

DataCoin: Dataset Sazonal para Estratégias Quantitativas com Bitcoin

José Jeovane R. Cordeiro¹, Arlino H. Magalhães de Araújo¹, Guilherme A. Avelino¹

¹Universidade Federal do Piauí (UFPI)

Campus Universitário Ministro Petrônio Portella – Bairro Ininga
CEP: 64.049-550 – Teresina – PI

jose.cordeiro@ifma.edu.br, {arlino, gaa}@ufpi.edu.br

Abstract. *Bitcoin price forecasting poses a significant challenge due to its high volatility and the influence of multiple nonlinear and non-stationary factors. Traditional time series and econometric models often fail to capture such complexities. To address this issue, this study proposed the construction of an enriched and multifactorial dataset. The analysis of the dataset revealed consistent seasonal patterns, indicating the presence of recurring effects in Bitcoin's behavior. These findings provide relevant insights for the development of more robust quantitative strategies and predictive models. The dataset and source code will be made publicly available, encouraging reuse by researchers, investors, and developers.*

Resumo. *A previsão de preços do Bitcoin representa um desafio significativo, dada sua elevada volatilidade e a influência de múltiplos fatores não lineares e não estacionários. Modelos tradicionais de séries temporais e econometria frequentemente falham em capturar tais complexidades. Para enfrentar esse problema, este trabalho propôs a construção de um dataset enriquecido e multifactorial. A análise do dataset revelou padrões sazonais consistentes, indicando a presença de efeitos recorrentes no comportamento do Bitcoin. Esses achados oferecem subsídios relevantes para o desenvolvimento de estratégias quantitativas e modelos preditivos mais robustos. O dataset e o código-fonte serão disponibilizados publicamente, incentivando sua reutilização por pesquisadores, investidores e desenvolvedores.*

1. Introdução

No cenário dinâmico e em constante evolução das finanças digitais, o Bitcoin consolidou-se como a principal criptomoeda e a mais reconhecida, atraindo um interesse significativo por parte de empresas, consumidores, investidores e pesquisadores [Nayak et al. 2023]. Lançado em 2008 por um autor anônimo sob o pseudônimo de Satoshi Nakamoto, o Bitcoin revolucionou o sistema financeiro ao introduzir uma moeda digital descentralizada. Sua crescente popularidade e adoção em larga escala tornaram-no um instrumento de investimento proeminente, com capitalização de mercado que supera os das demais criptomoedas.

Uma característica marcante do Bitcoin é sua extrema volatilidade de preços, evidenciada por flutuações rápidas e expressivas [Nayak et al. 2023]. Essa instabilidade, embora proporcione oportunidades para ganhos expressivos, torna a previsão dos preços

do Bitcoin uma tarefa complexa e desafiadora. Fatores externos – como condições econômicas e políticas, comportamento do investidor, atividades sociais, dados da rede *blockchain* e inovações tecnológicas – interferem no preço do Bitcoin, dificultando a quantificação e a modelagem precisa dessas influências [Raman et al. 2024]. Adicionalmente, sua natureza não estacionária impõe desafios adicionais, tornando as metodologias analíticas padrão menos eficazes [Rafi et al. 2023]. Todo esse contexto contribui para que modelos de previsão tradicionais, como os baseados em séries temporais e em econometria, frequentemente apresentem limitações para capturar as complexas dependências não lineares e os múltiplos fatores que influenciam o preço do Bitcoin [Guo et al. 2022].

Frente a esse cenário, o presente estudo propõe a construção de um *dataset* abrangente e multifatorial, com o objetivo de oferecer suporte à tomada de decisões por investidores, além de servir como base para o desenvolvimento de modelos preditivos. Ao fornecer essa base de dados, o presente trabalho busca ampliar a compreensão sobre os momentos em que o preço do Bitcoin tende a subir ou cair, contribuindo, assim, para uma tomada de decisão mais informada. Dessa forma, espera-se facilitar a identificação dos momentos mais adequados para comprar, vender ou manter (HODL) os ativos, mitigando riscos e otimizando o retorno sobre o investimento. O *dataset* e todo o código-fonte estão disponíveis publicamente no GitHub e podem ser acessados pelo seguinte link: <https://github.com/jeovanereges/DSW2025>.

Este trabalho está organizado em seções que descrevem o problema em estudo, os procedimentos metodológicos adotados, os resultados obtidos e as soluções propostas. A Seção 2 apresenta a fundamentação teórica para melhor compreensão do estudo. Na Seção 3, são discutidos os principais trabalhos correlatos. A Seção 4 descreve os procedimentos de aquisição e processamento dos dados. Na Seção 5 são apresentados os *insights* inicialmente obtidos. A Seção 6 descreve os possíveis cenários de utilização do *dataset*. Por fim, a Seção 7 apresenta as conclusões do estudo e trabalhos futuros.

2. Referencial Teórico

Esta seção aborda os fundamentos teóricos que sustentam o desenvolvimento do *dataset* proposto.

