

HENRIQUE PENNA BARBOSA DINIZ

ANÁLISE DE ESTRATÉGIAS E MÉTODOS PARA PREDIÇÃO DE
FUNDOS DE INVESTIMENTOS IMOBILIÁRIOS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

Orientador: Fabrício A. Silva

VIÇOSA - MINAS GERAIS
2024

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

D583a
2024

Diniz, Henrique Penna Barbosa, 1997-
Análise de estratégias e métodos para predição de fundos de
investimentos imobiliários / Henrique Penna Barbosa Diniz. –
Viçosa, MG, 2024.
1 dissertação eletrônica (79 f.): il. (algumas color.).

Orientador: Fabrício Aguiar Silva.
Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Informática, 2024.
Referências bibliográficas: f. 77-79.
DOI: <https://doi.org/10.47328/ufvbbt.2024.201>
Modo de acesso: World Wide Web.

1. Aprendizado do computador. 2. Investimentos
imobiliários - Processamento de dados. 3. Análise de séries
temporais. I. Silva, Fabrício Aguiar, 1981-. II. Universidade
Federal de Viçosa. Departamento de Informática. Programa de
Pós-Graduação em em Ciência da Computação. III. Título.

CDD 22. ed. 006.331

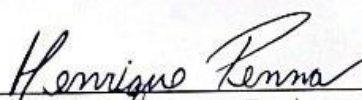
HENRIQUE PENNA BARBOSA DINIZ

**ANÁLISE DE ESTRATÉGIAS E MÉTODOS PARA PREDIÇÃO DE
FUNDOS DE INVESTIMENTOS IMOBILIÁRIOS**

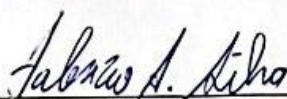
Dissertação apresentada à Universidade
Federal de Viçosa, como parte das exigências
do Programa de Pós-Graduação em Ciência
da Computação, para obtenção do título de
Magister Scientiae.

APROVADA: 08 de março de 2024.

Assentimento:



Henrique Penna Barbosa Diniz
Autor



Fabrício Aguiar Silva
Orientador

Agradecimentos

Aos meus pais, cuja dedicação e sacrifícios proporcionaram a oportunidade de estudar na Universidade Federal de Viçosa.

Aos meus amigos de Viçosa e Pará de Minas, cujo apoio incondicional e incentivo foram pilares durante esta jornada acadêmica.

À UFV, por oferecer um ambiente propício ao aprendizado e à pesquisa, além de todas as oportunidades concedidas.

Ao Professor Fabrício, pela orientação acadêmica precisa e pela confiança depositada em meu trabalho.

E, principalmente, à minha irmã, por seu papel crucial, oferecendo revisões de texto, sugestões valiosas e um constante incentivo, contribuindo significativamente para a conclusão deste trabalho.

A todos mencionados e a todos que de alguma forma contribuíram, meu mais sincero agradecimento.

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de Minas Gerais – Brasil (FAPEMIG) – Código de Financiamento 001.

Resumo

Diniz, Henrique, M.Sc., Universidade Federal de Viçosa, março de 2024. **Análise de Estratégias e Métodos para Predição de Fundos de Investimentos Imobiliários.** Orientador: Fabrício A. Silva.

Este trabalho aborda a previsão da variação do preço das cotas de Fundos de Investimento Imobiliário (FIIs), uma questão de relevância crescente para investidores e gestores de portfólio. Diante da complexidade desse desafio e da necessidade de tomar decisões informadas no mercado imobiliário, esta pesquisa explora estratégias inovadoras de modelagem, contrastando métodos tradicionais de séries temporais com abordagens de aprendizado de máquina. Um aspecto inovador foi o enriquecimento dos dados com informações adicionais sobre os imóveis subjacentes aos FIIs, buscando entender como características específicas influenciam sua valorização. Os resultados demonstraram que, em muitos casos, os modelos de aprendizado de máquina superaram as técnicas tradicionais de análise de séries temporais, especialmente na previsão de tendências de médio prazo. Além disso, este estudo investigou a viabilidade de utilizar um classificador para determinar a estratégia mais eficiente na predição, considerando os resultados obtidos com os modelos ARIMA e XGBoost. A análise revelou que diferentes fundos exibem comportamentos distintos ao longo do tempo, e a escolha do algoritmo adequado pode variar conforme as características individuais de cada fundo.

Palavras-chave: Fundo de Investimento Imobiliário, Aprendizado de Máquina, Séries Temporais

Abstract

Diniz, Henrique, M.Sc., Universidade Federal de Viçosa, June, 2021. **Analysis of Strategies and Methods for Predicting Real Estate Investment Trusts**. Advisor: Fabrício A. Silva.

This work focuses on predicting the variation in the price of Real Estate Investment Trusts (REITs), an increasingly relevant issue for investors and portfolio managers. Given the complexity of this challenge and the need to make informed decisions in the real estate market, this research explores innovative modeling strategies, contrasting traditional time series methods with machine learning approaches. An innovative aspect was the enrichment of data with additional information about the underlying properties of the REITs, seeking to understand how specific characteristics influence their appreciation. The results demonstrated that, in many cases, machine learning models outperformed traditional time series analysis techniques, especially in forecasting medium-term trends. Additionally, this study investigated the feasibility of using a classifier to determine the most efficient strategy in prediction, considering the results obtained with ARIMA and XGBoost models. The analysis revealed that different funds exhibit distinct behaviors over time, and the choice of the appropriate algorithm may vary depending on the individual characteristics of each fund.

Keywords: Real Estate Investment Fund, Machine Learning, Time Series

Lista de ilustrações

Figura 1 – Estrutura da seção de Metodologia	20
Figura 2 – Modelagens utilizadas no contexto da estratégia E1. Na modelagem geral, os mesmos atributos são usados para criar modelos para todos os fundos. Na modelagem categórica, cada categoria utiliza um conjunto de atributos diferente. E na modelagem personalizada, cada fundo utiliza um conjunto de atributos diferente para criar o modelo.	35
Figura 3 – <i>Walk forward validation</i> (KINOSHITA; MENEGHETTI; BIANCHI, 2022).	36
Figura 4 – Funcionamento do classificador de modelos	37
Figura 5 – Estratégia Leave-One-Out Cross-Validation no contexto de FIIs	41
Figura 6 – Estratégias utilizadas para treinamento dos classificadores	41
Figura 7 – Intervalo de confiança dos erros para 1 mês por estratégia	44
Figura 8 – Intervalo de confiança dos erros para 6 meses por estratégia	45
Figura 9 – Intervalo de confiança dos erros para 1 mês por estratégia da categoria papel	46
Figura 10 – Intervalo de confiança dos erros para 6 meses por estratégia da categoria papel	47
Figura 11 – Intervalo de confiança dos erros para 1 mês por estratégia da categoria híbrido	48
Figura 12 – Intervalo de confiança dos erros para 6 meses por estratégia da categoria híbrido	49
Figura 13 – Intervalo de confiança dos erros para 1 mês por estratégia da categoria tijolo	50
Figura 14 – Intervalo de confiança dos erros para 6 meses por estratégia da categoria tijolo	51
Figura 15 – Intervalo de confiança dos erros para 1 mês por estratégia	55
Figura 16 – Intervalo de confiança dos erros para 1 mês por estratégia	56
Figura 17 – Boxplot da variação do “CloseChange”	58
Figura 18 – intervalo de confiança dos erros de cada estratégia para o curto prazo .	75
Figura 19 – intervalo de confiança dos erros de cada estratégia para o médio prazo .	76

Lista de tabelas

Tabela 1 – Alterações nos nomes das colunas do <i>DataFrame</i> de endereços	24
Tabela 2 – Desempenho médio dos modelos	40
Tabela 3 – Bibliotecas utilizadas e suas versões	42
Tabela 4 – Erros em geral para 1 mês por estratégia	43
Tabela 5 – Erros em geral para 6 meses por estratégia	44
Tabela 6 – Erros da categoria papel para 1 mês por estratégia	46
Tabela 7 – Erros da categoria papel para 6 meses por estratégia	47
Tabela 8 – Erros da categoria híbrido para 1 mês por estratégia	47
Tabela 9 – Erros da categoria híbrido para 6 meses por estratégia	48
Tabela 10 – Erros da categoria tijolo para 1 mês por estratégia	49
Tabela 11 – Erros da categoria tijolo para 6 meses por estratégia	50
Tabela 12 – MAE de cada categoria para 1 mês por estratégia	52
Tabela 13 – MAE de cada categoria para 6 meses por estratégia	52
Tabela 14 – Características dos modelos da estratégia E2 para o curto prazo	54
Tabela 15 – Características dos modelos da estratégia E2 para o médio prazo	54
Tabela 16 – Tabela de erros da estratégia E2 para 1 mês	55
Tabela 17 – Tabela de erros da estratégia E2 para 6 meses	56
Tabela 18 – Siglas dos atributos mais utilizados nos modelos da estratégia E1	60
Tabela 19 – Melhores modelos para categoria Papel (1 Mês)	60
Tabela 20 – Melhores modelos para categoria Papel (6 Meses)	61
Tabela 21 – Melhores modelos para categoria Híbrido (1 Mês)	61
Tabela 22 – Melhores modelos para categoria Híbrido (6 Meses)	61
Tabela 23 – Melhores modelos para categoria Tijolo (1 Mês)	62
Tabela 24 – Melhores modelos para categoria Tijolo (6 Meses)	62
Tabela 25 – Melhores modelos para cada fundo de papel individualmente no curto prazo	65
Tabela 26 – Melhores modelos para cada fundo de papel individualmente no médio prazo	67
Tabela 27 – Melhores modelos para cada fundo híbrido individualmente no curto prazo	68
Tabela 28 – Melhores modelos para cada fundo híbrido individualmente no médio prazo	68
Tabela 29 – Melhores modelos para cada fundo de tijolo individualmente no curto prazo	71
Tabela 30 – Melhores modelos para cada fundo de tijolo individualmente no médio prazo	73

Tabela 31 – Comparativa dos resultados (Curto prazo)	73
Tabela 32 – Comparativa dos resultados (Médio prazo)	74
Tabela 33 – Erros médios obtidos por cada estratégia no curto prazo	75
Tabela 34 – Erros médios obtidos por cada estratégia no médio prazo	76

Sumário

1	Introdução	12
1.1	Justificativa	13
1.2	Objetivo Geral	14
1.3	Questões de Pesquisa	15
1.4	Organização do Texto	15
2	Trabalhos Relacionados	16
3	Metodologia	20
3.1	Coleta dos Dados	20
3.1.1	FIIIs listados na bolsa de valores	21
3.1.2	Dividendos pagos ao longo do tempo	21
3.1.3	Preço de fechamento ao longo do tempo	22
3.1.4	Endereços dos imóveis dos FIIIs	23
3.1.5	Valor do índice IFIX ao longo do tempo	24
3.1.6	Índices IGP-M, Selic e IPCA ao longo do tempo	25
3.1.7	Desdobramentos de FIIIs ocorridos ao longo do tempo	26
3.2	Descrição dos dados	27
3.3	Preparação dos dados	28
3.3.1	Enriquecimento dos imóveis	30
3.4	Métodos	32
3.4.1	Estratégia 1 (E1): Um modelo por fundo	32
3.4.2	Estratégia 2 (E2): Um modelo geral para todos os fundos	33
3.4.3	Modelagem Personalizada	34
3.5	Treinamento e Validação	36
3.6	Classificador de Modelos	36
3.6.1	Extração de atributos	37
3.6.2	Treinamento do modelo	38
4	Resultados e Discussões	42
4.1	Q1: Modelos de aprendizado de máquina podem melhorar os resultados se comparados com métodos tradicionais de séries temporais, como ARIMA?	43
4.2	Q2: Separar os fundos em diferentes categorias (papel, tijolo, híbrido) ajuda a melhorar os resultados?	45
4.3	Q3: Utilizar dados enriquecidos dos imóveis dos fundos ajuda a melhorar os modelos?	52
4.4	Q4: É possível fazer uma previsão de médio prazo (variação daqui a 6 meses) com erros aceitáveis, se comparado com a previsão de curto prazo (1 mês)?	57

4.5	Q5: As redes neurais (e.g., LSTM) apresentam boas perspectivas para previsão de fundos imobiliários?	58
4.6	Q6: Criar um modelo individual para cada fundo é melhor do que um modelo geral que considera todos os fundos de uma vez?	59
4.7	Q7: Houveram padrões de atributos mais frequentes nos modelos mais eficazes?	59
4.7.1	Categoria Papel	60
4.7.2	Categoria Híbrido	61
4.7.3	Categoria Tijolo	61
4.7.4	Discussão	62
4.8	Q8: É eficaz utilizar uma modelagem personalizada para cada fundo imobiliário?	63
4.8.1	Fundos de Papel	63
4.8.1.1	Curto prazo	63
4.8.1.2	Médio prazo	65
4.8.2	Fundos de Híbridos	67
4.8.2.1	Curto prazo	67
4.8.2.2	Médio prazo	68
4.8.3	Fundos de Tijolo	69
4.8.3.1	Curto prazo	69
4.8.3.2	Médio prazo	71
4.8.4	Discussão	73
4.9	Q9: É possível criar um classificador de modelos com erros aceitáveis? . . .	74
4.9.0.1	Curto prazo	75
4.9.0.2	Médio prazo	76
5	Conclusões e Trabalhos Futuros	77
	Referências	79

1 Introdução

Os investimentos são a espinha dorsal da economia global, agindo como motores que impulsionam o crescimento, a inovação e a estabilidade financeira. Seja em mercados ascendentes ou em períodos de turbulência, os investimentos alimentam o crescimento econômico e pessoal, permitindo que tanto indivíduos quanto instituições construam riqueza ao longo do tempo. Enquanto alguns preferem a segurança dos títulos de renda fixa, outros buscam o potencial de crescimento oferecido pela renda variável, como ações e fundos de investimento imobiliário, explorando as nuances do mercado para alavancar seu capital.

De acordo com dados divulgados pela B3 (a bolsa de valores oficial do Brasil), o número de pessoas que investem em renda variável teve uma alta de 43% entre junho de 2019 e junho de 2021 (B3, 2021). Os dados demonstram o maior interesse dos brasileiros em investimentos, o que resulta na maior busca por recomendações de ativos. Segundo a ANBIMA (Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais) (ANBIMA, 2018), renda variável é todo tipo de investimento que você não consegue dimensionar os ganhos ou perdas. Alguns exemplos são: ações, fundos de investimento, ouro e câmbio.

A adoção de tecnologias avançadas, como inteligência artificial e análise de dados, tem transformado o cenário dos investimentos. A análise de grandes volumes de dados, aliada a ferramentas de análise preditiva e algoritmos de aprendizado de máquina, pode ajudar a revelar tendências de mercado e avaliar o desempenho de ativos de renda variável. Isso permite que investidores ou gestores de fundos identifiquem rapidamente as mudanças nas condições do mercado e tomem decisões embasadas em dados, resultando em uma gestão de investimento mais estratégica e fundamentada.

Na literatura e no mercado financeiro já existem modelos preditivos para diferentes ativos, como ações internacionais (KIM; JEONG; GHANI, 2014), ações nacionais (BRANDÃO, 2021), moedas digitais (HELFENSTEIN, 2022), dentre outros (FREITAS; REIS, 2022). Porém, os fundos de investimento imobiliários (FIIs) são ativos que têm crescido bastante nos últimos anos e ainda precisam de modelos que consigam otimizar a tomada de decisões, imprimindo dinamismo e eficiência às operações.

Os FIIs podem ser definidos como condomínios de investimento no Brasil - ou seja, aglomeram investimentos de diversas pessoas em torno de uma mesma finalidade - cujo objetivo é aplicar recursos em produtos ou empreendimentos imobiliários, incluindo alguns de seus derivativos. Dessa forma, os FIIs têm como propósito oferecer aos investidores um ambiente de investimento ancorado em ativos com expectativa de estabilidade de valor

no longo prazo, além de aplicar recursos em produtos capazes de produzir renda mensal regular a longo prazo (JR, 2011). É importante ressaltar que, diferentemente dos REITs encontrados em outros países, os FIIs são regulamentados de acordo com as leis brasileiras e seguem as especificidades do mercado imobiliário nacional.

O investimento em imóveis, historicamente apreciado pelos brasileiros, está passando por uma transformação com a ascensão desses fundos. Os FIIs, quando comparados à aquisição direta de imóveis físicos, têm evidenciado vantagens notáveis, tais como tributação favorável e despesas mais acessíveis (SANTOS, 2019). No entanto, mesmo diante de características constantes e conservadoras em alguns fundos de investimento, os modelos atuais ainda enfrentam dificuldades em prever com precisão qual será o rendimento desses fundos (LINS, 2020).

Motivados por expectativas de contínuo crescimento da construção civil, os FIIs se tornaram uma opção de investimento atraente, principalmente por estarem atrelados a imóveis (STEFFEN, 2015). Sua regulamentação no Brasil foi feita em 1993, com a promulgação da Lei nº 8.668/93. Nos termos da mesma, a partir de então foram instituídos os Fundos de Investimento Imobiliário, sem personalidade jurídica, caracterizados pela comunhão de recursos captados por meio do Sistema de Distribuição de Valores Mobiliários, na forma da Lei nº 6.385, de 7 de dezembro de 1976, destinados a aplicação em empreendimentos imobiliários. No entanto, foi apenas a partir de 2010 que os fundos imobiliários passaram a despertar mais atenção de investidores devido a publicação da Instrução Normativa CVM nº 472, que alterou a forma de negociação dos mesmos, uma vez que os FIIs tornaram-se negociáveis na bolsa de valores, o que ampliou significativamente o acesso dos investidores a esse tipo de investimento.

Apesar de suas atrativas vantagens, também existem pontos de impasse para o investidor, como a previsão da evolução de seus preços. Para melhorar essa previsão, além de técnicas tradicionais baseadas em séries temporais, o uso de Inteligência Artificial tem crescido bastante nos últimos anos, tendo sido agregada a tradicionais análises econômicas. Porém, ainda são poucos os trabalhos que investigam o uso destas novas ferramentas de previsão para FIIs, que poderão trazer mais previsibilidade e segurança no momento de escolha do investimento.

1.1 Justificativa

De acordo com o portal do investidor¹, um Fundo de Investimento é constituído por uma comunhão de recursos sob a forma de condomínio. Este fundo é destinado a aplicação em ativos financeiros, sendo assim uma forma de investimento coletivo, reunindo muitos investidores e resultando em investimentos conjuntos no mercado financeiro

¹ <https://www.investidor.gov.br/menu/Menu_Investidor/fundos_investimentos/introducao>

([PORTALDOINVESTIDOR, 2021](#)). Já o Fundo de Investimento Imobiliário (FII) pode ser definido como a comunhão de recursos a serem usados em aplicação de ativos no mercado imobiliário. O fundo é constituído por uma instituição financeira específica, chamada de administrador. A instituição também é responsável pela captação de recursos feita pela venda das cotas ([B3, 2021](#)).

Uma grande dificuldade para aquele que começa a investir nos FIIs é a escolha de quais fundos estarão presentes em sua carteira de investimentos, já que, com o grande número de possibilidades (atualmente são 455² no Brasil, sendo que na época da coleta dos dados para este trabalho eram 287), se torna muito difícil saber quais serão os com maior potencial de rentabilidade, e que ao mesmo tempo, quais se alinham com o perfil do indivíduo.

Assim, constata-se uma escassez de estudos que tenham empreendido uma análise aprofundada acerca da previsão da rentabilidade futura de fundos imobiliários. A ausência de uma base substancial de pesquisa nessa área destaca a singularidade deste trabalho, que emerge como um dos pioneiros a inaugurar uma discussão mais aprofundada sobre o tema.

Desta forma, esta dissertação não apenas atende à necessidade imediata de investidores que buscam orientação na escolha de fundos imobiliários para inclusão em suas carteiras de investimentos, mas também serve como um catalisador para novas discussões e investigações, enriquecendo o panorama acadêmico e prático dessa temática para futuros trabalhos acadêmicos.

1.2 Objetivo Geral

O propósito fundamental desta dissertação consiste em analisar variadas estratégias para desenvolver modelos de previsão de valor de fundos imobiliários (FIIs). O escopo de investigação abrange considerações sobre a escolha de técnicas apropriadas, métodos eficazes de treinamento de modelos, a relevância da inclusão de dados relativos aos imóveis físicos, e a eficácia das técnicas de séries temporais. Além disso, este trabalho visa estabelecer diretrizes iniciais que possam orientar pesquisadores futuros a contribuir e avançar no campo da previsão de FIIs. Neste contexto, a presente pesquisa não apenas contribui para o entendimento aprofundado do desempenho de diferentes estratégias de previsão de FIIs, mas também estabelece bases sólidas para futuras investigações, fomentando o avanço contínuo neste campo acadêmico e prático.

² <https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/fundos-de-investimentos/fii/fiis-listados/>

1.3 Questões de Pesquisa

Nesta dissertação, serão abordadas as seguintes questões de pesquisa:

- Q1: Modelos de aprendizado de máquina podem melhorar os resultados se comparados com métodos tradicionais de séries temporais, como ARIMA?
- Q2: Separar os fundos em diferentes categorias (papel, tijolo, híbrido) ajuda a melhorar os resultados?
- Q3: Utilizar dados enriquecidos dos imóveis dos fundos ajuda a melhorar os modelos?
- Q4: É possível fazer uma previsão de médio prazo (variação daqui a 6 meses) com erros aceitáveis, se comparado com a previsão de curto prazo (1 mês)?
- Q5: As redes neurais (e.g., LSTM) apresentam boas perspectivas para previsão de fundos imobiliários?
- Q6: Criar um modelo individual para cada fundo é melhor do que um modelo geral que considera todos os fundos de uma vez?
- Q7: Houve padrões de atributos mais frequentes nos modelos mais eficazes?
- Q8: É eficaz utilizar uma modelagem personalizada para cada fundo imobiliário?
- Q9: É possível criar um classificador de modelos com erros aceitáveis?

