



INSTITUTO BRASILEIRO DE ENSINO, DESENVOLVIMENTO E PESQUISA

CURSO DE ECONOMIA

**PREVISÃO DE RETORNOS DO BITCOIN: UMA ABORDAGEM COMPARATIVA
ENTRE MODELOS ARIMA E LSTM INTEGRADOS COM DADOS DE TENDÊNCIAS DO
GOOGLE**

WILLIAN PINHEIRO TORMIN

Brasília - DF
2024

WILLIAN PINHEIRO TORMIN

Previsão de Retornos do Bitcoin: Uma Abordagem Comparativa Entre Modelos ARIMA e LSTM Integrados com Dados de Tendências do Google

Trabalho de Conclusão de Curso apresentado ao Programa de Graduação em Economia, do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP), como parte dos requisitos para obtenção do título de Bacharel em Economia.

Aprovado em: 04 / 12 / 24

Banca Examinadora:

Prof. DANIEL TAVARES DE CASTRO – Professor Orientador
Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP)

Prof. ALEXANDRE MAGNO
Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP)

Prof. CLAUDIO CARDOSO FLORES
Pontifícia Universidade Católica do Rio de Janeiro (PUCRJ)

Brasília - DF
2024

LISTA DE ABREVIATURAS E SIGLAS

AIC – Akaike Information Criterion

ARIMA – Autoregressive Integrated Moving Average

ARIMAX – Autoregressive Integrated Moving Average with Exogenous Variables

ASIC – Application-Specific Integrated Circuit

BTC – Bitcoin

CNN – Convolutional Neural Network

dApps – Decentralized Applications

DeFi – Decentralized Finance

ETH – Ether (criptomoeda nativa da plataforma Ethereum)

GA – Genetic Algorithm

GRU – Gated Recurrent Unit

LSTM – Long Short-Term Memory

MAE – Mean Absolute Error

MAPE – Mean Absolute Percentage Error

NLP – Natural Language Processing

PoS – Proof of Stake

PoW – Proof of Work

RMSE – Root Mean Square Error

S&P500 – Standard & Poor's 500

SVM – Support Vector Machine

UTC – Coordinated Universal Time

VIX – Volatility Index

Sumário

1. Introdução.....	2
1.1 Contextualização do Tema	3
1.2 Problema de Pesquisa.....	6
1.3 Hipóteses da Pesquisa.....	7
1.3.1 Hipótese Principal:	7
1.3.2 Hipóteses Secundárias:.....	7
1.4 Objetivos Geral e Específicos.....	7
1.4.1 Objetivo Geral	7
1.4.2 Objetivos Específicos	7
1.5 Delimitação do Escopo do Estudo.....	8
1.6 Justificativa do Tema.....	9
1.7 Organização do Estudo.....	10
2. Fundamentação Teórica.....	11
2.1 Modelos Tradicionais de Previsão Financeira.....	11
2.2 Avanços em Previsão com Machine Learning	12
2.3 Comparações de Modelos e Abordagens Híbridas.....	12
2.4 Variáveis Exógenas na Previsão de Preços de Criptomoedas	13
2.5 Implicações para Investidores e Pesquisas Futuras	13
3. Metodologia.....	14
3.1 Método de Abordagem.....	14
3.2 Tipo de Pesquisa.....	14
3.3 Coleta de Dados.....	14
3.4 Preparação e Pré-processamento de Dados	15
3.5 Construção dos Modelos	16
3.5.1 Modelo ARIMA	16
3.5.2 Modelo LSTM.....	16
3.5.3 Modelo LSTM integrado ao PyTrends.....	17
3.6 Comparação de Capacidade dos Modelos.....	18
3.7 Treinamento dos Modelos	18
3.7.1 Treinamento do ARIMA	18
3.7.2 Treinamento dos Modelos LSTM.....	19
3.8 Distinção dos Dados	19
3.8.1 Conjuntos de Teste	19
3.9 Critérios de Avaliação dos Modelos.....	20
4. Resultados e Discussão.....	20

4.1 Análise Exploratória dos Dados	20
4.2 Resultados dos Modelos	21
5. Conclusões e Recomendações	23
5.1 Conclusões Gerais	23
5.2 Implicações Práticas	24
5.3 Limitações do Estudo	25
5.4 Recomendações para Pesquisas Futuras	25
6. Referências Bibliográficas.....	28

Resumo

A previsão de retornos financeiros é um dos maiores desafios enfrentados por economistas, especialmente em mercados altamente voláteis como o de criptomoedas. Este estudo buscou compreender como diferentes abordagens preditivas, representadas pelos modelos ARIMA e LSTM, podem capturar a complexidade dos retornos do Bitcoin. Partindo da premissa de que fatores externos, como o comportamento coletivo expresso nas tendências de busca, podem influenciar os retornos, investigou-se se a inclusão dessas variáveis, melhora a capacidade preditiva dos modelos. Os resultados mostraram que o ARIMA, apesar de simples, apresentou maior estabilidade em métricas como RMSE e MAE. Por outro lado, o LSTM demonstrou maior habilidade para capturar padrões não lineares, ainda que tenha mostrado limitações em sua precisão percentual. A inclusão de variáveis externas, como dados do Google Trends, não gerou uma melhora significativa nos resultados, sugerindo que ajustes mais robustos são necessários para explorar plenamente o potencial dessas informações.

Palavras-chave: Bitcoin, Previsão Financeira, Modelos ARIMA, Redes Neurais LSTM, Tendências de Busca.

Abstract

Forecasting financial returns is one of the greatest challenges faced by economists, particularly in highly volatile markets such as cryptocurrencies. This study aimed to understand how different predictive approaches, represented by ARIMA and LSTM models, can address the complexity of Bitcoin returns. Based on the premise that external factors, such as collective behavior expressed through search trends, influence prices, this research investigated whether the inclusion of these variables improves model predictive capabilities. Results indicated that ARIMA, despite its simplicity, showed greater stability in metrics such as RMSE and MAE. Conversely, LSTM exhibited superior ability to capture nonlinear patterns, though it displayed limitations in its percentage accuracy. The integration of exogenous variables, such as Google Trends data, did not significantly enhance the results, suggesting the need for more robust adjustments to fully explore the potential of this information. This study contributes to the understanding of how behavioral and technical factors can be integrated into predictive modeling, highlighting the opportunities and challenges inherent to disruptive financial markets.

Keywords: Bitcoin, Financial Forecasting, ARIMA Models, LSTM Neural Networks, Search Trends.

1. Introdução

Segundo Karameshinova, E. A. (2023), “Pela primeira vez na história, graças aos recentes avanços na tecnologia blockchain, foi possível transferir valor de qualquer ponto do mundo sem o uso de um intermediário. A este fenômeno deu-se o nome de Criptomoeda.”

A rápida aceitação das criptomoedas desde o início dos anos 2000 é impulsionada por uma combinação de avanços tecnológicos e mudanças na percepção social sobre valor e utilidade na economia, e como resultado, o ambiente das criptomoedas se caracteriza por sua dinâmica e volatilidade, com flutuações de preços abruptas e imprevisíveis. O Bitcoin, sendo a primeira e mais proeminente criptomoeda, exemplifica essas características, tornando-se um ativo de grande interesse especulativo, além de servir como uma reserva de valor para alguns investidores.

A emergência das criptomoedas representa uma inovação sem precedentes na história financeira global, possibilitada pelos avanços recentes na tecnologia do blockchain, o funcionamento das criptomoedas é sustentado por esta tecnologia, que pode ser entendida como um livro-razão digital distribuído. Imagine uma cadeia composta de blocos, onde cada bloco contém um conjunto de transações recentes e uma conexão criptográfica ao bloco anterior. Essa ligação forma uma cadeia cronológica, que não pode ser alterada sem o consenso de todos os participantes da rede, garantindo transparência e segurança. Por exemplo, se alguém tentasse modificar uma transação antiga, isso quebraria a cadeia, e o sistema rapidamente identificaria a tentativa de fraude. A descentralização do blockchain é uma característica essencial: nenhuma entidade única controla a rede, o que reduz custos e aumenta a eficiência, especialmente quando comparado a sistemas bancários convencionais.

Antes do surgimento das criptomoedas, qualquer transação financeira internacional exigia a mediação de instituições financeiras centralizadas, que regulavam o fluxo de capital, adicionando camadas de custo e complexidade, agora, graças à arquitetura descentralizada do blockchain, o dinheiro pode ser movimentado de maneira direta, rápida e segura, de forma que cada transação é validada e armazenada em um registro público e imutável, acessível a todos.

O conceito de criptomoeda, exemplificado pelo Bitcoin, foi idealizado em 2008 por Satoshi Nakamoto, que vislumbrou a criação de um sistema de dinheiro eletrônico peer-to-peer. Isso significa que as transações poderiam ser realizadas diretamente entre duas partes, sem a necessidade de um intermediário central para verificar ou registrar a operação, abrindo caminho para uma nova classe de ativos financeiros que se destacaram por sua complexidade e pelo potencial de transformar o sistema financeiro tradicional.

