



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

LUCAS GOMES AIRES

DEU TRADER

**CAMPINA GRANDE - PB
2022**

LUCAS GOMES AIRES

DEU TRADER

**Trabalho de Conclusão Curso
apresentado ao Curso Bacharelado em
Ciência da Computação do Centro de
Engenharia Elétrica e Informática da
Universidade Federal de Campina
Grande, como requisito parcial para
obtenção do título de Bacharel em
Ciência da Computação.**

Orientador : Herman Martins Gomes

CAMPINA GRANDE - PB

2022

LUCAS GOMES AIRES

DEU TRADER

**Trabalho de Conclusão Curso
apresentado ao Curso Bacharelado em
Ciência da Computação do Centro de
Engenharia Elétrica e Informática da
Universidade Federal de Campina
Grande, como requisito parcial para
obtenção do título de Bacharel em
Ciência da Computação.**

BANCA EXAMINADORA:

**Herman Martins Gomes
Orientador – UASC/CEEI/UFCG**

**Hyggo Oliveira De Almeida
Examinador – UASC/CEEI/UFCG**

**Francisco Vilar Brasileiro
Professor da Disciplina TCC – UASC/CEEI/UFCG**

Trabalho aprovado em: 02 de Setembro de 2022.

CAMPINA GRANDE - PB

ABSTRACT

The cryptocurrency market has gained a lot of visibility over the last few years. According to Forbes, Brazil is within the top 5 countries with the largest number of investors in the world, reaching 10 million investors in 2022. Despite the large number of investors, the crypto market represents a high degree of risk.

This work proposes a predictive method for loss mitigation in the cryptocurrency market using machine learning techniques. To this end, a tool based on mathematical and graphic indicators was developed to support decisions to buy and sell the digital asset. The Gala/USDT currency was chosen for carrying out experiments, the Python language was used as support for the creation and manipulation of the Data Frame and in the end we obtained an accuracy of around 90% in the test set that represented 30 % of the Data Frame.

Deu Trader

Lucas Gomes Aires
UFCG - Universidade Federal de
Campina Grande
lucas.aires@ccc.ufcg.edu.br

RESUMO

O mercado de criptomoedas ganhou muita visibilidade nos últimos anos. Segundo a Forbes, o Brasil está entre os 5 países com maior número de investidores no mundo, chegando a 10 milhões de investidores¹ em 2022. Apesar do grande número de investidores, o mercado de criptomoedas representa um alto grau de risco.

Este trabalho propõe um método preditivo para mitigação de perdas no mercado de criptomoedas utilizando técnicas de aprendizado de máquina. Para tanto, foi desenvolvida uma ferramenta baseada em indicadores matemáticos e gráficos para apoiar as decisões de compra e venda do ativo digital. A moeda Gala/USDT foi escolhida para a realização dos experimentos, utilizou-se a linguagem Python [9], como apoio para criação e manipulação do *Data Frame* e no final obtivemos uma precisão em torno de 90% no conjunto de testes que representou 30% do *Data Frame*.

Keywords

Criptomoedas, técnicas de predição, machine learning.

1. INTRODUÇÃO

As criptomoedas podem ser usadas com as mesmas finalidades do dinheiro físico em si. Às três principais funções são servir como meio de troca, facilitando as transações comerciais; reserva de valor, para a preservação do poder de compra no futuro; e ainda como unidade de

conta, quando os produtos são precificados e o cálculo econômico é realizado em função dela². Além das funções citadas acima, é possível adquirir lucros em estratégias de compra e venda em determinado período, ou seja, através de operações de *Trader*. Para afirmarmos que o *trader* deu certo deve haver lucro na operação, que por sua vez é pautada em análises de variáveis de mercado, na experiência do operador em selecionar e analisar um conjunto de indicadores matemáticos, nos quais buscam sinalizar uma possível alta ou queda do ativo. Esse projeto investiga o uso de modelos preditivos para suporte aos investidores nas tomadas de decisões.

2. FUNDAMENTAÇÃO TEÓRICA

Nesta seção, são tecidas considerações sobre o funcionamento do mercado de criptomoedas e dos algoritmos de aprendizagem de máquina utilizados neste trabalho.