2.1. O Bitcoin e a Tecnologia Blockchain

Diferentemente das moedas fiduciárias – como o dólar, o euro e o real –, que são emitidas por bancos centrais, o Bitcoin é criado e gerenciado por uma rede descentralizada de computadores distribuídos globalmente [Bâra et al. 2024]. As transações são realizadas diretamente entre os usuários por meio de uma rede *peer-to-peer* (P2P), eliminando a necessidade de intermediários [Muminov et al. 2024]. A base tecnológica que viabiliza esse funcionamento é a *blockchain* [Yang et al. 2024], que, segundo Wang et al., pode ser compreendida como um livro-razão público e seguro, no qual todas as transações são registradas de forma transparente e imutável [Wang et al. 2025]. Nesta rede, cada “bloco” de informações é criptograficamente vinculado ao anterior por meio de um *hash* (código único), formando uma “cadeia” [Yang et al. 2024]. Segundo Sujatha et al., essa inovação impulsionou o Bitcoin a se tornar um dos ativos de investimento mais proeminentes, com um valor que cresceu exponencialmente ao longo dos anos [Sujatha et al. 2021].

2.2. Métodos e Abordagens para a Previsão de Preços do Bitcoin

Ao longo dos anos, pesquisadores têm explorado diversas metodologias para enfrentar os desafios da previsão de preços do Bitcoin, desde modelos estatísticos tradicionais até técnicas avançadas de Inteligência Artificial (IA). Os modelos estatísticos tradicionais são fundamentados em princípios matemáticos e estatísticos, e são frequentemente empregados na análise de séries temporais. Entre os principais modelos utilizados, destacam-se: ARIMA [Waheed et al. 2020], GARCH [Raman et al. 2024], Prophet [Yenidog̃an et al. 2018] e modelos de regressão [S et al. 2025]. Já os modelos baseados em *Machine Learning* (ML) e *Deep Learning* (DL) buscam superar as limitações dos modelos estatísticos tradicionais na captura de padrões não lineares e complexos. Entre os principais modelos utilizados, destacam-se: Redes Neurais Artificiais (ANN) [Sujatha et al. 2021], LSTMs (*Long Short-Term Memory*) [Wu et al. 2018], Aprendizagem por Reforço (RL) [Azamjon et al. 2023] e Sistemas Fuzzy [Hajek and Olej 2023].

3. Trabalhos Relacionados

A crescente popularidade do Bitcoin como ativo de investimento global, aliada à sua notória volatilidade, tem impulsionado um vasto corpo de pesquisas voltadas à previsão de seus preços. Compreender as complexas dinâmicas que influenciam o seu preço é essencial tanto para investidores quanto para reguladores, considerando suas características únicas de descentralização e a ausência de controle central [Nayak et al. 2023]. A literatura existente emprega uma variedade de abordagens, que podem ser categorizadas em modelos estatísticos tradicionais, modelos de ML, modelos de DL e análises que incorporam fatores externos, como sentimento de mercado e dados *on-chain*. No entanto, cada uma dessas abordagens apresenta limitações significativas.

3.1. Modelos Estatísticos Tradicionais

Historicamente, a previsão de séries temporais financeiras tem-se fundamentado em modelos estatísticos. Entre eles, o ARIMA é amplamente reconhecido por sua eficácia na análise de dados temporais, sendo frequentemente aplicado à modelagem de tendências de curto prazo e padrões sazonais nos preços do Bitcoin [De Leon et al. 2022]. Pesquisas iniciais, como a de Kristoufek, aplicaram o modelo ARIMA na previsão do preço do Bitcoin, destacando a relevância dos dados históricos no processo preditivo [Kristoufek 2015]. Sutiksono et al. [Sutiksono et al. 2018] também aplicaram o modelo ARIMA em conjunto com o indicador α -Sutte para prever os preços históricos do Bitcoin. O modelo GARCH também tem sido amplamente utilizado para modelar a volatilidade e identificar padrões de agrupamento (*clustering*) em séries temporais de criptomoedas [Waseem and Singh 2024]. Uma limitação desses modelos é a incapacidade de capturar padrões não lineares, os quais são característicos de mercados voláteis [S et al. 2025]. Além disso, a natureza não estacionária e não linear do Bitcoin dificulta a manutenção das suposições de distribuição de dados inerentes aos modelos estatísticos [Rafi et al. 2023].

3.2. Modelos de Machine Learning

Diante da crescente complexidade e volatilidade do mercado de criptomoedas, pesquisadores têm recorrido a técnicas de ML como alternativa para superar as limitações dos modelos estatísticos. Algoritmos como *Support Vector Machines* (SVM), *Random Forest*

