

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Luis Felipe Nunes Reis

**Influência da movimentação de baleias sobre a
cotação do Bitcoin**

Uberlândia, Brasil

2025

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Luis Felipe Nunes Reis

**Influência da movimentação de baleias sobre a cotação
do Bitcoin**

Trabalho de conclusão de curso apresentado
à Faculdade de Computação da Universidade
Federal de Uberlândia, como parte dos requi-
sitos exigidos para a obtenção título de Ba-
charel em Sistemas de Informação.

Orientador: Prof. Dr. Ivan da Silva Sendin

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Sistemas de Informação

Uberlândia, Brasil

2025

Luis Felipe Nunes Reis

Influência da movimentação de baleias sobre a cotação do Bitcoin

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Sistemas de Informação.

Trabalho aprovado. Uberlândia, Brasil, 23 de setembro de 2025:

Prof. Dr. Ivan da Silva Sendin
Orientador

Prof. Dr. Diego Nunes Molinos

Prof. Dr. Leonardo Muttoni

Uberlândia, Brasil

2025

Agradecimentos

Agradeço primeiramente a Deus, pela iluminação durante o longo caminho da graduação, a minha mãe Rejane, que já se encontra no descanso eterno, minha maior inspiração de vida, a Julia, o amor da minha vida e minha maior motivadora para não desistir ao longo desses anos, aos meus pais Gian e Julio, por todos os ensinamentos sobre como ser um homem, a FACOM e seus professores, em especial, meu orientador Professor Dr. Ivan, pelos ensinamentos técnicos, a família e amigos, em especial, Dr. João Victor Vilela, pela base sólida que vocês são para mim e por fim a mim mesmo, pelos erros e acertos, mas sobre tudo, a vontade de ser algo maior. Um conselho para quem ler essa monografia: estude e entenda as cinco leis do ouro.

Resumo

O Bitcoin é a pioneira e principal criptomoeda disponível para comércio nos tempos atuais, cada vez mais investidores dão atenção a esse ativo, seja criando reservas da mesma em suas carteiras, seja visando entender melhor como tudo isso funciona. Como na bolsa de valores, onde temos os grandes investidores também conhecidos como "Tubarões", na Blockchain temos carteiras, ou endereços, que detêm grande quantidade da moeda, conhecidas como "Baleias". Neste trabalho, analisamos as transações dessas carteiras, somado as variações da cotação do Bitcoin, para tentar entender se o primeiro, exerce algum tipo de influência no segundo. Estipulamos a janela de 1 de Agosto de 2021 até 6 de Agosto de 2022 e montamos duas bases com dados desse período, uma com as transações, outra com as cotações. Diante desses dados, aplicamos duas técnicas de Análise de dados e *Machine Learning*, Regressão Linear e Regressão Logística para avaliar a hipótese levantada. De primeira mão, não foi possível identificar nenhuma relação linear ou direta entre essas duas variáveis, com coeficientes de relação quase zerados entre elas. Porém, após treinar um modelo, utilizando Regressão Logística, tivemos resultados mais satisfatórios. Onde foi possível notar que alguns fatores como o preço de fechamento, com um coeficiente alto de aproximadamente 3.80, e o montante transacionado pelas Baleias — o ponto central desse estudo — exercem influência positiva, de aproximadamente 0.10, na probabilidade de um retorno positivo no dia, ou seja, o preço de fechamento maior que o de abertura.

Palavras-chave: Bitcoin, Baleias, Cotação, Análise de Dados, Regressão Logística

Lista de ilustrações

Figura 1 – Cotação do Bitcoin do final de 2014 até meados de 2020 Fonte: (INVESTIDOR10, 2025)	12
Figura 2 – Volume transacionado por baleias por dia. Fonte: O Autor	20
Figura 3 – Preço de Fechamento (USD) vs Montante de Transações (BTC). Fonte: O Autor	20
Figura 4 – Volatilidade comparada com o Volume de transações de baleias. Fonte: O Autor	21
Figura 5 – HeatMap das Variáveis do Estudo. Fonte: O Autor	22
Figura 6 – Curva ROC. Fonte: O Autor	30

Lista de tabelas

Tabela 1 – Estatísticas gerais das transações observadas	19
Tabela 2 – Coeficientes do Modelo	28

Lista de abreviaturas e siglas

API	Application Programming Interface (Interface de Programação de Aplicação)
UTXO	Unspent Transaction Output (Saida de transação ainda não gasta)
BTC	Bitcoin
ROC	Receiver Operating Characteristic (Curva Característica de Operação do Receptor)
P2P	Rede Peer to Peer
AUC	Area under the curve (Área sob a Curva)
TPR	True Positive Rate (Taxa de verdadeiros positivos)
FPR	False Positive Rate (Taxa de falsos positivos)

Sumário

1	INTRODUÇÃO	9
1.1	Objetivos	10
1.1.1	Objetivo Geral	10
1.1.2	Objetivos Específicos	10
2	REVISÃO BIBLIOGRÁFICA	11
2.1	Bitcoin: o que é e como funciona	11
2.2	O que são baleias e como identificá-las	13
2.3	Análise das Cotações do Bitcoin	13
2.4	Régressão Logística	14
2.5	Trabalhos relacionados	15
3	DESENVOLVIMENTO	17
3.1	Base de Cotações do Bitcoin	17
3.2	Base de Transações de Baleias	18
3.3	Observações iniciais	19
3.3.1	Soma do montante de BTCS transacionados por dia	19
3.3.2	Preço de Fechamento vs Montante de Transações	19
3.3.3	Volatilidade x Volume de transações	21
3.3.4	Correlação entre as variáveis do Estudo	22
3.4	Aplicação da Régressão Logística no Estudo	23
3.4.1	Normalização das variáveis	24
3.4.2	Separação em Treino e Teste	24
3.4.3	Treino e Avaliação do modelo	25
4	RESULTADOS	27
4.1	Acurácia do Modelo	27
4.2	Coeficientes do Modelo	27
4.2.1	Curva ROC e AUC	28
5	CONCLUSÃO	31
	REFERÊNCIAS	33

1 Introdução

Em meados de 2008, Satoshi Nakamoto publicou o artigo "Bitcoin, uma moeda digital baseada em rede peer-to-peer" (NAKAMOTO; BITCOIN, 2008), os detalhes de como a moeda, a *blockchain* e os fatores técnicos envolvidos funcionam, não é o fator a ser descrito nesse momento e sim como ao longo dos anos o Bitcoin ganhou força no mercado de *exchanges*. No ano de 2020, o primeiro ano da pandemia do COVID-19, a moeda apresentou uma valorização de 170% (NEWS, 2020). Essa valorização notória pode estar relacionada ao fato da moeda ser deflacionária como previsto por Nakamoto. A moeda parará de ser emitida quando atingir 21 milhões de unidades em 2140. As oscilações no preço estão ligadas a clássica lei de mercado, oferta e procura, como também eventos globais e segurança de transações (INFOMONEY, 2023). No mercado de ações, um fator que causa oscilações, além dos já citados, é quando grandes carteiras fazem movimentações, como no caso da venda de parte de seus ativos da *Apple* por *Warren Buffet* (SANTOS, 2024). Considerado por muitos, o maior investidor da história, dono da filosofia de sempre comprar e nunca vender, a venda de metade da sua participação na companhia fundada por Steve Jobs, trouxe uma queda na confiança dos demais investidores, o que acarretou na queda do preço do ativo. A hipótese deste trabalho é que as movimentações de grandes investidores em Bitcoin, as chamadas 'baleias' (KHARIF, 2017), exercem influência significativa nas variações de preço da criptomoeda.

A análise do impacto das grandes movimentações de mercado no preço do Bitcoin é um tema de grande relevância tanto para a sociedade quanto para a academia (GANDAL et al., 2018). Para a sociedade, essa pesquisa pode auxiliar na tomada de decisões de investimento e na compreensão dos riscos associados às criptomoedas. Para a academia, esse trabalho traz relevância sobre dar força a ideia que o Bitcoin, apesar de ser um ativo completamente diferente de ativos do mercado financeiro, também está relacionado a perspectivas do mercado, especulações e fatores extra técnicos. Isso abre margem pra diversas pesquisas sobre previsões do valor do Bitcoin, popularidade da moeda, entre outras coisas.

A ideia deste trabalho é montar uma base de dados com as transações das Baleias dentro da *Blockchain* e uma segunda base dos valores de negociação do Bitcoin, definir um intervalo de tempo relevante e analisar as variações do preço do ativo. Com esses dados e, utilizando técnicas de análise de dados e *machine learning*, podemos chegar em uma conclusão sobre se afinal, as movimentações das baleias tem ou não relação com a variação do preço do Bitcoin.