Essas questões orientarão a estrutura e a análise realizada ao longo deste estudo.

1.4 Organização do Texto

Os capítulos subsequentes serão distribuídos da seguinte maneira:

- **Capítulo 2 - Trabalhos Relacionados:** Explora trabalhos anteriores relacionados à previsão de rentabilidade/valor de ativos, com foco nos FIIs.
- **Capítulo 3 - Metodologia:** Detalha a metodologia utilizada, incluindo a coleta e preparação dos dados, descrição dos dados, métodos empregados, e estratégias de treinamento e validação.
- **Capítulo 4 - Resultados e Discussões:** Apresenta as respostas para as perguntas específicas, melhores modelos por categoria, modelagem personalizada, discussões e classificação de modelos.
- **Capítulo 5 - Conclusões e Trabalhos Futuros:** Sumariza as conclusões do estudo e sugere possíveis direções para pesquisas futuras.

2 Trabalhos Relacionados

A revisão da literatura existente revela um crescente interesse na previsão de desempenho de ativos financeiros, impulsionado pelo aumento significativo no número de investidores na renda variável, conforme indicam dados recentes da B3. A literatura atual abrange uma variedade de modelos preditivos aplicados a diferentes classes de ativos, desde ações e índices econômicos até moedas digitais. Contudo, observa-se uma lacuna significativa no que diz respeito aos FIIs, um segmento em expansão no mercado financeiro brasileiro.

No contexto de previsão de índices e variáveis macroeconômicas, (KIM; JEONG; GHANI, 2014) utilizaram um método de mineração de opiniões de texto para analisar notícias para prever altas e quedas no KOSPI (*Korea Composite Stock Price Index*). (REZENDE, 2021), destacou a utilização de modelos de linguagem para prever o movimento do IBOVESPA, principal indicador de desempenho das ações negociadas na B3. A pesquisa explorou a eficácia de modelos baseados em redes neurais na análise de notícias financeiras para a predição do mercado de ações. Em comparação, o presente estudo concentra-se principalmente em modelos estatísticos para atingir os objetivos propostos.

Ainda sobre o IBOVESPA, (MESQUITA; OLIVEIRA; PEREIRA, 2024) trabalharam com a predição desse índice utilizando redes neurais recorrentes (RNNs), especificamente GRU e LSTM, para prever diferentes horizontes temporais. Já (CARVALHO, 2019) propôs um modelo econométrico para prever o comportamento futuro do índice IFIX, principal indicador de medição do desempenho dos FIIs, utilizando três modelos distintos: PLS, ARIMA e Espaço-Estado com Filtro de Kalman. (COSTA et al., 2023) utilizaram modelos de aprendizado de máquina, como Random Forest e Multilayer Perceptron Regressor, para analisar variáveis explicativas para fenômenos econômicos, como a taxa Selic, crescimento do PIB, taxa de câmbio e inflação. Embora tenham utilizado atributos semelhantes, a abordagem adotada no presente estudo concentra-se em enriquecer a análise do valor futuro dos FIIs com alguns desses atributos, aspecto não explorado por eles.

Quanto à utilização em moedas digitais, (LEITE et al., 2023) abordaram a aplicação de *Deep Learning*, especificamente o modelo LSTM, na previsão do preço e tendência do Bitcoin, visando auxiliar traders na tomada de decisões. Este estudo, por outro lado, não se limita à predição de uma única série temporal, mas sim de centenas, o que aumenta significativamente a complexidade do cenário analisado.

Explorando o cenário das previsões de séries temporais no mercado de ações, (MARTINEZ; PEREIRA, 2020) investigou a viabilidade do Aprendizado Profundo por

Reforço (DRL) no contexto específico do *day trade* em ações. Em uma linha similar de pesquisa, (BRANDÃO, 2021) e (MACHADO; ASSIS; PEREIRA, 2020) desenvolveram frameworks para prever tendências no mercado de ações, empregando técnicas de aprendizado de máquina, incluindo o aprendizado por reforço profundo. Ambas as abordagens apresentaram resultados positivos. No entanto, este estudo contrasta com essas pesquisas, uma vez que se concentra em modelos estatísticos para prever exclusivamente o desempenho de FIIs.

No trabalho de (ANTÔNIO; LIMA; JUNIOR, 2015) buscou-se verificar se é possível obter retornos extraordinários, superiores aos oferecidos por uma carteira de mercado, com o acompanhamento das recomendações de ações emitidas por analistas do mercado de capitais brasileiro. Diversos estudos foram capazes de detectar evidências de que as recomendações dos analistas são capazes de afetar o comportamento dos preços das ações como (DIMSON; MARSH, 1984), (BRAV; LEHAVY, 2003) e (MOSHIRIAN; NG; WU, 2009). Em contrapartida, este estudo direciona sua atenção para uma outra classe de ativos de renda variável.

No que diz respeito aos fundos imobiliários, (GABRIEL; POST; ROGERS, 2015) propuseram uma categorização dos *Real Estate Investment Trusts* (REITs), denominados como os fundos imobiliários nos Estados Unidos, e dos Fundos de Investimento Imobiliário brasileiros com base em diferentes medidas de desempenho, abrangendo períodos pré, durante e pós-crise financeira de 2008. (GERTSEKOVICH et al., 2019) abordaram as condições de formação de uma carteira diversificada com uma abundância de Fundos de Investimento Imobiliários russos de diversos setores, incluindo propriedades residenciais, hipotecas e mistas. Outros autores que também abordaram a recomendação de FIIs são (LAURIA, 2020) e (SANTOS, 2023).

Também no contexto de recomendação, (ROZÁRIO, 2022) destacou a utilização de indicadores de risco e a análise da distribuição média de dividendos para a recomendação de FIIs. Em seguida, (UTANI; SILVA; FARIA, 2022) direcionaram o foco para métricas imobiliárias específicas, como Dividend Yield, taxa de vacância, desempenho comparativo com o índice IFIX, localização e liquidez dos imóveis, avaliando a estabilidade e o retorno potencial dos investimentos. Posteriormente, (BARBOZA et al., 2022) examinou a eficácia da gestão de carteiras de FIIs utilizando algoritmos, concluindo que esta abordagem supera as estratégias tradicionais de mercado e se adapta melhor às suas dinâmicas. No entanto, ao contrário desses estudos, o foco desta pesquisa recai sobre a predição do desempenho, em vez de abordar recomendações específicas de investimento.

Já com foco nas predições, (FREIRE, 2021) aplicou três redes neurais com diferentes arquiteturas, usando dados abrangentes de 20 anos para prever preços de REITs. Posteriormente, (PAULUS; LAUTENSCHLAEGER; SCHÄFERS, 2023) analisou cerca de 4 milhões de tweets, feitos ao longo de 10 anos, utilizando análise de sentimento para

prever retornos de REITs, descobrindo que o sentimento social é um indicador crucial para o mercado. Por fim, no mesmo ano, (ZHANG; LI; LIEW, 2023) concentrou-se em otimizar a precisão da previsão de preços de REITs, experimentando com vários tamanhos de amostra e parâmetros de algoritmos de aprendizado de máquina. Embora suas análises tenham fornecido insights valiosos sobre os REITs, é essencial destacar que essas investigações foram realizadas exclusivamente dentro do contexto desses instrumentos financeiros, deixando de abranger a dinâmica do mercado imobiliário brasileiro como um todo.

Também no contexto de previsões, (ZANANDREA et al., 2018) complementaram essa abordagem de segmentação, propondo o desenvolvimento de um modelo para avaliar a rentabilidade dos FIIs, analisando dados de prospectos e relatórios mensais de 82 FIIs listados na BM & FBOVESPA no período de janeiro de 2017 a outubro de 2017. O modelo utilizou análise de regressão para identificar variáveis estatisticamente significativas que impactaram na rentabilidade dos fundos. O trabalho contribuiu para o entendimento do mercado de FIIs no Brasil e sugeriu possíveis direções para futuras pesquisas, como a inclusão de variáveis macroeconômicas. A quantidade de FIIs analisados e o período de estudo foram significativamente superiores nesta dissertação. Além disso, destaca-se a diversidade de modelos e estratégias testadas, sendo o diferencial a utilização dos imóveis físicos dos fundos imobiliários para extrair características e avaliar sua relevância na predição.

Em um cenário mais recente, (JUNIOR, 2023) utilizou dados de 120 FIIs entre fevereiro de 2019 e dezembro de 2021 para desenvolver um modelo de aprendizado de máquina. Sua abordagem transformou séries temporais em imagens visuais compactas, oferecendo uma representação gráfica intuitiva do comportamento dos FIIs. Posteriormente, cada uma dessas imagens foi rotulada com as categorias “Manter”, “Comprar” ou “Vender”. Embora tenha alcançado resultados notáveis, é relevante observar que a presente pesquisa se destaca ao abordar uma variedade de estratégias e ao incorporar um aprofundamento mais significativo dos atributos, uma dimensão que não foi explorada pelo autor citado.

Assim, conforme exposto neste capítulo, o tema apresenta fundamental importância tendo em vista o aumento do interesse em investimentos, sobretudo fundos imobiliários, tanto em nível nacional quanto global. Apesar de o assunto ter ganhado cada vez mais espaço nas mídias sociais, conversas entre pessoas e na academia, ainda são inexistentes e incompletas ferramentas de investimento - especialmente automatizadas - que atraiam a atenção e confiabilidade do investidor. Com o desenvolvimento deste trabalho, visa-se aumentar a quantidade de investidores em FIIs de modo transparente e interessante para o usuário, que muitas vezes tem seus investimentos não apenas como uma fonte de renda, mas também uma área de interesse. Saber em qual investimento seu dinheiro está alocado se torna não apenas um estímulo ao investimento, mas também uma importante

ferramenta de desenvolvimento econômico.

3 Metodologia

Neste capítulo, serão delineados os procedimentos adotados para a coleta, pré-processamento e limpeza dos dados relativos aos fundos imobiliários. Além disso, serão detalhados os critérios utilizados para selecionar os modelos e algoritmos de análise preditiva, assim como para escolher as características relevantes a serem utilizadas na análise. Será apresentada uma descrição completa da validação dos resultados, incluindo a definição dos critérios utilizados para a avaliação do desempenho dos modelos. Por fim, será proposta a criação de um classificador capaz de determinar o melhor modelo para realizar a predição de um FII. A estrutura desta seção pode ser vista na Figura 1.

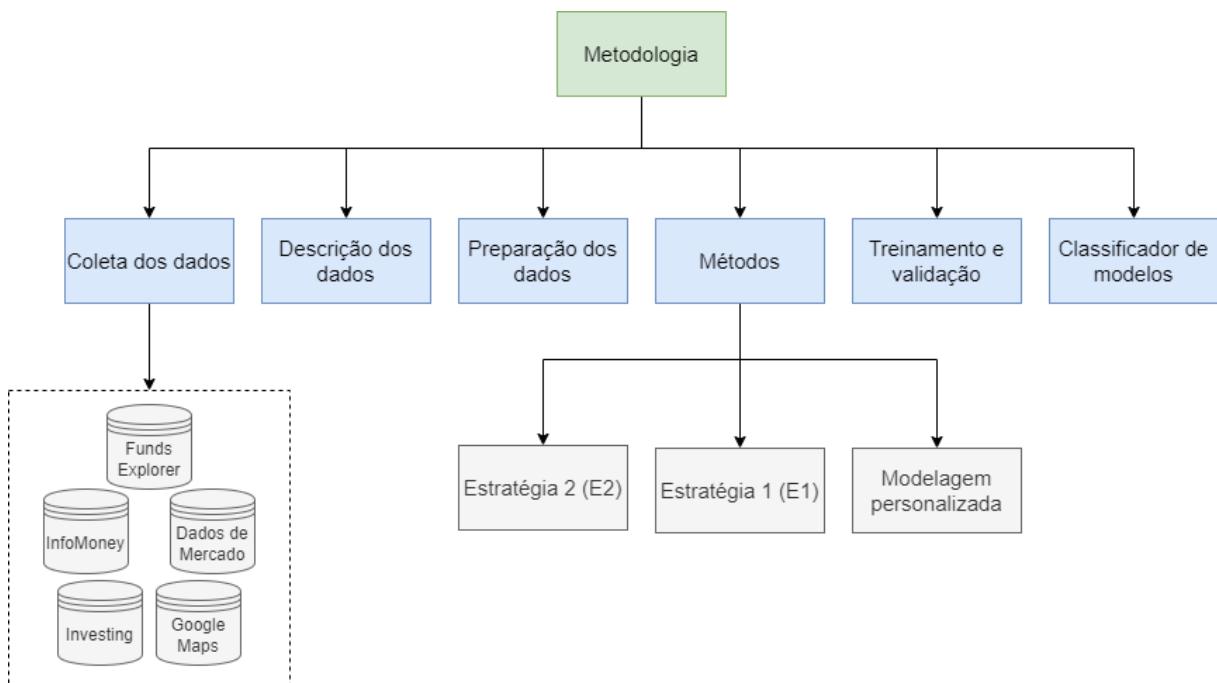


Figura 1 – Estrutura da seção de Metodologia

3.1 Coleta dos Dados

A obtenção de dados foi realizada por meio de técnicas de *web scraping* e APIs, provenientes de múltiplas fontes. Para essa coleta, optou-se por selecionar um período recente, visando incorporar um conjunto mais abrangente de FIIs na análise e evitando possíveis vieses decorrentes de dados muito antigos na predição de informações atuais. Com o intuito de assegurar uma análise abrangente e considerando que os primeiros resultados emergiram no primeiro trimestre de 2023, deliberou-se abarcar dados até dezembro de 2022, garantindo, assim, a abrangência de uma sazonalidade anual completa. Diante disso, estabeleceu-se a data de início em janeiro de 2021 e a data de fim em dezembro de

2022, proporcionando cerca de 2 anos de dados para nutrir o treinamento dos modelos.

3.1.1 FIIs listados na bolsa de valores

A extração de dados de fundos de investimento foi conduzida na aba ranking do site Funds Explorer¹. Nesse processo, uma requisição da página é feita e os dados de tabela HTML resultantes são convertidos em uma estrutura manipulável por meio da biblioteca Pandas².

O *DataFrame* resultante é manipulado para remover linhas cuja coluna 'Setor' possui valor ausente ou linhas cuja coluna 'Ticker' apresentem duplicatas. Em seguida, as colunas 'Ticker' e 'Setor' são convertidas para o tipo categórico.

Finalmente, o *DataFrame* resultante é ordenado pela coluna 'Ticker' e ajustado para incluir apenas as colunas 'Ticker', 'Setor', e 'Quantidade Ativos'.

3.1.2 Dividendos pagos ao longo do tempo

A relação de dividendos pagos por fundo ao longo do tempo foi obtida na aba *funds* do site *Funds Explorer*³. Nesse processo, para cada FII existente, foram realizados os passos abaixo para a obtenção de um *DataFrame* contendo todos os dividendos obtidos.

Para facilitar a demonstração, o fundo imobiliário **ALZR11** será utilizado como exemplo. Primeiramente, uma requisição é feita adicionando o ticker do FII ao fim da URL, da seguinte forma: <https://www.fundsexplorer.com.br/funds/ALZR11>

Em seguida, fazendo uso da biblioteca BeautifulSoup⁴, voltada para a realização de scraping em dados HTML e XML no ambiente Python, extrai-se o conteúdo da página obtido como resposta à requisição. Logo após, são empregadas as seguintes expressões regulares:

- “labels”:\[.*?\]: para obter todos os meses dos dividendos pagos
- “data”:\[.*?\]: para obter todos os valores dos dividendos pagos

Posteriormente, esses dados são convertidos do formato JSON para listas utilizáveis. As datas são então convertidas de “MM/AAAA” para “AAAA-MM-01 00:00:00”, para garantir que sejam datas válidas no padrão ISO 8601 (norma internacional para representação de data e hora).

Ao concluir o processo, um *DataFrame* Pandas é gerado, consolidando informações relativas aos dividendos do fundo, limitando-se aos dividendos posteriores à data mínima

¹ <https://www.fundsexplorer.com.br/ranking>

² <https://pandas.pydata.org/>

³ <https://www.fundsexplorer.com.br/funds/>

⁴ <https://beautiful-soup-4.readthedocs.io/en/latest/>

especificada. A fim de assegurar a consistência nos cálculos subsequentes de atributos derivados e prevenir erros de divisão por zero, foi adotada a prática de substituir valores nulos de dividendos por uma quantia simbólica de 1 centavo (R\$ 0,01).

3.1.3 Preço de fechamento ao longo do tempo

A relação do preço de fechamento por fundo ao longo do tempo foi obtida através do site InfoMoney. Nesse processo, para cada FII existente, foram realizados os passos abaixo, para a obtenção de um *DataFrame* contendo todos os preços de fechamentos.

Para facilitar a demonstração, o fundo imobiliário **ALZR11** será utilizado como exemplo. Primeiramente, uma requisição é feita adicionando os parâmetros *Ticker*, *DataInicio* e *DataFim*, da seguinte forma: <<https://fii-api.infomoney.com.br/api/v1/fii/cotacao/historico/grafico?Ticker=ALZR11&DataInicio=01-01-2021&DataFim=31-12-2022>>

A resposta da API será semelhante ao exemplo de JSON fornecido abaixo. É importante notar que, na situação real, os registros abrangerão datas até 31/12/2022:

```
1 {
2   "cnpj": "28.737.771/0001-85",
3   "ticker": "ALZR11",
4   "dataValor": [
5     {
6       "data": "04-01-2021T00:00:00",
7       "valor": 128.75
8     },
9     {
10      "data": "05-01-2021T00:00:00",
11      "valor": 128.13
12    },
13    {
14      "data": "06-01-2021T00:00:00",
15      "valor": 127.99
16    },
17  ]
18 }
```

A propriedade *dataValor* desse JSON é então convertida em um *DataFrame* e a coluna *Ticker* é criada, replicando o valor da propriedade *ticker* para todos os registros.

Por fim, a coluna *valor* é renomeada para *Close* e todos os valores R\$0,00 dessa coluna são substituídos por R\$0,01, para garantir a consistência nos cálculos subsequentes, sem impactar nos resultados.

3.1.4 Endereços dos imóveis dos FIIs

A relação de endereços dos imóveis por fundo foi também obtida pelo site InfoMoney. Nesse processo, para cada FII existente, foram realizados os passos abaixo, para a obtenção de um *DataFrame* contendo todos os endereços obtidos. A relação de endereços dos imóveis por fundo foi também obtida via API disponibilizada pelo site InfoMoney. Nesse processo, para cada FII existente, foram realizados os passos abaixo, para a obtenção de um *DataFrame* contendo todos os endereços obtidos. A relação de endereços dos imóveis por fundo foi também obtida via API disponibilizada pelo site InfoMoney. Nesse processo, para cada FII existente, foram realizados os passos abaixo, para a obtenção de um *DataFrame* contendo todos os endereços obtidos.

Para facilitar a demonstração, o fundo imobiliário **ALZR11** será utilizado como exemplo. Primeiramente, uma requisição é feita adicionando o ticker do FII ao fim da URL, da seguinte forma: <https://fi-api.infomoney.com.br/api/v1/propertie/ALZR11>

A resposta da API será semelhante ao exemplo de JSON fornecido abaixo:

```
1 {
2   "ticker": "ALZR11",
3   "property": [
4     {
5       "type": "Prédio Corporativo",
6       "name": "Imóvel Decathlon",
7       "datePurchase": null,
8       "percentagePartic": 0.0,
9       "valueGrossLeasableArea": 8750.0,
10      "state": "SP",
11      "city": "São Paulo",
12      "address": "Avenida Duquesa de Goiás 381",
13      "googleMapsLink": "",
14      "percentVacancy": 0.0,
15      "percent90DayDelinquency": 0.0,
16      "percentFii": 9.32,
17      "tenantSector": [
18        {
19          "type": "Comercial",
20          "percentPropertyRevenue": 100.0,
21          "percentTenantToFiiRevenue": 9.32
22        }
23      ]
24    }
25  ]
26 }
```

A propriedade *property* desse JSON é então convertida em um *DataFrame* e as

seguintes colunas são criadas: *Ticker*, replicando o valor da propriedade *ticker* para todos os registros, *Latitude*, recebendo *NaN* como valor padrão e *Longitude*, também recebendo *NaN* como valor padrão.

Na sequência, a coluna *tenantSector* é eliminada e os nomes das colunas do *DataFrame* são alterados de acordo com a Tabela 1.

Nome novo	Nome antigo
Tipo	type
Nome	name
DataCompra	datePurchase
ValorAreaBrutaLocavel	valueGrossLeasableArea
Estado	state
Cidade	city
Endereco	address
GoogleMapsLink	googleMapsLink
PercentualPartic	percentagePartic
PecentualVacancia	percentVacancy
PercentualInadimplencia90Dias	percent90DayDelinquency
PercentualFii	percentFii

Tabela 1 – Alterações nos nomes das colunas do *DataFrame* de endereços

3.1.5 Valor do índice IFIX ao longo do tempo

Outro dado que também foi obtido pelo site InfoMoney foi a relação do valor do IFIX ao longo do tempo.