(RF) e *Multi-Layer Perceptron* (MLP) têm sido amplamente aplicados nesse contexto. Zhang *et al.* [Zhang et al. 2019] aplicaram o algoritmo *Random Forest* e demonstraram sua elevada eficácia na identificação de relações intrincadas e não lineares em dados de criptomoedas. Greaves *et al.* analisaram a *blockchain* do Bitcoin com o objetivo de prever seu preço utilizando os algoritmos *Support Vector Machines* (SVM) e *Artificial Neural Networks* (ANN), relatando uma acurácia de 55% na previsão da direção do preço ao empregar uma ANN regular [Greaves and Au 2015]. Um estudo conduzido por Mallqui e Fernandes [Mallqui and Fernandes 2019] previu as taxas de câmbio do Bitcoin – incluindo os preços máximo, mínimo e de fechamento – utilizando os modelos *Artificial Neural Networks* (ANN), *Recurrent Neural Networks* (RNN) e *Support Vector Machines* (SVM), sendo este último o que apresentou os melhores resultados. Apesar de sua capacidade de capturar não linearidades, muitos modelos de ML – especialmente o *Random Forest* – ainda demandam uma engenharia de atributos extensiva, como a criação de variáveis defasadas, para lidar com a estrutura temporal dos dados, o que torna o processo demorado e complexo [S et al. 2025].

3.3. Modelos de Deep Learning

As redes neurais LSTM, as *Gated Recurrent Units* (GRU) e as Redes Neurais Convolucionais (CNN) têm ganho destaque por sua capacidade de lidar com séries temporais e de modelar relações não lineares, tornando-as particularmente adequadas para contextos de alta volatilidade, como o do Bitcoin [Raman et al. 2024]. Modelos híbridos que combinam LSTM com GRU (LSTM-GRU) ou com CNN (CNN-LSTM) têm sido propostos na literatura com o objetivo de capturar dependências de longo e curto prazo, além de aprimorar a precisão preditiva [Hajare et al. 2025]. Outros estudos demonstram que as redes LSTM e suas variantes – como *Bidirectional LSTM* (Bi-LSTM) e GRU – superam outros métodos em termos de precisão na previsão do preço do Bitcoin [Vysotska et al. 2024]. No entanto, de modo geral, muitos estudos baseados em DL concentram-se na previsão de curto prazo, enquanto as previsões de longo prazo permanecem desafiadoras devido à crescente complexidade e à elevada volatilidade das séries temporais [Iqbal et al. 2024]. Além disso, a interpretabilidade constitui um desafio significativo nos modelos de DL. Embora essas abordagens apresentem elevado desempenho preditivo, frequentemente não fornecem explicações claras sobre os fatores que levaram a uma determinada previsão de preço – o que é essencial para a tomada de decisões estratégicas por parte dos investidores [Waseem and Singh 2024].

3.4. Análise de Sentimento e Mídia Social

A crescente atenção dada às mídias sociais tem sido identificada como um fator crucial na flutuação dos preços do Bitcoin [Sattarov et al. 2020]. Pesquisas recentes têm explorado o uso de dados textuais provenientes de diferentes plataformas com o objetivo de extrair informações de sentimento e prever a volatilidade dos preços [Ren et al. 2025]. O estudo de Sattarov *et al.* indica a existência de uma correlação entre o sentimento positivo expresso no X (antigo Twitter) e o preço do Bitcoin, especialmente em horizontes de curto prazo [Sattarov et al. 2020]. O modelo Twitter-roBERTa, por exemplo, superou os métodos tradicionais de análise de sentimento, alcançando uma acurácia de 0,90 [Shang 2025]. Vale ressaltar que a análise textual tradicional requer um conjunto predefinido de palavras-chave, que pode variar entre diferentes contextos culturais e linguísticos, e tende a ser imprecisa sem o suporte de técnicas de ML. Além disso, essas técnicas

exigem grandes volumes de dados para treinamento e demandam elevado custo computacional. Além disso, Muminov *et al.* destacam que a extração de dados a partir de um conjunto limitado de plataformas pode não capturar toda a complexidade da dinâmica de sentimento que impulsiona o mercado mais amplo [Muminov *et al.* 2024].

3.5. Dados On-chain e Fatores Exógenos

Além das mídias sociais, a transparência da tecnologia *blockchain* permite a análise de dados *on-chain*, que fornecem *insights* relevantes sobre o funcionamento da rede Bitcoin, como o volume de transações, a atividade da rede e o comportamento dos mineradores [Sujatha *et al.* 2021]. Indicadores como a taxa de *hash*, a dificuldade de mineração e os custos de transação também têm demonstrado correlação com os preços do Bitcoin [Azamjon *et al.* 2023]. Além disso, fatores macroeconômicos – como taxas de juros e a probabilidade de recessão – também têm sido explorados como variáveis exógenas, demonstrando potencial para aprimorar a capacidade preditiva dos modelos [Mahfooz and Phillips 2024]. Porém, o uso de dados *on-chain* apresenta desafios significativos, principalmente devido ao volume massivo e ao rápido crescimento das informações registradas em *blockchains* públicas, o que dificulta o processamento eficiente e a extração de detalhes cruciais [Azamjon *et al.* 2023].

4. Aquisição e Processamento dos Dados

O conjunto de dados descrito neste artigo foi construído a partir de informações coletadas por meio da biblioteca *yfinance*¹, que fornece acesso a dados históricos do mercado financeiro por meio da API da plataforma Yahoo Finance. Esta plataforma é amplamente utilizada por investidores e entusiastas de finanças por fornecer informações sobre cotações de ações, notícias de mercado, relatórios financeiros, entre outros recursos. Além disso, ela disponibiliza uma ampla variedade de ativos que podem ser acessados por meio da biblioteca *yfinance*, incluindo aproximadamente 106.332 ações, 21.196 ETFs, 9.294 contratos futuros, 80.017 índices, 24.926 fundos mútuos e 4.019 pares de moedas [Bagul *et al.* 2022].