Para realizar tal atividade, se utilizou a linguagem Python (PYTHON, 2025), tanto

para a obtenção dos dados, quanto na parte de geração de relatórios, com auxílio das bibliotecas *Pandas*, *SkLearn*, *Seaborn*, *NumPy* e *MatPlot Lib*. Para construir a primeira base de dados com as transações das baleias, o script criado integra com as APIs disponibilizadas pela ([BLOCKCHAIN.COM, 2025](#)) e ([SERVICES., 2025](#)). A segundo base, sobre a variação do preço do Bitcoin para cada dia analisado, foi utilizada a base de dados *Bitcoin OHLCV* ([BAHAGHIGHAT, 2024](#)). Esses dados foram transformados em dois arquivos *json*, um para transações, outro para cotações.

Com esses dois *datasets* prontos, o estudo empregou o uso de duas abordagens, a primeira utilizando regressão linear, para verificar se existe uma relação direta das movimentações com a cotação da moeda. A segunda abordagem consiste em utilizar regressão logística, criando uma variável *target*, que é o retorno da cotação do dia, ou seja, se o preço subiu ou caiu, junto dela, algumas outras variáveis preditoras, que foram utilizadas para treinar o modelo. Com o modelo treinado foi possível visualizar, quais dessas variáveis tem um efeito positivo para a variável alvo, ou seja, aumentam a possibilidade de um retorno positivo e quais tem efeito negativo.

1.1 Objetivos

1.1.1 Objetivo Geral

O objetivo deste trabalho é analisar o impacto da movimentação das carteiras Baleias na cotação do Bitcoin.

1.1.2 Objetivos Específicos

Empregar duas abordagens estatísticas com o objetivo de analisar o comportamento das variáveis selecionadas em relação à cotação do BTC. Inicialmente, aplicar a regressão linear com a finalidade de identificar possíveis relações lineares entre as variáveis independentes e o preço do ativo.

Na sequência, adotar o modelo de regressão logística, tendo como variável dependente o retorno diário do BTC. As variáveis preditores (*features*) serão o preço de fechamento, a máxima e mínima diárias, além do volume de transações realizado pelas baleias, este último sendo de maior interesse para a investigação. O treinamento do modelo permite estimar os coeficientes associados a cada variável, possibilitando a interpretação da magnitude e direção (positiva ou negativa) da influência de cada preditor sobre a variável de interesse.

2 Revisão Bibliográfica

2.1 Bitcoin: o que é e como funciona

Criado em meados de 2008, por Satoshi Nakamoto, o Bitcoin foi a pioneira do que hoje se conhece como criptomoeda (NAKAMOTO; BITCOIN, 2008). Com uma proposta descentralizada e livre de burocracias governamentais, há até quem o considere Anárquico, pois, baseando em conceitos anarquistas clássicos, como os introduzidos por Pierre-Joseph Proudhon, em sua obra *O Princípio Federativo* (PROUDHON, 1863), onde o autor declara que a ordem social deveria surgir de nada além de transações e trocas.

Moedas oficiais, são de propriedade de seus respectivos países, além de impostos e taxas estatais em transações com as mesmas, contudo o Bitcoin, é de propriedade de um endereço, e as transações dizem respeito apenas aos dois endereços envolvidos, sem intervenção de qualquer orgão regulador ou governo (SCOTT, 2016).

A moeda digital, organizada em uma rede *P2P*, ou seja, os integrantes da rede não possuem um papel definido e podem atuar tanto como utilizador de recurso, utilizando a rede para participar de uma transação, ou servir como mecanismo realizador dessas transações, através da mineração (ANTONOPoulos, 2014). Em seus primeiros anos, o Bitcoin não teve uma aceitação expressiva do grande público, seja profissionais da área de *exchange*, investidores, ou cidadãos comuns. Porém, se tornou popular principalmente em transações entre pessoas que buscavam o anonimato da mesma, pelas mais variadas razões. De pouco em pouco, o Bitcoin foi se tornando cada vez mais conhecido, por consequência, tornando-se também, mais valorizado, até que tivemos o primeiro grande salto, no fim de 2017, chegando a custar quase 20 mil dólares, conforme pode ser visto na Figura 1.

Para os olhos de alguns investidores e estudiosos do Bitcoin, a moeda representa a única forma de se proteger contra a inflação da economia do país que se vive e também, possuir algo que seja totalmente seu, pois moedas estatais, ativos financeiros, propriedades, entre outros bens comuns, estão atrelados ao Estado onde se encontram, o que não acontece para o Bitcoin (SCOTT, 2016).

Do ponto de vista técnico, o funcionamento do Bitcoin é sustentado por um modelo de transações baseado em *UTXO* (*Unspent Transaction Output*), que difere significativamente dos modelos de conta utilizados em sistemas financeiros tradicionais. Em vez de manter saldos associados a contas, como ocorre em bancos ou no *Ethereum*¹, o Bitcoin utiliza saídas de transações não gastas como unidades fundamentais.

¹ Criptomoeda fundada em 2015 usando a tecnologia blockchain.

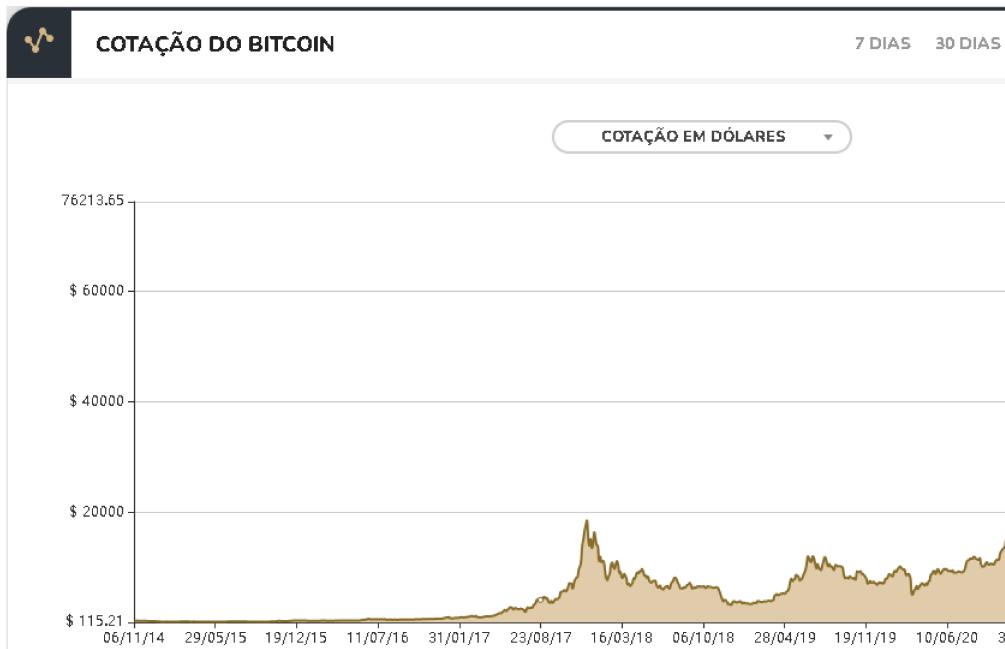


Figura 1 – Cotação do Bitcoin do final de 2014 até meados de 2020 Fonte: ([INVESTIDOR10, 2025](#))

Cada transação no Bitcoin consome um ou mais UTXOs como entrada e cria novos UTXOs como saída. Assim, o "saldo" de um endereço é, na verdade, o somatório de todos os UTXOs associados a ele. Isso garante maior rastreabilidade e segurança, além de favorecer a paralelização na validação das transações ([ANTONOPoulos, 2014](#)).

Os **endereços Bitcoin** são identificadores únicos gerados a partir de chaves públicas criptográficas, codificados geralmente no formato *Base58Check*², e são utilizados como destino (ou origem) de transações. Cada endereço representa a posse de uma ou mais chaves privadas, que permitem autorizar a movimentação de bitcoins vinculados àquele endereço.

O processo de envio de bitcoins envolve a *assinatura digital* de uma transação, provando a propriedade da chave privada correspondente ao endereço emissor. A rede valida essa transação, garantindo que os UTXOs utilizados não tenham sido gastos anteriormente (prevenindo o *double spending*) e atualizando o estado da blockchain ao incluir a transação em um bloco minerado.

² O *Base58Check* é um esquema de codificação utilizado em endereços de Bitcoin que combina a codificação Base58 com um checksum (verificação de integridade) para detectar erros de digitação ou cópia.