Primeiramente, é preciso configurar o *header* da requisição da seguinte forma:

```

1 {
2   "authority":"www.infomoney.com.br",
3   "accept":"application/json, text/javascript, */*; q=0.01",
4   "accept-language":"pt-BR,pt;q=0.9,en-US;q=0.8,en;q=0.7",
5   "content-type":"application/x-www-form-urlencoded; charset=UTF-8",
6   "origin":"https://www.infomoney.com.br",
7   "referer":"https://www.infomoney.com.br/cotacoes/b3/indice/ifix/historico/"
8 }
```

Na sequência, é feita a configuração do *body* da requisição da seguinte forma:

```

1 {
2   "page":0,
3   "numberItems":99999,
4   "initialDate":"2021-01-01",
5   "finalDate":"2022-12-31",
```

```

6   "symbol": "IFIX"
7   }

```

Logo na sequência, com o *header* e o *body* já configurados, é feita uma requisição POST para a seguinte URL: <https://www.infomoney.com.br/wp-json/infomoney/v1/quotes/history>.

A resposta da API será semelhante ao exemplo de JSON fornecido abaixo. É importante notar que, na situação real, os registros abrangerão datas até 31/12/2022:

```

1  [
2    [
3      {"display": "06/01/2022", "timestamp": "1641427200"},
4      "2.760",
5      "2.757",
6      "-0,10",
7      "2.751",
8      "2.770",
9      "n/d"
10   ],
11   [
12     {"display": "07/01/2022", "timestamp": "1641513600"},
13     "2.757",
14     "2.760",
15     "0,08",
16     "2.757",
17     "2.764",
18     "n/d"
19   ],
20 ]

```

Nesta resposta, os valores subsequentes ao objeto em cada array correspondem, respectivamente, aos seguintes rótulos: preço de abertura, preço de fechamento, variação, mínimo do dia, máximo do dia e volume de negociação.

Sendo assim, para construir o *DataFrame*, é utilizado, em cada posição do array, o campo *display* para obter a data do dia e o segundo valor numérico para extrair o preço de fechamento correspondente àquela data.

3.1.6 Índices IGP-M, Selic e IPCA ao longo do tempo

Os dados de inflação e taxa de juros foram obtidos a partir do site Dados de Mercado⁵. Para cada índice (IGP-M, Selic e IPCA), foi repetido o processo abaixo.

⁵ <https://www.dadosdemercado.com.br>

Para facilitar a demonstração, o índice **IGP-M** será utilizado como exemplo. Primeiramente, uma requisição é feita adicionando o índice em letra minúscula ao fim da URL, da seguinte forma: <https://www.dadosdemercado.com.br/indices/igp-m>

Os dados de tabela HTML resultantes são então convertidos em uma estrutura manipulável por meio da biblioteca Pandas. A tabela original é composta por doze colunas, representando cada mês do ano, e cada linha corresponde a um ano específico.

O código desenvolvido percorre todos os registros dessa tabela, gerando assim um novo *DataFrame*. Cada entrada desse novo *DataFrame* contém duas colunas fundamentais: a primeira, denominada *TimeStamp*, exibe o ano em análise seguido de um hífen e o número do mês correspondente; a segunda coluna apresenta o valor do índice convertido para formato decimal.

Dessa forma, a execução do código resulta na criação de uma tabela mais condensada e estruturada, facilitando análises e visualizações temporais mais eficientes.

3.1.7 Desdobramentos de FIIs ocorridos ao longo do tempo

Devido à elevada complexidade associada à realização de *web scraping* no site **Investing.com** e à baixa quantidade de dados no intervalo de tempo em questão, optou-se por agilizar a obtenção das informações de desdobramentos por meio de uma coleta manual. Essa coleta foi efetuada a partir da URL: <https://br.investing.com/stock-split-calendar/>.

É importante destacar que um desdobramento, assim como dito pela corretora Warren Investimentos⁶, refere-se a uma prática adotada por empresas que decidem dividir suas ações em múltiplos, alterando a relação entre o número de ações e o valor de cada uma. Essa medida visa tornar as ações mais acessíveis aos investidores, ao mesmo tempo em que mantém o valor total do investimento.

Os dados coletados foram organizados em um JSON, onde cada propriedade corresponde a um ativo específico e contém informações sobre o mês/ano em que ocorreu o desdobramento e qual é a sua proporção. Eis o exemplo do JSON utilizado:

```
1 {  
2   "BTCL11": [ "2023-01", 9 ],  
3   "CYCR11": [ "2022-10", 10 ],  
4   "EQIR11": [ "2022-09", 10 ],  
5   "VGIR11": [ "2022-09", 10 ],  
6   "GALG11": [ "2022-08", 10 ],  
7   "ARRI11": [ "2022-08", 10 ],  
8   "VIUR11": [ "2022-05", 10 ],  
9   "XPSF11": [ "2022-05", 10 ],
```

⁶ <https://warren.com.br/magazine/desdobramento-de-acoes/>

```
10  "VIFI11": [ "2022-04", 10 ],
11  "GAME11": [ "2022-03", 10 ],
12  "BLMR11": [ "2021-09", 10 ],
13  "MAXR11": [ "2021-04", 19 ],
14  "RMAI11": [ "2021-03", 10 ],
15  "FISC11": [ "2020-12", 10 ],
16  "PQAG11": [ "2020-11", 10 ]
17 }
```

3.2 Descrição dos dados

Durante a etapa de coleta, foram reunidos dados de 287 FIIs, dos quais foram obtidos um total de 1.091 registros de endereços de imóveis e 3.699 registros de histórico mensal de cotação e dividendos.

A seguir, serão apresentadas informações detalhadas sobre os principais atributos existentes até o momento:

- “Sector”: indica qual é o setor de atuação do fundo imobiliário em questão. Os setores contemplados em nosso conjunto de dados incluem: Títulos e Valores Mobiliários, Lajes Corporativas, Logística, Híbrido, Outros, Shoppings, Residencial, Hotel e Hospital.
- “Close”: representa o preço de fechamento de um ativo no mês e é um dos dados mais básicos que serão utilizados na análise. É importante observar que o preço de fechamento é uma variável-chave para avaliar o desempenho de um fundo imobiliário.
- “Dividends”: representa o valor em dinheiro distribuído aos acionistas de um Fundo de Investimento Imobiliário (FII) em particular durante o mês. Os dividendos são uma forma de distribuição desses rendimentos aos investidores, sendo um dos principais atrativos desses fundos.
- “Selic”: representa a taxa Selic no mês. A Selic é a taxa básica de juros da economia brasileira e é utilizada como referência para diversos investimentos, incluindo os fundos imobiliários.
- “IPCA”: indica o valor do IPCA no mês. O IPCA é um indicador de inflação utilizado como referência para diversos investimentos financeiros.
- “IGP-M”: representa o valor do IGP-M no mês. O IGP-M é um índice que mede a inflação de preços ao produtor e é utilizado como referência para reajustes de contratos de aluguel, por exemplo.

- “IFIX”: indica o valor do IFIX no mês. O IFIX é um índice que acompanha o desempenho médio dos fundos imobiliários negociados na B3.

Além disso, outro fator relevante durante o processo de predição foi a categorização dos fundos, os quais se dividem em três segmentos distintos:

Fundos de Papel: categoria de fundos de investimento imobiliário que têm como principal objetivo investir em ativos financeiros relacionados ao mercado imobiliário.

Fundos de Tijolo: são fundos de investimento que têm como objetivo investir em empreendimentos imobiliários físicos, como edifícios comerciais, shoppings centers, galpões logísticos, entre outros.

Fundos Híbridos: são uma classe de ativos financeiros que combina tanto investimentos em imóveis físicos quanto em instrumentos financeiros.

3.3 Preparação dos dados

O primeiro passo no tratamento dos dados de fundos imobiliários é a remoção de fundos que não possuem registros em dezembro de 2022. Essa medida é importante para garantir que os dados utilizados estejam atualizados e reflitam a situação mais recente do mercado. Outro critério adotado no tratamento dos dados é a remoção de fundos imobiliários que não possuem pelo menos 12 registros de histórico, por não estarem ainda consolidados. Além disso, a escolha de um período de 12 meses para análise está fundamentada em considerações práticas, facilitando a comparação com índices de referência comuns, como o IFIX, que também são calculados com base em períodos de 12 meses.

Adicionalmente, são excluídos fundos imobiliários que apresentam lacunas nos dados, ou seja, aqueles que possuem algum intervalo ausente na janela de dados mensais. Essa prática é essencial para garantir que haja pelo menos algum valor registrado em todos os meses do intervalo, visto que a presença desses dados é crucial para uma análise precisa, e a ausência deles pode comprometer a qualidade dos resultados obtidos. Portanto, ressalta-se a importância de utilizar apenas dados completos e confiáveis.

Imediatamente a seguir, foi feito uso dos registros de desdobramentos obtidos para realizar o ajuste do valor de fechamento dos fundos. A título ilustrativo, um fundo que sofreu desdobramento na proporção de 1:10, terá todos os seus preços de fechamento, que fossem anteriores à data do referido evento, divididos por 10 a fim de prevenir a ocorrência de anomalias na variação do preço.

Na sequência, os seguintes atributos foram derivados para complementar as informações e melhorar a precisão do modelo:

- “DividendYield”: Resultado da divisão da coluna ‘Dividends’ pela coluna ‘Close’ e

multiplicando o resultado por 100.

- “DividendsChange”: Resultado da variação percentual da coluna ‘Dividends’ entre o último registro de um mês com o último registro do mês anterior.
- “CloseChange”: Resultado da variação percentual da coluna ‘Close’ entre o último registro de um mês com o último registro do mês anterior.
- “DividendYieldChange”: Resultado da variação percentual da coluna ‘DividendYield’ entre o último registro de um mês com o último registro do mês anterior.
- “DividendsChange6M”: Resultado da variação percentual da coluna ‘Dividends’ entre o último registro de um mês com o último registro de seis meses anteriores.
- “CloseChange6M”: Resultado da variação percentual da coluna ‘Close’ entre o último registro de um mês com o último registro de seis meses antes
- “DividendYieldChange6M”: Resultado da variação percentual da coluna ‘DividendYield’ o último registro de um mês com o último registro de seis meses antes.
- “SectorDividendsChangeMean”: Média do valor da coluna ‘DividendsChange’ do determinado mês de todos os FIIs de mesmo setor.
- “SectorCloseChangeMean”: Média do valor da coluna ‘CloseChange’ do determinado mês de todos os FIIs de mesmo setor.
- “SectorDividendYieldChangeMean”: Média do valor da coluna ‘DividendYieldChange’ do determinado mês de todos os FIIs de mesmo setor.

Após essa etapa, são excluídos da análise os FIIs que apresentaram variações mensais do preço de fechamento consideravelmente elevadas. A razão para essa exclusão está na instabilidade associada a fundos com flutuações intensas, tornando-os menos relevantes para a análise, pois estão mais suscetíveis a introduzir vieses ao modelo de predição. Portanto, FIIs que demonstraram uma variação mensal superior a 35% no atributo ‘CloseChange’ em qualquer mês foram removidos do conjunto de dados analisado.

Fundos imobiliários que não manifestaram variações em ‘CloseChange’ por pelo menos três meses consecutivos foram igualmente excluídos da análise. Essa decisão é fundamentada na alta probabilidade de o FII ter sido desativado ou apresentar inconsistências nos dados, uma vez que os FIIs são ativos voláteis e estão sujeitos a flutuações de forma constante.

Ao término do processo, dos 287 fundos imobiliários inicialmente identificados no período de execução da coleta, restaram 162 fundos para serem levados em consideração na análise.

3.3.1 Enriquecimento dos imóveis

Ao perceber que a maioria dos FIIs possui ativos físicos, levanta-se a possibilidade de que características específicas desses imóveis possam influenciar o preço de fechamento ou os dividendos. Diante desse cenário, realiza-se uma investigação mais aprofundada dessas características por meio de um processo de enriquecimento, com o objetivo de avaliar sua relevância nas predições em questão.

Inicialmente, procede-se com o enriquecimento dos dados de endereços por meio da obtenção das coordenadas geográficas dos imóveis. Para isso, utiliza-se a seguinte expressão regular “(?<=@)[-]*[\\d.]*,-[\\d.]*” para procurar por dois valores decimais separados por vírgula, imediatamente após o símbolo “@” na coluna que contém o link do Google Maps⁷ com a localização do imóvel, obtido na coleta dos endereços como descrito na seção 3.1.4.

Caso nenhum valor de coordenada geográfica seja obtido na etapa anterior, é realizada uma requisição para o Google Places⁸ com a inserção de todas as informações de endereço disponíveis, separadas por vírgula, após a última barra invertida da URL. Essa solicitação pode resultar na obtenção das coordenadas geográficas aproximadas do imóvel, que serão extraídas de acordo com o mesmo padrão mencionado anteriormente.

Para os casos onde ainda assim nenhuma coordenada tiver sido obtida, uma automação utilizando a ferramenta Selenium⁹ foi utilizada para tentar fazer a coleta individual das coordenadas faltantes. Para isso, a automação acessou cada uma das URLs dos endereços, aguardou o carregamento da página e após isso, aplicou a mesma expressão regular na URL carregada para a obtenção das coordenadas. Com isso, 98,53% dos endereços conseguiram ter suas coordenadas de latitude e longitude obtidas.

Em seguida, tendo em vista que certos fundos imobiliários são formados por imóveis físicos, foi possível expandir as informações disponíveis através das ferramentas MoreData (FIGUEIREDO et al., 2021) e Google Maps, disponíveis publicamente. Foram incluídas informações acerca da quantidade de estabelecimentos específicos situados num raio de 1 km das coordenadas dos imóveis, sendo: shoppings tradicionais, prédios corporativos, lajes corporativas, galpões, agências bancárias, centros logísticos, faculdades, lojas de rua, terrenos, shoppings temáticos, hotéis e hospitais. Essa etapa visou enriquecer os dados de um fundo com informações sobre as regiões em que os seus imóveis físicos estão localizados.

Com os dados aprimorados pelo Google Maps, foram coletadas informações sobre os estabelecimentos próximos a cada imóvel, incluindo não apenas a quantidade de cada tipo de estabelecimento, mas também suas avaliações médias no Google Places. Esses dados foram processados para fornecer os seguintes atributos:

⁷ <<https://www.google.com/maps>>

⁸ <<https://www.google.com/maps/place/>>

⁹ <<https://www.selenium.dev/>>

- “RestauranteTotal”: indica o número total de restaurantes que estão dentro do raio de 1 km do imóvel.
- “FarmáciaTotal”: indica o número total de farmácias que estão dentro do raio de 1 km do imóvel.
- “AcademiaTotal”: indica o número total de academias que estão dentro do raio de 1 km do imóvel.
- “EscolaTotal”: indica o número total de escolas que estão dentro do raio de 1 km do imóvel.
- “SupermercadoTotal”: indica o número total de supermercados que estão dentro do raio de 1 km do imóvel.
- “PostoPolicialTotal”: indica o número total de postos policiais que estão dentro do raio de 1 km do imóvel.
- “HotelTotal”: indica o número total de hotéis que estão dentro do raio de 1 km do imóvel.
- “EstaçãoDeMetrôTotal”: indica o número total de estações de metrô que estão dentro do raio de 1 km do imóvel.
- “RestauranteAverageRating”: indica a nota média no google places dos restaurantes que estão dentro do raio de 1 km do imóvel.
- “FarmáciaAverageRating”: indica a nota média no google places das farmácias que estão dentro do raio de 1 km do imóvel.
- “AcademiaAverageRating”: indica a nota média no google places das academias que estão dentro do raio de 1 km do imóvel.
- “EscolaAverageRating”: indica a nota média no google places das escolas que estão dentro do raio de 1 km do imóvel.
- “SupermercadoAverageRating”: indica a nota média no google places dos supermercados que estão dentro do raio de 1 km do imóvel.
- “PostoPolicialAverageRating”: indica a nota média no google places dos postos policiais que estão dentro do raio de 1 km do imóvel.
- “HotelAverageRating”: indica a nota média no google places dos hotéis que estão dentro do raio de 1 km do imóvel.
- “EstaçãoDeMetrôAverageRating”: indica a nota média no google places das estações de metrô que estão dentro do raio de 1 km do imóvel.

3.4 Métodos

Para a geração dos modelos de previsão de preços dos FIIs, foram utilizados três algoritmos distintos: ARIMA, XGBoost e *Long Short Term Memory* (LSTM). A seleção do ARIMA se baseou em sua ampla aceitação como uma técnica consagrada para previsão de séries temporais, uma escolha também adotada por (CARVALHO, 2019) ao buscar prever o índice IFIX. Apesar de existirem técnicas mais recentes, o ARIMA atende ao objetivo deste trabalho de avaliar o uso de um método de série temporal. O XGBoost foi escolhido por ser um algoritmo de aprendizado de máquina de destaque nos últimos anos, tendo alcançado bons resultados em diferentes problemas, assim como pode ser notado no trabalho de (SANTOS, 2022), que testou diversos algoritmos diferentes para a predição de ações. Ademais, a opção pela LSTM como técnica de redes neurais se justifica pela sua extensa utilização em questões associadas a investimentos, como destacado por (MESQUITA; OLIVEIRA; PEREIRA, 2020) ao buscar prever séries temporais no mercado de ações brasileiro.

Cada um desses algoritmos foi aplicado para gerar modelos com o objetivo de prever as variações dos preços dos fundos imobiliários com um horizonte temporal de 1 mês (curto prazo) e 6 meses (médio prazo). Essa escolha visa validar a aptidão do modelo em capturar tanto flutuações imediatas, ligadas a eventos específicos, quanto tendências de médio a longo prazo, associadas a fatores macroeconômicos.

Essa abordagem abrangente, combinada com diferentes conjuntos de atributos, proporciona modelos que permitem uma análise precisa e diversificada do comportamento dos preços dos FIIs. Ao oferecer previsões para múltiplos prazos, a pesquisa torna-se relevante para diversas estratégias de investimento, demonstrando a capacidade adaptativa do modelo diante de diferentes horizontes temporais.

Foram avaliadas diferentes combinações de atributos como entradas para os algoritmos. Essas combinações e a utilização de diferentes horizontes temporais permitiram obter uma ampla gama de modelos que possibilitam analisar e prever o comportamento dos preços dos fundos imobiliários com diferentes níveis de precisão e abrangência temporal.

Foram utilizadas algumas estratégias para a criação dos modelos, descritas a seguir.

3.4.1 Estratégia 1 (E1): Um modelo por fundo

A **Estratégia 1 (E1)** busca criar diferentes modelos, usando o histórico de dados de cada fundo, para realizar previsões personalizadas para cada um. Inicialmente, são selecionados os atributos a serem utilizados e, em seguida, para cada FII, é construído um modelo exclusivo que incorpora os atributos escolhidos e seus dados históricos específicos. A hipótese dessa estratégia é que cada fundo tem um comportamento distinto, e

portanto utilizar apenas seus dados passados para prever seu dado futuro, poderia acabar acarretando numa melhor previsão.

Além disso, comparou-se a eficácia de técnicas de aprendizado de máquina (XGBoost), com um método tradicional de série temporal (ARIMA), em dois cenários distintos: curto (1 mês) e médio (6 meses) prazo. O primeiro cenário objetivou prever a variação do preço de fechamento de um ativo no próximo mês, enquanto o segundo previu a variação desse mesmo preço daqui a seis meses.

O modelo XGBoost utilizado por esta estratégia foi configurado com uma seleção específica de hiperparâmetros: o número de estimadores (`n_estimators`), fixado em 4000, o número de rodadas de parada precoce (`early_stopping_rounds`), estabelecido como 100, a taxa de aprendizado (`learning_rate`), ajustada para 0.1, a profundidade máxima da árvore (`max_depth`), definida como 6 para controlar a complexidade do modelo, o peso mínimo em uma folha (`min_child_weight`), mantido em 1, e a fração de colunas a serem amostradas para construção de cada árvore (`colsample_bytree`), estabelecida em 0.4. Os valores restantes dos hiperparâmetros foram mantidos como padrão.

Já o modelo ARIMA foi utilizado em conjunto com a extensão SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with eXogenous regressors) para realizar previsões em séries temporais. O modelo SARIMAX foi configurado com uma ordem de (1,1,1), representando as ordens da parte autoregressiva (AR), da parte de diferenciação (I) e da parte de média móvel (MA), respectivamente. Além disso, foi estabelecido um limite máximo de iterações em 1000 para o processo de otimização.

Por fim, foi avaliado se a previsão feita separadamente por categoria de fundo (Papel, Tijolo ou Híbrido) apresentava resultados superiores àquela realizada de maneira geral, com os fundos de todas as categorias juntos.

3.4.2 Estratégia 2 (E2): Um modelo geral para todos os fundos

A **Estratégia 2 (E2)** visa criar um modelo que se baseia nos dados de histórico de todos fundo para realizar previsões. Nesse sentido, foram criados modelos com diferentes combinações de atributos, escolhidos empiricamente. A hipótese dessa estratégia é que a evolução dos preços dos fundos possuem comportamentos similares, e um único modelo seria capaz de representar esse comportamento.

Ademais, utilizou-se de duas técnicas de aprendizado, sendo elas o XGBoost e LSTM. Vale ressaltar que o ARIMA não é utilizado nesta estratégia, pois os alvos são vários valores, um para cada fundo. Os modelos foram construídos como na estratégia anterior para prever em um horizonte temporal de 1 e 6 meses.