Enquanto o mercado de criptomoedas oferece inovação e possibilidades disruptivas, o mercado de ações permanece um componente fundamental da economia global, sendo amplamente regulado e mais previsível. Diferente das criptomoedas, o mercado de ações oferece um ambiente relativamente estável para a alocação de capital e a geração de riqueza, embora também seja influenciado por fatores econômicos e políticos. Ambos os mercados, no entanto, compartilham uma característica: a sensibilidade a eventos externos. Seja a volatilidade

do Bitcoin diante de mudanças regulatórias ou a reação do índice S&P 500 a uma crise econômica, a dinâmica de preços é uma função de múltiplos fatores, incluindo os sociais.

Nesse contexto de mercados interconectados, as redes sociais têm emergido como uma força poderosa, moldando a percepção pública e influenciando comportamentos de investimento. Plataformas como Twitter, Reddit e Facebook servem como canais rápidos e eficazes para a disseminação de informações, desde análises de especialistas até rumores não verificados. Por exemplo, uma única postagem viral sobre uma atualização tecnológica no Ethereum pode desencadear uma onda de compra, enquanto um boato de regulamentação mais rigorosa pode precipitar uma venda em massa de Bitcoin. O impacto das redes sociais é tão significativo que se tornou um fator a ser considerado nas análises de mercado, seja para ativos tradicionais ou para criptomoedas, dada sua capacidade de amplificar sentimentos e influenciar movimentos de preço.

1.1 Contextualização do Tema

Desde a criação da internet em 1969, o mundo tem experimentado um progresso tecnológico sem precedentes, especialmente em áreas que revolucionaram a comunicação e o processamento de informações. Essas inovações não apenas transformaram os aspectos cotidianos da vida moderna, mas também redefiniram profundamente a estrutura da economia global.

No ano de 2024, o mercado de criptomoedas continua a alcançar novos patamares, refletindo sua crescente relevância na economia global. A capitalização de mercado total das criptomoedas atingiu aproximadamente US\$ 2,4 trilhões, consolidando sua posição como um importante segmento financeiro que atrai a atenção de investidores institucionais, empresas multinacionais e até mesmo governos de diversos países. Este crescimento é impulsionado por uma combinação de adoção tecnológica, reconhecimento como uma classe de ativos alternativos e mudanças macroeconômicas que favorecem a diversificação de investimentos.

Um exemplo marcante desse avanço foi o desempenho do Bitcoin, que atingiu um valor recorde de US\$ 73.777 em 14 de março de 2024. Esse aumento de preço foi impulsionado por uma série de fatores, incluindo a adoção crescente por parte de grandes instituições financeiras, o interesse renovado de investidores buscando proteção contra a inflação e a implementação de novas políticas regulatórias que deram maior legitimidade ao uso das criptomoedas em mercados tradicionais. Com o Bitcoin mantendo sua posição como a criptomoeda mais valiosa, sua dominância no mercado representou mais de 50% da capitalização total, destacando sua importância no ecossistema cripto. Além do Bitcoin, o setor de criptomoedas abrange mais de 10.000 ativos digitais diferentes, cada um com suas próprias características e inovações tecnológicas.

Figura 1 - Histórico de preços do Bitcoin (2018-2024)



Fonte: Elaborado pelo autor com dados extraídos do Yahoo Finance (2024).

O gráfico apresenta o histórico dos preços de fechamento ajustados do Bitcoin, em dólares americanos (USD), no período de 2018 a 2024. A linha azul retrata as variações diárias no preço de fechamento ajustado, evidenciando a alta volatilidade característica do ativo.

” O mercado de criptomoedas tem sido resiliente às crises económicas, políticas e relacionadas com a pandemia globais. Embora a economia global estivesse preparada para uma recuperação saudável da cadeia de abastecimento e das crises pandémicas, a recente invasão russa da Ucrânia levou a um interesse renovado na utilização de criptomoedas para o comércio, como moedas de refúgio seguro, e para evitar sanções. Além disso, a inflação continua a atingir os ativos tradicionais, à medida que os investidores reagem à erosão dos valores da sua riqueza.” (Liu, Li, Nekhili, Sultan, 2023).

O Ethereum, por exemplo, continua a se destacar com a crescente adoção de suas aplicações descentralizadas (dApps) e o impacto positivo de sua transição para o mecanismo de prova de participação (Proof of Stake). No geral, o aumento da capitalização do mercado cripto demonstra o interesse contínuo e a confiança dos investidores em uma classe de ativos que, apesar de sua volatilidade, oferece oportunidades únicas em um mundo financeiro cada vez mais digitalizado e descentralizado.

A previsão dos retornos de ativos financeiros sempre foi um desafio significativo para economistas e analistas de mercado, mas o surgimento das criptomoedas elevou essa complexidade a um novo patamar. Ao contrário dos ativos financeiros tradicionais, como ações e títulos, as

criptomoedas são influenciadas por uma ampla gama de fatores não lineares, que incluem avanços tecnológicos, regulamentações em constante evolução, sentimentos de mercado amplificados pelas redes sociais e eventos macroeconômicos globais.

Entre os principais fatores que afetam a formação dos retornos das criptomoedas, temos:

- Sentimento do Mercado: Notícias, rumores e discussões em plataformas como Twitter e Reddit podem causar oscilações rápidas e intensas, afetando diretamente o comportamento dos investidores e, conseqüentemente, os retornos.

- Desenvolvimentos Tecnológicos: Atualizações no protocolo das criptomoedas, como o "Merge" da rede Ethereum, têm impactos diretos na percepção de valor e na adoção dessas tecnologias.

- Regulamentações: Mudanças nas políticas de governos ao redor do mundo, sejam elas favoráveis ou restritivas, afetam a confiança no mercado e a demanda por esses ativos.

- Eventos Macroeconômicos: Situações globais, como crises financeiras ou políticas monetárias expansionistas, podem aumentar o apelo das criptomoedas como ativos de refúgio ou investimentos especulativos.

À medida que as criptomoedas ganham mais relevância no cenário financeiro global, a análise dos fatores que determinam os retornos desses ativos torna-se cada vez mais complexa e desafiadora. Modelos tradicionais, como ARIMA, frequentemente falham em capturar a natureza não linear e volátil dos mercados de criptomoedas. Por essa razão, modelos avançados, como redes neurais Long Short-Term Memory (LSTM), são particularmente eficazes em lidar com séries temporais financeiras.

A integração de dados de tendências de busca é uma abordagem emergente que reconhece o papel das buscas online na formação do sentimento de mercado. O índice do Google Trends sintetiza o interesse público sobre determinados temas e fornece insights valiosos para a previsão de movimentos no mercado financeiro. Um aumento no volume de buscas por "Bitcoin", por exemplo, pode refletir tanto o crescente interesse dos investidores quanto suas preocupações, funcionando como um proxy para o comportamento coletivo. Estudos como os de Kristoufek (2013) e Preis, Moat e Stanley (2013) indicam que esses picos de interesse geralmente precedem flutuações significativas nos retornos, sugerindo que a percepção pública afeta diretamente o mercado. Incorporar estas informações em modelos preditivos pode melhorar a precisão das previsões, uma vez que capta a influência do comportamento das massas nos retornos das criptomoedas.

O processo de mineração do Bitcoin é fundamental para manter a segurança e integridade da rede. A mineração envolve a resolução de problemas matemáticos complexos, usando hardware especializado, como ASICs (Application-Specific Integrated Circuits). Esses problemas são tão difíceis que exigem um enorme poder computacional, e o minerador que resolve o problema primeiro, valida um bloco de transações, que é então adicionado ao blockchain. Esse método, conhecido como

prova de trabalho (Proof of Work, PoW), garante a descentralização da rede, mas consome uma quantidade significativa de energia, gerando preocupações ambientais. O consumo de energia associado à mineração de Bitcoin tem sido criticado por sua escala, que muitas vezes é comparada ao consumo energético de países inteiros.

Para abordar essas limitações, outras criptomoedas adotaram métodos alternativos de consenso. Por exemplo, o Ethereum, lançado em 2015 por Vitalik Buterin, introduziu funcionalidades revolucionárias ao conceito de blockchain, como os contratos inteligentes (smart contracts). Contratos inteligentes são scripts que executam automaticamente os termos de um contrato quando as condições pré-definidas são atendidas, sem a necessidade de intermediários. Essa inovação expandiu as possibilidades de aplicação da tecnologia blockchain, permitindo o desenvolvimento de uma vasta gama de aplicações descentralizadas (dApps), que abrangem desde finanças descentralizadas (DeFi) até jogos e sistemas de governança descentralizada.

Em 2022, o Ethereum passou por uma atualização significativa chamada "The Merge", na qual migrou do modelo de prova de trabalho (PoW) para o modelo de prova de participação (Proof of Stake, PoS). Diferentemente da mineração de Bitcoin, onde mineradores competem para resolver problemas matemáticos, o 'PoS' seleciona validadores com base na quantidade de criptomoedas que eles possuem e estão dispostos a "travar" como garantia. O token nativo da rede Ethereum, o Ether (ETH), desempenha múltiplas funções. Ele é usado para pagar taxas de transação na rede, atuar como moeda de troca dentro do ecossistema Ethereum e serve como um ativo especulativo, cujo valor é influenciado por fatores diversos. Esses fatores incluem avanços tecnológicos, como a implementação de novas funcionalidades no blockchain, mudanças nas regulamentações, a crescente adoção de dApps, e o comportamento geral do mercado.