4.1. Extração e Agregação dos Dados

Os dados utilizados no experimento correspondem à série temporal diária do par BTC-USD, abrangendo o período de janeiro de 2017 a maio de 2025. Os dados extraídos contemplam as colunas de abertura, máxima, mínima e fechamento para cada dia. Após a coleta, os dados foram agregados em frequência mensal por meio da técnica de *resample*, utilizando agregações específicas: o valor de abertura foi definido como o primeiro valor registrado no mês, o valor de fechamento como o último, e os valores de máxima e mínima como os extremos observados no período. Essa transformação teve como objetivo gerar uma representação consolidada e mais robusta dos movimentos mensais do ativo. A Figura 1 apresenta o fluxo geral do processo de extração e enriquecimento dos dados.

¹Yfinance: <https://pypi.org/project/yfinance/>

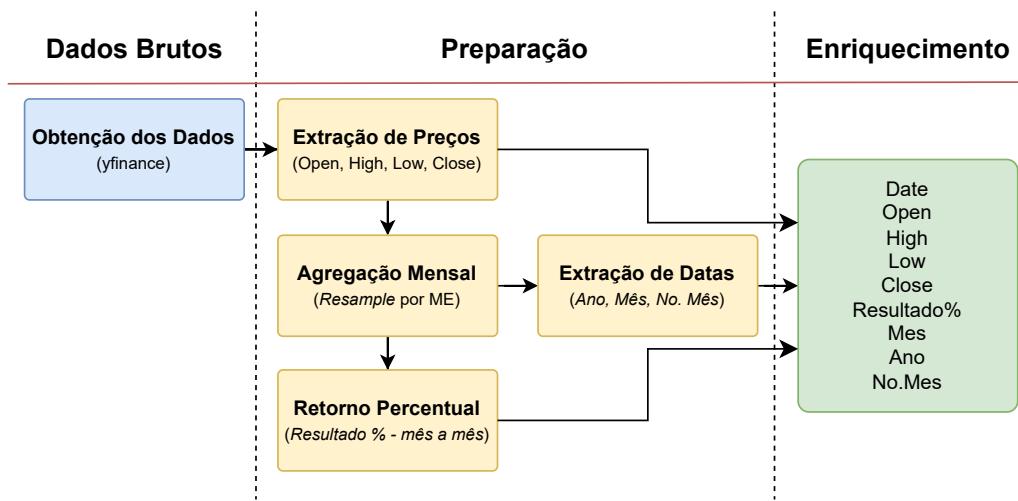


Figura 1. Fluxo Geral de Aquisição e Enriquecimento dos Dados.

4.2. Enriquecimento

A partir da extração dos dados, foi possível derivar novas colunas com base nas variáveis originais, conforme descrito a seguir:

- **Resultado%:** Coluna que representa a variação percentual do preço de fechamento de um mês para o subsequente, fornecendo uma estimativa direta do desempenho mensal do ativo. Essa métrica é calculada por meio da equação apresentada na Figura 2.

$$\text{Resultado\%}_t = \left(\frac{\text{Close}_t - \text{Close}_{t-1}}{\text{Close}_{t-1}} \right) \times 100$$

Figura 2. Equação para Cálculo da Variação Percentual.

- **Mes:** Coluna extraída a partir do índice temporal da série, indicando o mês correspondente a cada observação (em formato numérico de 1 a 12). Essa informação permite a análise sazonal dos retornos.
- **Ano:** Coluna semelhante à anterior, responsável por indicar o ano correspondente a cada observação. Essa variável serve de base para agrupamentos e comparações entre diferentes períodos.
- **No.Mes:** Coluna derivada do índice temporal, representando o mês no formato de duas casas decimais (por exemplo, ‘01’ para janeiro e ‘12’ para dezembro). Essa variável é especialmente útil para a organização e visualização de dados em tabelas dinâmicas e gráficos ordenados por mês.

Ao término do processamento, os dados foram organizados em um *DataFrame* contendo nove colunas: as cinco primeiras resultantes da etapa de coleta e as quatro últimas derivadas por transformação.

4.3. Volume de Dados

Foram processados dados históricos do par BTC-USD ao longo de aproximadamente 8,5 anos, compreendendo o período entre janeiro de 2017 e maio de 2025. Após a etapa de pré-processamento, obteve-se um total de 157 registros mensais únicos, cada um representando os dados consolidados de preços para um determinado mês, incluindo os valores de abertura, máxima, mínima, fechamento e a variação percentual mensal.