O modelo XGBoost utilizado por essa estratégia foi configurado com diversos hiperparâmetros específicos: o número de estimadores (`n_estimators`), ajustado para 5000,

o número de rodadas de parada precoce (`early_stopping_rounds`), definido como 10, a profundidade máxima da árvore (`max_depth`), limitada a 3 para controlar a complexidade do modelo, o número de núcleos de CPU utilizados (`n_jobs`), configurado para usar todos os núcleos disponíveis (-1), a função de objetivo (`objective`), estabelecida como `'reg:squarederror'` para otimização do erro quadrático médio, a taxa de aprendizado (`learning_rate`), definida em 0.01, o peso mínimo em uma folha (`min_child_weight`), mantido em 1, a fração de instâncias de treinamento a serem amostradas aleatoriamente (`subsample`), estabelecida em 0.8, e o parâmetro de regularização L2 (`reg_lambda`), definido como 0. Os valores restantes dos hiperparâmetros foram mantidos como padrão.

Já o modelo LSTM foi configurado com três camadas LSTM, cada uma com um número específico de unidades: 60, 48 e 24, respectivamente. A inclusão de múltiplas camadas LSTM permite ao modelo aprender representações mais complexas e capturar relações temporais de longo prazo nos dados. Adicionalmente, uma camada densa com um neurônio foi empregada para produzir a saída final do modelo. A função de perda adotada foi o erro médio quadrático, otimizado através do algoritmo Adam com uma taxa de aprendizado de 0.001. Durante o treinamento, foram executadas 100 épocas, sendo empregado um callback de parada antecipada para mitigar o risco de overfitting, monitorando o RMSE e interrompendo o treinamento se não houvesse melhora após 40 épocas.

Inicialmente, um conjunto de modelos foi criado com várias combinações possíveis de atributos (C1). Para avaliar a importância dos atributos físicos dos imóveis, um novo conjunto de modelos foi criado, adicionando essas informações aos modelos anteriores (C2). Por fim, os modelos foram treinados e avaliados separando a base de dados por categoria (Papel, Tijolo e Híbrido).

3.4.3 Modelagem Personalizada

Na estratégia E1 usando o algoritmo XGBoost, uma mesma combinação de atributos foi utilizada para prever todos os fundos imobiliários de um dado agrupamento de FIIs. Já na modelagem personalizada discutida nesta seção, a meta foi identificar o melhor modelo (no caso, melhor conjunto de atributos) possível para cada fundo imobiliário, considerando cada combinação de atributos individualmente. Ao final, a média dos RMSEs obtidos foi analisada para avaliar a relevância dessa estratégia.

Com o intuito de facilitar a compreensão, a figura 2 foi desenvolvida para ilustrar as diferenças entre a Modelagem Personalizada e as outras abordagens empregadas na estratégia E1, considerando um contexto de exemplo envolvendo seis FIIs arbitrários.

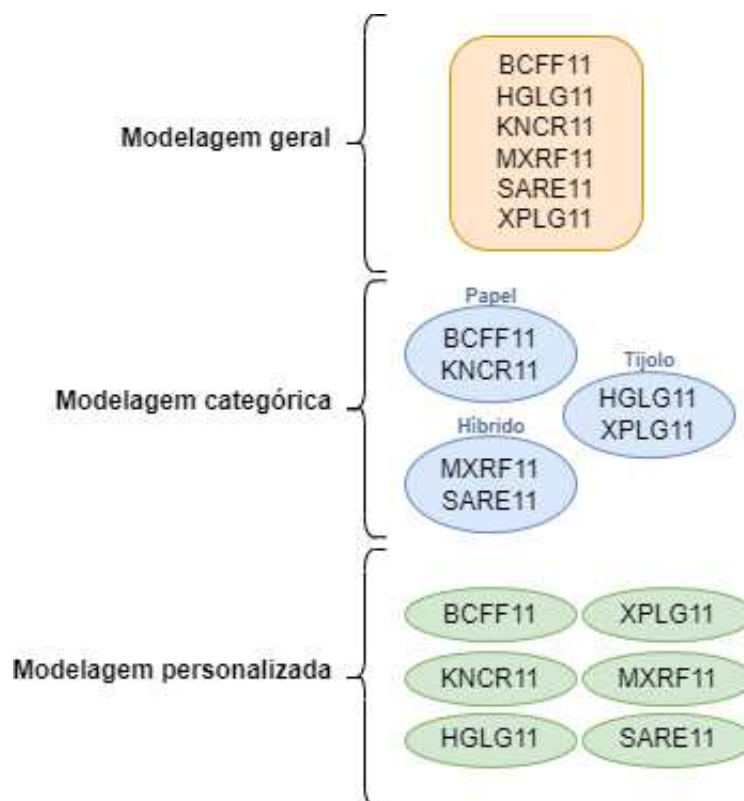


Figura 2 – Modelagens utilizadas no contexto da estratégia E1. Na modelagem geral, os mesmos atributos são usados para criar modelos para todos os fundos. Na modelagem categórica, cada categoria utiliza um conjunto de atributos diferente. E na modelagem personalizada, cada fundo utiliza um conjunto de atributos diferente para criar o modelo.

Em síntese, na modelagem geral, identificaram-se os atributos que compunham o modelo apresentando os melhores resultados médios para todos os fundos analisados. Em outras palavras, uma única combinação de atributos foi empregada em todos os modelos individuais. Na modelagem categórica, realizou-se o mesmo procedimento, entretanto, os FIIs foram divididos em três grupos distintos, resultando na seleção de três combinações diferentes de atributos, uma para cada categoria. Por fim, na modelagem personalizada, uma combinação exclusiva de atributos foi atribuída a cada um dos fundos imobiliários, viabilizando a obtenção do modelo resultante da melhor combinação possível para cada fundo.

Com esse propósito, foram conduzidas 848 previsões, abrangendo uma ampla variedade de combinações de atributos, visando identificar a configuração mais eficaz para cada fundo. Os resultados dessas análises serão minuciosamente examinados na seção 4.8, onde serão discutidos em detalhes.

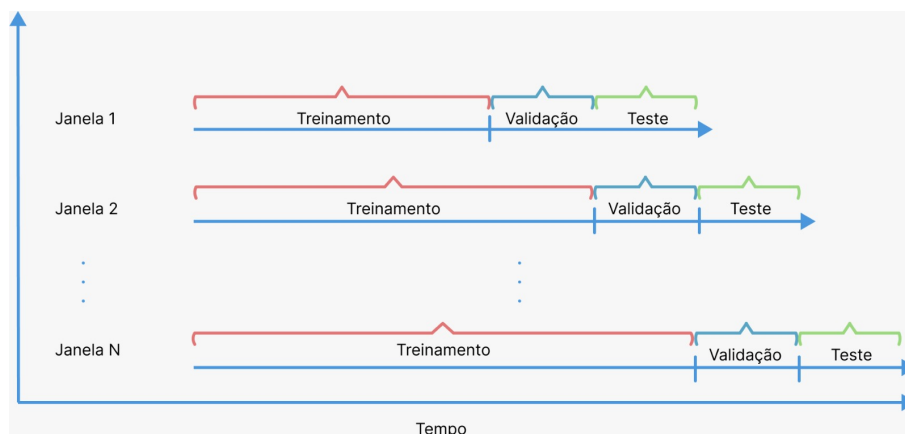


Figura 3 – *Walk forward validation* (KINOSHITA; MENEGHETTI; BIANCHI, 2022).

3.5 Treinamento e Validação

Para a realização do treinamento e validação dos modelos criados nas seções 3.4.1, 3.4.2 e 3.4.3, adotou-se a metodologia de *Walk Forward*, ilustrada na figura 3, na qual o modelo é treinado em uma janela de dados no tempo, validado em outra e testado numa janela menor imediatamente subsequente à de validação. Inicialmente, os dados de treino compõem 60% da base, os dados de validação são compostos pelos dados situados no mês seguinte aos dados de treino e os dados de teste no mês seguinte aos dados de validação. Após a primeira janela, os dados de validação são incorporados aos dados de treino. Para a validação, são utilizados os dados de teste da janela passada. Por fim, para o subgrupo de teste são utilizados os dados do mês subsequente ao mês de teste da janela passada. Com o decorrer dos experimentos, as janelas deslizam cronologicamente até o fim da base de dados.

Foram utilizados as métricas de erro absoluto, raiz do erro quadrático médio (RMSE) e erro absoluto médio (MAE) para avaliar as diferentes configurações de treinamento. Vale ressaltar que não foram feitos ajustes finos nos hiper-parâmetros dos algoritmos, o que é uma limitação deste trabalho deixada como trabalhos futuros.

Para além da metodologia adotada, vale destacar que nenhum dos modelos submetidos ao procedimento de *Walk Forward* passou por uma otimização dos hiperparâmetros. Essa decisão foi motivada pela consideração da vasta quantidade de modelos gerados. Diante dessa ampla gama de possibilidades, o ajuste fino individualizado para cada modelo seria uma tarefa extremamente árdua e demandaria um esforço considerável.

3.6 Classificador de Modelos

A criação de um modelo de classificação foi explorada para discernir a estratégia mais eficaz na previsão do preço de fechamento de cada FII de forma personalizada. Esta abordagem metodológica visa identificar o modelo que produzirá a previsão mais precisa,

minimizando o erro, entre as técnicas ARIMA e XGBoost, utilizando os dados obtidos por meio da modelagem personalizada, detalhada na seção 3.4.3.

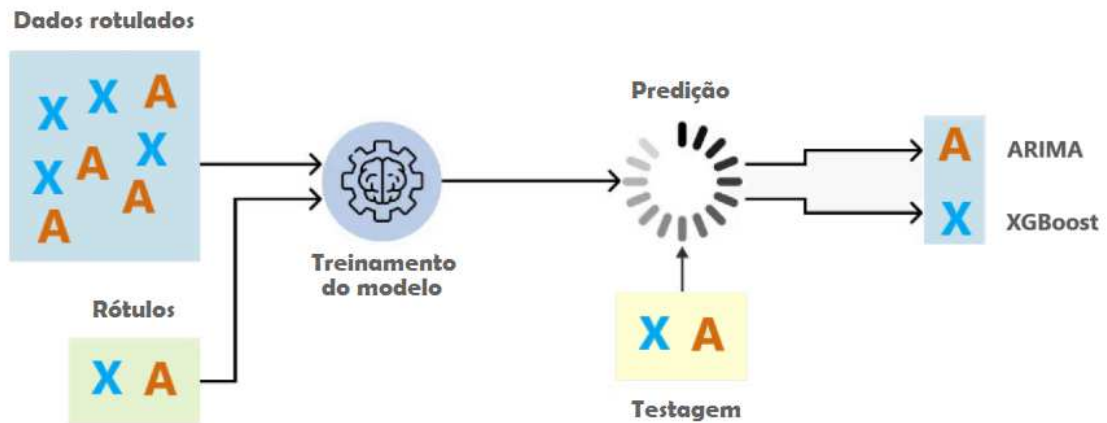


Figura 4 – Funcionamento do classificador de modelos

3.6.1 Extração de atributos

O primeiro passo realizado foi a extração de características relevantes que pudessem ser utilizadas para treinar o classificador proposto. Para esse fim, foram selecionadas as colunas fortemente relacionadas aos fundos imobiliários individualmente: “CloseChange”, “DividendsChange”, “DividendYieldChange”, “Dividends”, “Close” e “DividendYield”. Em seguida, para cada uma dessas colunas, uma série de cálculos e extrações de características foram realizados. As características extraídas incluem:

- **Variância dos dados:** A variância é uma medida de dispersão que indica o quão distantes os valores de uma série estão da média. Quanto maior a variância, maior é a dispersão dos valores em relação à média, indicando possivelmente maior volatilidade nos preços.
- **Transformada de Fourier:** A transformada de Fourier é aplicada aos dados para análise da frequência dos padrões presentes. O coeficiente de fractalidade é calculado como a razão entre a média dos quadrados dos componentes de frequência (exceto o componente de frequência zero) e a média total dos quadrados. Esse coeficiente fornece *insights* sobre a complexidade e auto-similaridade dos dados.
- **Teste de Dickey-Fuller Aumentado (ADF):** O teste de Dickey-Fuller Aumentado (ADF) é utilizado para avaliar a estacionariedade dos dados e identificar tendências. O resultado inclui estatísticas como o valor do teste, o valor-p, o número de *lags*, o número de observações, e valores críticos para diferentes níveis de significância (1%, 5%, 10%).

- **Estatísticas Descritivas:** Estatísticas descritivas tradicionais, como média, desvio padrão, mínimo, percentis de 25%, mediana, percentis de 75% e máximo, são calculadas para oferecer uma visão geral das características numéricas da série temporal.
- **Autocorrelação:** A autocorrelação é calculada para verificar a correlação entre os valores passados e presentes da série temporal, auxiliando na identificação de possíveis padrões temporais na evolução dos preços.
- **Média Móvel de Janela 3:** A média móvel de janela 3 é obtida calculando a média dos valores em uma janela deslizante de tamanho 3, permitindo a análise de tendências de curto prazo na série temporal.
- **Volatilidade (Desvio Padrão) de Janela 5:** A volatilidade, medida pelo desvio padrão, é calculada em uma janela deslizante de tamanho 5, fornecendo uma medida da variação dos preços e ajudando a capturar mudanças na volatilidade ao longo do tempo.

Essas características foram selecionadas com base na sua relevância teórica e potencial para contribuir com a previsão do preço de fechamento dos fundos imobiliários. A extração dessas características proporciona uma representação quantitativa abrangente dos dados, capturando diferentes aspectos e propriedades que podem ser úteis para o treinamento do modelo classificador. Tais características fornecem informações essenciais sobre o comportamento passado dos fundos imobiliários, permitindo ao modelo aprender padrões e realizar previsões mais precisas e eficientes.

3.6.2 Treinamento do modelo

A variável alvo da predição, denominada “arima”, foi escolhida como o atributo alvo para o treinamento do modelo. Esse atributo possui o valor *True* quando o modelo ARIMA obteve o menor RMSE para o respectivo fundo imobiliário e *False* caso contrário.

Para o treinamento do modelo, foram selecionados 14 diferentes algoritmos de classificação: “MLPClassifier”, “KNeighborsClassifier”, “SVC”, “GaussianNB”, “DecisionTreeClassifier”, “GradientBoostingClassifier”, “LogisticRegression”, “RandomForestClassifier”, “AdaBoostClassifier”, “GradientBoostingRegressor”, “XGBClassifier”, “LGBMClassifier”, “ExtraTreesClassifier” e “BaggingClassifier”. Com o objetivo de facilitar a avaliação de desempenho dos modelos, utilizou-se o MLFlow para armazenar as métricas de todos os modelos treinados.

Para realizar o treinamento do modelo, foi empregada uma abordagem de validação cruzada de tamanho 10. A partir dos resultados obtidos pela validação cruzada, foram obtidas as seguintes métricas de avaliação de desempenho:

- Acurácia média.
- Valor médio de verdadeiros negativos (VN) (instâncias que foram corretamente classificadas como negativas pelo modelo).
- Valor médio de verdadeiros positivos (VP) (instâncias que foram corretamente classificadas como positivas pelo modelo).
- Valor médio de falsos positivos (FP) (instâncias que foram erroneamente classificadas como positivas pelo modelo, quando na verdade eram negativas).
- Valor médio de falsos negativos (FN) (instâncias que foram erroneamente classificadas como negativas pelo modelo, quando na verdade eram positivas).
- Valor médio de predições corretas da matriz de confusão (verdadeiros negativos e verdadeiros positivos).
- Valor médio de predições incorretas da matriz de confusão (falsos negativos e falsos positivos).

Essas métricas proporcionaram uma avaliação quantitativa do desempenho dos modelos de classificação no contexto da predição do valor do atributo “arima”, permitindo uma comparação entre os diferentes modelos utilizados e fornecendo *insights* valiosos sobre a eficácia de cada modelo.

Após o treinamento de todos os modelos, devido aos valores médios da matriz de confusão, foi possível observar que a maioria dos modelos apresentava um viés em direção à classificação *False* para o atributo “arima”, uma vez que a 72,10% dos dados de treino estavam representados nessa categoria. Isso resultou em uma valorização significativa dos modelos capazes de equilibrar a acurácia geral com o número de acertos na categoria *True*, conforme indicado pelo valor médio de verdadeiros positivos.

Dessa forma, para a seleção dos modelos com maior potencial, aplicou-se o seguinte critério de filtro: uma acurácia média superior a 70% e um número médio de verdadeiros positivos superior a 6 durante a validação. Esses parâmetros foram estabelecidos com base em uma análise detalhada dos resultados, evidenciando que modelos que atendiam a esses critérios demonstraram consistentemente um desempenho robusto. A escolha desses limiares foi orientada pela necessidade de garantir não apenas uma boa acurácia média, mas também uma capacidade significativa de identificar corretamente instâncias positivas. Com base nesse filtro, foram mantidos apenas os seguintes modelos: “AdaBoostClassifier”, “LGBMClassifier” e “XGBClassifier”.

Após a seleção dos modelos com maior potencial, foram conduzidos procedimentos de otimização dos hiperparâmetros para cada um deles. Ao final dessa etapa, dois modelos foram escolhidos para prosseguir no estudo:

- Modelo 1 (M1): `XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1, colsample_bytree=1, gamma=0.4, learning_rate=0.1, max_delta_step=0, max_depth=3, min_child_weight=1, missing=None, n_estimators=50, n_jobs=1, nthread=None, objective='binary:logistic', random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None, silent=True, subsample=1)` utilizando os atributos extraídos: “CloseChange”, “DividendsChange”, “DividendYieldChange”, “Dividends”, “Close”, “DividendYield”
- Modelo 2 (M2): `XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1, colsample_bytree=1, gamma=0.4, learning_rate=0.1, max_delta_step=0, max_depth=3, min_child_weight=3, missing=None, n_estimators=50, n_jobs=1, nthread=None, objective='binary:logistic', random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None, silent=True, subsample=1)` utilizando os atributos extraídos: “CloseChange”, “DividendYieldChange”, “DividendYield”

O modelo M1 foi o que apresentou a mais alta acurácia média. Já o modelo M2 foi o que obteve o maior valor médio de verdadeiros positivos, entre os modelos que possuíam a maior acurácia. O desempenho médio desses modelos é apresentado na tabela 24, onde “Acc” representa a acurácia média, “VN” os verdadeiros negativos, “VP” os verdadeiros positivos, “FP” os falsos positivos e “FN” os falsos negativos:

Modelo	Acc	VN	VP	FP	FN
M1	0,73	40,80	5,30	11,80	6,10
M2	0,73	39,00	7,10	10,40	7,50

Tabela 2 – Desempenho médio dos modelos

Para garantir uma avaliação imparcial do modelo de classificação, foi adotada a estratégia de *Leave-One-Out Cross-Validation* (LOOCV). Essa técnica é amplamente reconhecida no campo de aprendizado de máquina e ciência de dados por sua eficácia em avaliar o desempenho de modelos estatísticos quando há uma quantidade limitada de dados disponíveis. Um aspecto crucial da LOOCV é a utilização completa dos dados disponíveis durante o treinamento do modelo, minimizando assim o viés e otimizando a generalização.

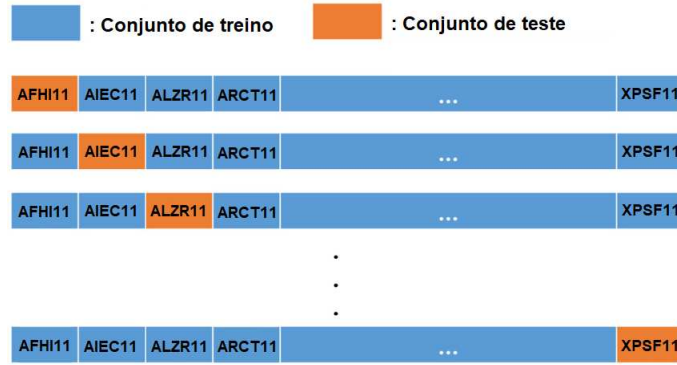


Figura 5 – Estratégia Leave-One-Out Cross-Validation no contexto de FIIs

Diferentemente da validação cruzada tradicional, onde os dados são geralmente divididos em conjuntos de treinamento e teste, na LOOCV, o modelo é treinado repetidamente utilizando todos os dados de FIIs, exceto um, que é mantido de fora em cada iteração. Essa estratégia é aplicada a todos os Fundos Imobiliários disponíveis. Tal abordagem garante que o modelo seja testado em conjuntos de dados nunca antes vistos durante o treinamento, proporcionando assim uma avaliação mais robusta do seu desempenho e capacidade de generalização.

Além disso, é importante ressaltar que os resultados foram divididos em duas estratégias. Na primeira estratégia (S1), foram utilizados os dados mensais e semestrais para prever ambas as sazonalidades. Na segunda estratégia (S2), apenas os dados mensais foram utilizados para prever a variação mensal, enquanto somente os dados semestrais foram utilizados para prever a variação semestral.

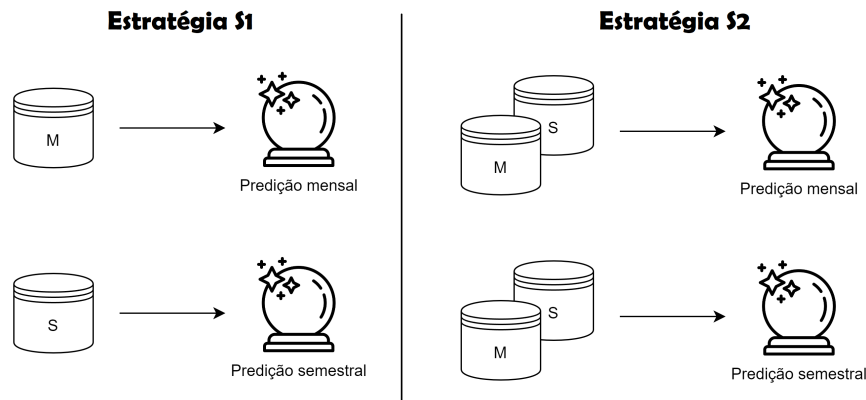


Figura 6 – Estratégias utilizadas para treinamento dos classificadores

4 Resultados e Discussões

Este capítulo apresenta os resultados visando responder às perguntas elaboradas na seção 1.3, usando como base os melhores modelos encontrados com diferentes combinações de atributos. A fim de fornecer um contexto transparente e garantir a reprodutibilidade, foi detalhado o ambiente de experimentação utilizado.