2.2 O que são baleias e como identificá-las

No contexto do mercado financeiro tradicional, o termo *shark* (tubarão) é frequentemente utilizado para designar investidores ou instituições com elevado patrimônio e significativa capacidade de movimentação de capital (ALIZADEH et al., 2025). Analogamente, no ecossistema das criptomoedas — especialmente na rede do Bitcoin — adota-se o termo *whale* (baleia) para se referir a endereços que detêm grandes volumes da moeda digital. Essas baleias podem ser identificadas por meio da análise das transações registradas na *blockchain*, mais especificamente nos campos de *inputs* e *outputs* do modelo UTXO, onde movimentações de grande escala evidenciam a atuação desses agentes.

As Baleias são indivíduos ou entidades que possuem uma quantidade substancial de Bitcoin, geralmente superior a 1.000 BTC (TASCA; HAYES; LIU, 2018) e segundo levantamento feito em 2017, na época as Baleias detinham 40% do mercado (KHARIF, 2017). Devido à sua capacidade de realizar transações de grande volume, essas entidades podem exercer uma influência significativa sobre o mercado de criptomoedas, podendo impactar a dinâmica de preços e contribuir para a volatilidade do mercado (HERREMANS; LOW, 2022).

A identificação das Baleias e a compreensão de seus padrões de negociação são essenciais para a observar sua influencia no mercado (KHARIF, 2017).

Um dos principais fatores que possibilitam a identificação das baleias de Bitcoin é a análise dos dados da blockchain do Bitcoin. Por meio do histórico de transações e da distribuição das participações em Bitcoin, é possível identificar os endereços mais ativos e influentes na rede. Essa abordagem, permite a caracterização dos principais agentes da economia do Bitcoin, incluindo as baleias.

2.3 Análise das Cotações do Bitcoin

A cotação do Bitcoin é determinada de forma descentralizada e dinâmica, refletindo exclusivamente as forças de oferta e demanda nos diversos mercados em que é negociado. Diferentemente das moedas fiduciárias, como Real e Dólar, cuja taxa de câmbio pode ser influenciada por bancos centrais e políticas monetárias, o Bitcoin não possui um emissor central ou órgão regulador responsável por sua estabilização. Seu preço é formado nos chamados mercados de câmbio de criptomoedas (*exchanges*), onde compradores e vendedores realizam negociações em tempo real.

O valor do ativo flutua conforme o volume, liquidez e interesses dos participantes do mercado global (GANDAL et al., 2018). Como não existe uma cotação oficial única, diferentes plataformas podem apresentar variações de preço em determinados momentos. Essa estrutura aberta e globalizada torna a cotação do Bitcoin altamente sensível a fatores

como eventos macroeconômicos, regulação, adoção institucional (ALONSO et al., 2023) e sentimento de mercado.

Nesse trabalho, uma análise sobre as movimentações de grandes detentores do ativo, conhecidos como *whales*, demonstra a influência desses agentes na cotação.

2.4 Regressão Logística

A regressão logística é um modelo estatístico amplamente utilizado para modelar a probabilidade condicional de ocorrência de um evento binário (variável resposta categórica com dois níveis), dado um conjunto de variáveis explicativas. Trata-se de um caso particular da família dos modelos lineares generalizados (GLMs), onde a função de ligação utilizada é a logit (JR; LEMESHOW; STURDIVANT, 2013), definida como:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = \mathbf{x}^\top \boldsymbol{\beta}, \quad (2.1)$$

em que $p = P(Y = 1 \mid \mathbf{x})$ representa a probabilidade condicional do evento de interesse ocorrer, $\mathbf{x} = (1, x_1, \dots, x_k)^\top$ é o vetor de preditores (com intercepto) e $\boldsymbol{\beta}$ é o vetor de coeficientes a ser estimado.

A função inversa da logit, conhecida como função logística ou sigmoide, mapeia os valores da combinação linear dos preditores para o intervalo $(0, 1)$, assegurando uma interpretação probabilística da resposta:

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad \text{com } z = \mathbf{x}^\top \boldsymbol{\beta}. \quad (2.2)$$

A estimativa dos parâmetros $\boldsymbol{\beta}$ é realizada via o método da máxima verossimilhança, através da maximização da função de verossimilhança associada à distribuição Bernoulli:

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i}, \quad (2.3)$$

ou, de forma equivalente, pela log-verossimilhança:

$$\log \mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^n [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]. \quad (2.4)$$

Um dos principais atrativos do modelo logístico está na interpretação dos coeficientes em termos da razão de chances (odds ratio). O valor de e^{β_j} representa a variação multiplicativa na razão de chances associada a um incremento unitário na variável x_j , mantendo-se os demais preditores constantes.

Conforme destacado em (JR; LEMESHOW; STURDIVANT, 2013), a regressão logística é especialmente adequada para situações onde a variável resposta é dicotômica e os pressupostos da regressão linear clássica não são válidos. Além de seu uso tradicional em epidemiologia e ciências sociais, a regressão logística tem ganhado relevância em problemas de classificação binária em aprendizado de máquina, onde é frequentemente empregada como baseline por sua robustez e interpretabilidade (PENG; LEE; INGERSOLL, 2002).

2.5 Trabalhos relacionados

Essa sessão apresenta alguns trabalhos que tem relações com o tema deste trabalho, ou fizeram algo que se assemelha quanto ao objetivo do mesmo.

Hathidara, Atavale e Chaudhary (2024) propõem investigar a correlação entre o conteúdo de postagens na rede social Twitter e a variação no preço do Bitcoin ao longo do tempo. O estudo parte da hipótese de que menções à criptomoeda por usuários de alta influência — como políticos, artistas e figuras públicas — podem exercer impacto significativo sobre o mercado, promovendo flutuações no valor do ativo digital.

O principal objetivo dos autores é desenvolver um modelo preditivo baseado em aprendizado de máquina, capaz de antecipar a direção ou magnitude das variações no preço do Bitcoin, com base no conteúdo textual de tweets. Para isso, os pesquisadores construíram um modelo de regressão treinado com dados coletados entre os anos de 2016 e 2019 — período no qual o valor do Bitcoin apresentou variações superiores a US\$ 19.000.

Apesar do esforço analítico e da aplicação de técnicas modernas de mineração de texto e modelagem estatística, os resultados obtidos revelaram desempenho insatisfatório. As métricas de avaliação, especificamente a curva ROC (Receiver Operating Characteristic), indicaram que a acurácia dos modelos treinados ficou muito próxima de 0,5, o que sugere uma capacidade de predição comparável ao acaso. Em outras palavras, os modelos demonstraram baixa capacidade discriminativa, não sendo eficazes para prever com confiabilidade o comportamento futuro do preço do Bitcoin com base exclusivamente nas informações extraídas dos tweets.

Fratrič et al. (2022), os pesquisadores partiram da premissa que fraudes cometidas entre 2017 e 2018 foram responsáveis pela variação do preço do Bitcoin durante esse período. Utilizando dados empíricos desse período, criaram uma cópia do ambiente do Bitcoin, diferentes perfis de investidores, e simularam algumas ações que foram cometidas no mundo real. Os pesquisadores constataram que a queda no preço e o volume de anomalias foram devidas à fraudes cometidas pelo perfis de fraudadores criados na simulação. Sua relação com esse futuro trabalho é o fato do comportamento das pessoas ligadas ao Bitcoin, ter influência na variação do preço.

[Herremans e Low \(2022\)](#) apresentam um método de predição de volatilidade do preço do Bitcoin, utilizando *deep learning* com dados sobre carteiras baleias, extraídos principalmente da Blockchain e de *tweets* de alerta de baleias. Após criados os algoritmos de previsão e alguns meses de testes, os pesquisadores integraram os resultados de previsão com várias estratégias de negociação de referência simples e obtiveram lucro, provando a eficácia do método proposto. O trabalho desenvolve uma maneira de prever o preço do Bitcoin, similar com o proposto desse futuro trabalho, utilizando informações sobre carteiras baleias.

[Filho \(2019\)](#) utiliza técnicas de aprendizagem supervisionada treinando Rede Neural Perceptron Multicamadas, utilizando dados de tweets, com o objetivo de estabelecer relações entre o sentimento desses tweets e os movimentos do preço do Bitcoin. Após 3 meses de testes, o autor obteve uma precisão de 84% tendo como fonte de dados apenas dados do Twitter, o que superou a precisão de 72% do modelo utilizando dados do Mercado. O trabalho traz várias técnicas de como lidar com esses dados, tal como diferentes resultados se modificando a janela de tempo de experimento, o que será fonte valiosa de informação para esse futuro trabalho.

