. doi:10.1109/cis.2017.00139
- [19] Malkiel, B.G. (2005), Reflections on the Efficient Market Hypothesis: 30 Years Later. *Financial Review*, 40: 1-9. <https://doi.org/10.1111/j.0732-8516.2005.00090.x>
- [20] McGowan, Michael J. "The rise of computerized high frequency trading: use and controversy." *Duke L. & Tech. Rev.* 9 (2009): 1.
- [21] Shuang Yao, Zhan Zhang, Weiqiang Huang. Heterogeneous Investors. no. 3, p. 2418-2421, 2011.
- [22] Baur, Dirk G., and Thomas Dimpfl. "Realized bitcoin volatility." *SSRN* 2949754 (2017): 1-26.
- [23] ALEXANDRE MACEDO, Jose; THEODORO OLIVEIRA CAMARGO, Luis; CESAR BRANDAO DE OLIVEIRA, Humberto; EDUARDO DA SILVA, Luiz; MENEZES SALGADO, Ricardo. An intelligent decision support system to investment in the stock market. *IEEE Latin America Transactions*, vol. 11, no. 2, p. 812-819, 2013. <https://doi.org/10.1109/TLA.2013.6533971>.
- [24] HOLLAND, Frank. Cryptocurrency prices fall in December, and investors blame omicron, climate change. *CNBC*, 2021. Available at: <https://www.cnn.com/2021/12/29/cryptocurrency-prices-fall-in-december-and-investors-blame-omicron-climate-change.html>. Access on: 25-mar-2022.
- [25] HAJRIC, Vildana. Bitcoin Declines to Lowest Level Since December's Flash Crash. *Bloomberg*, 2022. Available at: <https://www.bloomberg.com/news/articles/2022-01-05/bitcoin-declines-to-lowest-level-since-december-s-flash-crash>. Access on: 26-mar-2022.
- [26] PONCIANO, Jonathan. 'Looking Ugly': Crypto Market Crash Intensifies After \$300 Billion Sell-Off—How Low Can Bitcoin Prices Go?. *Forbes*, 2022. Available at: <https://www.forbes.com/sites/jonathanponciano/2022/01/10/looking-ugly-crypto-market-crash-intensifies-after-300-billion-sell-off-how-low-can-bitcoin-prices-go/>

[0-billion-sell-off-how-low-can-bitcoin-prices-go/?sh=9acf96aa11bd](https://www.forbes.com/sites/jonathanponciano/2022/01/10/looking-ugly-crypto-market-crash-intensifies-after-300-billion-sell-off-how-low-can-bitcoin-prices-go/). Access on: 26-mar-2022.



Gustavo Moreira Penna is graduating in Computer Science at Universidade Federal de Alfenas (UNIFAL-MG). Currently works at the development of computer solutions. Acts in the artificial intelligence and information security areas, with emphasis on the financial market.



Ricardo Menezes Salgado has graduated (2002) in Mathematics at Universidade Federal de Viçosa (UFV), master's degree (2004) and doctorate degree (2009) in Electrical Engineering at Universidade Estadual de Campinas (Unicamp). Currently he's a core teacher in Computer Science at Universidade Federal de Alfenas (UNIFAL-MG) in the intelligent systems and data analysis areas. He's the leader of the research team at the Laboratory of computational intelligence (LInC), studying areas such as time series predictions and computational intelligence. He's a permanent teacher at the program of Masters and applied statistics and biometry at UNIFAL-MG, in which he develops research based on time series predictions and computational intelligence.