Os experimentos foram conduzidos em uma máquina equipada com um processador AMD Ryzen 5 2600 Six-Core 3.40 GHz, 16GB de memória RAM e uma placa de vídeo RX 580 GDDR5 1750 MHz com 4GB de VRAM. O Google Colab com Python 3 e CPU foi empregado em determinadas etapas do processo. Todas as análises foram realizadas utilizando as bibliotecas informadas na tabela 3.

Para rastreamento e organização dos experimentos, o MLflow¹ 2.6.0 em nuvem foi utilizado, hospedado pelo DagsHub², onde foram registradas informações detalhadas sobre cada configuração de modelo, hiperparâmetros e métricas associadas.

Biblioteca	Versão
mlflow	2.9.2
matplotlib	3.7.1
mysqlclient	2.2.1
numpy	1.23.5
pandas	1.5.3
pmdarima	2.0.4
psycopg2-binary	2.8.6
pyngrok	7.0.5
pystan	3.8.0
regex	2023.6.3
requests	2.31.0
seaborn	0.12.2
sqlalchemy	2.0.23
statsmodels	0.14.1
unidecode	1.3.7
xgboost	2.0.3

Tabela 3 – Bibliotecas utilizadas e suas versões

¹ <<https://mlflow.org/>>

² <<https://dagshub.com/>>

4.1 Q1: Modelos de aprendizado de máquina podem melhorar os resultados se comparados com métodos tradicionais de séries temporais, como ARIMA?

A pertinência dessa questão reside na revelação potencial da importância dos atributos utilizados, uma vez que o ARIMA não leva em consideração variáveis adicionais. Caso os modelos de aprendizado de máquina superem o ARIMA, isso sugere a relevância dos atributos, incentivando uma investigação mais profunda para otimizar a precisão dos modelos.

Para isso, foram realizados testes extensivos, explorando diversos modelos e atributos. A análise englobou tanto o curto (1 mês) quanto o médio prazo (6 meses), proporcionando uma compreensão abrangente das complexidades envolvidas.

Na análise de curto prazo, os resultados apresentados na tabela 4 revelam que o ARIMA e o XGboost (para a estratégia E2) apresentaram desempenho superior em métricas diferentes. Ao analisar os intervalos de confiança na figura 7, não é possível afirmar com certeza qual método é superior, dada a sobreposição dos intervalos. No entanto, ao considerar 95% dos erros, o ARIMA demonstra uma média menor, mas uma variabilidade significativamente maior em comparação ao E2 - XGboost.

Modelo	RMSE	MAE
E1 - Xgboost	0.064434	0.045758
E1 - Arima	0.056331	0.039462
E2 - Xgboost	0.054430	0.041151
E2 - LSTM	0.060005	0.042082

Tabela 4 – Erros em geral para 1 mês por estratégia

Já na análise de médio prazo, conforme apresentado na tabela 5, o XGboost da estratégia E2 supera os outros modelos em ambas as métricas. A figura 8 confirma essa superioridade, pois não há sobreposição nos intervalos de confiança em comparação com ARIMA e LSTM. Embora haja uma pequena sobreposição com E1 - Xgboost, o menor valor médio favorece a preferência por E2 - XGboost.

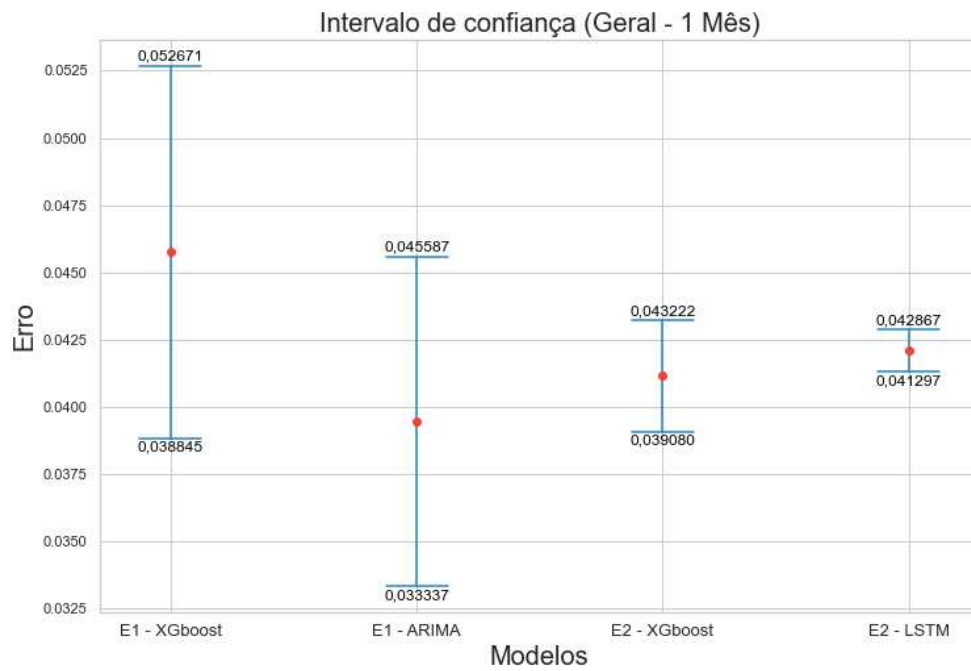


Figura 7 – Intervalo de confiança dos erros para 1 mês por estratégia

Modelo	RMSE	MAE
E1 - Xgboost	0.133602	0.080815
E1 - Arima	0.150348	0.104505
E2 - Xgboost	0.082978	0.063845
E2 - LSTM	0.097009	0.073377

Tabela 5 – Erros em geral para 6 meses por estratégia

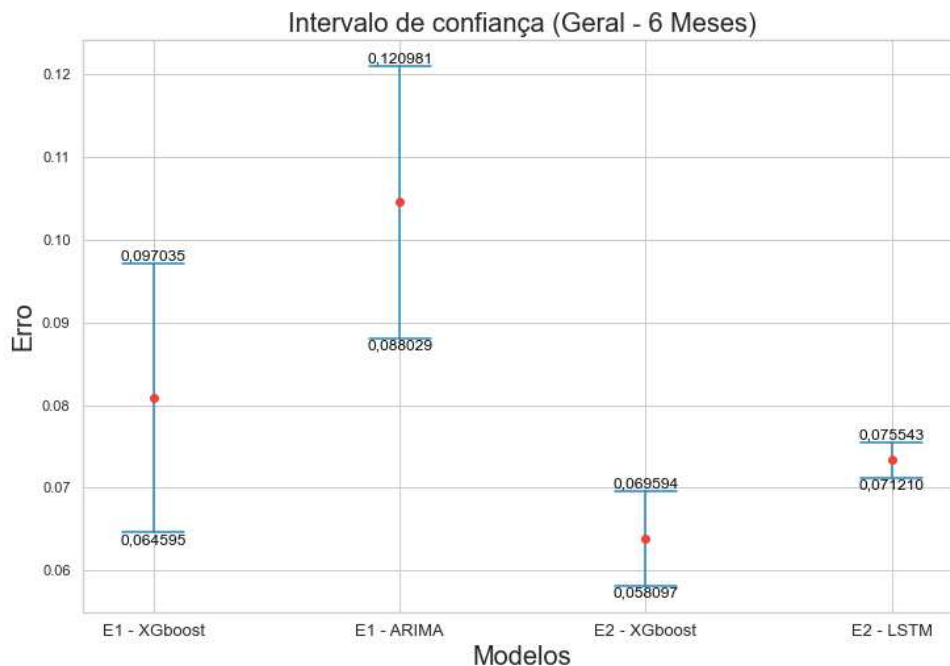


Figura 8 – Intervalo de confiança dos erros para 6 meses por estratégia

No curto prazo, o ARIMA consegue competir com os modelos de aprendizado de máquina, muito provavelmente devido à capacidade desse método em capturar padrões sazonais e comportamentos temporais mais imediatos. Sua habilidade em lidar eficazmente com a volatilidade de curto prazo pode explicar seu desempenho comparativamente forte nas previsões para um período de 1 mês.

No contexto de médio prazo, a superioridade da estratégia E2 utilizando aprendizado de máquina, especificamente o XGBoost, pode ser atribuída à sua capacidade de capturar relações não-lineares complexas e ajustar-se a padrões mais sutis e de longo alcance. Enquanto o ARIMA pode enfrentar desafios ao lidar com mudanças graduais ao longo do tempo, o XGBoost destaca-se na adaptação a essas nuances, proporcionando uma previsão mais precisa em um horizonte de 6 meses.

4.2 Q2: Separar os fundos em diferentes categorias (papel, tijolo, híbrido) ajuda a melhorar os resultados?

A importância dessa investigação surge da diversidade nas características dos FIIs em diferentes categorias. Ao considerar treinamentos de nicho por categoria, utiliza-se dados mais homogêneos durante o processo de treinamento. Além disso, a possibilidade de otimização dos atributos é explorada, assumindo que fundos de uma mesma categoria compartilham relações fortes com características similares. Essa abordagem visa aprimorar

rar a precisão das predições, adaptando-as às particularidades específicas de cada tipo de FII.

Nesse contexto, foram conduzidos testes abrangentes, investigando uma variedade de modelos e atributos. A análise foi segmentada em três grupos distintos, cada um contendo FIIs de uma categoria específica: papel, tijolo e híbrido. Essa abordagem abraçou tanto o curto quanto o médio prazo, oferecendo, assim, uma visão abrangente das complexidades inerentes aos diferentes tipos de FIIs.

No curto prazo, a estratégia E2 - Xgboost obteve resultados superiores na categoria papel e no RMSE da categoria híbrido, como evidenciado nas tabelas 6 e 8. Em contrapartida, o ARIMA demonstrou desempenho superior na categoria tijolo e no MAE da categoria híbrido, como indicado nas tabelas 8 e 10.

Já no médio prazo, durante o treinamento dos fundos considerando apenas a categoria específica, a estratégia E2 - XGboost apresentou os menores erros em todas as categorias, como pode ser observado nas Tabelas 7, 9 e 11.

Modelo	RMSE	MAE
E1 - Xgboost	0.052190	0.036700
E1 - Arima	0.045126	0.032468
E2 - Xgboost	0.043549	0.029513
E2 - LSTM	0.053412	0.037991

Tabela 6 – Erros da categoria papel para 1 mês por estratégia

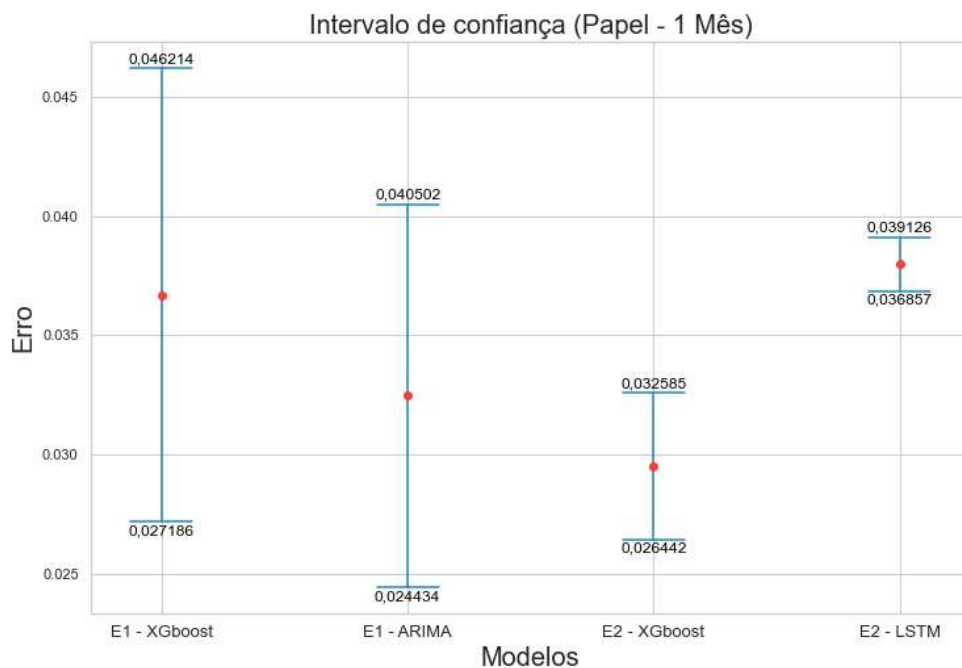


Figura 9 – Intervalo de confiança dos erros para 1 mês por estratégia da categoria papel

Modelo	RMSE	MAE
E1 - Xgboost	0.080742	0.059437
E1 - Arima	0.118805	0.090949
E2 - Xgboost	0.075478	0.051876
E2 - LSTM	0.078749	0.056676

Tabela 7 – Erros da categoria papel para 6 meses por estratégia

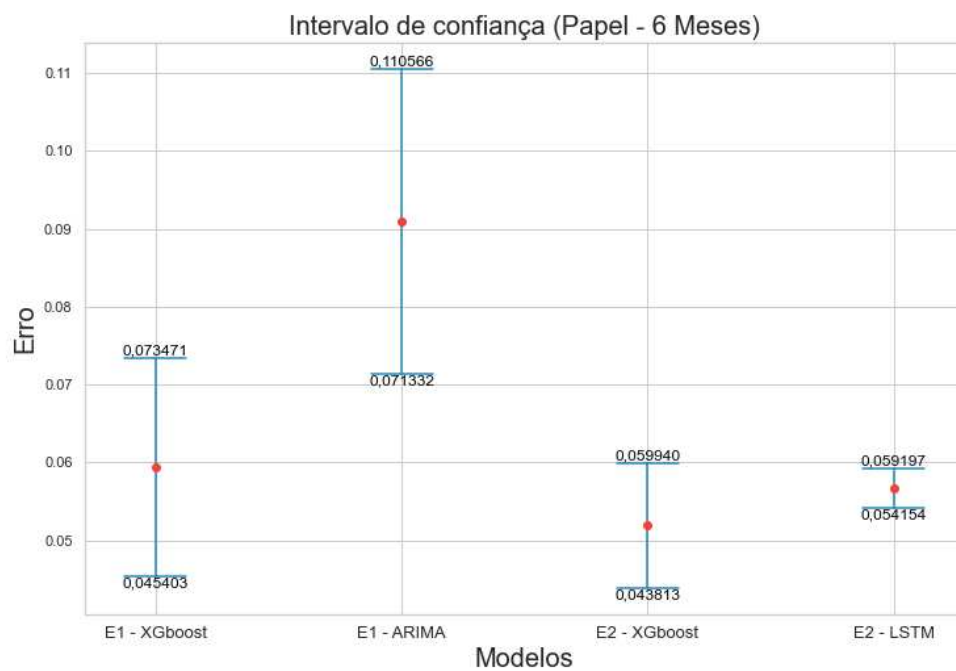


Figura 10 – Intervalo de confiança dos erros para 6 meses por estratégia da categoria papel

Modelo	RMSE	MAE
E1 - Xgboost	0.071832	0.050244
E1 - Arima	0.062200	0.043555
E2 - Xgboost	0.057927	0.044485
E2 - LSTM	0.076081	0.058558

Tabela 8 – Erros da categoria híbrido para 1 mês por estratégia

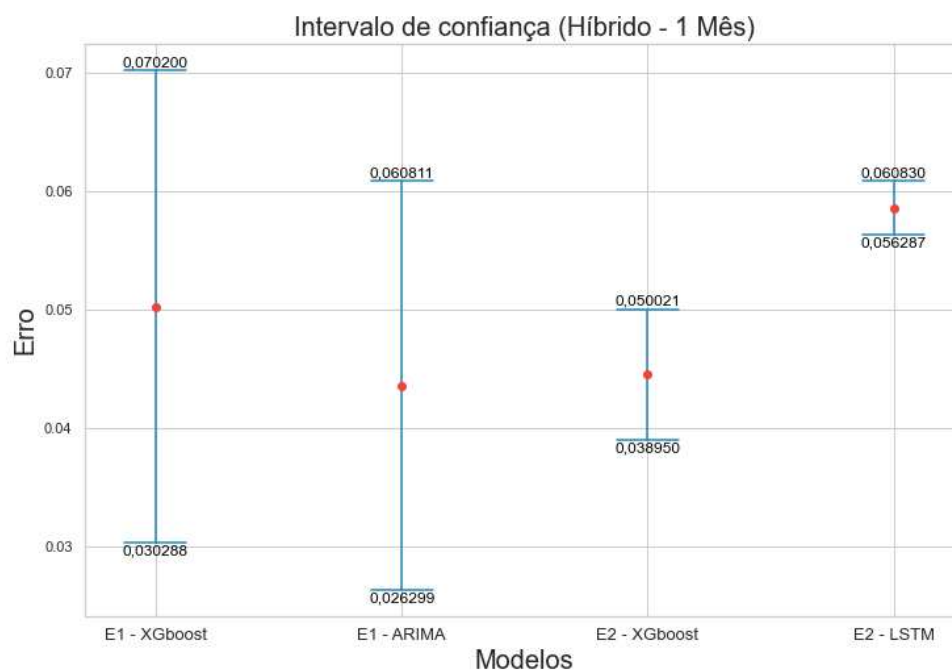


Figura 11 – Intervalo de confiança dos erros para 1 mês por estratégia da categoria híbrido

Modelo	RMSE	MAE
E1 - Xgboost	0.107010	0.079168
E1 - Arima	0.156162	0.108825
E2 - Xgboost	0.079148	0.062825
E2 - LSTM	0.088195	0.070385

Tabela 9 – Erros da categoria híbrido para 6 meses por estratégia

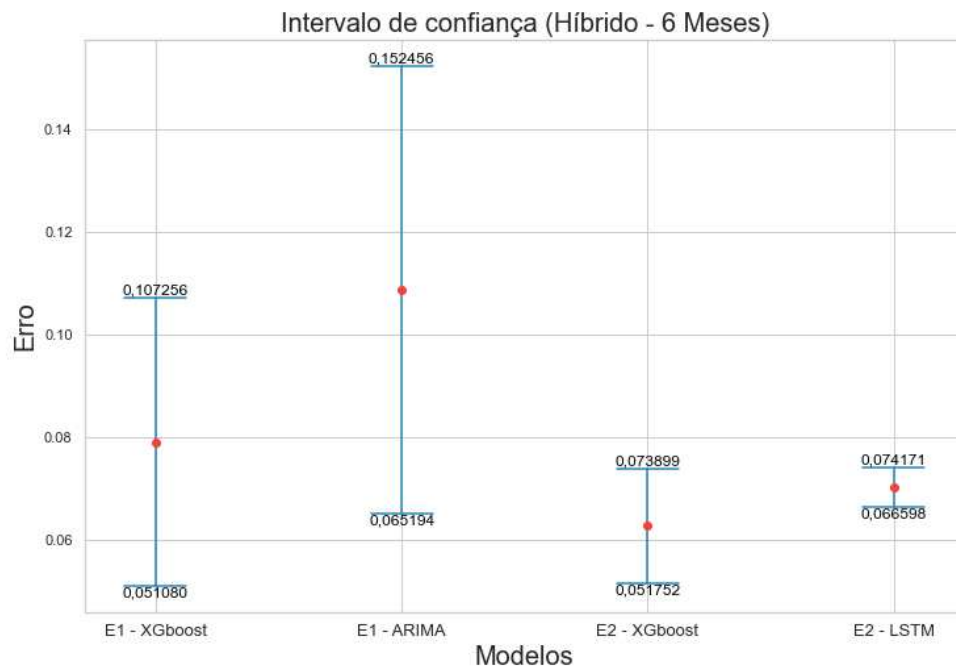


Figura 12 – Intervalo de confiança dos erros para 6 meses por estratégia da categoria híbrido

Modelo	RMSE	MAE
E1 - Xgboost	0.068239	0.049951
E1 - Arima	0.057989	0.041208
E2 - Xgboost	0.065266	0.047619
E2 - LSTM	0.069958	0.050513

Tabela 10 – Erros da categoria tijolo para 1 mês por estratégia

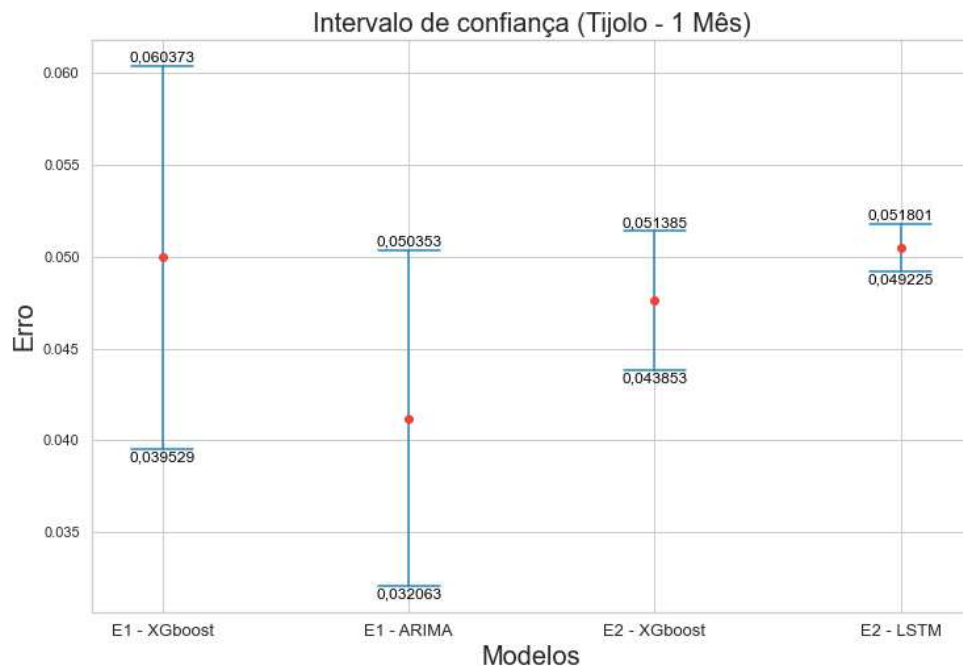


Figura 13 – Intervalo de confiança dos erros para 1 mês por estratégia da categoria tijolo