Um dos mais simples processos de compra de criptomoedas requer alguns passos: (1) primeiramente deve-se criar uma conta em uma *exchange* — corretora de criptomoedas; (2) em seguida, é necessário que ocorra uma transferência de dinheiro para conta criada; e (3) por fim. Faz-se a compra da moeda digital³. Antes da concretização da compra, é possível realizar uma análise gráfica da cotação na moeda naquele instante da operação. Existem inúmeros

1

<https://forbes.com.br/forbes-money/2022/05/mercado-de-cripto-do-brasil-ja-e-um-dos-cinco-maiores-do-mundo/#:~:text=Mais%20de%2010%20milh%C3%B5es%20de.n%C3%BAmero%20n%C3%A3o%20para%20de%20crescer&text=O%20Brasil%20est%C3%A1%20entre%20os.%2C%20EUA%2C%20R%C3%BAssia%20e%20Nig%C3%A9ria.>

2 <https://www.infomoney.com.br/guias/criptomoedas/>

3

<https://www.infomoney.com.br/guias/criptomoedas/#:~:text=%C3%89%20poss%C3%ADvel%20comprar%20cotas%20de.uma%20das%20formas%20mais%20simples>

tipos de gráficos, o mais comum e utilizado é o de *Candle*. O gráfico de *Candle* condensa diversas informações a respeito da moeda num instante de tempo pré determinado do gráfico (15 m, 1h, 2h, etc). Entre as principais informações condensadas, destacam-se as cotações de abertura e fechamento da moeda, as cotações máxima e mínima alcançadas naquele intervalo de tempo e se houve valorização ou desvalorização em um período de tempo anterior.



Figura 1. Gráfico de *Candle* da moeda Gala - 15m

Apesar de ser visto como um investimento fácil, é preciso observar que a simples compra aleatória de um ativo, traz muitos riscos devido à grande volatilidade do mercado. Podem ocorrer valorizações e desvalorizações em um curto tempo. Esta volatilidade é o que torna o mercado de criptomoedas tão atrativo, potencializando as possibilidades tanto de lucros quanto de perdas rápidas.

Quando um ativo é comprado, a compra fica registrada de forma criptografada e com elevado nível de segurança na blockchain.

No tocante aos algoritmos de aprendizagem de máquina, foram utilizados os seguintes neste trabalho:

- *Random Forest*: é um algoritmo de aprendizagem supervisionada que cria uma floresta de múltiplas árvores de decisão, cada uma treinada sobre partições aleatórias do conjunto de treinamento. A “floresta” é também vista como uma combinação (*ensemble*) de árvores de decisão, na maioria dos casos treinados com o método de *bagging*. [15] A ideia

principal do método de *bagging* é que a combinação dos modelos de aprendizado aumenta o resultado geral;

- *Gradient Boosting*: é uma técnica de aprendizado de máquina para problemas de regressão e classificação que produz um modelo de previsão na forma de um conjunto de modelos de previsão fracos, geralmente árvores de decisão [16];

- *AdaBoost*: é um algoritmo de aprendizado supervisionado do tipo boost. O *AdaBoost* combina um conjunto de funções simples de classificação, denominados classificadores fracos, para formar um classificador forte.[17]

3. METODOLOGIA

Para alcançar o alvo desta pesquisa, definiu-se como foco do estudo a criptomoeda Gala/Usdt, bem como foi definido o tempo de operação de 15 (quinze) minutos. Dentre os principais motivos para a escolha da Gala, estão: uma moeda com um bom tempo de mercado, ou seja, consolidada, sendo também uma das moedas que mais se valorizou em 2021. A opção de 15 m como intervalo de análise, deu-se por ser um dos tempos mais comuns a ser analisado pelos traders em suas operações. Fez-se necessário a criação de uma base de dados com as informações relacionadas aos candles, como, por exemplo: volume, horário de fechamento, valor do fechamento, etc.

Como não há nenhuma base pública de dados, foi necessária a extração da *exchange* Binance [1], que provê em sua api todas as informações e histórico de mercado das moedas que são comercializadas desde a inserção da moeda em sua base.