3 Desenvolvimento

Este capítulo descreve detalhadamente as etapas adotadas para investigar a possível relação entre grandes transações na blockchain do Bitcoin — conhecidas como movimentos de baleias — e a flutuação diária do preço do ativo. Inicialmente, foi estruturado um conjunto de dados contendo as cotações diárias do Bitcoin em um determinado período, incluindo informações como preço de abertura, fechamento, máximas, mínimas e volume negociado. Em seguida, foram utilizadas APIs especializadas em dados on-chain para extrair informações sobre transações com volumes superiores a 1000 BTC, associando essas movimentações aos respectivos dias de negociação. Com base nesses dados integrados, aplicou-se um modelo de regressão logística, com o objetivo de verificar se há correlação estatística significativa entre a atividade dessas grandes carteiras e a variação dos preços do Bitcoin. As seções seguintes apresentam, de forma detalhada, cada uma dessas etapas, desde a coleta e preparação dos dados até a modelagem e análise dos resultados obtidos. O código-fonte completo está disponível em [<https://github.com/LuisFelipeNuness/tcc-oficial>](https://github.com/LuisFelipeNuness/tcc-oficial).

3.1 Base de Cotações do Bitcoin

Para compreender o comportamento histórico do preço do Bitcoin, foram utilizados dados OHLCV (*Open, High, Low, Close, Volume* ou, respectivamente, preço de abertura, maior cotação do dia, menor cotação do dia, preço de fechamento e volume transacionado no dia) disponibilizados por (BAHAGHIGHAT, 2024), os quais fornecem informações diárias estruturadas sobre o mercado do ativo digital. Este tipo de dado é fundamental na análise de séries temporais financeiras, sendo amplamente utilizado tanto em estudos acadêmicos quanto em aplicações práticas no mercado financeiro.

A partir dessas variáveis básicas, foi possível derivar um conjunto de características financeiras que ampliam a capacidade de modelagem e interpretação do comportamento do preço do Bitcoin. Entre as variáveis derivadas, destacam-se:

- **Volatilidade:** calculada com base nas diferenças entre os preços máximos e mínimos diários (*High - Low*), servindo como proxy para o grau de incerteza e instabilidade do mercado em determinado dia;
- **Retorno diário:** definido pela diferença entre o preço de fechamento e o de abertura do mesmo dia (*Close - Open*), essa métrica reflete o ganho ou perda intradiária;

- **Retornos agregados:** consistem no somatório dos retornos diáridos acumulados nos últimos 3, 5 e 7 dias, respectivamente. Essas características são particularmente úteis para capturar tendências de curto prazo e identificar padrões de reversão ou continuidade.

Essas transformações são recorrentes na literatura sobre previsão de preços e análise de ativos financeiros, sendo frequentemente empregadas em modelos estatísticos e de aprendizado de máquina. Estudos de análise técnica já demonstraram a importância de utilizar retornos acumulados e volatilidade como variáveis explicativas na modelagem de preços de ações da bolsa de valores (EFFICIENCY, 1993), pode-se esperar uma importância similar, no que se tange a criptomoedas.

Além disso, a escolha de trabalhar com janelas móveis de retorno (3, 5 e 7 dias) permite incorporar a *memória de mercado*, capturando efeitos como *momentum* e *mean reversion*, amplamente discutidos na literatura de finanças comportamentais e análise técnica (EFFICIENCY, 1993).

3.2 Base de Transações de Baleias

A construção da base de transações, foi feita através da interação direta com a API da Blockstream <<https://blockstream.info/api>>, um recurso que fornece dados em tempo real e históricos da blockchain do Bitcoin, possibilitando a extração de identificadores de transações (txids), valores transacionados e endereços dos destinatários. A data da transação e os endereços que enviaram os montantes dessas mesmas são dados fundamentais para a análise desse trabalho, por isso, utilizou-se a API da Blockcypher <<https://api.blockcypher.com/v1/btc/main>> para complementar o conjunto de dados com esses detalhes.

As transações obtidas estão dentro do período de 1 de Agosto de 2021 até 6 de Agosto de 2022. As APIs possuem limitações para utilização em modo gratuito, por isso o recorte de aproximadamente 1 ano.

A integração desses conjuntos de dados distintos — dados de preço e volume do Bitcoin juntamente com registros de transações de Baleias — constitui a base para uma exploração da interação entre as cotações da moeda e o volume desse tipo de transação. Esse processo estabelece o cenário para a aplicação de técnicas analíticas, com o intuito de revelar padrões ocultos e dependências no ecossistema do Bitcoin. Alguns dados sobre a base na Tabela 1.

Tabela 1 – Estatísticas gerais das transações observadas

Métrica	Valor
Dias observados	371
Número de blocos verificados	54.501
Total de transações de baleias	5.219
Média de transações por dia	≈ 14,06
Média de transações por bloco	≈ 0,09

3.3 Observações iniciais

Com as bases criadas, se tornou possível traçar algumas análises sobre esses dados, que possam trazer algumas respostas e exposições de comportamentos que sejam produto das relações causadas pelas cotações diárias e transações da moeda. Para isso, foi utilizado a linguagem *Python* e suas bibliotecas, *json* para formatação dos dados, *pandas* para criação dos *DataFrames* utilizando esses dados, *matplotlib* e *seaborn* para visualização gráfica das análises.

3.3.1 Soma do montante de BTCs transacionados por dia

A primeira análise importante para o estudo, seria o montante de BTC transacionado por dia, para identificar dias que sejam mais relevantes que outros, como por exemplo, o dia com maior montante, que rompeu a barreira dos 400 mil BTC em 18/06/2022 e menor montante, em 18/04/2022, onde houve apenas uma transação com mais de 1000 BTC, precisamente 1003.98.

A Figura 2 demonstra que a maioria dos dias não rompe a barreira de 200000 BTC em negociações. Após essa faixa, vai se diminuindo gradualmente o volume, onde temos apenas 12 dias entre 200000 e 300000, e apenas 8 dias entre 300000 e 400000.

3.3.2 Preço de Fechamento vs Montante de Transações.

Na Figura 3, temos o eixo **Data**, representando os dias do estudo e temos dois eixos perpendiculares. **Preço de Fechamento (USD)**, onde temos a linha azul para mostrar a evolução desse dado e **Montante de Transações (BTC)** para representar o montante de BTC utilizado nas transações desse período.

É possível observar que, no dia 18/06/2022, onde ocorreu o maior montante de transações em BTC, superior a 400 mil BTCs negociados, houve também o menor preço de fechamento de todo o período, U\$ 18.956,80. No dia seguinte, houve uma redução de aproximadamente 200 mil no montante de BTCs transacionados naquele dia e também houve alta no fechamento comparado ao dia anterior, fechando o dia em U\$ 20.552,82.