4.4. Dicionário de Dados

A Tabela 1 apresenta o dicionário de dados do conjunto histórico de preços do par BTC-USD. Cada registro no *dataset* corresponde a um consolidado mensal, contendo os preços de abertura, máxima, mínima e fechamento do ativo, além de colunas derivadas, como a variação percentual e os marcadores temporais (mês e ano). Todos os campos foram padronizados em tipos compatíveis com formatos de arquivos CSV e Excel, utilizando tipagens numéricas, textuais e temporais. O objetivo deste dicionário é apoiar pesquisadores, analistas de mercado e desenvolvedores na correta compreensão da estrutura da base de dados, promovendo transparência e incentivando sua reutilização adequada.

| Campo | Tipo | Descrição |
|------------|---------|--|
| Date | Data | Data de referência da observação mensal (último dia do mês). |
| Open | Real | Valor de abertura do BTC-USD no primeiro dia útil do mês. |
| High | Real | Maior valor de cotação do BTC-USD durante o mês. |
| Low | Real | Menor valor de cotação do BTC-USD durante o mês. |
| Close | Real | Valor de fechamento do BTC-USD no último dia útil do mês. |
| Resultado% | Real | Variação percentual do preço de fechamento em relação ao mês anterior. |
| Mes | Inteiro | Mês da observação (valor numérico de 1 a 12). |
| Ano | Inteiro | Ano da observação (quatro dígitos). |
| No.Mes | Texto | Representação do mês em formato de duas casas (ex: “01”, “12”). |

Tabela 1. Dicionário de Dados do *Dataset*.

5. Análise Descritiva e Exploratória

Esta seção apresenta uma análise descritiva e exploratória da distribuição dos retornos mensais do par BTC-USD, considerando os recortes temporais por ano e por mês. O objetivo é identificar padrões de comportamento, variações significativas ao longo do tempo e possíveis tendências sazonais que possam influenciar a tomada de decisões de investimento.

5.1. Análise Mensal

A Figura 3 apresenta a variação percentual mensal do par BTC-USD, calculada ao longo do período de janeiro de 2017 a maio de 2025, possibilitando a identificação de padrões sazonais de desempenho e de flutuações características do mercado de criptomoedas. Os dados revelam meses com ganhos superiores a 20% e perdas próximas de -5%, o que reforça o caráter especulativo do Bitcoin.

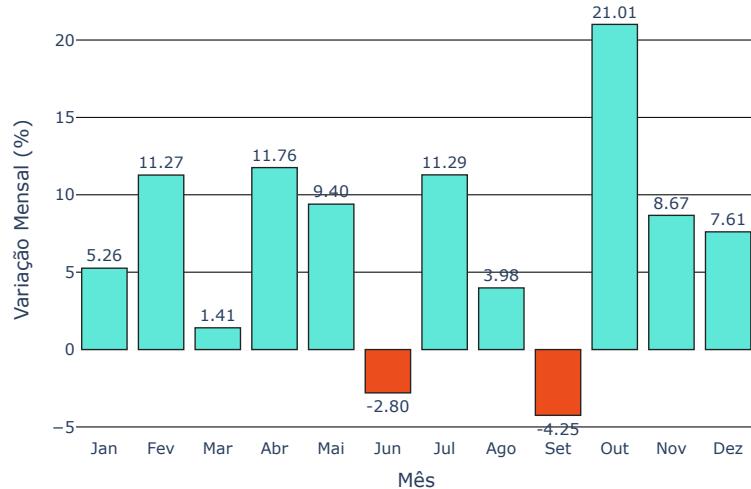


Figura 3. Média Mensal da Variação (%).

A Figura 3 também sugere uma possível tendência de recuperação no último trimestre do ano, contrastando com um desempenho mais fraco nos meses centrais, como junho e setembro. Tais padrões podem refletir ciclos econômicos, divulgação de relatórios financeiros, sazonalidade da liquidez institucional e variações no comportamento dos investidores. Esses resultados podem subsidiar estratégias de *market timing*, nas quais investidores ajustam suas posições com base em tendências sazonais observadas.

5.2. Análise Anual e Mensal

A Tabela 2 sintetiza o desempenho mensal e anual do par BTC-USD no período de janeiro de 2017 a maio de 2025, oferecendo uma visão longitudinal da volatilidade e dos ciclos de valorização e desvalorização do Bitcoin. A análise revela uma variação significativa entre os anos, com destaque para os fortes ganhos observados em 2017, 2020 e 2023. O ano de 2017 se destaca como o recordista, apresentando diversos meses com variações superiores a 50%. Por outro lado, os anos de 2018 e 2022 registraram perdas expressivas, refletindo os efeitos da mudança de ciclo econômico, de políticas monetárias restritivas e de eventos macroeconômicos adversos.