Para a extração dos dados, fez-se necessário o desenvolvimento de um backend com node.js [2] juntamente com o framework do Axios [3] para realização das requisições para API da Binance. Como banco de dados utilizou-se o MongoDB [4] por sua fácil integração com node.js e fácil visualização por meio do MongoDB Atlas[5]. Utilizou-se também o framework Mongoose [6] para realizar a conexão entre a aplicação e o banco de dados. Após a aquisição dos dados,

procedeu-se ao processamento, o qual envolveu a ordenação e o cálculo dos indicadores com o auxílio da biblioteca do *Technical Indicators* [7].

Com esta biblioteca é possível obter os valores dos indicadores (rsi, ema, vwap, etc.) de maneira rápida e com pouco esforço.

Foi criada uma rota de exportação dos arquivos com a utilização do framework express [8] em formato de csv. Por fim, todo o código desenvolvido neste trabalho foi adicionado a um repositório no github⁴.

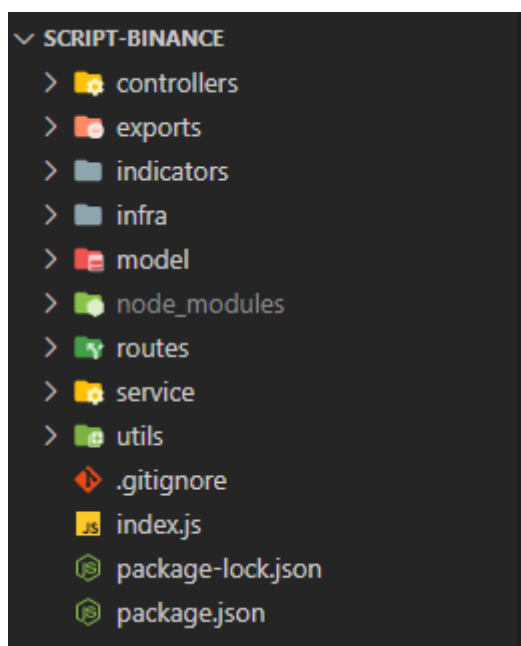


Figura 2 - Backend

A linguagem de programação Python [9] foi selecionada para construção do *Data Frame*. O data frame pode ser visto como uma tabela da base de dados, em que cada linha corresponde a um registro da tabela. Ele se fez necessário para possibilitar a análise exploratória dos dados, manipulação e organização.

Utilizando-se a biblioteca Pandas [10] é possível realizar manipulações de conteúdo, criar atributos a partir de outros já existentes e salvar em um novo *Data Frame*, ou mesmo substituir o existente.

Após a criação do Data Frame, foi feita uma ordenação dos dados e um tratamento que retirou as colunas que apresentavam poucos dados. Este tratamento fez-se necessário, pois, a grande

disparidade de dados entre os indicadores tornaram o tratamento da lógica de lucro mais onerosa.

O valor chave para predição foi inserido na coluna de closes, que é efetivamente o valor da moeda após determinado período anteriormente selecionado.

Foi escolhida uma janela, de início, arbitrariamente, para realização da predição. Estipulou-se também a margem de lucro inicialmente em um valor de 1.5% a cada intervalo de período, utilizando um modelo binário de escolha no qual o resultado foi inserido no Data frame.

O Data Frame foi dividido na proporção de 70% para treinamento e 30% para o teste, num esquema conhecido como método Holdout [11], que consiste em dividir os dados em dois grupos, geralmente separa 2/3 dos dados para treino e 1/3 para testes. Após a separação dos dados do data frame, com o uso da biblioteca Sklearn [12], foram selecionados três algoritmos de aprendizagem de máquina. Tratam-se de algoritmos de uso geral, que se adaptaram facilmente aos dados da pesquisa, além de serem conhecidos como bons preditores.

Para escolha das entradas de teste, foi considerada a quantidade de dados obtidos através do cálculo dos indicadores, além dos indicadores mais relevantes utilizados pelos *Traders*.