Modelo	RMSE	MAE
E1 - Xgboost	0.162330	0.090787
E1 - Arima	0.171944	0.128760
E2 - Xgboost	0.082338	0.064103
E2 - LSTM	0.090440	0.071980

Tabela 11 – Erros da categoria tijolo para 6 meses por estratégia

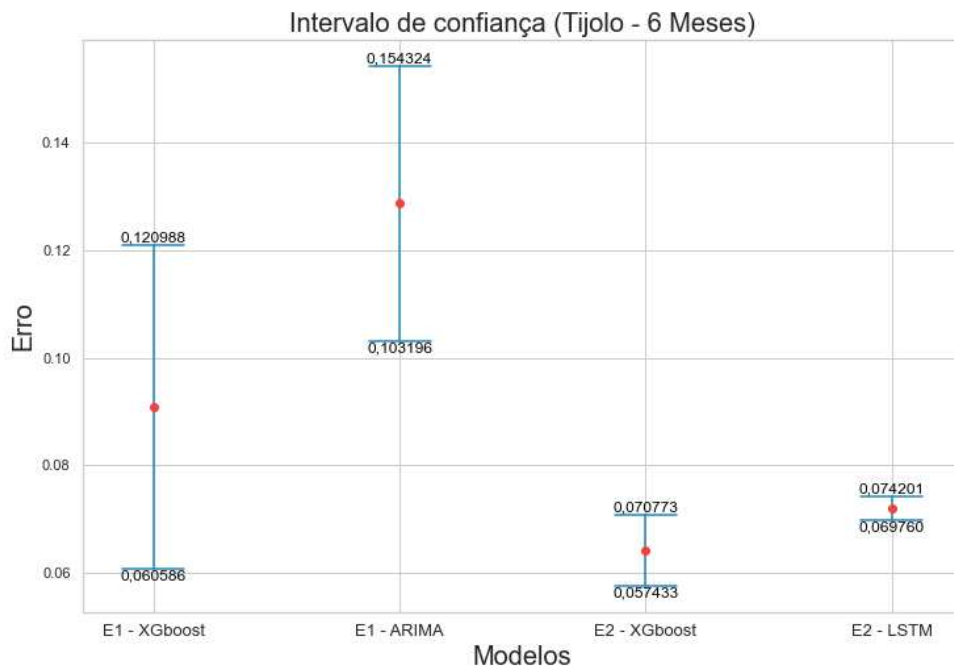


Figura 14 – Intervalo de confiança dos erros para 6 meses por estratégia da categoria tijolo

Para avaliar o impacto positivo da segmentação dos fundos em diversas categorias nos resultados, foi realizada uma comparação de desempenho ao treinar os modelos sem a utilização de categorizações e, posteriormente, ao incorporar essas categorias durante o treinamento. Este processo foi conduzido com o auxílio dos intervalos de confiança dos erros de todos os modelos testados, assegurando uma base estatística sólida para a análise.

A preferência pelo MAE como critério de comparação encontra sua justificativa na sinergia com os intervalos de confiança. Essa decisão estratégica baseia-se na capacidade do MAE de representar a média absoluta dos erros, proporcionando uma avaliação estável e resistente a discrepâncias extremas nos dados. A combinação do MAE com intervalos de confiança simplifica a comparação entre modelos, estabelecendo uma análise consistente e confiável.

Para o curto prazo, conforme evidenciado na tabela 12, destaca-se a importância da implementação de um modelo específico dedicado à categoria “papel”. Nesse contexto, observa-se uma redução significativa de aproximadamente 25,21% no erro em comparação com a abordagem do modelo geral. No entanto, para as categorias “híbrido” e “tijolo”, essa estratégia não se mostrou vantajosa, já que a utilização de um modelo geral para o treinamento e previsão de todos os fundos resultou em um erro inferior em comparação aos modelos específicos dessas categorias.

Modelo	MAE Geral	MAE Papel	MAE Híbrido	MAE Tijolo
E1 - Xgboost	0.045758	0.036700	0.050244	0.049951
E1 - Arima	0.039462	0.032468	0.043555	0.041208
E2 - Xgboost	0.041151	0.029513	0.044485	0.047619
E2 - LSTM	0.042082	0.037991	0.058558	0.050513

Tabela 12 – MAE de cada categoria para 1 mês por estratégia

Para o médio prazo, conforme evidenciado na tabela 13, destaca-se a importância da implementação de um modelo específico exclusivo para a categoria “papel”. Nesse contexto, observa-se uma redução significativa de aproximadamente 18,75% no erro em comparação com a abordagem do modelo geral. Entretanto, para a categoria “tijolo”, essa estratégia não se mostrou vantajosa, visto que a utilização de um modelo que abrange o treinamento e previsão de todos os fundos resultou em um erro inferior em comparação ao modelo específico da categoria. Quanto à categoria “híbrido”, houve uma diminuição pouco significativa de aproximadamente 1,60% no erro em comparação com a abordagem do modelo geral. Nesse contexto, a escolha de considerar isso relevante para o cenário específico do cientista de dados fica a critério do mesmo.

Modelo	MAE Geral	MAE Papel	MAE Híbrido	MAE Tijolo
E1 - Xgboost	0.080815	0.059437	0.079168	0.090787
E1 - Arima	0.104505	0.090949	0.108825	0.128760
E2 - Xgboost	0.063845	0.051876	0.062825	0.064103
E2 - LSTM	0.073377	0.056676	0.070385	0.071980

Tabela 13 – MAE de cada categoria para 6 meses por estratégia

A partir dessa análise, observa-se que a categorização dos fundos demonstra benefícios ao realizar previsões para FIIs da categoria papel em ambas as sazonalidades. No entanto, essa abordagem apresenta-se com vantagens limitadas na previsão semestral de FIIs da categoria híbrido e mostra-se desvantajosa nas demais situações (predição mensal de híbridos, predição mensal de tijolos e predição semestral de tijolos).

4.3 Q3: Utilizar dados enriquecidos dos imóveis dos fundos ajuda a melhorar os modelos?

Neste tópico, será apresentada uma comparação entre modelos da estratégia **E2** que utilizam as informações dos imóveis físicos e os que não utilizaram. O intuito é avaliar se a inclusão de características físicas dos imóveis geridos por FIIs resulta em melhorias nos erros obtidos durante o processo de predição ou não. Essa comparação não faz sentido para a estratégia E1, pois quando um modelo é criado por fundo, todos os atributos de imóveis serão idênticos para todas as entradas.

É importante ressaltar que, entre os resultados da estratégia **E2** mostrados nas seções 4.1 e 4.2, alguns já podem ter incluído os atributos dos imóveis, enquanto outros não. Nas análises anteriores, apenas o erro de predição foi considerado, independentemente de o modelo utilizar dados de imóveis ou não, já que o objetivo era identificar o modelo com o menor erro entre todos os treinados. No entanto, nesta seção, o foco é observar o desempenho específico dos modelos que utilizaram dados de imóveis.

Para avaliar o impacto da inclusão desses dados, será comparado o desempenho dos 6 melhores modelos encontrados (4 usando XGBoost e 2 usando LSTM) considerando o conjunto C1 (atributos sem considerar os imóveis), com o seu equivalente no conjunto C2 (atributos acrescidos das informações dos imóveis). A hipótese desta análise é que as características da região em que os imóveis físicos gerenciados pelos fundos estão localizados são importantes para a evolução dos preços dos fundos. Este é o primeiro trabalho que considera essa análise, e uma contribuição importante deste estudo.

Com o objetivo de aprimorar a clareza na interpretação dos dados, os modelos foram nomeados e descritos na Tabela 14, para o curto prazo, e na Tabela 15, para o médio prazo. É relevante observar que a inclusão do prefixo “XGB” denota a utilização do algoritmo XGBoost, enquanto aqueles que iniciam com “LSTM” indicam a aplicação do modelo de *Long Short-Term Memory* (LSTM). Além disso, qualquer designação finalizada com “A” implica a incorporação de dados imobiliários provenientes do MoreData, ao passo que aqueles finalizados com “B” estão associados a informações de imóveis obtidas do Google Maps. Essa nomenclatura específica facilitará a identificação e compreensão dos métodos empregados em cada análise.

Nome	Propriedades	Imóveis
XGB_1	Propriedades <i>close</i> e <i>closeChange</i>	-
XGB_1A	Propriedades <i>close</i> , <i>closeChange</i> com dados de imóveis	MoreData
XGB_1B	Propriedades <i>close</i> , <i>closeChange</i> com dados de imóveis	Google Maps
XGB_2	Todas as propriedades	-
XGB_2A	Todas as propriedades com dados de imóveis	MoreData
XGB_2B	Todas as propriedades com dados de imóveis	Google Maps
XGB_3	Todas as propriedades e suas versões com delay	-
XGB_3A	Todas as propriedades e suas versões com delay com dados de imóveis	MoreData
XGB_3B	Todas as propriedades e suas versões com delay com dados de imóveis	Google Maps
XGB_4	Principais propriedades e suas versões com delay	-
XGB_4A	Principais propriedades e suas versões com delay com dados de imóveis	MoreData

XGB_4B	Principais propriedades e suas versões com delay com dados de imóveis	Google Maps
LSTM_1	Todas as propriedades e delay	-
LSTM_1A	Todas as propriedades e delay com dados de imóveis	MoreData
LSTM_1B	Todas as propriedades e delay com dados de imóveis	Google Maps
LSTM_2	Principais propriedades e delay	-
LSTM_2A	Principais propriedades e delay com dados de imóveis	MoreData
LSTM_2B	Principais propriedades e delay com dados de imóveis	Google Maps

Tabela 14 – Características dos modelos da estratégia E2 para o curto prazo

Nome	Propriedades	Imóveis
XGB_1	Propriedade <i>close</i>	-
XGB_1A	Propriedade <i>close</i> com dados de imóveis	MoreData
XGB_1B	Propriedade <i>close</i> com dados de imóveis	Google Maps
XGB_2	Todas as propriedades	-
XGB_2A	Todas as propriedades com dados de imóveis	MoreData
XGB_2B	Todas as propriedades com dados de imóveis	Google Maps
XGB_3	Todas as propriedades e suas versões com delay	-
XGB_3A	Todas as propriedades e suas versões com delay com dados de imóveis	MoreData
XGB_3B	Todas as propriedades e suas versões com delay com dados de imóveis	Google Maps
XGB_4	Principais propriedades e suas versões com delay	-
XGB_4A	Principais propriedades e suas versões com delay com dados de imóveis	MoreData
XGB_4B	Principais propriedades e suas versões com delay com dados de imóveis	Google Maps
LSTM_1	Todas as propriedades e delay	-
LSTM_1A	Todas as propriedades e delay com dados de imóveis	MoreData
LSTM_1B	Todas as propriedades e delay com dados de imóveis	Google Maps
LSTM_2	Principais propriedades e delay	-
LSTM_2A	Principais propriedades e delay com dados de imóveis	MoreData
LSTM_2B	Principais propriedades e delay com dados de imóveis	Google Maps

Tabela 15 – Características dos modelos da estratégia E2 para o médio prazo

Nessa perspectiva, tendo em vista a predição com horizonte temporal de 1 mês, foi possível observar, com a tabela 16, uma melhora em apenas 3 dos 6 modelos para as métricas de RMSE e MAE. Entretanto, para o horizonte de 6 meses houve uma melhora

geral nas duas métricas avaliadas, sendo que apenas um modelo apresentou um aumento nos erros, como pode ser visto na tabela 17.

Modelo	RMSE	MAE	Lower	Upper
XGB_1	0.056973	0.040585	0.038261	0.04291
XGB_1A	0.057352	0.040953	0.038619	0.043287
XGB_1B	0.057513	0.041012	0.038668	0.043356
XGB_2	0.060335	0.044585	0.042222	0.046948
XGB_2A	0.060673	0.044900	0.042527	0.047272
XGB_2B	0.059706	0.043794	0.041435	0.046154
XGB_3	0.057324	0.041399	0.039094	0.043704
XGB_3A	0.057750	0.041735	0.039414	0.044055
XGB_3B	0.057689	0.041647	0.039326	0.043968
XGB_4	0.055232	0.042214	0.040144	0.044285
XGB_4A	0.054430	0.041151	0.03908	0.043222
XGB_4B	0.054543	0.041433	0.039371	0.043495
LSTM_1	0.076923	0.056754	0.055801	0.057707
LSTM_1A	0.073565	0.052421	0.051473	0.053368
LSTM_1B	0.069954	0.050526	0.049637	0.051414
LSTM_2	0.061866	0.044120	0.043324	0.044917
LSTM_2A	0.062736	0.044783	0.043976	0.045589
LSTM_2B	0.060005	0.042082	0.041297	0.042867

Tabela 16 – Tabela de erros da estratégia E2 para 1 mês

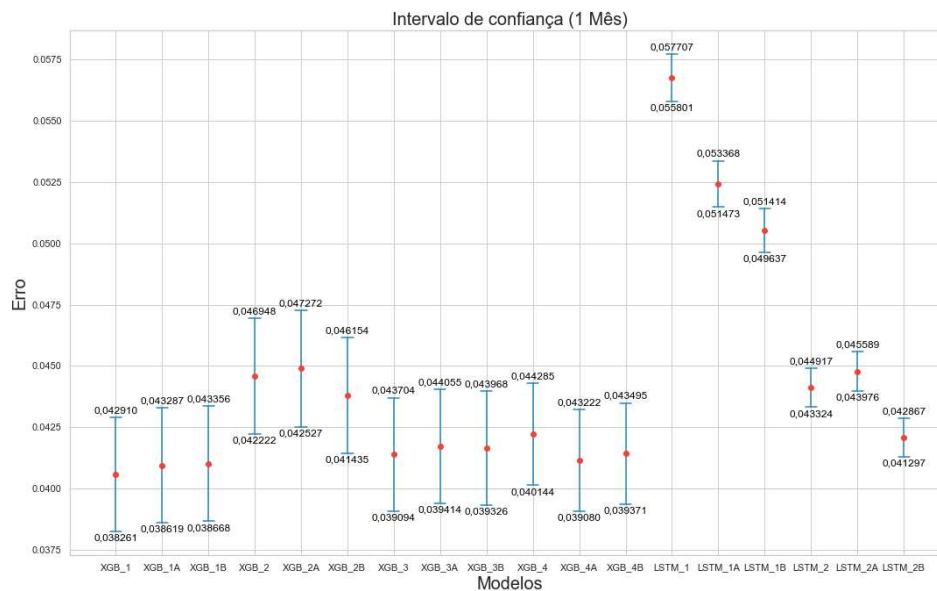


Figura 15 – Intervalo de confiança dos erros para 1 mês por estratégia

Para o curto prazo, como pode ser visto na tabela 16, a inclusão de dados de imóveis revelou-se relevante apenas para o modelo LSTM, já que não houve sobreposição do

intervalo de confiança (Figura 15) entre o melhor modelo e aquele que não utilizou os dados enriquecidos de imóveis. Entretanto, para o XGBoost, todos os intervalos de confiança se sobrepuseram, resultando em uma média de erros dividida igualmente. Metade dos casos apresentou uma média menor ao não utilizar dados imobiliários, enquanto a outra metade obteve melhorias ao incorporar esses dados.

Modelo	RMSE	MAE	Lower	Upper
XGB_1	0.091063	0.068062	0.061501	0.074623
XGB_1A	0.087407	0.065073	0.058744	0.071402
XGB_1B	0.084810	0.063292	0.057169	0.069414
XGB_2	0.086372	0.066843	0.06091	0.072775
XGB_2A	0.084629	0.065122	0.05926	0.070983
XGB_2B	0.085548	0.066260	0.060392	0.072129
XGB_3	0.084545	0.065044	0.059187	0.070902
XGB_3A	0.082978	0.063845	0.058097	0.069594
XGB_3B	0.084073	0.065302	0.059559	0.071045
XGB_4	0.089387	0.067649	0.061313	0.073986
XGB_4A	0.086922	0.065262	0.059035	0.071489
XGB_4B	0.085235	0.064265	0.058192	0.070337
LSTM_1	0.120721	0.092112	0.089448	0.094775
LSTM_1A	0.112130	0.086190	0.083742	0.088639
LSTM_1B	0.106521	0.081331	0.078983	0.083679
LSTM_2	0.097009	0.073377	0.07121	0.075543
LSTM_2A	0.098347	0.073986	0.071774	0.076197
LSTM_2B	0.108509	0.083890	0.081541	0.08624

Tabela 17 – Tabela de erros da estratégia E2 para 6 meses

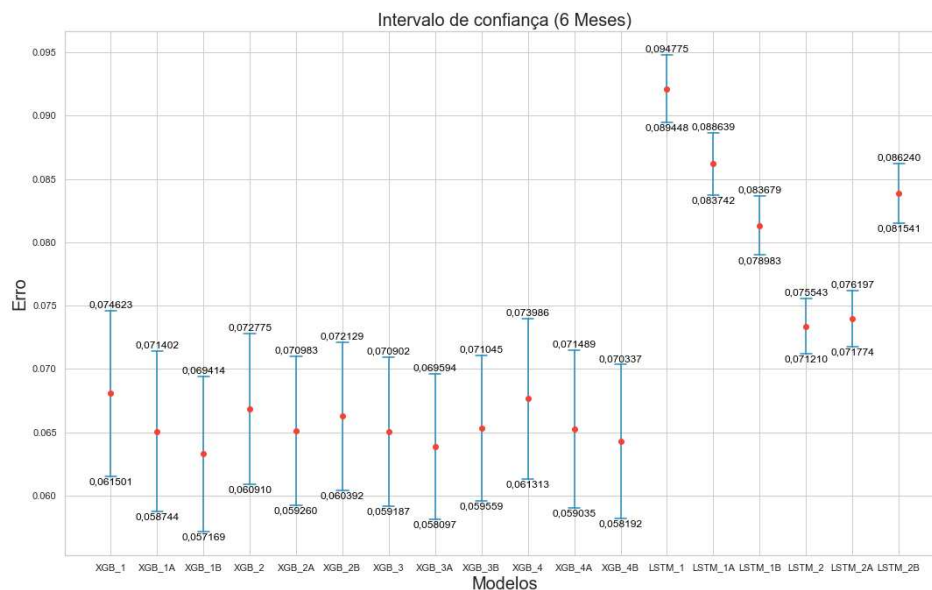


Figura 16 – Intervalo de confiança dos erros para 1 mês por estratégia

No médio prazo, conforme destacado na tabela 17, observa-se sobreposição do intervalo de confiança (Figura 16) em todos os cenários, exceto no caso do modelo LSTM_1, onde a versão com dados de imóveis apresentou um erro superior, e não ocorreu sobreposição dos intervalos de confiança. No entanto, para todos os modelos XGBoost que utilizaram enriquecimento de imóveis, a inclusão de dados imobiliários resultou em uma média de erro geral inferior.

A análise revelou uma variedade significativa nos resultados de diferentes modelos e estratégias. Enquanto alguns modelos XGBoost se beneficiaram da inclusão de dados imobiliários em ambos os horizontes de previsão, o modelo LSTM apresentou um cenário mais complexo, com vantagem em curto prazo e desvantagem em médio prazo. Isso sugere que a natureza da influência dos dados imobiliários, além de depender de diversos fatores, também pode depender da arquitetura específica do modelo.

A disparidade de desempenho entre os modelos mais simples (XGB_1) e os mais complexos (XGB_3, XGB_4) aponta para a necessidade crítica de alcançar um equilíbrio preciso entre a complexidade do modelo e a quantidade de dados imobiliários incorporados. A complexidade excessiva desses modelos pode resultar em subutilização de informações disponíveis. Nesse sentido, em trabalhos futuros, é essencial explorar a identificação do ponto ótimo em termos de complexidade. Isso permitirá que o modelo atinja seu desempenho máximo ao integrar de maneira eficaz os dados imobiliários, otimizando assim as previsões de fundos imobiliários.

4.4 Q4: É possível fazer uma previsão de médio prazo (variação daqui a 6 meses) com erros aceitáveis, se comparado com a previsão de curto prazo (1 mês)?

Ao comparar os modelos de previsão de curto prazo (1 mês) com os de médio prazo (6 meses), torna-se claro que os modelos destinados a horizontes temporais mais próximos exibem métricas de erro notavelmente inferiores, como destacado nas tabelas 4 e 5. Essa disparidade pode ser atribuída a desafios específicos associados aos modelos de previsão de 6 meses.

A limitação na quantidade de dados disponíveis para treinamento emerge como um desafio significativo. A escassez de informações ao longo do tempo pode prejudicar a capacidade do modelo de aprender padrões e tendências relevantes para horizontes temporais mais extensos. Além disso, a maior variabilidade nos valores a serem previstos, conforme evidenciado pelo gráfico de caixa na figura 17, apresenta-se como um obstáculo adicional para os modelos de médio prazo. A ampla dispersão dos dados para o horizonte de 6 meses, representada pela barra laranja, contrasta nitidamente com a menor variabilidade

observada para 1 mês (barra azul).