Temos uma situação que demonstra como a valorização do Bitcoin sofre influência

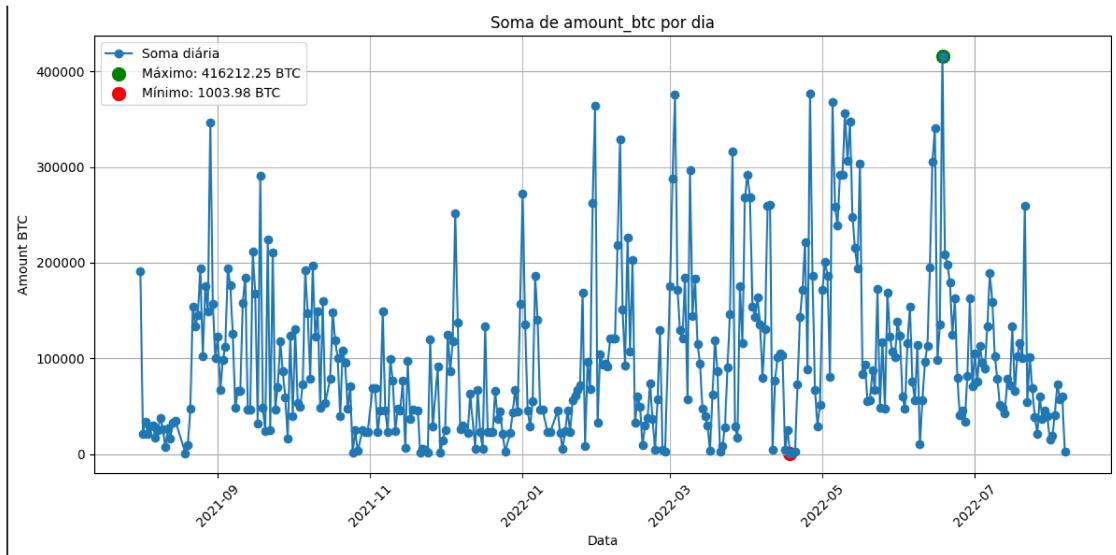


Figura 2 – Volume transacionado por baleias por dia. Fonte: O Autor

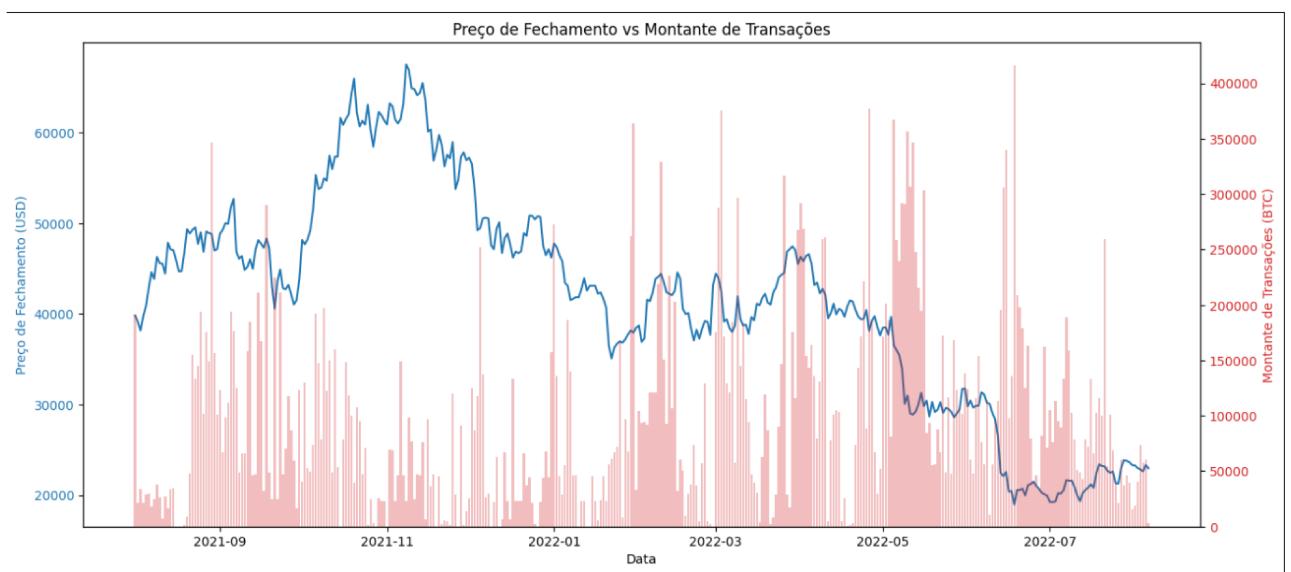


Figura 3 – Preço de Fechamento (USD) vs Montante de Transações (BTC). Fonte: O Autor

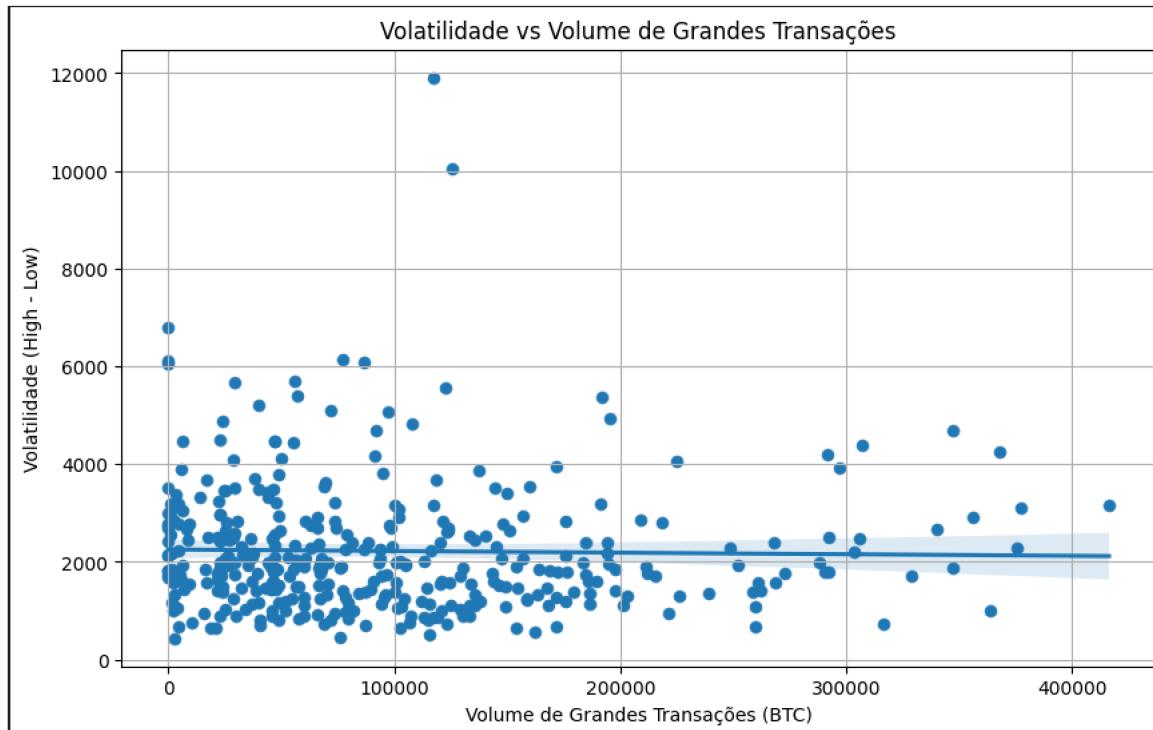


Figura 4 – Volatilidade comparada com o Volume de transações de baleias. Fonte: O Autor

pelo montante negociado da moeda, por ter sua proposta como uma moeda deflacionária, onde a mesma foi projetada para ser um recurso escasso, logo seu valor está atrelado a raridade (TASCA; HAYES; LIU, 2018), um dia com muitas trocas da moeda, pode ferir a esse princípio e causar uma queda no valor, como a queda registrada no dia 18 de junho. No dia seguinte, tivemos a valorização, como também um montante menor transacionado, na Blockchain.

3.3.3 Volatilidade x Volume de transações

A Figura 4 apresenta a relação da Volatilidade de cada dia comparado com o volume de transações. A Volatilidade é representada pela equação:

$$\text{Volatilidade}(dia) = \text{High}(dia) - \text{Low}(dia).$$

A análise gráfica da relação entre a volatilidade e o volume de transações de Baleias (movimentações superiores a 1000 BTC) revela a ausência de uma correlação linear significativa entre essas variáveis. Conforme evidenciado na Figura 4, a linha de regressão apresenta inclinação praticamente nula, sugerindo que variações no volume dessas transações não implicam, de forma direta, em alterações proporcionais na volatilidade do ativo. A alta dispersão dos dados, principalmente em níveis baixos e intermediários de volume, reforça essa conclusão, indicando que dias com grande movimentação por parte de baleias não necessariamente estão associados a maior oscilação de preços.

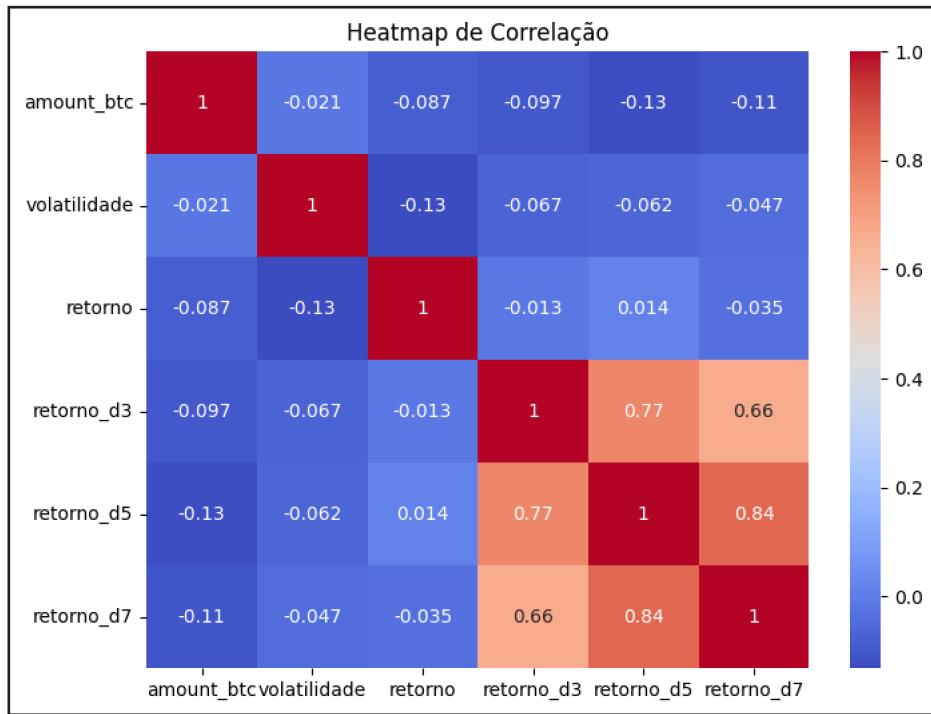


Figura 5 – HeatMap das Variáveis do Estudo. Fonte: O Autor

3.3.4 Correlação entre as variáveis do Estudo

Os dados levantados para construção das bases empíricas desse Estudo, são apresentados em seis variáveis, são elas:

1. **Amount_BTC**: Montante transacionado por Baleias naquele dia;
2. **Volatilidade**: Maior cotação subtraído da menor cotação do dia;
3. **Retorno**: Fechamento subtraído da abertura do dia;
4. **Retorno_D3**: Retorno dos últimos 3 dias acumulados;
5. **Retorno_D5**: Retorno dos últimos 5 dias acumulados;
6. **Retorno_D7**: Retorno dos últimos 7 dias acumulados.