| Ano | Jan | Fev | Mar | Abr | Mai | Jun | Jul | Ago | Set | Out | Nov | Dez |
|------|--------|--------|--------|--------|--------|--------|-------|--------|--------|-------|--------|--------|
| 2017 | 0.00 | 21.60 | -9.17 | 25.76 | 69.63 | 8.50 | 15.90 | 63.58 | -7.75 | 49.09 | 58.21 | 38.33 |
| 2018 | -27.80 | 1.73 | -32.93 | 32.51 | -18.90 | -14.55 | 21.49 | -9.55 | -5.85 | -4.65 | -36.41 | -6.83 |
| 2019 | -7.61 | 11.48 | 6.50 | 30.33 | 60.25 | 26.15 | -6.76 | -4.51 | -13.88 | 10.92 | -17.72 | -4.97 |
| 2020 | 29.98 | -8.03 | -25.13 | 34.48 | 9.27 | -3.41 | 23.92 | 3.16 | -7.67 | 27.79 | 42.41 | 47.77 |
| 2021 | 14.18 | 36.31 | 30.53 | -1.98 | -35.35 | -6.14 | 18.79 | 13.31 | -7.16 | 40.03 | -7.03 | -18.77 |
| 2022 | -16.89 | 12.24 | 5.43 | -17.18 | -15.70 | -37.77 | 17.95 | -14.09 | -3.08 | 5.48 | -16.23 | -3.62 |
| 2023 | 39.84 | 0.03 | 23.03 | 2.78 | -7.00 | 11.97 | -4.09 | -11.29 | 4.00 | 28.55 | 8.78 | 12.07 |
| 2024 | 0.75 | 43.72 | 16.56 | -15.00 | 11.30 | -7.13 | 3.10 | -8.74 | 7.39 | 10.87 | 37.36 | -3.13 |
| 2025 | 9.61 | -17.61 | -2.16 | 14.12 | 11.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Tabela 2. Resultado Médio (%) por Ano e Mês.

No recorte mensal, especialmente em 2017, os meses de maio (69,63%), agosto (63,58%), outubro (49,09%), novembro (58,21%) e dezembro (38,33%) marcaram o auge da valorização no ciclo de alta pré-2018. Em 2021, os ganhos registrados em fevereiro (36,31%), março (30,53%) e outubro (40,03%) indicaram uma forte tendência de alta antes da reversão observada em 2022. O ano de 2023 destacou-se por uma recuperação consistente, com resultados positivos em nove dos doze meses, sugerindo o início de um possível novo ciclo de valorização. Já os primeiros cinco meses de 2025 apontam para uma valorização acumulada de 9,61%, com destaque para abril (14,12%) e maio (11,07%), o que pode indicar um início de ano promissor.

Nota-se que o mês de fevereiro se destaca como um dos mais favoráveis ao longo da série histórica, apresentando variações mensais positivas em sete dos nove anos analisados. Exceções ocorreram apenas em 2020 (-8,03%) e 2025 (-17,61%), o que ressalta a força sazonal de fevereiro como um período recorrente de recuperação ou impulso de alta. Da mesma forma, o mês de julho apresentou desempenho amplamente positivo, com variações negativas registradas apenas em 2019 (-6,76%) e 2023 (-4,09%), o que indica uma tendência de valorização no início do segundo semestre. Outro destaque importante é o mês de outubro, que apresentou variação negativa apenas em 2018 (-4,65%), e forte desempenho em anos como 2017 (49,09%), 2020 (27,79%), 2021 (40,03%) e 2023 (28,55%), consolidando-se como um dos meses mais robustos para o ativo.

Em contraste, alguns meses apresentam tendência predominantemente negativa. O mês de junho, por exemplo, registrou variações positivas apenas em 2017 (8,5%), 2019 (26,15%) e 2023 (11,97%), apresentando desempenho negativo nos demais anos, o que sugere cautela por parte dos investidores nesse período. Agosto, por sua vez, apresentou variações positivas apenas em 2017 (63,58%), 2020 (3,16%) e 2021 (13,31%), registrando desempenhos negativos nos demais anos. Por fim, o mês de setembro encerrou no campo positivo apenas em 2023 (4,0%) e 2024 (7,39%), sendo historicamente marcado por perdas ao longo dos demais períodos analisados.

Esses dados reforçam a presença de um componente sazonal relevante no comportamento mensal do Bitcoin, evidenciando meses e anos com recorrência de ganhos – como fevereiro, julho e outubro – e outros historicamente associados a perdas, como junho, agosto e setembro, especialmente em anos de baixa. Tais padrões são valiosos para a construção de estratégias quantitativas e modelos preditivos baseados em fatores temporais, contribuindo para decisões mais informadas em contextos de investimento e análise de risco. Além disso, compreender esses padrões pode subsidiar decisões estratégicas de alocação temporal de capital no mercado de criptoativos.

6. Disponibilização e Utilização

6.1. Disponibilização dos Dados e Código

O conjunto de dados gerado neste estudo, assim como o código-fonte utilizado para sua extração, agregação e enriquecimento, será disponibilizado publicamente em um repositório no GitHub². O *dataset* contempla informações mensais consolidadas do par BTC-USD, abrangendo o período de janeiro de 2017 a maio de 2025. As informações contidas no *dataset* incluem os preços de abertura, máxima, mínima e fechamento, a variação percentual mensal e marcadores temporais, como mês e ano de referência. No repositório

²Repositório: <https://github.com/jeovanereges/DSW2025>

também estará disponível um *script* em Python, desenvolvido de forma modular e comentada, o que facilita sua reutilização e adaptação para outros ativos financeiros acessíveis por meio da API pública do Yahoo Finance, via biblioteca *yfinance*.