1.2 Problema de Pesquisa

O mercado de criptomoedas apresenta uma dinâmica altamente volátil e sensível a uma série de fatores exógenos, especialmente o Bitcoin, que tem se mostrado desafiador para previsão. Diante dessa complexidade, surge a necessidade de desenvolver modelos preditivos eficazes, que possam capturar a natureza multifacetada desses fatores e fornecer previsões mais assertivas.

Entretanto, parece haver uma lacuna na literatura quanto à comparação de modelos tradicionais de séries temporais, como o ARIMA, e modelos mais modernos baseados em redes neurais, como o LSTM, especialmente quando integrados a dados de tendências do Google.

Assim, o problema de pesquisa que se pretende responder é: **Quais são as diferenças de desempenho entre os modelos ARIMA e LSTM na previsão dos retornos do Bitcoin,**

considerando a integração de dados de tendências do Google? Esta questão se desdobra em um esforço para entender qual dos modelos apresenta maior precisão, menor erro e maior capacidade de captura da dinâmica do mercado quando complementado por variáveis externas, como as de interesse popular pelo ativo.

1.3 Hipóteses da Pesquisa

1.3.1 Hipótese Principal:

A inclusão de dados de tendências de busca (como os obtidos via Google Trends) em modelos de redes neurais LSTM melhora significativamente a precisão das previsões de retornos do Bitcoin, em comparação com modelos que usam apenas dados históricos financeiros.

1.3.2 Hipóteses Secundárias:

- Dados de tendências de busca servem como indicadores de movimentos de retornos no mercado de criptomoedas, proporcionando um ganho de precisão na previsão.
- A combinação de variáveis externas com dados financeiros históricos em modelos LSTM resulta em um modelo mais robusto, capaz de ajustar-se melhor a mudanças súbitas no mercado.

1.4 Objetivos Geral e Específicos

1.4.1 Objetivo Geral

O objetivo geral deste estudo é investigar o impacto da incorporação de dados de tendências de busca em modelos de redes neurais LSTM na previsão dos retornos do Bitcoin, com o intuito de aumentar a precisão e a robustez das previsões em um mercado notoriamente volátil.

1.4.2 Objetivos Específicos

Analisar o efeito das variáveis externas: Investigar como dados de tendências de busca afetam a precisão dos modelos de previsão dos retornos, comparando o desempenho com modelos que utilizam apenas dados históricos financeiros.

Desenvolver e treinar modelos LSTM: Construir modelos de redes neurais LSTM que incorporem múltiplas fontes de dados, como preços históricos e os dados de busca.

Avaliar a precisão dos modelos preditivos: Aplicar métricas como Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), para medir o desempenho dos modelos.

Comparar modelos com e sem variáveis exógenas: Realizar uma análise comparativa para identificar se a inclusão de dados de tendências, resulta em um modelo mais eficaz.

Propor recomendações para aprimoramento: Baseado nos resultados, sugerir melhorias e futuras direções de pesquisa para o desenvolvimento de modelos mais eficientes no contexto de previsão de preços de criptomoedas.

1.5 Delimitação do Escopo do Estudo

O presente estudo delimita-se a analisar os fatores determinantes que influenciam os retornos das criptomoedas, com um foco especial na aplicação de modelos de redes neurais Long Short-Term Memory (LSTM) e na incorporação de dados externos, como as tendências de busca. Para assegurar a precisão e relevância da análise, o escopo do estudo será definido da seguinte forma:

Período de Análise: A pesquisa abrangerá o intervalo de janeiro de 2018 até outubro de 2024. Essa escolha se justifica por incluir uma série de eventos significativos, como grandes flutuações de preço, atualizações tecnológicas relevantes e mudanças regulatórias que impactaram o mercado de criptomoedas. O período selecionado também permite uma análise robusta dos padrões de comportamento do mercado, incluindo fases de alta e baixa volatilidade.

Criptomoedas Focadas: A análise será restrita ao Bitcoin, que foi selecionado por ser a principal criptomoeda do mercado, representando aproximadamente 45% da capitalização total das criptomoedas (COINMARKETCAP, 2024). O Bitcoin não só é a primeira criptomoeda a ser lançada, estabelecendo as bases para o ecossistema atual, mas também continua a ser a mais negociada e reconhecida globalmente, influenciando de forma significativa o movimento de todo o mercado de criptomoedas. Embora outras criptomoedas, como Ethereum, Litecoin e Tron, também sejam relevantes devido às suas inovações e adoções específicas, o foco deste estudo será exclusivamente no Bitcoin, por ser a criptomoeda de maior capitalização de mercado.

Variáveis Externas: Este estudo integrará dados de tendências de busca (provenientes de ferramentas como Google Trends). A análise dessa variável ajudará a entender como o interesse público e o sentimento geral afetam os movimentos de preços.

Modelo de Previsão LSTM: A pesquisa usará redes neurais LSTM para modelar séries temporais complexas, com foco em capturar a não linearidade e a alta volatilidade dos preços das criptomoedas. O modelo será treinado com dados financeiros históricos combinados com variáveis externas, para avaliar o impacto dessas informações adicionais na precisão das previsões.

Revisão de Literatura: A base teórica será construída a partir de artigos científicos e estudos publicados internacionalmente, com foco em pesquisas que exploram o uso de aprendizado profundo

e modelos preditivos em mercados financeiros. A literatura revisada oferecerá suporte à metodologia e ajudará a contextualizar os achados do estudo dentro de um panorama global.

1.6 Justificativa do Tema

Na revisão de literatura, Fang (2022, p. 3) realiza uma pesquisa multidisciplinar abrangente, destacando que "this paper provides a comprehensive survey of 146 cryptocurrency trading papers, across different academic disciplines such as finance and economics, artificial intelligence and computer science". Complementando essa análise, Nazareti Reddy (2023, p. 11) revisa a literatura sobre a previsão de preços de criptomoedas específicas, afirmando que "this section reviews the literature on predicting prices of specific cryptocurrencies like Bitcoin, Ethereum, Litecoin, Monero, Dogecoin, and many more".

A volatilidade característica das criptomoedas, aliada a uma série de fatores tecnológicos, sociais e econômicos, torna sua análise uma tarefa complexa que demanda abordagens inovadoras. A utilização de modelos de aprendizado profundo, como redes neurais Long Short-Term Memory (LSTM), é especialmente pertinente para lidar com a natureza não linear e dinâmica do comportamento de retornos das criptomoedas.

A escolha de investigar o impacto de variáveis externas, como tendências de busca, se justifica pela crescente influência dessas informações no mercado financeiro. Dados de tendências de busca refletem o interesse público em tempo real, o que pode antecipar movimentos de mercado. A integração dessas variáveis com modelos LSTM representa uma abordagem inovadora e ainda pouco explorada na literatura, contribuindo significativamente para o campo de previsão de retornos financeiros.

Este trabalho também é relevante pela sua contribuição prática. Investidores, analistas financeiros e gestores de portfólio poderão utilizar os achados da pesquisa para desenvolver estratégias de investimento mais informadas e robustas, minimizando riscos em um ambiente de alta volatilidade. Além disso, as descobertas podem servir como base para o desenvolvimento de regulamentações mais eficazes, ao demonstrar como fatores externos, como o interesse público, impactam os retornos das criptomoedas.

Do ponto de vista acadêmico, o estudo aborda lacunas importantes na literatura atual, especialmente no que se refere ao uso combinado de aprendizado profundo e variáveis exógenas para previsão de retornos. Enquanto a maioria dos estudos existentes foca em análises financeiras tradicionais ou em modelos estatísticos lineares, esta pesquisa propõe uma metodologia que captura a complexidade inerente ao mercado de criptomoedas. Assim, ela contribui para o avanço teórico e

metodológico na análise de ativos financeiros emergentes, ampliando a compreensão de como diferentes fatores interagem para influenciar o comportamento dos retornos.

Em suma, a relevância deste trabalho está em sua capacidade de integrar tecnologia, análise de dados e finanças para oferecer insights valiosos em um campo de pesquisa que continua a evoluir rapidamente. Ao aplicar técnicas modernas de aprendizado de máquina e explorar variáveis que são cada vez mais influentes no comportamento do mercado, este estudo se posiciona como uma contribuição para a prática no setor financeiro.

1.7 Organização do Estudo

Este trabalho está estruturado em cinco partes principais, organizadas da seguinte forma:

Introdução: A primeira parte apresenta uma explicação dos conceitos centrais relacionados às criptomoedas, incluindo a tecnologia blockchain e a importância do Bitcoin no contexto financeiro atual. Além disso, são discutidos os desafios da previsão dos retornos das criptomoedas e a relevância de modelos preditivos avançados, como redes neurais Long Short-Term Memory (LSTM).

Fundamentação Teórica: A segunda parte oferece uma revisão da literatura relevante ao tema. São abordadas pesquisas nacionais e internacionais que sustentam o estudo, discutindo o uso de modelos de aprendizado profundo na previsão de ativos financeiros e o impacto de variáveis externas, como dados de tendências de busca, na modelagem preditiva. Também são analisadas as características do mercado de criptomoedas e a volatilidade inerente desses ativos, que justificam o uso de abordagens inovadoras na previsão de retornos.