Os primeiros resultados de testes foram significativamente bons. Atingiu-se a entre 70% a 90% de acurácia na predição dos dados de teste. Foi realizado o processo de adaptação dos algoritmos, para maior efetividade e por fim, realizados os testes no modelo que apresentou, assim como o treino, elevado índice de aproveitamento. Entretanto, foi percebida a presença de ruído nos dados, o que fez com que a predição fosse feita incorretamente.

Para solucionar o problema foi necessário a implementação da técnica de suavização da média móvel [13], comumente utilizada em aplicações que envolvem séries temporais.

A suavização é uma técnica aplicada às séries temporais para remover a variação refinada entre etapas de tempo.

⁴ <https://github.com/lucasaires/script-binance>

A esperança de suavização é remover o ruído e expor melhor o sinal dos processos causais subjacentes. As médias móveis são um tipo simples e comum de suavização usado na análise de séries temporais e na previsão de séries temporais.

O cálculo de uma média móvel envolve a criação de uma nova série em que os valores são compostos pela média das observações brutas na série temporal original.

Uma média móvel requer que seja especificado um parâmetro chamado de largura da janela. Isso define o número de observações brutas usadas para calcular o valor da média móvel.

A parte “móvel” na média móvel refere-se ao fato de que a janela definida pela largura da janela é deslizada ao longo da série temporal para calcular os valores médios na nova série.

Existem dois tipos principais de média móvel que são usados: média móvel centrada e média móvel [14].

Com auxílio da função `rolling()` do `pandas`, que agrupa uma janela e retorna o valor médio, foi realizada a suavização dos dados da tabela de close e inserida uma nova coluna do Data Frame.

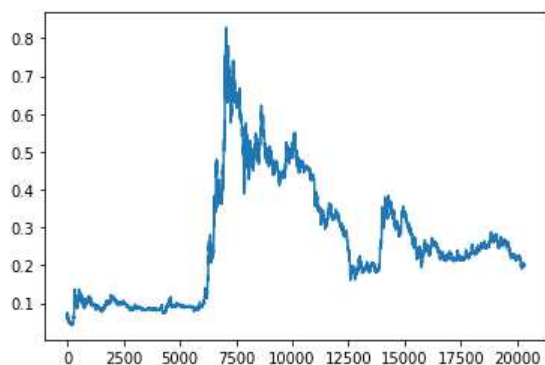


Figura 2 - Gráfico de closes antes da suavização

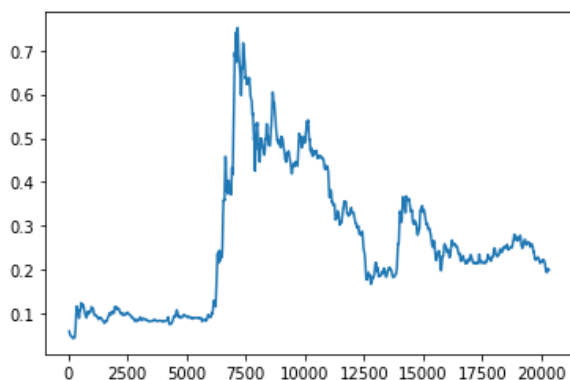


Figura 3 - Gráfico de closes após a suavização.

As Figuras 2 e 3 mostram um comparativo antes e depois da suavização da coluna de closes. A coluna esta, que representa o valor de fechamento da moeda. É de fácil percepção a mudança em relação à densidade da curva. Pelo mesmo motivo já citado anteriormente foi aplicada a mesma estratégia para todos os indicadores utilizados na predição.

4. RESULTADOS E DISCUSSÃO

Ao processar os algoritmos foi verificada uma elevada taxa de acerto. Para maior evidência de efetividade nos resultados, foi computada a Matriz de Confusão para cada algoritmo.