Consequentemente, é prudente concluir que, embora os modelos de previsão de curto prazo demonstrem maior precisão, os desafios associados aos modelos de 6 meses, como a escassez de dados e a variabilidade aumentada, contribuem para métricas de erro mais elevadas. Assim, ao buscar previsões de médio prazo com erros aceitáveis, é crucial considerar esses desafios inerentes e explorar estratégias aprimoradas de modelagem e coleta de dados para melhorar a eficácia das previsões em horizontes temporais mais longos.

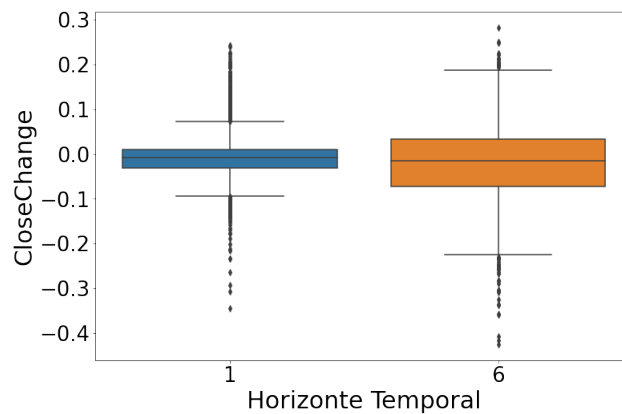


Figura 17 – Boxplot da variação do “CloseChange”

4.5 Q5: As redes neurais (e.g., LSTM) apresentam boas perspectivas para previsão de fundos imobiliários?

As redes neurais, notadamente as Long Short-Term Memory (LSTM), são amplamente reconhecidas por sua habilidade em capturar padrões temporais e relações não-lineares em conjuntos de dados. A investigação do uso dessas redes na previsão de fundos imobiliários oferece uma oportunidade única para obter *insights* valiosos sobre a dinâmica complexa desses ativos e como fatores diversos podem interagir e influenciar os resultados financeiros.

A análise comparativa entre modelos LSTM e XGBoost revelou que, de modo geral, os modelos LSTM apresentaram resultados inferiores, conforme demonstrado nas Tabelas 16 e 17. Esse desempenho inferior pode ser atribuído a diversos fatores, como a complexidade do comportamento dos fundos imobiliários, que pode não ser totalmente capturada pelas estruturas de redes neurais.

Ao contrastar os resultados com o modelo ARIMA, observou-se que os modelos LSTM apresentaram desempenho inferior no curto prazo, mas superaram nas previsões de médio prazo. Essa discrepância pode ser explicada pela natureza adaptativa das redes neurais, que podem se beneficiar de padrões mais complexos e tendências a longo prazo.

Contudo, é crucial ressaltar que, por se tratar de uma rede neural, melhorias substanciais podem ser alcançadas em trabalhos futuros. Isso inclui a exploração de bases de dados mais extensas em termos de período e uma otimização mais aprofundada das configurações das camadas da rede.

Em conclusão, os resultados indicam que, com as condições atuais, os modelos LSTM não se destacaram como a escolha preferencial para a previsão de fundos imobiliários. No entanto, essa conclusão é preliminar e sujeita a revisões com a realização de estudos mais aprofundados e aprimoramentos nas abordagens de modelagem. A complexidade inerente aos fundos imobiliários e a natureza dinâmica dos mercados financeiros sugerem que a busca por modelos preditivos eficazes requer uma abordagem contínua e iterativa, com adaptações constantes para enfrentar os desafios específicos deste domínio.

4.6 Q6: Criar um modelo individual para cada fundo é melhor do que um modelo geral que considera todos os fundos de uma vez?

Nesta análise, o ARIMA foi desconsiderado por não estar relacionado com essa questão. Ao confrontar as estratégias E1 e E2, evidenciadas nas Tabelas 4 e 5, emerge uma conclusão contundente em favor da estratégia E2.

Os modelos individuais (E1) apresentaram desempenho inferior em ambas as sazonalidades, enquanto a abordagem E2, que considera conjuntamente todos os fundos, demonstrou uma habilidade mais robusta na identificação de tendências. A superioridade da estratégia E2 sugere que treinar um modelo geral é mais eficiente para previsões financeiras. Essa escolha estratégica não apenas ressalta a capacidade do modelo global em capturar padrões complexos de mercado, mas também aponta para uma eficácia notável na interpretação de sinergias entre diferentes fundos.

Portanto, os dados analisados indicam que a estratégia de adotar um modelo geral (E2) é uma opção mais sólida, proporcionando previsões mais abrangentes e precisas em comparação com modelos individuais (E1).

4.7 Q7: Houveram padrões de atributos mais frequentes nos modelos mais eficazes?

Na estratégia E1, em contraposição à estratégia E2, um grande número de modelos de aprendizado de máquina foi criado. Diante dessa multiplicidade de modelos, nesta seção, surge a oportunidade de analisar os resultados de curto e médio prazo gerados por eles, com o intuito de extrair *insights* sobre os atributos mais relevantes.

Com o objetivo de melhorar a clareza e organização dos dados, foi criada a tabela 18, contendo as siglas e seus respectivos resumos de significado. A tabela inclui apenas os atributos que foram identificados nos melhores modelos de cada categoria. Observa-se que o atributo *CloseChange* foi ocultado da análise, uma vez que é o atributo alvo da predição e sempre estará presente em todos os modelos.

Coluna	Sigla	Resumo
Close	CL	Preço de fechamento
DividendsChange	DC	Variação do preço dos dividendos
DividendYield	DY	Valor do <i>dividend yield</i>
DividendYieldChange	DYC	Variação do <i>dividend yield</i>
SectorCloseChangeMean	SCC	Variação média do preço de fechamento do setor
SectorDividendsChangeMean	SDC	Variação média dos dividendos do setor
Selic	SEL	Taxa Selic
IFIX	IFIX	Índice IFIX
IPCA	IPCA	Índice IPCA
IGP-M	IGPM	Índice IGP-M

Tabela 18 – Siglas dos atributos mais utilizados nos modelos da estratégia E1

4.7.1 Categoria Papel

Na predição de curto prazo, os atributos que mais apareceram nos melhores modelos foram: *DividendsChange* (5 vezes), *IFIX* (5 vezes), *IGP-M* (4 vezes), *SectorCloseChangeMean* (3 vezes), *IPCA* (3 vezes), *DividendYield* (3 vezes), *DividendYieldChange* (1 vez).

Modelo	RMSE	MAE
XGBoost: DC, IFIX, IPCA, SCC	0,052190	0,036700
XGBoost: DC, IFIX, IGPM, IPCA, SCC	0,052202	0,036721
XGBoost: DC, IFIX, IGPM, IPCA	0,052474	0,037564
XGBoost: DC, IFIX, IGPM, DY, DYC	0,052477	0,037334
XGBoost: DC, IFIX, IGPM, DY, SCC	0,052632	0,037165

Tabela 19 – Melhores modelos para categoria Papel (1 Mês)

Já na predição de médio prazo, os atributos que mais apareceram nos melhores modelos foram: *DividendYieldChange* (5 vezes), *SectorCloseChangeMean* (5 vezes), *IPCA* (5 vezes), *IGP-M* (4 vezes), *IFIX* (3 vezes), *Selic* (3 vezes)

Modelo	RMSE	MAE
XGBoost: DYC, SCC, IPCA, IGPM, IFIX	0,080742	0,059437
XGBoost: DYC, SCC, SEL, IPCA, IGPM, IFIX	0,080958	0,058652

XGBoost: DYC, SCC, IPCA, IGPM	0,081011	0,059058
XGBoost: DYC, SCC, SEL, IPCA, IGPM	0,081267	0,058658
XGBoost: DYC, SCC, SEL, IPCA, IFIX	0,081403	0,058958

Tabela 20 – Melhores modelos para categoria Papel (6 Meses)

4.7.2 Categoria Híbrido

Na predição de curto prazo, os atributos que mais apareceram nos melhores modelos foram: IPCA (5 vezes), IGP-M (5 vezes), SectorCloseChangeMean (4 vezes), IFIX (4 vezes), DividendsChange (3 vezes), Close (1 vezes), SectorDividendsChangeMean (1 vezes)

Modelo	RMSE	MAE
XGBoost: DC, SCC, IPCA, IGPM, IFIX	0,071832	0,050244
XGBoost: SCC, IPCA, IGPM, IFIX	0,072078	0,051555
XGBoost: DC, SCC, IPCA, IGPM	0,072462	0,050651
XGBoost: DC, IPCA, IGPM, IFIX	0,072639	0,050308
XGBoost: CL, SDC, SCC, IPCA, IGPM, IFIX	0,072694	0,051317

Tabela 21 – Melhores modelos para categoria Híbrido (1 Mês)

Já na predição de médio prazo, os atributos que mais apareceram nos melhores modelos foram: Close (5 vezes), SectorDividendsChangeMean (5 vezes), IPCA (5 vezes), IFIX (3 vezes), SectorCloseChangeMean (2 vezes), IGP-M (2 vezes)

Modelo	RMSE	MAE
XGBoost: CL, SDC, SCC, IPCA, IFIX	0,107010	0,079168
XGBoost: CL, SDC, IPCA, IFIX	0,107403	0,077878
XGBoost: CL, SDC, IPCA, IGPM	0,107481	0,079665
XGBoost: CL, SDC, SCC, IPCA	0,107854	0,079305
XGBoost: CL, SDC, IPCA, IGPM, IFIX	0,108242	0,078782

Tabela 22 – Melhores modelos para categoria Híbrido (6 Meses)

4.7.3 Categoria Tijolo

Na predição de curto prazo, os atributos que mais apareceram nos melhores modelos foram: IPCA (5 vezes), IFIX (5 vezes), SectorCloseChangeMean (4 vezes), IGP-M (4 vezes), SectorDividendsChangeMean (4 vezes), Selic (2 vezes), DividendsChange (1 vezes), DividendYield (1 vezes)

Modelo	RMSE	MAE
XGBoost: SCC, SEL, IPCA, IGPM, IFIX	0,068239	0,049951
XGBoost: DC, SDC, IPCA, IFIX	0,068526	0,051082
XGBoost: DY, SDC, SCC, IPCA, IGPM, IFIX	0,068574	0,050198
XGBoost: SDC, SCC, IPCA, IGPM, IFIX	0,068619	0,049739
XGBoost: SDC, SCC, SEL, IPCA, IGPM, IFIX	0,068652	0,050558

Tabela 23 – Melhores modelos para categoria Tijolo (1 Mês)

Já na predição de médio prazo, os atributos que mais apareceram nos melhores modelos foram: DividendsChange (5 vezes), SectorDividendsChangeMean (5 vezes), SectorCloseChangeMean (5 vezes), Selic (5 vezes), IGP-M (3 vezes), IPCA (2 vezes), IFIX (2 vezes)

Modelo	RMSE	MAE
XGBoost: DC, SDC, SCC, SEL, IPCA, IFIX	0,162330	0,090787
XGBoost: DC, SDC, SCC, SEL	0,162395	0,091274
XGBoost: DC, SDC, SCC, SEL, IGPM, IFIX	0,162477	0,090562
XGBoost: DC, SDC, SCC, SEL, IPCA, IGPM	0,162478	0,090941
XGBoost: DC, SDC, SCC, SEL, IGPM	0,162586	0,091229

Tabela 24 – Melhores modelos para categoria Tijolo (6 Meses)

4.7.4 Discussão

Com base nos resultados obtidos, foram identificados os atributos mais relevantes, observando que cada um deles apareceu em mais de 60% dos melhores modelos:

- **IPCA (25 de 30 modelos):** O índice, frequentemente utilizado como indicador de inflação no Brasil, sugere que os movimentos inflacionários podem ter um impacto significativo nas previsões dos fundos imobiliários. A variação do IPCA influencia diretamente os custos associados aos investimentos em imóveis, como manutenção, impostos, e outros fatores que afetam o desempenho dos fundos.
- **SectorCloseChange (23 de 30 modelos):** A recorrência deste atributo indica que as tendências de variação nos preços de fechamento dos setores específicos são relevantes para as previsões. Variações médias nos preços de fechamento dos setores podem refletir mudanças macroeconômicas ou eventos específicos que impactam todo o setor, influenciando assim o desempenho dos fundos imobiliários associados.
- **IFIX (22 de 30 modelos):** A presença frequente do IFIX indica que o desempenho geral do mercado de fundos imobiliários é um fator crucial nas previsões. O

IFIX, índice que reflete o comportamento médio dos fundos imobiliários listados na Bolsa de Valores, sugere que a dinâmica do mercado como um todo é um indicador relevante para prever o comportamento dos fundos individuais.

- **IGP-M (22 de 30 modelos):** A recorrência do IGP-M sugere que as variações gerais nos preços do mercado têm impacto nas previsões. Assim como o IPCA, o IGP-M abrange uma variedade de setores econômicos, indicando que movimentos macroeconômicos amplos podem influenciar o desempenho dos fundos imobiliários.

4.8 Q8: É eficaz utilizar uma modelagem personalizada para cada fundo imobiliário?

Conforme discutido nas seções anteriores, esta seção tem como objetivo investigar uma nova abordagem e verificar sua eficácia em relação à modelagens mais abrangentes. A partir dos resultados obtidos, foram destacadas as melhores combinações de atributos para cada FII individual de cada uma das categorias, para os intervalos de curto e médio prazo.

4.8.1 Fundos de Papel

4.8.1.1 Curto prazo

Para o intervalo de um mês, os modelos que apresentaram o menor RMSE para cada um dos fundos de papel são apresentados na tabela 25.

Fundo	Modelo	RMSE	MAE
AFHI11	ARIMA	0,020794	0,024493
ARRI11	XGBoost: CL, DY, DC, DY, SCC, SEL	0,015718	0,019993
BARI11	ARIMA	0,015775	0,019882
BBFO11	XGBoost: DY, DC, DY, IGPM, IFIX	0,050578	0,057102
BCFF11	XGBoost: CL, DY, DC, SDC, SCC, SEL, IPCA	0,036317	0,045355
BCIA11	ARIMA	0,029428	0,033106
BCRI11	XGBoost: CL, DY, DC, IPCA, IGPM	0,014214	0,016649
BICR11	ARIMA	0,029402	0,034220
BLMR11	XGBoost: CL, DY, IPCA, IFIX	0,034174	0,053146
BPFF11	XGBoost: DC, SCC, SEL, IPCA	0,022676	0,025192
CPTS11	ARIMA	0,014032	0,017784
CRFF11	ARIMA	0,031371	0,039322
CVBI11	ARIMA	0,017338	0,019911
DEVA11	XGBoost: CL, DY, DC, SEL	0,012336	0,015231

FEXC11	XGBoost: CL, DY, SDC, SCC, SEL	0,027126	0,035123
FLCR11	ARIMA	0,012289	0,015808
GCFF11	XGBoost: CL, DY, DC, DY, IGPM, IFIX	0,021720	0,030310
GCRI11	ARIMA	0,006828	0,007619
HABT11	XGBoost: DY, SDC, SEL, IGPM, IFIX	0,027652	0,042479
HFOF11	ARIMA	0,024508	0,033175
HGCR11	XGBoost: DY, DY, SCC, IPCA, IGPM	0,009419	0,011679
HGFF11	ARIMA	0,030328	0,033582
HGIC11	ARIMA	0,008811	0,013441
HREC11	ARIMA	0,010121	0,012081
HSAF11	XGBoost: CL, DY, DC, DY, IGPM	0,010391	0,015057
IRDM11	XGBoost: SDC, SCC, SEL, IPCA, IGPM, IFIX	0,015107	0,017973
JPPA11	XGBoost: DY, SEL, IPCA, IGPM	0,014522	0,017192
KFOF11	XGBoost: CL, DC, SDC, IPCA	0,053320	0,063174
KNCR11	XGBoost: SDC, SCC, SEL, IPCA, IFIX	0,012345	0,014616
KNHY11	XGBoost: DY, SEL, IGPM, IFIX	0,012794	0,015827
KNIP11	ARIMA	0,022879	0,025774
KNSC11	XGBoost: CL, DY, DY, IGPM, IFIX	0,012355	0,015117
MCCI11	ARIMA	0,012773	0,017981
MGCR11	XGBoost: DY, SCC, SEL, IGPM	0,029549	0,037843
MGFF11	XGBoost: DC, DY, SEL, IPCA, IGPM	0,029502	0,043146
MORE11	ARIMA	0,036582	0,051628
NAVT11	ARIMA	0,045362	0,055698
OUFF11	XGBoost: CL, DY, SDC, IPCA	0,025999	0,032095
OUIP11	XGBoost: CL, SDC, IPCA, IGPM	0,013521	0,014893
OURE11	XGBoost: DY, DC, SDC, SEL	0,007949	0,009320
PLRI11	XGBoost: DC, DY, SDC, IGPM	0,037808	0,047850
PORD11	XGBoost: DY, SDC, SEL, IGPM	0,010149	0,014560
RBFF11	XGBoost: CL, SDC, SEL, IPCA, IFIX	0,032451	0,037023
RBHY11	XGBoost: DC, SCC, IPCA, IFIX	0,014700	0,018109
RBRF11	XGBoost: DC, SCC, SEL, IPCA	0,033682	0,043424
RBRR11	ARIMA	0,021520	0,026889
RBRY11	ARIMA	0,020278	0,024080
RDPD11	XGBoost: DC, SDC, IGPM, IFIX	0,068939	0,108477
RECR11	XGBoost: DY, DC, DY, SCC, SEL, IGPM	0,025544	0,028299
RFOF11	XGBoost: DY, SDC, IGPM, IFIX	0,024316	0,033396
RRCI11	XGBoost: CL, SDC, IPCA, IGPM	0,021559	0,025015
RVBI11	XGBoost: DC, DY, SDC, SEL, IGPM	0,022053	0,024806

RZAK11	XGBoost: CL, DY, SDC, SCC	0,021814	0,024562
SADI11	XGBoost: DY, DC, SDC, SCC, SEL, IGPM	0,006133	0,007358
VCJR11	ARIMA	0,021559	0,028649
VGHF11	XGBoost: DY, DC, DY, SDC, SCC, IPCA	0,017573	0,021528
VGIR11	XGBoost: DY, DC, DY, SDC, SEL, IPCA, IGPM	0,015280	0,016918
VIFI11	ARIMA	0,036761	0,045997
VOTS11	ARIMA	0,021728	0,026825
VRTA11	XGBoost: DY, DC, DY, IFIX	0,009889	0,011859

Tabela 25 – Melhores modelos para cada fundo de papel individualmente no curto prazo

Para a predição de 1 mês, o RMSE médio ficou em **0,029668**, enquanto o melhor RMSE obtido havia sido **0,043549** (estratégia E2 - XGBoost)

4.8.1.2 Médio prazo

Para o intervalo de seis meses, os modelos que apresentaram o menor RMSE para cada um dos fundos de papel são apresentado na tabela 26.