Metodologia: A terceira parte descreve a metodologia adotada no estudo, detalhando as etapas de coleta, tratamento e análise dos dados. Inclui-se a descrição das fontes de dados utilizadas (como dados financeiros históricos e tendências do Google), a justificativa para a escolha dos modelos ARIMA e LSTM, e o detalhamento do processo de construção dos modelos. Também são explicadas as métricas de avaliação aplicadas (RMSE, MAE, MAPE) e o processo de integração das variáveis externas.

Resultados e Discussão: A quarta parte apresenta os resultados obtidos com os modelos preditivos, comparando o desempenho dos modelos ARIMA e LSTM, com e sem a inclusão de variáveis externas. A discussão é feita à luz das hipóteses formuladas, destacando os pontos fortes e limitações de cada abordagem. Além disso, são discutidas as implicações dos resultados para a prática no setor financeiro.

Considerações Finais: A última parte resume as principais contribuições do estudo, ressaltando sua relevância prática. São apresentadas as conclusões sobre o impacto da incorporação de dados de

tendências de busca na previsão dos retornos do Bitcoin e são sugeridas possíveis direções para futuras investigações, como a inclusão de novas variáveis exógenas e o teste de diferentes abordagens de aprendizado de máquina. Além disso, são discutidos os benefícios dos achados para investidores, reguladores e a academia, destacando a importância de métodos avançados de previsão para o mercado das criptomoedas.

2. Fundamentação Teórica

2.1 Modelos Tradicionais de Previsão Financeira

Os modelos estatísticos tradicionais, como o **Modelo Autorregressivo Integrado de Médias Móveis (ARIMA)**, são utilizados para a análise de séries temporais financeiras devido à sua estrutura relativamente simples, que combina três componentes principais: autorregressão (AR), integração (I) e médias móveis (MA). O ARIMA é eficaz em capturar padrões lineares, mas carece de capacidade para lidar com não linearidades, uma característica marcante nos preços das criptomoedas, que são altamente voláteis e influenciados por múltiplos fatores exógenos.

Karameshinova (2023) aplicou o modelo ARIMA para prever os preços do Bitcoin, considerando variáveis como Ethereum, ouro, petróleo, o índice S&P500 e o VIX (índice de volatilidade). A pesquisa utilizou dados de junho de 2019 a maio de 2023 e empregou testes de correlação e causalidade de Granger para avaliar as relações entre essas variáveis. Apesar do uso da função ‘auto_arima()’ para a automatização do processo de modelagem, o estudo mostrou que o ARIMA teve dificuldades em capturar completamente o impacto das variáveis externas, mesmo após ajustes manuais. O modelo final identificado, ARIMA (4,1,2), foi relativamente eficaz, mas apresentou limitações, sugerindo que métodos mais sofisticados poderiam ser necessários para lidar com a complexidade dos dados.

O estudo de **Corrêa (2023)** comparou o ARIMA com o modelo **PROPHET**, uma ferramenta de previsão desenvolvida pelo Facebook que é mais flexível para capturar tendências sazonais e efeitos de longo prazo. Corrêa utilizou dados históricos de fechamento diário do Bitcoin de 2014 a 2018 e implementou transformações, como a logaritmização e a diferenciação, para tornar a série estacionária. A análise revelou que o PROPHET foi mais eficiente em prever oscilações de preço em períodos mais longos, enquanto o ARIMA se mostrou adequado apenas para previsões de curto prazo, com perda de precisão em períodos mais voláteis.

2.2 Avanços em Previsão com Machine Learning

A complexidade e a natureza não lineares dos mercados de criptomoedas levaram à adoção de técnicas mais avançadas de Machine Learning e Deep Learning. **Basher (2022)** realizou um estudo comparativo entre modelos econométricos tradicionais e algoritmos de aprendizado de máquina, usando dados de preços do Bitcoin e variáveis macroeconômicas como taxas de juros, inflação e volatilidade de mercado. O autor implementou modelos como logit e florestas aleatórias, comparando sua eficácia em prever a direção dos preços em horizontes de 5, 10 e 20 dias. O estudo destacou que as florestas aleatórias, um algoritmo baseado em ensembles de árvores de decisão, superaram os modelos logit, alcançando precisões de até 85% para previsões de longo prazo. Basher também observou que a volatilidade do mercado era o preditor mais influente, seguida por taxas de juros e inflação, sugerindo a relevância de fatores econômicos no comportamento do Bitcoin.

2.3 Comparações de Modelos e Abordagens Híbridas

Os avanços em aprendizado profundo têm se mostrado particularmente promissores na previsão de preços de criptomoedas. **Mendes (2022)** comparou o desempenho do **LSTM (Long Short-Term Memory)** com o **ARIMA**, aplicando esses modelos em dados diários de preços do Bitcoin de 2016 a 2019. Utilizando rotinas em Python, o estudo dividiu os dados em conjuntos de treinamento (83,5%) e teste (16,5%) e avaliou a precisão dos modelos com métricas como **RMSE** e **MAE**. Os resultados mostraram que o **LSTM** superou significativamente o **ARIMA**, melhorando a precisão das previsões em até 94%. **Mendes (2022)** também destacou que o **LSTM** é capaz de capturar padrões temporais complexos devido às suas células de memória e mecanismos de portas, que permitem a retenção de informações relevantes ao longo do tempo.

Norde (2023) expandiu essa análise ao explorar o uso de **LSTM** e **SVM (Support Vector Machines)** na previsão de preços de criptomoedas como Bitcoin, Ethereum, Cardano e BinanceCoin. O estudo implementou técnicas de pré-processamento, como normalização e imputação de valores ausentes, e comparou o desempenho dos modelos com métricas como Erro Médio Absoluto (EMA) e Coeficiente de Determinação (R^2). A pesquisa revelou que o **LSTM** era mais adequado para prever criptomoedas como ADA e BNB, enquanto o **SVM** mostrou um desempenho superior para Bitcoin e Ethereum. Norde também enfatizou a importância de escolher a janela temporal correta, observando que janelas de 90 dias forneceram previsões mais estáveis.

Seabe, Moutsinga e Pindza (2023) conduziram uma análise aprofundada de três arquiteturas de redes neurais recorrentes: **LSTM**, **GRU (Gated Recurrent Unit)** e **Bi-Directional LSTM (Bi-LSTM)**. O estudo utilizou dados diários de preços do Bitcoin, Ethereum e Litecoin de 2018 a 2023, aplicando

técnicas de pré-processamento como MinMaxScaling para normalizar os dados. Os resultados mostraram que o Bi-LSTM, que processa informações em ambas as direções temporais, obteve o melhor desempenho, com menores valores de RMSE e MAPE. O estudo também sublinhou a importância do ajuste fino de hiperparâmetros, como o tamanho do lote e o número de épocas, para otimizar os modelos.

2.4 Variáveis Exógenas na Previsão de Preços de Criptomoedas

A incorporação de variáveis exógenas, como dados de **tendências de busca** e **análise de sentimentos**, tem se mostrado uma estratégia eficaz para melhorar a precisão preditiva. **Braga (2023)** investigou a utilidade dos dados do Google Trends na previsão de preços do Bitcoin com redes LSTM. O estudo utilizou palavras-chave como "Bitcoin" e "Crypto" e comparou o desempenho de modelos univariados e multivariados de LSTM. Braga concluiu que a inclusão dessas variáveis aumentou a precisão das previsões, especialmente em períodos de tendência de queda, com melhores resultados para janelas de 10 dias.

Omole e Enke (2024) introduziram modelos híbridos, como o **CNN-LSTM**, que combinam redes convolucionais para extração de características com redes LSTM para modelagem temporal. O estudo utilizou dados on-chain e implementou técnicas de seleção de características, como Boruta e GA (Genetic Algorithm), para otimizar a performance dos modelos. O CNN-LSTM mostrou maior precisão preditiva, com uma taxa de acerto de 82,44% na previsão da direção dos preços do Bitcoin. Além disso, os autores testaram diferentes estratégias de trading e descobriram que a estratégia long-and-short gerou um retorno anual de 6654%, destacando o potencial prático das previsões baseadas em aprendizado profundo.

2.5 Implicações para Investidores e Pesquisas Futuras

Os achados desses estudos têm implicações significativas para investidores e formuladores de políticas. **Seabe, Moutsinga e Pindza (2023)** sugerem que o uso de técnicas avançadas, como o Bi-LSTM, pode oferecer uma vantagem competitiva no mercado de criptomoedas. No entanto, a pesquisa também ressalta que a precisão dos modelos pode ser afetada pela alta volatilidade e pelas mudanças bruscas no mercado, o que requer uma abordagem de modelagem adaptativa.

Além disso, a integração de variáveis exógenas, como o sentimento de mercado e o volume de buscas, destaca a importância de uma abordagem mais abrangente na análise preditiva. **Omole e Enke (2024)** enfatizam que, embora os modelos baseados em aprendizado profundo sejam promissores, é

essencial considerar fatores externos, como eventos globais e mudanças regulatórias, para aprimorar ainda mais a precisão das previsões.