Na Figura 5, observa-se que as correlações entre Amount_BTC, Volatilidade e Retorno com as demais características são muito fracas. Retorno_D3, Retorno_D5 e Retorno_D7, possuem correlação maior entre si, porém não possui muita relevância, dado que Retorno_D3 é parte de Retorno_D5 e Retorno_D7, e Retorno_D5 é parte de Retorno_D7.

3.4 Aplicação da Regressão Logística no Estudo

Diante da ausência de correlação linear significativa entre as variáveis analisadas, optou-se por adotar um modelo alternativo de análise estatística: a regressão logística. Enquanto a regressão linear busca estimar relações contínuas entre variáveis, a regressão logística permite modelar a probabilidade de ocorrência de um evento binário, nesse caso, se as movimentações das baleias, junto com os dados das cotações do preço do ativo, para cada um dos dias observados, formando assim um conjunto de características, influenciam na variação do preço do BTC.

Para se criar um modelo de regressão logística, é necessário definir uma variável alvo binária (*target*), que retorna o valor 1, se a variável da base de dados, **retorno** é maior que 0, e 0 se não.

Listagem 3.1 – Criando a variável Target

```
# Variavel alvo binaria: 1 se retorno >= 0, 0 caso contrario
df[ 'target' ] = ( df[ 'retorno' ] >= 0 ). astype( int )
```

Após criada a variável *target*, é necessário criar as variáveis preditoras, que são responsáveis por fornecer os fatores que influenciam essa variável dependente. São elas:

1. **Amount_BTC**: Montante transacionado por Baleias naquele dia;
2. **Volatilidade**: Maior cotação subtraído da menor cotação do dia;
3. **Retorno_ontem**: Fechamento subtraído da abertura do dia anterior;
4. **Retorno_D3**: Retorno dos últimos 3 dias acumulados;
5. **Retorno_D5**: Retorno dos últimos 5 dias acumulados;
6. **Retorno_D7**: Retorno dos últimos 7 dias acumulados;
7. **Low**: Cotação Mínima do dia;
8. **High**: Cotação Máxima do dia;
9. **Open**: Cotação de abertura do dia;
10. **Close**: Cotação de fechamento do dia.

Assim, é possível definir no modelo, as variáveis de carácter preditoras e a variável dependente (*target*)

Listagem 3.2 – Criando as variáveis preditoras e definindo o papel de todas as variáveis envolvidas no estudo

```
features = [ 'low', 'high', 'open', 'close', 'volatilidade',
'retorno_ontem', 'retorno_d3', 'retorno_d5', 'retorno_d7',
'amount_btc']
```

```
X = df[features]
y = df['target']
```

3.4.1 Normalização das variáveis

Após a definição das variáveis preditoras utilizadas no modelo, foi realizada a etapa de normalização dos dados por meio do método *StandardScaler* (LEARN, 2025e), que transforma os valores de cada *feature* para uma distribuição com média zero e desvio padrão igual a um. Essa etapa é importante para garantir que todas as variáveis contribuam de forma equilibrada para o modelo. Sem essa padronização, variáveis com magnitudes maiores poderiam dominar o processo de aprendizado, comprometendo a qualidade das previsões.

Listagem 3.3 – Normalizando as variáveis preditoras

```
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

3.4.2 Separação em Treino e Teste

Com os dados preparados e as variáveis normalizadas, foi realizada a separação entre os dados de treino e de teste, utilizando a função *train_test_split* (LEARN, 2025f), com 70% das amostras destinadas ao treinamento do modelo e os 30% restantes reservados para a etapa de teste. Essa divisão é essencial para avaliar a capacidade de generalização do modelo, ou seja, sua performance diante de dados que não foram utilizados durante o aprendizado. A definição de uma semente aleatória (*random_state=100*) garante a reprodutibilidade do experimento, permitindo que os resultados possam ser replicados nas mesmas condições.

Listagem 3.4 – Separando os dados em treino e teste

```
# Separar em 70% treino, 30% teste (test_size=0.3)
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y,
```

```
test_size=0.3, random_state=100)
```

3.4.3 Treino e Avaliação do modelo

Feito tudo isso mencionado nas subseções anteriores, é chegada a hora de treinar o modelo e obter os resultados do estudo aplicando Regressão Logística, utilizando a biblioteca *LogisticRegression* do *scikit-learn* (LEARN, 2025b). Para garantir a convergência do algoritmo durante o ajuste dos coeficientes, o parâmetro *max_iter* foi ampliado para 100.000 iterações, prevenindo possíveis interrupções por número insuficiente de ciclos de otimização. O modelo foi então treinado com os dados *X_train* (variáveis independentes) e *y_train* (variável dependente binária), permitindo que ele aprendesse os padrões que relacionam as características observadas à ocorrência ou não do evento de interesse.

Listagem 3.5 – Treinando o modelo

```
model = LogisticRegression(max_iter=100000)
model.fit(X_train, y_train)
```

Após o treinamento do modelo, foi realizada a etapa de previsão utilizando o conjunto de teste (*X_test*). Essa etapa consiste em aplicar o modelo treinado para prever os resultados da variável alvo (*y_pred*) com base nas variáveis independentes que o modelo ainda não havia visto. O objetivo é simular como o modelo se comportaria em um cenário real, avaliando sua capacidade de generalização diante de novos dados.

Listagem 3.6 – Prevendo no conjunto de teste

```
y_pred = model.predict(X_test)
```

Em seguida, foi calculada a acurácia do modelo por meio da função *accuracy_score* (LEARN, 2025a), que compara os valores previstos (*y_pred*) com os valores reais observados (*y_test*). A acurácia representa a proporção de previsões corretas realizadas pelo modelo em relação ao total de observações do conjunto de teste. Quanto maior o valor da acurácia, melhor o desempenho do modelo na tarefa de classificação.

Listagem 3.7 – Calculando acurácia do modelo

```
acuracia = accuracy_score(y_test, y_pred)
```

Foi realizada a previsão das probabilidades associadas à classe positiva (classe 1) utilizando o método *predict_proba* (LEARN, 2025b), o que permite uma análise mais aprofundada do desempenho do modelo. Com essas probabilidades, foi possível construir a curva ROC (*Receiver Operating Characteristic*), com a função *roc_curve* (LEARN,

2025d), que relaciona a taxa de verdadeiros positivos (TPR) com a taxa de falsos positivos (FPR) em diferentes limiares de classificação. A partir da curva ROC, foi calculada a métrica AUC (Área Sob a Curva), por meio da função `roc_auc_score` (LEARN, 2025c). O valor de AUC fornece uma medida robusta da capacidade discriminativa do modelo, indicando o quanto bem ele consegue separar as classes. Quanto mais próximo de 1 for o valor de AUC, melhor é o desempenho do modelo na distinção entre os casos positivos e negativos.

Listagem 3.8 – Previsão das probabilidades associadas à classe positiva e Curva ROC

```
# Prever probabilidades
y_probs = model.predict_proba(X_train)[:, 1]  # Probabilidades da
classe 1 (retorno >= 0)