6.2. Possíveis Cenários para Utilização

O conjunto de dados e o código-fonte disponibilizados neste estudo apresentam ampla aplicabilidade em diversos contextos do mercado financeiro, especialmente em análises quantitativas, modelagem preditiva e estratégias de investimento baseadas em dados históricos. A seguir, enumeram-se alguns cenários potenciais de aplicação:

- **Tomada de Decisão Estratégica:** Investidores e demais partes interessadas podem utilizar os dados históricos e os padrões identificados para identificar momentos oportunos de ampliação de portfólio ou realização de lucros, com base em análises quantitativas e em sinais de mercado derivados de ciclos temporais e tendências recorrentes.
- **Desenvolvimento de Modelos Preditivos:** Pesquisadores e desenvolvedores podem utilizar o *dataset* para o treinamento e a validação de modelos de previsão de preços do Bitcoin, empregando técnicas de ML, DL ou abordagens híbridas que incorporem variáveis temporais.
- **Análise Sazonal e Comportamental:** Analistas de mercado podem explorar os padrões mensais e anuais para identificar comportamentos recorrentes de valorização e desvalorização, subsidiando decisões mais informadas sobre a alocação temporal de capital ao longo do ano.
- **Expansão para Outros Ativos:** A abordagem adotada para a construção e análise da base de dados permite que os usuários ajustem facilmente os parâmetros para coletar dados de outros criptoativos ou ativos tradicionais – como ações, índices, *commodities* e moedas cambiais –, viabilizando tanto a reproduzibilidade dos experimentos quanto o desenvolvimento de portfólios multiativos.

7. Conclusões e Trabalhos Futuros

O presente trabalho apresentou uma solução computacional voltada para a construção e disponibilização de um conjunto de dados enriquecido, com foco na análise e na previsão de preços do Bitcoin. Por meio da biblioteca *yfinance*, foi possível extrair, transformar e organizar dados históricos do par BTC-USD em frequência mensal, gerando variáveis derivadas, como a variação percentual do preço de fechamento, além de marcadores temporais que possibilitam análises sazonais e estudos quantitativos aprofundados.

A análise exploratória evidenciou a existência de padrões mensais recorrentes de valorização e desvalorização, indicando a influência de fatores sazonais no comportamento do Bitcoin. Esses achados reforçam a utilidade do *dataset* proposto, tanto para o desenvolvimento de modelos preditivos mais robustos quanto para a formulação de estratégias de investimento orientadas por dados históricos. Além disso, a modularidade do código-fonte e a estrutura clara da base de dados favorecem sua reutilização em distintos contextos, incluindo o ensino, a pesquisa acadêmica e aplicações práticas.

Como trabalhos futuros, propõe-se: (a) a expansão do *dataset* com variáveis exógenas, como indicadores macroeconômicos, dados *on-chain* e métricas de sentimento

extraídas de mídias sociais, (b) a replicação do estudo para outros criptoativos ou ativos financeiros tradicionais, ampliando a análise comparativa entre comportamentos sazonais e (c) a aplicação e comparação de diferentes modelos de previsão, incluindo abordagens híbridas.

Referências

- [Azamjon et al. 2023] Azamjon, M., Sattarov, O., and Cho, J. (2023). Forecasting bitcoin volatility through on-chain and whale-alert tweet analysis using the q-learning algorithm. *IEEE Access*.
- [Bagul et al. 2022] Bagul, J., Warkhade, P., Gangwal, T., and Mangaonkar, N. (2022). Arima vs lstm algorithm – a comparative study based on stock market prediction. In *2022 5th International Conference on Advances in Science and Technology (ICAST)*.
- [Bâra et al. 2024] Bâra, A., Georgescu, I. A., Oprea, S.-V., and Cristescu, M. P. (2024). Exploring the dynamics of brent crude oil, s&p500 and bitcoin prices amid economic instability. *IEEE Access*.
- [De Leon et al. 2022] De Leon, L. G. N., Gomez, R. C., Tacal, M. L. G., Taylor, J. V., Nojor, V. V., and Villanueva, A. R. (2022). Bitcoin price forecasting using time-series architectures. In *2022 International Conference on ICT for Smart Society (ICISS)*.
- [Greaves and Au 2015] Greaves, A. and Au, B. (2015). Using the bitcoin transaction graph to predict the price of bitcoin. *No data*, 8:416–443.
- [Guo et al. 2022] Guo, H., Gao, K., Yu, Y., Liu, Y., and Fu, L. (2022). A diluted bitcoin-dollar-gold mean prediction scheme based on periodic prediction method. *IEEE Access*.
- [Hajare et al. 2025] Hajare, R., Puri, C., and Gote, P. M. (2025). Bitcoin price prediction using deep learning models. In *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSDL)*.
- [Hajek and Olej 2023] Hajek, P. and Olej, V. (2023). Hierarchical intuitionistic tsk fuzzy system for bitcoin price forecasting. In *2023 IEEE International Conference on Fuzzy Systems (FUZZ)*.
- [Iqbal et al. 2024] Iqbal, M., Iqbal, A., Alshammary, A., Ali, I., Maghrabi, L. A., and Usman, N. (2024). Sell or hodl cryptos: Cryptocurrency short-to-long term projection using simultaneous classification-regression deep learning framework. *IEEE Access*.
- [Kristoufek 2015] Kristoufek, L. (2015). What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis. *PLOS ONE*.
- [Mahfooz and Phillips 2024] Mahfooz, A. and Phillips, J. L. (2024). Conditional forecasting of bitcoin prices using exogenous variables. *IEEE Access*.
- [Mallqui and Fernandes 2019] Mallqui, D. C. and Fernandes, R. A. (2019). Predicting the direction, maximum, minimum and closing prices of daily bitcoin exchange rate using machine learning techniques. *Applied Soft Computing*.
- [Muminov et al. 2024] Muminov, A., Sattarov, O., and Na, D. (2024). Enhanced bitcoin price direction forecasting with dqn. *IEEE Access*.
- [Nayak et al. 2023] Nayak, S. C., Das, S., Dehuri, S., and Cho, S.-B. (2023). An elitist artificial electric field algorithm based random vector functional link network for cryptocurrency prices forecasting. *IEEE Access*.
- [Rafi et al. 2023] Rafi, M., Mirza, Q. A. K., Sohail, M. I., Aliasghar, M., Aziz, A., and Hameed, S. (2023). Enhancing cryptocurrency price forecasting accuracy: A feature