3. Metodologia

3.1 Método de Abordagem

Este estudo adota uma abordagem dedutiva, partindo de teorias previamente estabelecidas para desenvolver e aplicar modelos preditivos no contexto do mercado de criptomoedas. A metodologia combina métodos estatísticos tradicionais e técnicas avançadas de aprendizado de máquina, com o objetivo de melhorar a precisão das previsões de retornos de criptomoedas, especificamente do Bitcoin. A implementação de redes neurais Long Short-Term Memory (LSTM) em conjunto com dados exógenos, como tendências de busca, exemplifica a integração de abordagens clássicas e modernas. Essa combinação permite explorar tanto as propriedades temporais dos dados financeiros quanto os efeitos de interesse popular, refletidos nas tendências de busca.

3.2 Tipo de Pesquisa

A pesquisa é quantitativa e descritiva, focada na análise numérica de grandes volumes de dados financeiros e de tendências de busca para prever o comportamento dos retornos do Bitcoin. A combinação de aprendizado profundo com variáveis externas permite a modelagem de padrões complexos, não capturados por modelos lineares tradicionais, proporcionando uma análise mais robusta e preditiva. Esta abordagem quantitativa é ideal para identificar correlações e testar a eficácia de diferentes metodologias preditivas.

3.3 Coleta de Dados

- **Dados de Preços de Criptomoedas:** Os preços históricos e ajustados do Bitcoin (BTC-USD) foram obtidos do site Yahoo Finance, abrangendo o período de 1º de janeiro de 2018 até o final de outubro de 2024.
- **Dados de Tendências de Busca:** Utilizando a API PyTrends, foram coletadas informações de tendências de busca para o termo "Bitcoin", desde janeiro de 2018 até outubro de 2024. Essa abordagem de granularidade diária permite capturar o interesse público flutuante em torno do Bitcoin, refletindo a percepção e comportamento dos investidores ao longo do tempo.

3.4 Preparação e Pré-processamento de Dados

O pré-processamento dos dados foi uma etapa essencial para garantir a integridade e a qualidade do conjunto de informações utilizado no estudo. Inicialmente, os dados históricos de preços do Bitcoin foram extraídos e ajustados ao fuso horário UTC, de modo a eliminar qualquer discrepância temporal que pudesse comprometer a análise. A seguir, realizamos a conversão e o alinhamento de datas para assegurar que os registros estivessem sincronizados cronologicamente. Além disso, aplicamos uma verificação de integridade para identificar e tratar valores inconsistentes ou ausentes, evitando possíveis distorções que pudessem impactar os resultados.

Com o objetivo de enriquecer o conjunto de dados, integramos os preços históricos do Bitcoin com as tendências de busca extraídas do Google Trends. Essa combinação foi realizada cuidadosamente para preservar a consistência entre as variáveis, utilizando técnicas como o preenchimento de valores ausentes com o método (fillna). Este método foi escolhido para tratar valores ausentes por sua simplicidade, eficiência e capacidade de preservar a continuidade do conjunto de dados sem introduzir ruído ou perda de informações relevantes. Essa etapa foi crucial para evitar a perda de informações importantes, garantindo que o conjunto de dados permanecesse completo e confiável para análise. A inclusão dos dados de busca foi concebida para capturar como as variações no interesse público podem influenciar os movimentos dos retornos do Bitcoin, permitindo explorar potenciais correlações de forma mais robusta.

Antes de introduzir os dados nos modelos preditivos, foi realizada uma etapa de normalização utilizando o método **MinMaxScaler** da biblioteca Scikit-Learn, que escalonou os valores para o intervalo de 0 a 1. Essa técnica foi escolhida porque as redes neurais do tipo LSTM são altamente sensíveis a diferenças de escala entre as variáveis. A normalização não apenas promoveu a estabilidade do treinamento, mas também aumentou a eficiência computacional, permitindo que o modelo lidasse de forma mais precisa com a natureza complexa e não linear dos retornos das criptomoedas.

O conjunto de dados foi então dividido em duas partes: 80% dos dados foram destinados ao treinamento do modelo e 20% ao teste. Essa divisão é amplamente reconhecida na literatura como uma prática padrão em aprendizado de máquina, sendo projetada para garantir que o modelo aprenda com um volume suficiente de dados, enquanto reserva uma parcela separada para avaliar seu desempenho em condições inéditas. Durante o treinamento, o modelo ajustou seus parâmetros com base no conjunto maior, enquanto o conjunto de teste permaneceu isolado, permitindo uma avaliação justa de sua capacidade de generalização.

Todo o processo de pré-processamento foi planejado para assegurar que os dados estivessem devidamente preparados para suportar os modelos de redes neurais LSTM, ao mesmo tempo em que mantivemos a integridade e a consistência das informações.

3.5 Construção dos Modelos

Neste estudo, utilizamos três modelos distintos para prever os retornos do Bitcoin, explorando tanto abordagens tradicionais quanto modernas. O primeiro modelo aplicado foi o ARIMA, uma técnica clássica de séries temporais que busca capturar padrões lineares nos dados históricos. Em seguida, implementamos um modelo LSTM tradicional, que utiliza redes neurais avançadas para identificar padrões não lineares e dinâmicos nos dados financeiros, oferecendo uma abordagem mais sofisticada para lidar com a volatilidade das criptomoedas.

Além destes, desenvolvemos um terceiro modelo LSTM, agora integrado com dados de tendências de busca obtidos através do PyTrends. Esse modelo foi projetado para considerar, além dos dados históricos de preços, o interesse público em torno do Bitcoin, representado pelo volume de buscas, com o objetivo de enriquecer a análise preditiva.

3.5.1 Modelo ARIMA

O Modelo Autorregressivo Integrado de Médias Móveis (ARIMA) é amplamente utilizado para análise de séries temporais financeiras devido à sua estrutura relativamente simples, que combina três componentes principais: autorregressão (AR), integração (I) e médias móveis (MA). Neste estudo, o ARIMA foi utilizado como referência por ser um método clássico de previsão de séries temporais, adequado para capturar padrões lineares. No entanto, o ARIMA possui limitações na captura de não linearidades, que são frequentemente observadas nos retornos do Bitcoin, tornando-se menos eficaz em períodos de alta volatilidade ou quando fatores exógenos têm impacto significativo.

3.5.2 Modelo LSTM

O modelo LSTM (Long Short-Term Memory) foi construído utilizando a biblioteca TensorFlow, com o objetivo de explorar a capacidade do aprendizado profundo em capturar padrões não lineares e dependências temporais prolongadas nos dados financeiros. A arquitetura do modelo incluiu duas camadas LSTM, cada uma com 50 unidades:

- A primeira camada foi configurada para retornar sequências, permitindo que a camada subsequente tenha acesso a toda a informação temporal, o que é essencial para capturar padrões complexos e relações sequenciais ao longo do tempo.

- A segunda camada LSTM fornece uma saída consolidada, representando a informação mais relevante extraída ao longo de todo o período analisado.
- A última camada LSTM não retorna sequências, mas fornece uma saída consolidada, representando a informação mais relevante extraída ao longo de todo o período analisado.

Ao final, foi adicionada uma camada densa (fully connected) com uma unidade, responsável pela previsão dos retornos do Bitcoin. Esse modelo visa capturar padrões essenciais e foi treinado apenas com os dados financeiros históricos do Bitcoin, oferecendo uma linha de base para comparar o impacto da inclusão de variáveis externas.

3.5.3 Modelo LSTM integrado ao PyTrends

O modelo LSTM com PyTrends foi construído de forma distinta do LSTM tradicional, para incorporar de maneira eficaz as variáveis externas provenientes das tendências de busca do Google. Diferentemente do modelo LSTM tradicional, o LSTM integrado com PyTrends foi estruturado para trabalhar com um conjunto de dados multivariado, que incluía tanto as séries temporais dos retornos do Bitcoin quanto os dados das tendências de busca. A preparação dos dados envolveu escalonamento separado para os dados financeiros e a variável externa, seguido pela sua concatenação, garantindo que cada tipo de dado contribuísse de forma adequada para a modelagem. A arquitetura do modelo incluiu duas camadas LSTM configuradas para lidar com dados sequenciais:

- A primeira camada LSTM contém 64 unidades e está configurada para retornar sequências, seguida por uma camada de dropout com taxa de 20% para reduzir o risco de overfitting.
- A segunda camada LSTM contém 32 unidades e também é seguida por uma camada de dropout com taxa de 20%, para aumentar a capacidade de generalização do modelo.

Por fim, uma camada densa com uma unidade foi adicionada para realizar a previsão dos retornos. O modelo foi compilado utilizando o otimizador Adam, com uma taxa de aprendizado ajustada para 0,0005, e treinado por 100 épocas com um tamanho de lote de 16. Tais configurações foram escolhidas e testadas, visando maximizar o impacto das informações exógenas, permitindo ao modelo identificar padrões ocultos que pudessem não ser evidentes apenas pelos dados financeiros.