A matriz de confusão é uma tabela que mostra as frequências de classificação para cada classe do modelo. As taxas constituintes de uma matriz de confusão estão detalhadas a seguir:

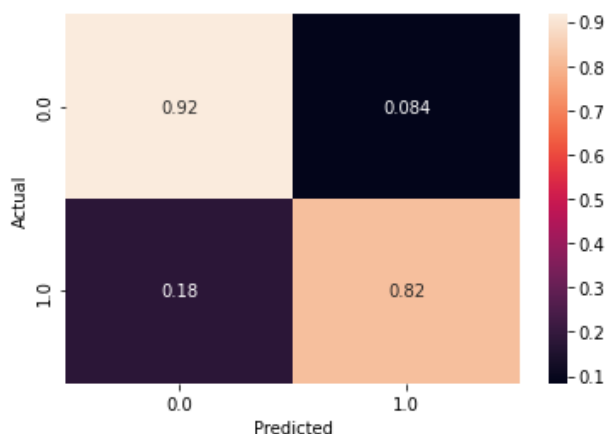
- **Verdadeiro positivo (*true positive* — TP):** ocorre quando no conjunto real, a classe que estamos buscando foi prevista corretamente.
- **Falso positivo (*false positive* — FP):** ocorre quando no conjunto real, a classe que estamos buscando prever foi prevista incorretamente.
- **Falso verdadeiro (*true negative* — TN):** ocorre quando no conjunto real, a classe que não estamos buscando prever foi prevista corretamente.
- **Falso negativo (*false negative* — FN):** ocorre quando no conjunto real, a classe que não estamos buscando prever foi prevista incorretamente.[14]

TP	FP
FN	TN

Figura 4 - Exemplo de matriz de confusão.

Foram obtidos os seguintes resultados:

Random Forest:

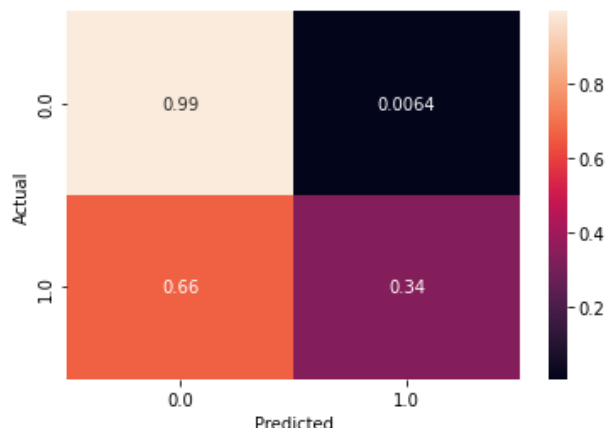


Resultado do algoritmo:
0.911107459743674
91.11

	precision	recall	f1-score	support
0.0	0.99	0.92	0.95	5769
1.0	0.35	0.82	0.49	317
accuracy			0.91	6086
macro avg	0.67	0.87	0.72	6086
weighted avg	0.96	0.91	0.93	6086

Figura 5 - Matriz de confusão Random Forest

Gradient Boost:

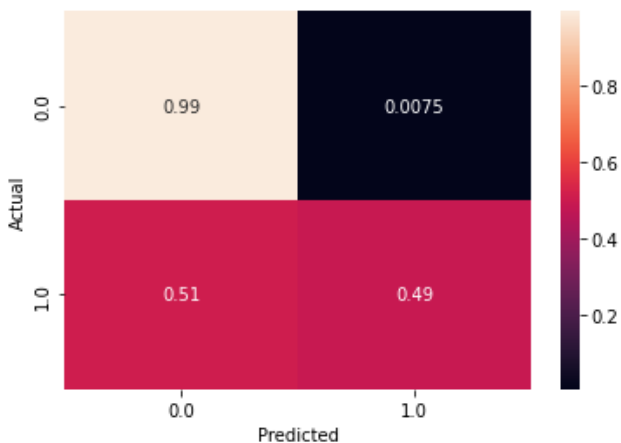


Resultado do algoritmo:
0.9594150509365758
95.94

	precision	recall	f1-score	support
0.0	0.96	0.99	0.98	5769
1.0	0.74	0.34	0.46	317
accuracy			0.96	6086
macro avg	0.85	0.67	0.72	6086
weighted avg	0.95	0.96	0.95	6086

Figura 7 - Matriz de confusão Gradient Boost

AdaBoost:



Resultado do algoritmo:
0.9664804469273743
96.65

	precision	recall	f1-score	support
0.0	0.97	0.99	0.98	5769
1.0	0.78	0.49	0.60	317
accuracy			0.97	6086
macro avg	0.88	0.74	0.79	6086
weighted avg	0.96	0.97	0.96	6086

Figura 6 - Matriz de confusão AdaBoost

Analisando as matrizes foi possível fazer algumas inferências a respeito dos resultados:

O algoritmo que demonstrou melhor acurácia foi o AdaBoost, seguido pelo Gradiente Boost e por fim Random Forest. Entretanto, essa acurácia é efeito apenas do resultado do primeiro quadrante (TP - Verdadeiro positivo) mostrado na Figura 4, nosso real interesse é a predição da diagonal principal, neste caso o algoritmo que apresentou melhor desempenho foi o Random Forest. Mesmo apresentando o menor índice de acurácia, comparado ao AdaBoost e o Gradiente Boost que apresentaram [0,99; 0,49] e [0,99; 0,34] respectivamente em sua diagonal principal, teve elevado percentual de assertividade [0,92; 0,82], o Random Forest comete menos erros, na predição do momento de possível alta da moeda, tornando-o o melhor.

O algoritmo que ocorreu mais erros foi o Gradient Boost, no qual ele apontou 66% de erros, na tentativa de predição de alta.

5. CONCLUSÃO

Pode-se concluir que, através dos resultados deste TCC, é possível obter elevadas taxas de acerto na predição comportamental da moeda Gala/Usdt no período gráfico de 15m, com um valor de lucro fixado em 1,5%. Este resultado evidencia que a ferramenta desenvolvida poderia ser empregada como suporte às decisões estratégicas de compra e venda de criptoativos.

Contudo, é importante ressaltar que, apesar de termos um possível caso de sucesso, são necessárias mais análises em cenários diferenciados para entendermos melhor quão poderosa é a ferramenta e o possível índice de lucratividade.

Pode-se ainda instigar o maior desenvolvimento da ferramenta, através de um maior estruturação do backend, hospedando em algum serviço de cloud como, por exemplo, AWS, e uma construção de um frontend integrado com a própria api da binance, automatizando a compra e venda de qualquer criptoativo.

6. REFERÊNCIAS

- [1] Binance- <https://binance-docs.github.io/apidocs/spot/en/#introduction> .
- [2] Node.js- <https://nodejs.org>.
- [3] Axios- <https://axios-http.com/ptbr/docs/intro>
- [4] MongoDB- <https://www.mongodb.com/>
- [5] MongoDB Atlas- <https://www.mongodb.com/cloud/atlas>.
- [6] Mongoose ODM - Object modeling for Node.js- <https://mongoosejs.com/>
- [7] Technical Indicators- <https://github.com/anandanand84/technicalindicators#readme>
- [8] Express- Node.js web application framework. <https://expressjs.com>
- [9] Python- <https://www.python.org>
- [10] Pandas- <https://pandas.pydata.org>
- [11] Qasem A Al-Radaideh, Emad M Al-Shawakfa, and Mustafa I Al-Najjar.2006. Mining student data using decision trees. In International Arab Conference on Information

Technology (ACIT'2006), Yarmouk University, Jordan. 4–5.

- [12] Sklearn- <http://scikit-learn.org/stable>
- [13] Machine Learning Mastery- [Moving Average Smoothing for Data Preparation and Time Series Forecasting in Python](#)
- [14] Medium - [Entendendo o que é Matriz de Confusão com Python | by Emanuel G de Souza | Data Hackers | Medium](#)
- [15] Random Forest - <https://blogs.oracle.com/ai-and-datascience/post/random-forests-decision-trees-and-ensemble-methods-explained>
- [16] Gradient Boost - <https://medium.com/equal-lab/uma-breve-introducao-ao-algoritmo-de-machine-learning-gradient-boosting-utilizando-a-biblioteca-311285783099>]
- [17] AdaBoost- <https://machinelearningmastery.com/boosting-an-adaboost-for-machine-learning/>