- selection and weighting approach with bi-directional lstm and trend-preserving model bias correction. *IEEE Access*.
- [Raman et al. 2024] Raman, R., Kumar, V., Pillai, B. G., Rabadiya, D., Divekar, R., and Vachharajani, H. (2024). Forecasting bitcoin value with hybrid lstm-gru neural networks. In *2024 Second International Conference on Data Science and Information System (ICDSIS)*.
- [Ren et al. 2025] Ren, X., Jiang, W., Duan, K., and Mishra, T. (2025). Being an emotionally unaffected investor: Evidence from bitcoin. *IEEE Transactions on Engineering Management*.
- [S et al. 2025] S, S. A., D, R., and Devi, V. S. K. (2025). Forecasting bitcoin price trends: Integrated machine learning with market trends. In *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*.
- [Sattarov et al. 2020] Sattarov, O., Jeon, H. S., Oh, R., and Lee, J. D. (2020). Forecasting bitcoin price fluctuation by twitter sentiment analysis. In *2020 International Conference on Information Science and Communications Technologies (ICISCT)*.
- [Shang 2025] Shang, L. (2025). Sentiment-driven bitcoin price range forecasting: Enhancing cart decision trees with high-dimensional indicators and twitter dynamics. *IEEE Access*.
- [Sujatha et al. 2021] Sujatha, R., Mareeswari, V., Chatterjee, J. M., Mousa, A. A. A., and Hassanien, A. E. (2021). A bayesian regularized neural network for analyzing bitcoin trends. *IEEE Access*.
- [Sutiksno et al. 2018] Sutiksno, D. U., Ahmar, A. S., Kurniasih, N., Susanto, E., and Leiwakabessy, A. (2018). Forecasting historical data of bitcoin using arima and a-sutte indicator. *Journal of Physics: Conference Series*.
- [Vysotska et al. 2024] Vysotska, V., Smelyakov, K., Naumov, A., and Shtanko, V. (2024). A study of the effectiveness of using information technology based on artificial intelligence for bitcoin price forecasting. In *2024 IEEE 19th International Conference on Computer Science and Information Technologies (CSIT)*.
- [Waheed et al. 2020] Waheed, W., Shah, H., Jabreel, M., and Puig, D. (2020). Bitcoin price forecasting: A comparative study between statistical and machine learning methods. In *2020 2nd International Conference on Computer and Information Sciences (ICCIS)*.
- [Wang et al. 2025] Wang, M., Braslavski, P., Manevich, V., and Ignatov, D. I. (2025). Bitcoin ordinals: Bitcoin price and transaction fee rate predictions. *IEEE Access*.
- [Waseem and Singh 2024] Waseem, M. and Singh, S. N. (2024). Advanced predictive modelling for bitcoin and ethereum price dynamics. In *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*.
- [Wu et al. 2018] Wu, C.-H., Lu, C.-C., Ma, Y.-F., and Lu, R.-S. (2018). A new forecasting framework for bitcoin price with lstm. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*.
- [Yang et al. 2024] Yang, F., Qiao, Y., Bo, J., Ye, L., and Abedin, M. Z. (2024). Blockchain and digital asset transactions- based carbon emissions trading scheme for industrial internet of things. *IEEE Transactions on Industrial Informatics*.
- [Yenidoḡan et al. 2018] Yenidoḡan, I., Çayir, A., Kozan, O., Dāg, T., and Arslan, (2018). Bitcoin forecasting using arima and prophet. In *2018 3rd International Conference on Computer Science and Engineering (UBMK)*.
- [Zhang et al. 2019] Zhang, H., Nettleton, D., and Zhu, Z. (2019). Regression-enhanced random forests.