# Calcular curva ROC e AUC
fpr, tpr, thresholds = roc_curve(y_train, y_probs)
auc_score = roc_auc_score(y_train, y_probs)
```

Concluídas todas as etapas de preparação, modelagem e avaliação do desempenho, os resultados obtidos a partir da aplicação da regressão logística serão apresentados e analisados no próximo capítulo, intitulado **Resultados**.

4 Resultados

Neste capítulo, são apresentados os resultados obtidos a partir do experimento conduzido com o modelo de regressão logística. Após a realização das etapas de preparação dos dados, seleção das variáveis preditoras, normalização, separação em conjuntos de treino e teste, e treinamento do modelo, esta seção visa expor e analisar os principais indicadores e métricas extraídas.

4.1 Acurácia do Modelo

A acurácia alcançada pelo modelo, calculada pela função *accuracy_score* citada no capítulo anterior, foi de **80,73%**, isso indica que, em média, oito a cada dez previsões realizadas foram corretas, o que representa um desempenho satisfatório dentro do contexto do estudo (JR; LEMESHOW; STURDIVANT, 2013). Esse resultado demonstra que a regressão logística foi capaz de capturar padrões relevantes nos dados e distinguir de forma eficaz entre as classes previstas. Ela fornece uma visão inicial da efetividade geral do modelo, reforçando sua viabilidade para a tarefa proposta.

4.2 Coeficientes do Modelo

Para demonstrar os resultados deste estudo, é fundamental iniciar pela interpretação dos coeficientes do modelo de regressão logística, uma vez que eles representam a relação entre cada variável preditora e a probabilidade de ocorrência do evento analisado. Essa análise permite compreender não apenas o desempenho geral do modelo, mas também quais fatores exercem maior influência positiva ou negativa sobre o resultado. No modelo treinado, observou-se que a variável *close* apresentou o maior coeficiente positivo (**3.8057**), indicando forte associação com a classe positiva, ou seja, quanto maior o valor de fechamento, maior a probabilidade de ocorrência do evento, no caso desse trabalho, o *Retorno* positivo do dia. Em contrapartida, a variável *open* teve o coeficiente mais negativo (**-3.7421**), sugerindo uma influência inversa sobre o resultado. Os coeficientes podem ser melhor observados na Tabela 2

Listagem 4.1 – Coeficientes do Modelo

```
# Ver coeficientes do modelo (importancia das variaveis)
coeficientes = pd.DataFrame({
    'Variavel': features,
    'Coeficiente': model.coef_[0]
```

Tabela 2 – Coeficientes do Modelo

Coeficiente	Valor
close	3.805666
amount_btc	0.107892
low	0.060961
retorno_d3	0.049092
high	0.047174
retorno_ontem	-0.049547
retorno_d5	-0.102757
volatilidade	-0.103495
retorno_d7	-0.333667
open	-3.742053

```
})
print( " \nCoeficientes do Modelo : ")
print( coeficientes . sort_values( 'Coeficiente' , ascending=False ))
```

Coeficiente positivos ($\beta > 0$) aumentam a probabilidade de ($Y = 1$). Por exemplo, o montante transacionado pelas baleias, representado pela variável *amount_btc*, tem um coeficiente de ≈ 0.10 , um aumento no montante transacionado está associado a uma maior chance de retorno positivo. Coeficiente negativos ($\beta < 0$) diminuem a probabilidade de ($Y = 1$). Por exemplo, os retornos acumulados nos últimos 7 dias, representado pela variável *retorno_d7*, tem um coeficiente de ≈ -0.33 , logo sugerem maior chance de retorno negativo.

4.2.1 Curva ROC e AUC

Para complementar a avaliação do desempenho do modelo, foi gerada a Curva ROC, a partir das probabilidades previstas para a classe positiva (*y_probs*) no conjunto de treino.

Listagem 4.2 – Gerando curva ROC e calculando AUC

```
import matplotlib.pyplot as plt
from sklearn.metrics import roc_curve, roc_auc_score

# Prever probabilidades no conjunto de teste (nao apenas classes)
y_probs = model.predict_proba(X_train)[:, 1] # Probabilidades da classe 1 (retorno >= 0)

# Calcular curva ROC e AUC
fpr, tpr, thresholds = roc_curve(y_train, y_probs)
auc_score = roc_auc_score(y_train, y_probs)
```

```
# Plotar curva ROC
plt.figure(figsize=(8, 6))
plt.plot(fpr, tpr, color='blue', lw=2, label=f'Regressao
Logistica (AUC={auc_score:.2f})')
plt.plot([0, 1], [0, 1], color='gray', linestyle='--',
label='Classificador Aleatorio (AUC=0.5)')
plt.xlabel('Taxa de Falsos Positivos (FPR)')
plt.ylabel('Taxa de Verdadeiros Positivos (TPR)')
plt.legend()
plt.show()