Durante o desenvolvimento, foram realizados vários testes para identificar os melhores parâmetros para o modelo. Por exemplo, foram testados diferentes números de unidades nas camadas LSTM (variando de 32 a 128), diferentes taxas de dropout (de 10% a 50%), e tamanhos de lote (16, 32, 64). Esses experimentos ajudaram a ajustar a arquitetura e os hiperparâmetros do modelo para alcançar o melhor equilíbrio entre precisão e capacidade de generalização. Isso visou maximizar o

impacto das informações exógenas, permitindo ao modelo identificar padrões que pudessem não ser evidentes apenas pelos dados financeiros.

3.6 Comparação de Capacidade dos Modelos

A principal distinção entre os três modelos está na capacidade de captar padrões complexos e integrar variáveis externas, proporcionando diferentes perspectivas analíticas sobre os retornos do Bitcoin. O modelo ARIMA, amplamente utilizado para séries temporais financeiras, é eficaz na identificação de relações lineares, mas apresenta limitações significativas em contextos de alta volatilidade e não linearidades, como as encontradas nos retornos de criptomoedas. Por outro lado, o modelo LSTM tradicional, baseado em redes neurais recorrentes, é projetado para lidar com padrões não lineares e dinâmicas complexas, sendo mais adequado para mercados onde fatores como volatilidade e interdependências temporais são predominantes. Adicionalmente, o modelo LSTM integrado com (PyTrends) oferece uma abordagem mais abrangente ao incorporar variáveis externas, como dados de tendências de busca online, para enriquecer a análise preditiva. O uso de variáveis como o volume de buscas pelo termo "Bitcoin" permite explorar a relação entre o interesse público e os retornos do mercado, frequentemente antecipando movimentos que não seriam detectados apenas por dados financeiros históricos. Pesquisas anteriores, como as de **Preis, Moat e Stanley (2013)**, sugerem que picos no interesse de busca podem ser indicadores antecipados de volatilidade no mercado. Assim, este modelo híbrido combina a robustez das redes LSTM com insights comportamentais derivados das tendências de busca, tornando-se uma ferramenta poderosa para previsão.

3.7 Treinamento dos Modelos

3.7.1 Treinamento do ARIMA

O modelo ARIMA foi treinado utilizando os dados históricos dos retornos do Bitcoin, com a ordem do modelo definida por meio da análise de padrões de autocorrelação e autocorrelação parcial. Para automatizar o processo e identificar os parâmetros p , d e q que minimizassem o erro preditivo, utilizamos a função (auto_arima), que realiza uma busca iterativa baseada no critério de informação de Akaike (AIC). Durante o ajuste, várias configurações foram testadas, e os resultados indicaram que o modelo ARIMA (2,0,0) com intercepto, apresentou o menor (AIC), destacando-se como a melhor configuração.

3.7.2 Treinamento dos Modelos LSTM

O modelo foi treinado de maneira a garantir uma aprendizagem eficaz dos padrões presentes nos dados históricos e na variável externa:

- Hiperparâmetros do Treinamento:

Número de Épocas: Tanto o LSTM tradicional quanto o LSTM integrado com PyTrends foram treinados por 100 épocas. Esse número de iterações foi escolhido para garantir um equilíbrio entre o aprendizado dos padrões dos dados e a minimização de problemas de overfitting.

Tamanho do Lote: O tamanho do lote foi definido como 32, permitindo um bom compromisso entre velocidade de treinamento e estabilidade do gradiente durante a otimização.

- Função de Perda e Otimizador:

Função de Perda: Foi utilizada a função de erro quadrático médio (MSE) como métrica de perda, uma escolha adequada para problemas de regressão como a previsão de retornos.

Otimizador: O otimizador Adam foi escolhido devido à sua capacidade adaptativa de ajuste da taxa de aprendizado, facilitando uma convergência mais rápida e eficaz.

- Estratégia de Treinamento:

Janela Deslizante: Para capturar a dependência temporal nos dados, utilizou-se uma janela deslizante de 90 dias. Cada janela fornece ao modelo um histórico de três meses para prever o retorno do próximo dia. Essa técnica é eficaz para modelar a natureza sequencial dos dados financeiros. No caso do modelo LSTM integrado com PyTrends, a janela deslizante foi adaptada para incluir não apenas os dados financeiros, mas também as variações nas tendências de busca, permitindo capturar a interação dinâmica entre interesse público e retornos.

3.8 Distinção dos Dados

3.8.1 Conjuntos de Teste

Para validar a eficácia dos modelos, foi utilizado um conjunto de teste composto pelos retornos diários do Bitcoin a partir de outubro de 2024. Os dados do conjunto de teste foram previamente filtrados para remover valores ausentes e passaram pelo mesmo pré-processamento aplicado ao conjunto de treinamento, incluindo escalonamento e normalização por meio do método MinMaxScaler, garantindo que as variáveis permanecessem na mesma escala. Essa abordagem assegura consistência nos dados utilizados para treinamento e validação, permitindo uma comparação direta e confiável entre as previsões e os valores reais observados.

3.9 Critérios de Avaliação dos Modelos

A avaliação dos modelos LSTM e ARIMA foi realizada com base em três métricas principais, selecionadas por sua capacidade de medir diferentes aspectos do desempenho dos modelos preditivos:

Erro Quadrático Médio (RMSE): Utilizado para quantificar a médias das diferenças quadráticas média entre os valores previstos e os observados, proporcionando uma indicação da magnitude média dos erros. Esta métrica é particularmente útil por penalizar erros maiores de forma mais acentuada, sendo ideal para cenários em que desvios grandes devem ser rigorosamente considerados.

Erro Absoluto Médio (MAE): Empregado para calcular a média dos erros absolutos entre os valores previstos e os reais, oferecendo uma compreensão direta e intuitiva da precisão das previsões em termos absolutos. Esta métrica é menos afetada por outliers, sendo adequada para avaliar a precisão média de forma objetiva.

Erro Percentual Absoluto Médio (MAPE): Utilizado para medir o erro médio em termos percentuais, proporcionando uma perspectiva relativa dos desvios em relação aos valores observados. Esta métrica é valiosa para compreender o desempenho do modelo de maneira proporcional, facilitando a análise comparativa dos erros em diferentes contextos e escalas.

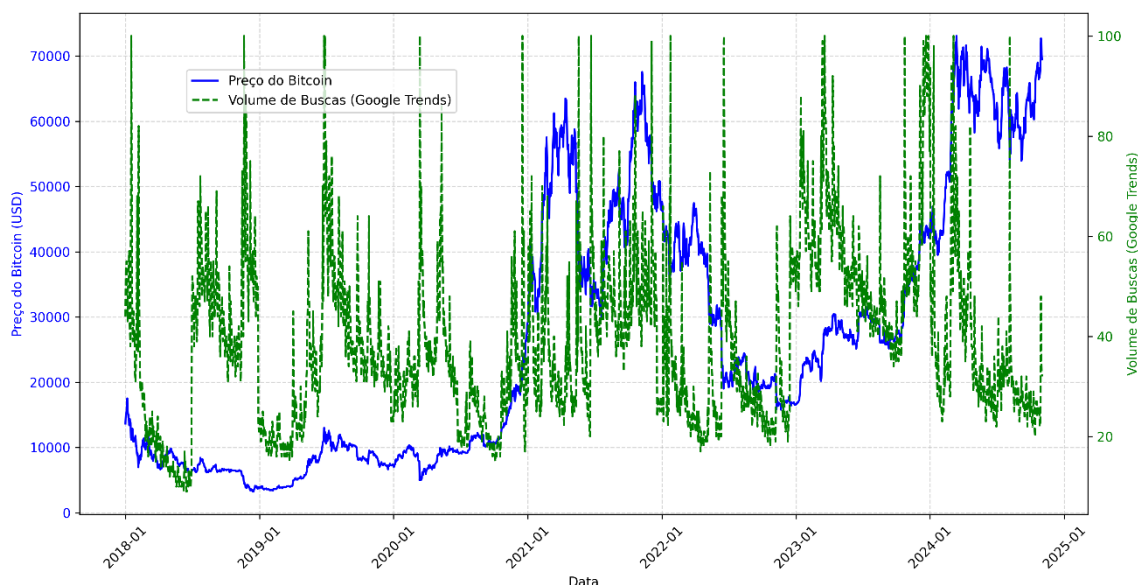
4. Resultados e Discussão

4.1 Análise Exploratória dos Dados

A análise inicial dos dados consistiu na inspeção visual e estatística dos preços históricos do Bitcoin e do interesse público no termo "Bitcoin", coletado por meio do Google Trends. Durante esse processo, foi possível observar possíveis padrões sazonais, picos de alta volatilidade e correlações potenciais entre o aumento das buscas por Bitcoin e as variações de preço. Essa análise permitiu identificar momentos de grande interesse público, que geralmente coincidem com picos de preço, sugerindo uma possível relação entre o comportamento dos investidores e a popularidade da criptomoeda.

Figura 2 - Volume de Tendências de Busca x Preço do Bitcoin (2018-2024)

Volume de Tendências de Busca x Preço do Bitcoin



Fonte: Elaborado pelo autor com dados extraídos do Yahoo Finance e Google Trends (2024).