print(f"\nAUC do modelo: {auc_score:.2%}")
```

A área sob a curva **AUC** foi calculada e resultou em um valor de 89,5%, indicando uma excelente capacidade de discriminação do modelo entre as classes. O gráfico da curva ROC também foi plotado, conforme a Figura 6, incluindo a comparação com um classificador aleatório (linha diagonal, com AUC = 0.5). Um AUC próximo de 1.0, sugere que o modelo possui alto poder preditivo, sendo eficaz em distinguir corretamente entre os casos positivos e negativos. Essa métrica complementa a análise da acurácia e reforça a robustez do modelo construído.

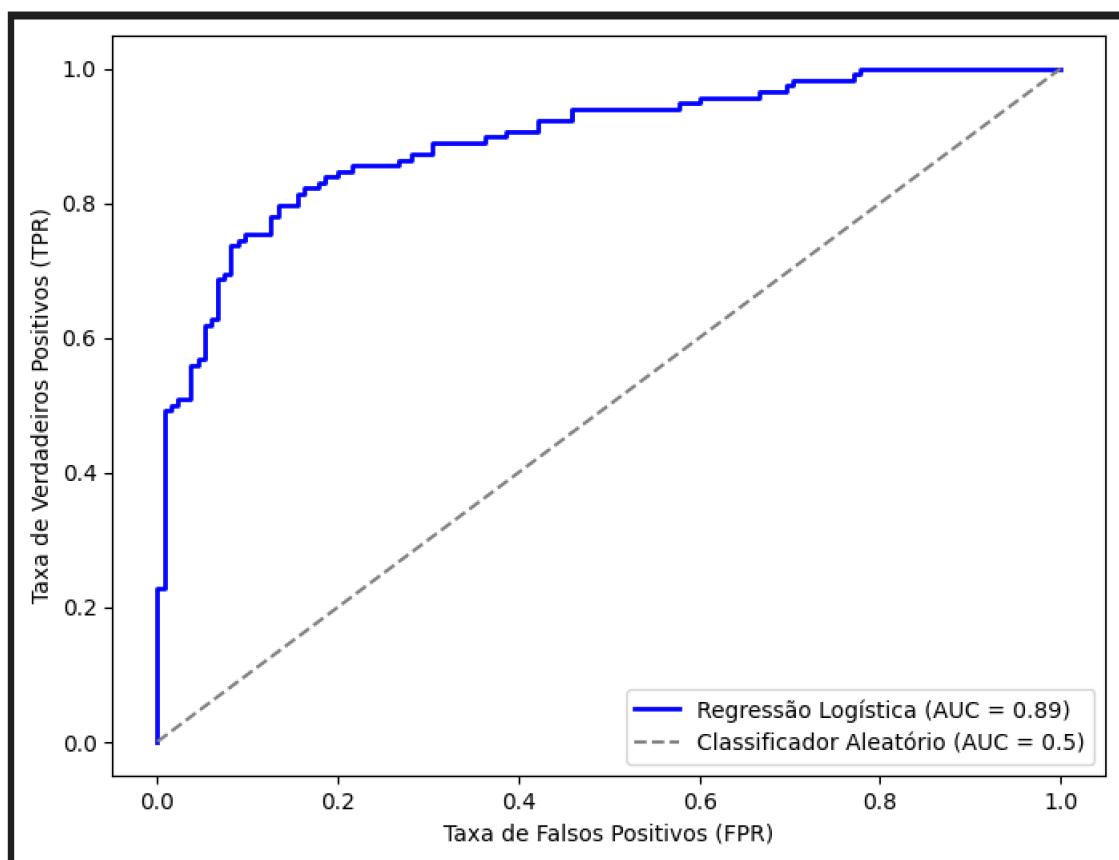


Figura 6 – Curva ROC. Fonte: O Autor

5 Conclusão

Inicialmente, o estudo buscou identificar, por meio de regressão linear, uma possível correlação direta entre o preço do Bitcoin e o montante transacionado diariamente pelas “baleias”. Entretanto, essa abordagem não evidenciou relações estatisticamente significativas. Diante disso, adotou-se a regressão logística, redefinindo o problema com a variável alvo representando o retorno diário (*retorno*) — atribuindo valor 1 quando positivo e 0 quando negativo — e utilizando como variáveis preditoras o montante transacionado por baleias (*amount_btc*), valores de abertura (*open*), fechamento (*close*), máximos (*high*) e mínimos (*low*) do dia, retorno do dia anterior (*retorno_ontem*), volatilidade do dia, ou seja, *high* subtraído de *low* (*volatilidade*), além dos retornos acumulados em 3, 5 e 7 dias (*retorno_d3*, *retorno_d5*, *retorno_d7*). Essa nova abordagem possibilitou investigar se tais variáveis exercem influência significativa sobre o comportamento do retorno diário do Bitcoin, oferecendo uma perspectiva mais adequada para a natureza binária do fenômeno analisado.

Os coeficientes do modelo nos permite ver, quais variáveis tem um impacto positivo ou aumentam a probabilidade de ($Y = 1$) e negativo ou diminuem a probabilidade de ($Y = 1$), isto é, quais aumentam as chances de um retorno positivo ou diminuem as chances. Dentre as que obtiveram um impacto positivo, A variável preditora *amount_btc*, apresentou um coeficiente positivo de 0,10 no modelo. Esse resultado indica que, mantendo-se as demais variáveis constantes, um aumento no volume transacionado por essas carteiras está associado a um aumento na probabilidade de o retorno diário do Bitcoin ser positivo. Tal achado é particularmente relevante, pois essa variável representa o cerne da questão investigada neste estudo: verificar se a atividade das baleias exerce influência significativa sobre o comportamento do retorno diário da criptomoeda. Embora o coeficiente não seja elevado, seu sinal positivo reforça a hipótese de que há uma relação entre o montante movimentado por esses grandes agentes de mercado e a direção do retorno observado.

Contudo, vale ressaltar que isso representa um aumento de probabilidade de ($Y = 1$) na variável *target* e que **Correlação** não é **Causação**, ou seja, coeficientes sugerem associação, não causalidade. E também, como as variáveis foram normalizadas, os coeficientes refletem importância relativa.

Devido as limitações das api's que disponibilizam os dados da Blockchain, construiu-se uma base de 1 ano de transações de Baleias, mas como qualquer estudo de Análise de Dados, quanto mais dados para o modelo melhor. Com isso, para trabalhos futuros, seria interessante tentar aumentar essa base para obter resultados mais acertivos, como também, criar um modelo que trabalhe com interações entre as variáveis, por exemplo,

amount_btc e *volatilidade*, juntos podem ter um efeito diferente. Testar outros algoritmos de Machine Learning, utilizar modelos como *Random Forest*, *Gradient Boosting*, *SVM* ou redes neurais para comparar desempenho com a regressão logística.

Referências

- ALIZADEH, P.; MOSTAFAIE, Z.; RAHIMI, M.; MORTAZAVI, M. et al. Modeling the impact of aggressive financial behavior on financial decision-making: Evidence from forex market investors. 2025. Citado na página 13.
- ALONSO, S. L. N.; FERNÁNDEZ, M. Á. E.; BAS, D. S.; RICO, C. P. El salvador: an analysis of the monetary integration law and the bitcoin law. **Brazilian Journal of Political Economy**, SciELO Brasil, v. 44, n. 1, p. 189–209, 2023. Citado na página 14.
- ANTONOPOULOS, A. M. **Mastering Bitcoin: unlocking digital cryptocurrencies.** [S.l.]: "O'Reilly Media, Inc.", 2014. Citado 2 vezes nas páginas 11 e 12.
- BAHAGHIGHAT, M. **Bitcoin OHLCV: Open, High, Low, and Close prices along with Volume of Bitcoin trades.** [S.l.]: Mendeley Data, V1, doi: 10.17632/kzcrwggp75.1, 2024. Citado 2 vezes nas páginas 10 e 17.
- BLOCKCHAIN.COM. **Site Oficial do Blockchain.com.** 2025. <<https://www.blockchain.com/explorer/api>>. Citado na página 10.
- EFFICIENCY, S. M. Returns to buying winners and selling losers: Implications for. **The Journal of Finance**, v. 48, n. 1, p. 65–91, 1993. Citado na página 18.
- FILHO, W. B. S. Prevendo o preço do bitcoin com redes neurais usando dados do twitter e de mercado. Universidade Federal de Uberlândia, 2019. Citado na página 16.
- FRATRIČ, P.; SILENO, G.; KLOUS, S.; ENGERS, T. van. Manipulation of the bitcoin market: an agent-based study. **Financial Innovation**, Springer, v. 8, n. 1, p. 60, 2022. Citado na página 15.
- GANDAL, N.; HAMRICK, J.; MOORE, T.; OBERMAN, T. Price manipulation in the bitcoin ecosystem. **Journal of Monetary Economics**, Elsevier, v. 95, p. 86–96, 2018. Citado 2 vezes nas páginas 9 e 13.
- HATHIDARA, A.; ATAVALE, G.; CHAUDHARY, S. Mining tweets to predict future bitcoin price. **arXiv preprint arXiv:2412.02148**, 2024. Citado na página 15.
- HERREMANS, D.; LOW, K. W. Forecasting bitcoin volatility spikes from whale transactions and cryptoquant data using synthesizer transformer models. **arXiv preprint arXiv:2211.08281**, 2022. Citado 2 vezes nas páginas 13 e 16.
- INFOMONEY. **Golpes com criptomoedas causam perda de R\$ 40 bi a 4 milhões de brasileiros em 5 anos.** 2023. <Https://www.infomoney.com.br/onde-investir/golpes-com-criptomoedas-causam-perda-de-r-40-bi-a-4-milhoes-de-brasileiros-em-5-anos-quem-ainda-cai>. Citado na página 9.
- INVESTIDOR10. **Cotação do Bitcoin (BTC).** 2025. <<https://investidor10.com.br/criptomoedas/bitcoin/>>. Citado 2 vezes nas páginas 5 e 12.
- JR, D. W. H.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression.** [S.l.]: John Wiley & Sons, 2013. Citado 3 vezes nas páginas 14, 15 e 27.

- KHARIF, O. **The Bitcoin Whales: 1,000 People Who Own 40 Percent of the Market.** 2017. <<https://www.bloomberg.com/news/articles/2017-12-08/the-bitcoin-whales-1-000-people-who-own-40-percent-of-the-market>>. Citado 2 vezes nas páginas 9 e 13.
- LEARN scikit. **Documentação accuracy_score.** 2025. <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html>. Citado na página 25.
- _____. **Documentação LogisticRegression.** 2025. <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html>. Citado na página 25.
- _____. **Documentação roc_auc_score.** 2025. <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html>. Citado na página 26.
- _____. **Documentação roc_curve.** 2025. <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html>. Citado na página 26.
- _____. **Documentação StandardScaler.** 2025. <<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>>. Citado na página 24.
- _____. **Documentação train_test_split.** 2025. <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html>. Citado na página 24.
- NAKAMOTO, S.; BITCOIN, A. A peer-to-peer electronic cash system. **Bitcoin.–URL: https://bitcoin.org/bitcoin.pdf**, 2008. Citado 2 vezes nas páginas 9 e 11.
- NEWS, B. **Como o Bitcoin atingiu valor recorde em meio à pandemia.** 2020. <<https://g1.globo.com/economia/noticia/2020/12/17/como-o-bitcoin-atingiu-valor-recorde-em-meio-a-pandemia.ghtml>>. Acesso em 17 de dezembro de 2020. Citado na página 9.
- PENG, C.-Y. J.; LEE, K. L.; INGERSOLL, G. M. An introduction to logistic regression analysis and reporting. **The journal of educational research**, Taylor & Francis, v. 96, n. 1, p. 3–14, 2002. Citado na página 15.
- PROUDHON, P.-J. The federative principle. **Poverty Is Theft**, p. 689–720, 1863. Citado na página 11.
- PYTHON. **Site Oficial do Python.** 2025. <<https://www.python.org/>>. Citado na página 9.
- SANTOS, M. **Ação da Apple amarga queda em Nova York após Warren Buffett reduzir participação pela metade.** 2024. <Https://www.seudinheiro.com/2024/bolsa-dolar/acao-da-apple-amarga-queda-em-nova-york-apos-warren-buffett-reduzir-participacao-pela-metade-mcss/>. Citado na página 9.
- SCOTT, B. **How can cryptocurrency and blockchain technology play a role in building social and solidarity finance?** [S.l.], 2016. Citado na página 11.
- SERVICES., B. W. **Site Oficial do blockcypher.com.** 2025. <<https://www.blockcypher.com/apis.html>>. Citado na página 10.

TASCA, P.; HAYES, A.; LIU, S. The evolution of the bitcoin economy: Extracting and analyzing the network of payment relationships. **The Journal of Risk Finance**, Emerald Publishing Limited, v. 19, n. 2, p. 94–126, 2018. Citado 2 vezes nas páginas [13](#) e [21](#).