A figura 2, ilustra a relação entre o preço do Bitcoin, representado pela linha azul (eixo esquerdo, em USD), e o volume de buscas pelo termo "Bitcoin" no Google Trends, representado pela linha verde pontilhada (eixo direito, escala de 0 a 100). O preço do Bitcoin mostra alta volatilidade ao longo do tempo, com picos significativos, especialmente em 2021, enquanto o volume de buscas reflete o interesse público, atingindo seu ápice em momentos de maior valorização da criptomoeda. Ambos os indicadores exibem variações relacionadas, sugerindo uma interação entre o comportamento do mercado e o interesse coletivo.

4.2 Resultados dos Modelos

Este estudo utilizou três modelos distintos para prever os retornos do Bitcoin: ARIMA, LSTM Tradicional e LSTM com a adição de uma variável externa (dados de busca). Cada modelo apresentou características e resultados específicos, conforme a Tabela 1, que compara o desempenho com base nas métricas utilizadas. Refletindo as diferentes abordagens de modelagem preditiva para o mercado das criptomoedas.

Tabela 1 – Comparação de desempenho dos modelos.

Modelos	RMSE	MAE	MAPE (%)
ARIMA	0.0199	0.0151	152.12

LSTM Tradicional	0.0194	0.0151	378.05
LSTM Integrado	0.0249	0.0173	280.47

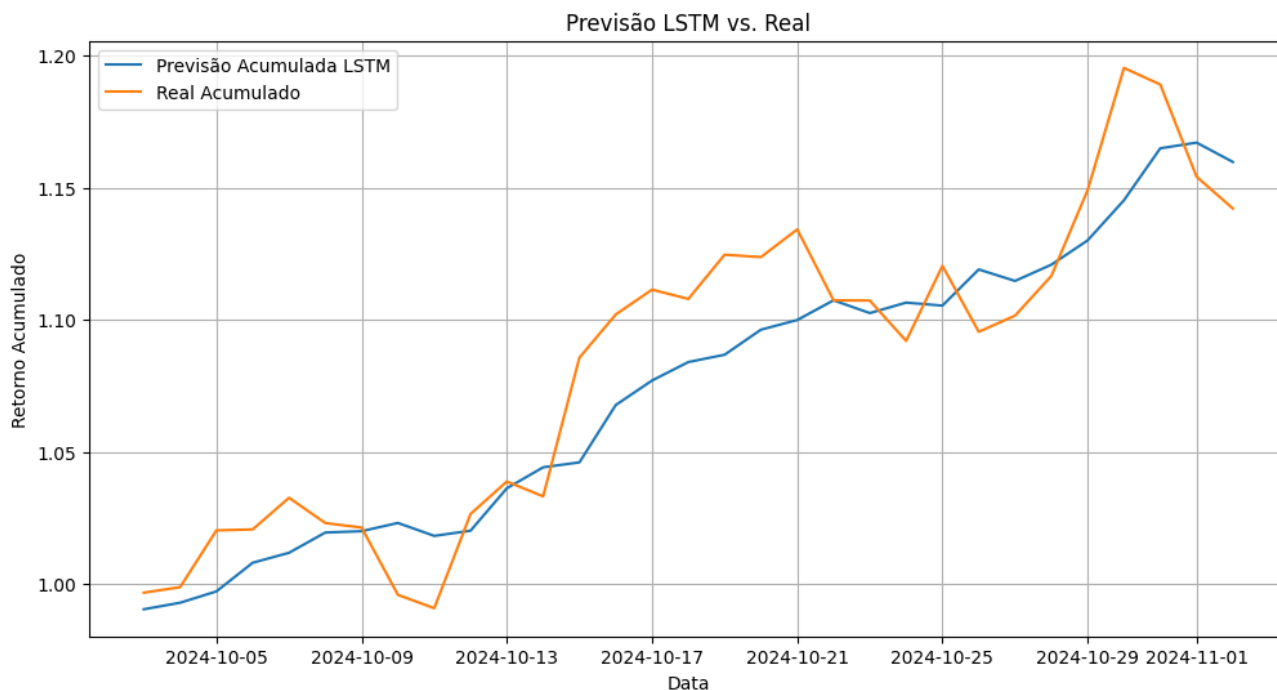
A integração de dados de busca, exige um ajuste rigoroso dos hiperparâmetros para evitar que o modelo capte ruídos em vez de padrões significativos. A falta de assertividade dos parâmetros usados, pode ter comprometido a capacidade do LSTM Integrado de aprender relações complexas entre a variável externa e o retorno do Bitcoin. Além disso, os dados de busca podem não apresentar uma relação suficientemente não linear com os retornos do Bitcoin para que o LSTM capture padrões úteis, especialmente neste mercado volátil.

O Modelo Autorregressivo Integrado de Médias Móveis (ARIMA) apresentou um RMSE de 0,0199 e um MAE de 0,0151, sugerindo uma boa capacidade preditiva em termos de erro absoluto médio e raiz do erro quadrático médio. Seu MAPE, de 152,12%, indica um erro percentual considerável, o que pode estar relacionado à volatilidade dos dados analisados.

O modelo LSTM tradicional obteve resultados levemente melhores em termos de RMSE (0,0194) e manteve um MAE igual ao do ARIMA (0,0151). No entanto, apresentou um MAPE significativamente mais elevado (378,05%), demonstrando que, apesar dos baixos valores absolutos dos erros, sua precisão percentual foi menor, possivelmente devido à sensibilidade do modelo à variabilidade dos dados.

Por outro lado, o LSTM com Dados de Busca, que integrou a variável externa, apresentou um RMSE mais alto (0,0249) e um MAE também superior (0,0173). Seu MAPE foi de 280,47%, indicando que a adição dos dados de busca não contribuiu para uma melhora significativa na precisão do modelo, mas trouxe um aumento no erro absoluto e percentual.

Figura 3 – Previsão LSTM x Retornos Reais (2024-2024)



Fonte: Elaborado pelo autor com dados extraídos do Yahoo Finance (2024).

O gráfico da figura 3, compara os retornos acumulados previstos pelo modelo LSTM (linha azul) com os retornos acumulados reais (linha laranja) durante o período de teste, de 5 de outubro a 1º de novembro de 2024. Enquanto a linha azul representa a soma dos retornos diários previstos pelo modelo com base nos padrões aprendidos, a linha laranja reflete os retornos acumulados reais observados no mercado. Essa comparação permite avaliar a capacidade do modelo de acompanhar a evolução dos retornos ao longo do tempo.

O gráfico evidencia que, embora o modelo LSTM tenha sido capaz de capturar tendências gerais nos retornos acumulados, ele apresenta limitações em cenários de alta volatilidade ou movimentos bruscos, o que é comum no mercado de criptomoedas. Essa observação destaca a importância de explorar melhorias no modelo, como a inclusão de variáveis adicionais ou o ajuste de hiperparâmetros, para melhorar sua capacidade preditiva em mercados voláteis.

5. Conclusões e Recomendações

5.1 Conclusões Gerais

Este estudo avaliou a eficácia dos modelos ARIMA, LSTM Tradicional e LSTM com Dados de Busca na previsão dos retornos do Bitcoin. Cada modelo apresentou pontos fortes e limitações

específicas, destacando diferenças importantes em relação à previsão de um mercado tão volátil como o de criptomoedas.

O modelo ARIMA demonstrou um bom desempenho em termos de erros absolutos, como indicado pelo RMSE e MAE, e é eficaz em capturar relações lineares nos dados. No entanto, seu MAPE elevado indica que ele tem limitações para lidar com a alta volatilidade, especialmente em termos de precisão percentual. Isso sugere que o ARIMA não é ideal para capturar movimentos não lineares e eventos imprevisíveis no mercado.

O modelo LSTM Tradicional apresentou um RMSE ligeiramente menor do que o ARIMA, mas o alto valor de MAPE demonstrou que, embora o modelo tenha capacidade para reconhecer padrões complexos e temporais, ele não foi suficientemente robusto em termos de precisão percentual. Sua sensibilidade à variabilidade dos dados indica que ele ainda enfrenta dificuldades em prever retornos com precisão durante momentos de oscilação intensa.

O modelo LSTM com Dados de Busca, por sua vez, integrou variáveis externas como o volume de buscas no Google Trends, mas não alcançou o desempenho esperado em termos de precisão. Embora tenha apresentado um RMSE mais alto e um MAE superior aos dos outros modelos, seu MAPE também não indicou uma melhora significativa na previsão dos retornos. Esses resultados sugerem que, apesar de os dados comportamentais agregarem valor à modelagem, eles não foram suficientes para superar as limitações do modelo em momentos de extrema volatilidade.

Portanto, os resultados indicam que a previsão dos retornos do Bitcoin continua sendo um desafio devido à sua natureza altamente volátil. A combinação de dados financeiros tradicionais com variáveis externas, como o interesse público, mostra potencial. O uso de dados de tendências de busca, apesar de fornecer insights valiosos, não capta integralmente outros fatores fundamentais, como grandes operações institucionais, notícias financeiras ou eventos geopolíticos. Isso sugere que as variáveis externas melhoram a precisão, mas não são suficientes para garantir previsões robustas em todos os cenários. A inclusão de variáveis comportamentais, oferece uma nova perspectiva para a modelagem preditiva de criptomoedas. No entanto, os desafios enfrentados pelo modelo LSTM com Dados de Busca destacam a necessidade de integrar outros tipos de dados e explorar abordagens híbridas mais avançadas para aumentar a robustez das previsões.

5.2 Implicações Práticas

Os resultados deste estudo apresentam implicações práticas para investidores e analistas financeiros interessados na previsão de preços de ativos digitais. A possibilidade de prever movimentos de preço com maior precisão pode auxiliar investidores a tomar decisões mais informadas,

permitindo a análise de comportamentos e sinais de mercado através de variáveis exógenas. A inclusão de dados de tendências de busca oferece uma nova perspectiva, ajudando a identificar padrões de interesse público que podem preceder variações de preço. Isso possibilita ajustes em estratégias de investimento, como a implementação de operações de hedge ou decisões de compra e venda mais assertivas em momentos de volatilidade.

5.3 Limitações do Estudo

Este estudo enfrentou algumas limitações importantes que devem ser consideradas ao interpretar os resultados. A principal limitação refere-se à alta volatilidade do mercado de criptomoedas, caracterizada por movimentos bruscos e imprevisíveis que os modelos de previsão, incluindo os LSTM, não conseguem capturar completamente. Mesmo o LSTM com Dados de Busca, que apresentou melhor desempenho, mostrou limitações em capturar eventos inesperados, como crises políticas ou grandes ataques a exchanges, que causam flutuações extremas nos preços. Esses eventos representam desafios, pois surgem abruptamente e afetam a eficácia de qualquer modelo preditivo que dependa apenas de dados históricos e variáveis exógenas.

Além disso, a inclusão da variável externa, como dados de tendências de busca, embora agregue valor ao modelo, não é suficiente para capturar a complexidade completa dos fatores que impactam o mercado de criptomoedas. Existem influências consideráveis provenientes de operações institucionais de grande porte, eventos geopolíticos significativos e outras informações financeiras, que não foram integradas neste modelo, limitando sua robustez, especialmente em momentos de crise. Dessa forma, o uso de dados de busca, enquanto útil, representa apenas uma dimensão do comportamento do mercado.

A qualidade dos dados também apresentou desafios. A coleta de dados sobre tendências de busca e preços históricos está sujeita a limitações de granularidade e de atualização em tempo real, o que pode impactar a precisão das previsões. Além disso, o ajuste dos hiperparâmetros dos modelos LSTM é um processo complexo que exige cuidados para evitar overfitting, especialmente devido à alta variabilidade dos dados financeiros, o que dificulta a generalização dos resultados.

5.4 Recomendações para Pesquisas Futuras

Para futuras pesquisas, sugere-se a ampliação da variável externa utilizada, incluindo indicadores macroeconômicos, análises de sentimentos em redes sociais e notícias financeiras em tempo real. Isso permitiria uma compreensão mais abrangente dos fatores que influenciam os retornos do Bitcoin e das criptomoedas em geral. A inclusão de dados de redes sociais, como Twitter e Reddit,

pode ajudar a capturar melhor o sentimento do mercado e antecipar oscilações inesperadas. As análises de sentimento oferecem insights sobre a percepção pública, o que pode ser especialmente útil para prever mudanças abruptas nos preços em resposta a eventos de impacto.

Além disso, outras variáveis exógenas importantes incluem índices de liquidez do mercado, que monitoram a facilidade de transação e podem indicar momentos de pressão de compra ou venda. Indicadores de volatilidade, como o índice de volatilidade implícita, também são úteis, pois refletem o nível de incerteza no mercado e podem antecipar períodos de maior flutuação de preços. Dados de transações on-chain, que monitoram o volume e o valor das transações na rede de Bitcoin, podem fornecer informações valiosas sobre o comportamento dos investidores e a movimentação de grandes quantias, sinalizando possíveis mudanças de tendência.

Fatores como dados de taxas de financiamento de contratos futuros e de derivativos de criptomoedas também são relevantes, pois indicam o nível de alavancagem e o sentimento de alta ou baixa entre os participantes do mercado. Indicadores de adoção tecnológica, como o número de novos endereços de carteira e o volume de transações diárias, ajudam a entender o crescimento e o uso da rede, que podem estar relacionados a aumentos ou quedas na demanda. Finalmente, variáveis relacionadas ao clima regulatório global, como anúncios de novas regulamentações ou restrições em mercados importantes, podem ter impactos significativos nos preços das criptomoedas e são fatores essenciais para capturar a complexidade do mercado.

Além disso, recomenda-se explorar modelos híbridos avançados, como a combinação de redes convolucionais (CNN) com LSTM, o que pode aprimorar a extração de características complexas dos dados, possibilitando a identificação de padrões mais sutis que impactam o preço. A aplicação de arquiteturas híbridas permite uma abordagem mais holística dos dados, combinando análises temporais e características visuais ou de padrões mais profundos, aumentando a precisão preditiva.

Outra recomendação é a implementação de técnicas de ajuste adaptativo e re-treinamento contínuo dos modelos, permitindo que eles se adaptem melhor às mudanças frequentes e abruptas das condições do mercado. O uso de técnicas de aprendizado contínuo possibilitaria que o modelo se adaptasse dinamicamente a novas informações, uma característica essencial em mercados de alta volatilidade como o de criptomoedas.

Por fim, explorar o impacto de eventos exógenos, como regulamentações ou crises políticas, pode ser valioso para entender como esses fatores influenciam o comportamento do mercado de criptomoedas. Isso ajudaria na elaboração de estratégias de mitigação de risco para investidores e possibilitaria uma análise mais detalhada de como esses eventos afetam os retornos de forma imediata e duradoura. Além disso, a implementação de técnicas de ajuste dinâmico de hiperparâmetros pode

melhorar a adaptabilidade dos modelos em tempo real, especialmente em mercados com rápida mudança de comportamento.

Em síntese, este estudo demonstra o potencial dos modelos LSTM na previsão de retornos de criptomoedas, mas sugere que a inclusão de variáveis adicionais e a utilização de modelos híbridos e de aprendizado adaptativo poderiam aprimorar a precisão e a aplicabilidade prática das previsões. Estudos futuros podem se beneficiar ao incorporar mais variáveis externas, incluindo dados de redes sociais, indicadores macroeconômicos e análises de sentimento, o que permitiria uma compreensão mais ampla dos fatores que afetam o mercado. A combinação com arquiteturas híbridas e o uso de re-treinamento contínuo são recomendações chave para melhorar a capacidade preditiva e adaptar os modelos a um ambiente de mercado em constante evolução.

6. Referências Bibliográficas

BASHER. *Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility*. 2022.

KARAMESHINOVA, E. A. *Análise de sensibilidade: Fatores que influenciam o preço das criptomoedas*. 2023. Dissertação (Mestrado) – Iscte - Instituto Universitário de Lisboa, Lisboa, 2023.

LIU, Yujun; LI, Zhongfei; NEKHILI, Ramzi; SULTAN, Jahangir. Forecasting cryptocurrency returns with machine learning. *Research in International Business and Finance*, v. 64, 2023.

NAZARETH, Noella; REDDY, Yeruva Venkata Ramana. *Financial applications of machine learning: A literature review*. Goa: Goa Business School, Goa University, 2023.

MENDES, João Filipe Batista. *Forecasting bitcoin prices: ARIMA vs LSTM*. 2020.

NORDE, Victor Hugo Werneck. *Previsões de preço de criptomoedas utilizando algoritmos de aprendizado de máquina*. Trabalho de Conclusão de Curso (Licenciatura em Matemática) – Universidade Federal de São Carlos, Sorocaba, 2023.

BRAGA, Rosabel Vieira. *Redes neurais LSTM e Google Trends aplicados para previsão de séries temporais do mercado financeiro no contexto de criptomoedas*. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Universidade Federal de Ouro Preto, João Monlevade, 2023.

CORRÊA, Adriano de Almeida. *Pré-processamento de dados para análise de séries temporais na previsão de valores de criptomoedas, utilizando ARIMA e PROPHET*. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Universidade Franciscana, Santa Maria, 2023.

SEABE, Phumudzo Lloyd; MOUTSINGA, Claude Rodrigue Bambe; PINDZA, Edson. *Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach. Fractal and Fractional*, 2023.

OMOLE, Oluwadamilare; ENKE, David. *Deep learning for Bitcoin price direction prediction: models and trading strategies empirically compared*. *Financial Innovation*, v.10, n.117, 2024.

KRISTOUFEK, L. *Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era*. *Scientific Reports*, v.3, p.3415, 2013.

PREIS, T.; MOAT, H. S.; STANLEY, H. E. *Quantifying trading behavior in financial markets using Google Trends*. *Scientific Reports*, v.3, p.1684, 2013.

YAHOO FINANCE. *Histórico de preços do Bitcoin (BTC-USD)*. Disponível em: <https://finance.yahoo.com>. Acesso em: 12 nov. 2024.

GOOGLE TRENDS. *Volume de busca por "Bitcoin"*. Disponível em: <https://trends.google.com>. Acesso em: 12 nov. 2024.

COINMARKETCAP. *Cryptocurrency Market Capitalizations*. 2024. Disponível em: <https://coinmarketcap.com>. Acesso em: 12 nov. 2024.

OPENAI. *ChatGPT* [recurso eletrônico]. Disponível em: <https://chat.openai.com/>. Acesso em: 12 nov. 2024.