Fundo	Modelo	RMSE	MAE
AFHI11	XGBoost: CL, DY, DY, SCC	0,066841	0,084028
ARRI11	ARIMA	0,130440	0,150047
BARI11	XGBoost: DY, DC, DY, IGPM	0,027525	0,029515
BBFO11	XGBoost: CL, DY, DC, DY, SDC, SCC, IGPM, IFIX	0,058585	0,091877
BCFF11	XGBoost: CL, DY, DC, SDC, SEL, IPCA	0,091432	0,120289
BCIA11	ARIMA	0,079334	0,117095
BCRI11	XGBoost: SDC, SCC, IGPM, IFIX	0,021398	0,024536
BICR11	XGBoost: CL, DY, DY, SDC	0,065157	0,073705
BLMR11	XGBoost: SDC, SCC, SEL, IPCA, IGPM, IFIX	0,068642	0,073454
BPFF11	XGBoost: DC, IPCA, IGPM, IFIX	0,054127	0,074317
CPTS11	XGBoost: DY, SDC, SCC, IPCA	0,008618	0,009838
CRFF11	XGBoost: DY, SEL, IPCA, IGPM	0,044161	0,050204
CVBI11	XGBoost: DY, SDC, IPCA, IFIX	0,016018	0,020608
DEVA11	XGBoost: DC, DY, SCC, SEL	0,023842	0,028907
FEXC11	ARIMA	0,060930	0,081035
FLCR11	ARIMA	0,016316	0,020627
GCFF11	XGBoost: Selic, IPCA, IGPM, IFIX	0,055113	0,062982
HABT11	XGBoost: CL, SDC, SEL, IPCA	0,047058	0,068405

HFOF11	XGBoost: DC, DY, SDC, SCC	0,087761	0,093149
HGCR11	XGBoost: CL, SDC, SCC, IPCA, IGPM	0,015669	0,026010
HGFF11	XGBoost: SCC, SEL, IPCA, IFIX	0,118615	0,145878
HREC11	XGBoost: CL, IPCA, IGPM, IFIX	0,019351	0,022261
HSAF11	XGBoost: SCC, IPCA, IGPM, IFIX	0,043474	0,043555
IRDM11	XGBoost: CL, DY, DC, SCC, SEL, IGPM	0,071207	0,083799
JPPA11	XGBoost: DY, SCC, SEL, IGPM	0,039051	0,040098
KFOF11	XGBoost: DY, DC, SCC, IFIX	0,097588	0,105360
KNCR11	XGBoost: DC, DY, SEL, IFIX	0,046800	0,052940
KNHY11	ARIMA	0,020178	0,023864
KNIP11	XGBoost: SCC, IPCA, IGPM, IFIX	0,022260	0,026099
KNSC11	ARIMA	0,030480	0,040545
MCCI11	XGBoost: CL, DY, SDC, SEL	0,012328	0,013659
MGCR11	XGBoost: CL, SDC, SEL, IGPM	0,017428	0,017549
MGFF11	XGBoost: SDC, SCC, IPCA, IFIX	0,026232	0,030075
MORE11	XGBoost: CL, DY, SDC, SCC, IFIX	0,066266	0,074560
NAVT11	XGBoost: DY, IPCA, IGPM, IFIX	0,021823	0,028996
OUFF11	XGBoost: SDC, SCC, SEL, IFIX	0,021636	0,024979
OUJP11	XGBoost: DY, DY, SDC, SCC, SEL, IGPM	0,065386	0,065390
OURE11	XGBoost: SCC, IPCA, IGPM, IFIX	0,022503	0,036965
PLRI11	XGBoost: DY, DC, DY, IGPM	0,114678	0,124968
PORD11	XGBoost: CL, DC, DY, IPCA, IGPM	0,024964	0,028641
RBFF11	XGBoost: DY, DC, DY, SEL, IGPM, IFIX	0,030656	0,035645
RBRF11	XGBoost: DY, DC, SCC, IPCA, IGPM, IFIX	0,008980	0,012849
RBRR11	XGBoost: CL, SCC, SEL, IPCA	0,070523	0,072561
RBRY11	XGBoost: DY, SDC, SCC, IFIX	0,041135	0,045093
RDPD11	ARIMA	0,069267	0,094859
RECR11	XGBoost: DC, DY, SDC, SCC, SEL, IGPM	0,021773	0,022135
RFOF11	XGBoost: DY, IPCA, IGPM, IFIX	0,070040	0,082081
RRCI11	XGBoost: DC, SDC, SCC, SEL, IPCA, IGPM, IFIX	0,053753	0,080509
RVBI11	XGBoost: DC, SDC, SCC, IGPM	0,023136	0,036582
RZAK11	XGBoost: SCC, SEL, IPCA, IFIX	0,037572	0,040595
SADI11	XGBoost: SCC, SEL, IGPM, IFIX	0,020594	0,024289
VCJR11	XGBoost: CL, SDC, IGPM, IFIX	0,040957	0,041188
VGHF11	XGBoost: DY, DC, DY, SDC, IFIX	0,051224	0,051539
VGIR11	XGBoost: CL, DC, SCC, IFIX	0,016634	0,017816
VIFI11	XGBoost: CL, DY, DC, DY	0,049916	0,050972
VOTS11	ARIMA	0,038993	0,051773

VRTA11	ARIMA	0,014649	0,015973
--------	-------	----------	----------

Tabela 26 – Melhores modelos para cada fundo de papel individualmente no médio prazo

Para a predição de 6 meses, o RMSE médio ficou em **0,056380**, enquanto o melhor RMSE obtido havia sido **0,075279** (estratégia E2 - LSTM)

4.8.2 Fundos de Híbridos

4.8.2.1 Curto prazo

Para o intervalo de um mês, os modelos que apresentaram o menor RMSE para cada um dos fundos híbridos é apresentado na tabela 27.

Fundo	Modelo	RMSE	MAE
ARCT11	XGBoost: CL, SEL, IGPM, IFIX	0,030189	0,038352
BREV11	XGBoost: DY, DY, SEL, IPCA	0,023944	0,026962
CPFF11	ARIMA	0,019646	0,024161
CXRI11	XGBoost: DC, SCC, SEL, IGPM, IFIX	0,020954	0,028125
FATN11	ARIMA	0,014076	0,025323
FLMA11	ARIMA	0,050357	0,064876
GALG11	ARIMA	0,023563	0,027585
HBRH11	ARIMA	0,044736	0,056186
HGRU11	XGBoost: DY, DY, SCC, IPCA	0,025618	0,031514
HOSI11	ARIMA	0,024822	0,032041
JSRE11	ARIMA	0,051042	0,064743
KNRI11	XGBoost: CL, DY, DC, DY, SDC, SCC, IFIX	0,018646	0,023030
MFAI11	ARIMA	0,043406	0,050294
MFII11	XGBoost: DY, DC, SDC, SCC, IPCA, IFIX	0,014871	0,017862
MXRF11	ARIMA	0,016790	0,019395
NEWL11	ARIMA	0,034579	0,039695
PATL11	ARIMA	0,044133	0,053556
PLCR11	XGBoost: CL, DY, DC, SDC, IGPM	0,021194	0,027053
QAMI11	XGBoost: DY, DC, DY, IPCA, IFIX	0,006886	0,007680
RBRD11	XGBoost: CL, SDC, SCC, SEL, IPCA	0,038780	0,058478
RBRS11	ARIMA	0,037018	0,041828
RECT11	XGBoost: DC, SEL, IPCA, IFIX	0,046671	0,063696
RELG11	XGBoost: CL, DY, IGPM, IFIX	0,039132	0,054231
RZTR11	XGBoost: CL, DY, SDC, SEL, IFIX	0,015557	0,017442
SARE11	XGBoost: DY, DY, SDC, SEL, IGPM	0,055302	0,082028

Tabela 27 – Melhores modelos para cada fundo híbrido individualmente no curto prazo

Para a predição de 1 mês, o RMSE médio ficou em **0,039998**, enquanto o melhor RMSE obtido havia sido **0,057927** (estratégia E2 - XGBoost)

4.8.2.2 Médio prazo

Para o intervalo de seis mês, os modelos que apresentaram o menor RMSE para cada um dos fundos híbridos é apresentado na tabela 28.

Fundo	Modelo	RMSE	MAE
ARCT11	XGBoost: DY, SDC, SEL, IGPM	0,111994	0,134306
BREV11	XGBoost: CL, SEL, IPCA, IFIX	0,110097	0,130209
CPFF11	XGBoost: DY, DY, SDC, SEL, IPCA, IFIX	0,035584	0,055647
CXRI11	XGBoost: DC, DY, SDC, IPCA	0,097892	0,109603
FATN11	XGBoost: DY, SEL, IPCA, IGPM	0,024244	0,024669
FLMA11	XGBoost: CL, DC, SCC, IPCA	0,054116	0,06909
GALG11	XGBoost: DC, DY, SCC, IGPM	0,020657	0,028772
HBRH11	XGBoost: CL, IPCA, IGPM, IFIX	0,080873	0,084891
HGRU11	XGBoost: DY, SCC, IPCA, IFIX	0,087816	0,088727
HOSI11	XGBoost: CL, DY, DC, DY, SCC, IPCA	0,030711	0,034882
JSRE11	XGBoost: CL, SDC, IPCA, IFIX	0,095607	0,109169
KNRI11	XGBoost: CL, DC, DY, SCC, SEL, IPCA	0,042747	0,057936
MFAI11	ARIMA	0,078370	0,093317
MFII11	XGBoost: DC, SDC, SCC, IPCA	0,037834	0,039506
MXRF11	XGBoost: DC, DY, SCC, IGPM	0,021448	0,026110
NEWL11	XGBoost: DY, DC, SEL, IGPM, IFIX	0,071240	0,076045
PATL11	XGBoost: DC, DY, SDC, SEL, IPCA, IFIX	0,073929	0,080912
PLCR11	ARIMA	0,043259	0,055963
QAMI11	XGBoost: SDC, SCC, IGPM, IFIX	0,035293	0,036009
RBRD11	XGBoost: CL, DY, DY, SDC, SEL	0,118916	0,131658
RBRS11	XGBoost: DC, SEL, IPCA, IGPM	0,025504	0,025928
RECT11	XGBoost: DC, DY, SDC, SCC, IGPM, IFIX	0,010874	0,010971
RELG11	XGBoost: DY, DY, SEL, IGPM	0,007324	0,009980
RZTR11	XGBoost: SDC, SCC, IPCA, IGPM, IFIX	0,019182	0,022136
SARE11	XGBoost: DY, DC, DY, IGPM, IFIX	0,014529	0,018227

Tabela 28 – Melhores modelos para cada fundo híbrido individualmente no médio prazo

Para a predição de 6 meses, o RMSE médio ficou em **0,063718**, enquanto o melhor RMSE obtido havia sido **0,079148** (estratégia E2 - XGBoost)

4.8.3 Fundos de Tijolo

4.8.3.1 Curto prazo

Para o intervalo de um mês, os modelos que apresentaram o menor RMSE para cada um dos fundos de tijolo é apresentado na tabela 29.

Fundo	Modelo	RMSE	MAE
AIEC11	XGBoost: CL, DC, SDC, SEL, IPCA, IFIX	0,038344	0,048167
ALZR11	XGBoost: CL, DC, SDC, SCC, IGPM, IFIX	0,014146	0,018109
BBFI11B	XGBoost: SCC, SEL, IPCA, IGPM	0,028790	0,036828
BBPO11	XGBoost: DY, SDC, SCC, IGPM	0,017190	0,023088
BBRC11	XGBoost: CL, DY, DC, DY, IPCA, IGPM, IFIX	0,028099	0,038797
BLMG11	ARIMA	0,013032	0,017951
BMLC11	XGBoost: DY, DY, SCC, IFIX	0,021826	0,025553
BNFS11	ARIMA	0,020803	0,027085
BRCO11	XGBoost: CL, DC, SEL, IGPM, IFIX	0,041553	0,056919
BTAL11	XGBoost: DY, DY, IPCA, IGPM, IFIX	0,036907	0,043670
BTLG11	XGBoost: DC, DY, SDC, IPCA	0,023763	0,031424
CBOP11	XGBoost: DC, DY, SDC, IGPM, IFIX	0,027107	0,044673
CEOC11	XGBoost: CL, SDC, SCC, SEL	0,039057	0,049975
CNES11	ARIMA	0,018437	0,023683
CXCO11	XGBoost: DY, DC, DY, IFIX	0,038695	0,048602
DRIT11B	XGBoost: CL, DC, DY, IGPM	0,067410	0,080008
EDFO11B	ARIMA	0,031040	0,033045
EDGA11	ARIMA	0,045789	0,057450
EURO11	ARIMA	0,022295	0,025332
FAED11	XGBoost: CL, DC, SCC, IPCA, IFIX	0,031710	0,037702
FCFL11	ARIMA	0,045622	0,050546
FIGS11	ARIMA	0,055440	0,071315
FIIP11B	XGBoost: DY, SDC, SEL, IFIX	0,048658	0,068298
FPAB11	XGBoost: DY, DY, SEL, IGPM	0,021604	0,026047
GGRC11	XGBoost: DY, SDC, SCC, IPCA, IGPM, IFIX	0,009726	0,011565
GTWR11	XGBoost: CL, DY, SDC, IFIX	0,032649	0,037468
HAAA11	ARIMA	0,019356	0,025586
HCTR11	ARIMA	0,029881	0,034823
HGBS11	ARIMA	0,051078	0,090990

HGLG11	XGBoost: DC, DY, SEL, IPCA, IFIX	0,010401	0,012001
HGPO11	XGBoost: DC, DY, SCC, IPCA, IFIX	0,017768	0,022494
HGRE11	ARIMA	0,039686	0,045134
HLOG11	ARIMA	0,015154	0,019653
HPDP11	ARIMA	0,041267	0,046452
HSLG11	ARIMA	0,050464	0,055332
HSML11	XGBoost: CL, DY, DC, DY, SCC, SEL, IPCA	0,026253	0,033804
HUSC11	XGBoost: CL, DY, DC, DY, SDC, IPCA	0,026383	0,037603
JFLL11	ARIMA	0,050284	0,057925
LGCP11	ARIMA	0,036510	0,043339
LUGG11	ARIMA	0,017397	0,024516
MALL11	XGBoost: CL, DY, DC, DY, IGPM, IFIX	0,010327	0,012528
MGHT11	ARIMA	0,024145	0,028933
NVHO11	XGBoost: DY, SDC, IGPM, IFIX	0,026071	0,029176
ONEF11	XGBoost: DC, DY, SDC, SEL, IPCA, IFIX	0,014360	0,019500
PATC11	XGBoost: DY, DY, SDC, IGPM	0,023456	0,027338
PLOG11	ARIMA	0,043050	0,048718
PQAG11	XGBoost: SDC, SEL, IPCA, IFIX	0,010922	0,012782
RBED11	ARIMA	0,032688	0,035069
RBRL11	XGBoost: CL, DY, DY, SDC	0,056150	0,072378
RBRP11	ARIMA	0,024371	0,031651
RBVA11	ARIMA	0,022071	0,027677
RCRB11	XGBoost: DY, SDC, SCC, IPCA	0,043925	0,057393
RECX11	ARIMA	0,051044	0,056283
RNGO11	ARIMA	0,033489	0,040395
SDIL11	ARIMA	0,035442	0,040237
SEQR11	XGBoost: DY, DC, DY, SDC, IPCA	0,021713	0,029083
SPTW11	XGBoost: CL, SCC, IPCA, IFIX	0,025683	0,034834
TEPP11	ARIMA	0,045155	0,059491
TGAR11	XGBoost: CL, DC, DY, SDC, SCC, SEL, IPCA, IGPM, IFIX	0,019424	0,024110
TRNT11	XGBoost: CL, SDC, SCC, IGPM, IFIX	0,035967	0,042150
TRXF11	ARIMA	0,021099	0,026726
URPR11	XGBoost: DY, DY, SCC, IPCA, IFIX	0,023428	0,044119
VGIP11	XGBoost: DY, DC, DY, SCC	0,020027	0,023366
VILG11	XGBoost: DC, SEL, IPCA, IGPM	0,028619	0,033912
VINO11	ARIMA	0,017130	0,025542
VISC11	XGBoost: DC, DY, SDC, SCC, IPCA, IGPM	0,018433	0,022639

VL0L11	XGBoost: CL, DY, IPCA, IGPM, IFIX	0,015208	0,017484
VSH011	XGBoost: CL, DY, DY, IPCA, IGPM	0,023204	0,032998
VSLH11	XGBoost: DC, DY, SEL, IPCA	0,017877	0,022751
VTLT11	XGBoost: DC, SDC, SCC, SEL, IPCA, IGPM	0,031451	0,037464
XPCI11	ARIMA	0,028449	0,032797
XPCM11	XGBoost: CL, SEL, IGPM, IFIX	0,096418	0,136718
XPIN11	ARIMA	0,039187	0,063519
XPLG11	XGBoost: DY, DY, SDC, IPCA	0,042505	0,048236
XPML11	ARIMA	0,024173	0,034447
XPPR11	XGBoost: CL, SDC, SEL, IFIX	0,049816	0,060703
XPSF11	ARIMA	0,035407	0,051282

Tabela 29 – Melhores modelos para cada fundo de tijolo individualmente no curto prazo

Para a predição de 1 mês, o RMSE médio ficou em **0,040233**, enquanto o melhor RMSE obtido havia sido **0,055105** (estratégia E1 - Arima)

4.8.3.2 Médio prazo

Para o intervalo de seis mês, os modelos que apresentaram o menor RMSE para cada um dos fundos de tijolo é apresentado na tabela 30.

Fundo	Modelo	RMSE	MAE
AIEC11	XGBoost: DC, SDC, SCC, IPCA	0,028878	0,033882
ALZR11	ARIMA	0,090628	0,103264
BBFI11B	XGBoost: SCC, SEL, IGPM, IFIX	0,074775	0,101386
BBPO11	XGBoost: DC, SDC, SCC, IGPM	0,091296	0,113342
BBRC11	XGBoost: DY, DY, SCC, IFIX	0,051065	0,064773
BMLC11	XGBoost: SDC, SCC, SEL, IPCA, IGPM, IFIX	0,074056	0,084370
BNFS11	XGBoost: DY, SCC, SEL, IPCA	0,015867	0,017190
BRCO11	XGBoost: CL, SCC, SEL, IPCA, IFIX	0,026391	0,032697
BTAL11	XGBoost: DY, SDC, SEL, IFIX	0,026208	0,034780
BTLG11	XGBoost: DC, DY, SDC, IGPM	0,023257	0,026840
CBOP11	XGBoost: DY, DC, SDC, SCC, IFIX	0,044318	0,065224
CEOC11	XGBoost: CL, DY, DY, SEL, IPCA	0,019753	0,023652
CNES11	XGBoost: DY, DY, IPCA, IFIX	0,127233	0,138838
CXCO11	XGBoost: DC, DY, SDC, SCC, SEL	0,047507	0,05017
DRIT11B	XGBoost: CL, DY, SDC, IGPM	0,070055	0,078769
EDFO11B	XGBoost: DC, DY, SCC, IPCA	0,033673	0,045729

EDGA11	XGBoost: CL, DC, SCC, SEL, IPCA, IGPM	0,016374	0,021287
EURO11	XGBoost: DY, SCC, IPCA, IFIX	0,054795	0,064215
FAED11	XGBoost: CL, DY, DC, SCC	0,057240	0,066441
FCFL11	XGBoost: CL, DY, SDC, SCC, IGPM	0,056408	0,077007
FIGS11	XGBoost: DY, SDC, SEL, IGPM	0,008273	0,010114
FIIP11B	XGBoost: CL, DY, SDC, SCC, SEL	0,006544	0,007057
FPAB11	XGBoost: DY, DC, SDC, SCC, SEL, IFIX	0,048024	0,050327
GGRC11	XGBoost: CL, SDC, IPCA, IGPM	0,059582	0,062483
GTWR11	XGBoost: SDC, IPCA, IGPM, IFIX	0,079335	0,094593
HAAA11	ARIMA	0,423229	0,858454
HCTR11	XGBoost: CL, SDC, SEL, IFIX	0,043265	0,054643
HGBS11	XGBoost: CL, DC, SCC, IFIX	0,099599	0,110525
HGLG11	XGBoost: CL, DY, DC, DY, SDC, IPCA, IFIX	0,033138	0,034316
HGPO11	XGBoost: SDC, SCC, SEL, IPCA	0,087910	0,090939
HGRE11	XGBoost: DY, SEL, IPCA, IGPM	0,031140	0,043484
HLOG11	XGBoost: DY, DC, DY, SEL	0,014217	0,015220
HPDP11	XGBoost: CL, DC, DY, SDC, SCC, SEL, IFIX	0,001097	0,001107
HSLG11	XGBoost: CL, DC, SDC, IGPM	0,090199	0,117531
HSML11	XGBoost: SDC, SCC, IPCA, IGPM	0,012366	0,015774
HUSC11	XGBoost: CL, SDC, SEL, IPCA	0,092685	0,099903
JFLL11	ARIMA	0,064841	0,077118
LGCP11	XGBoost: SDC, SEL, IGPM, IFIX	0,023346	0,025702
LUGG11	ARIMA	0,028185	0,031804
MALL11	XGBoost: DC, DY, IPCA, IGPM	0,024114	0,025285
MGHT11	XGBoost: SDC, SCC, SEL, IGPM	0,089084	0,090650
NVHO11	XGBoost: CL, DY, IPCA, IGPM, IFIX	0,080812	0,095104
ONEF11	XGBoost: CL, SDC, IPCA, IGPM, IFIX	0,005371	0,005703
PATC11	XGBoost: DC, SCC, SEL, IFIX	0,004449	0,006895
PLOG11	XGBoost: DC, SDC, IGPM, IFIX	0,052450	0,067543
PQAG11	ARIMA	0,029095	0,035427
RBED11	XGBoost: DY, SDC, SEL, IFIX	0,058762	0,075141
RBRL11	XGBoost: CL, DC, DY, SEL, IPCA, IFIX	0,034712	0,036705
RBRP11	ARIMA	0,057615	0,063576
RBVA11	ARIMA	0,036646	0,040892
RCRB11	XGBoost: SDC, SCC, SEL, IPCA	0,048270	0,049172
RECX11	XGBoost: DY, DC, DY, SDC, SEL	0,011871	0,014386
RNGO11	ARIMA	0,111740	0,125594
SDIL11	XGBoost: DY, DC, SDC, IPCA, IFIX	0,015407	0,015677

SEQR11	ARIMA	0,058957	0,083061
SPTW11	XGBoost: CL, DY, DC, IPCA	0,053931	0,091215
TEPP11	XGBoost: CL, DY, DC, IGPM	0,041149	0,050965
TGAR11	XGBoost: CL, SCC, IPCA, IGPM	0,054661	0,062817
TRNT11	XGBoost: DY, SDC, SEL, IFIX	0,055580	0,081531
TRXF11	XGBoost: DY, SCC, SEL, IGPM	0,022321	0,025665
URPR11	ARIMA	0,033550	0,049052
VGIP11	XGBoost: DY, SCC, IPCA, IGPM, IFIX	0,015329	0,019110
VILG11	XGBoost: CL, SDC, SCC, IPCA, IGPM	0,075132	0,080152
VINO11	ARIMA	0,020517	0,023402
VISC11	XGBoost: SDC, SCC, IPCA, IFIX	0,040509	0,042837
VLOL11	XGBoost: SCC, SEL, IPCA, IGPM	0,020743	0,023657
VSHO11	XGBoost: DY, DC, DY, SDC, IPCA, IFIX	0,039777	0,047770
VTLT11	XGBoost: DC, DY, SDC, SEL, IPCA	0,021237	0,025986
XPCI11	XGBoost: DY, DY, SEL, IFIX	0,047056	0,052646
XPCM11	XGBoost: CL, DY, DY, SCC, SEL	0,001902	0,002259
XPIN11	XGBoost: DC, DY, SDC, IGPM	0,047563	0,067067
XPLG11	XGBoost: CL, DY, SCC, SEL, IPCA	0,027655	0,036372
XPML11	XGBoost: CL, DC, SDC, IPCA, IGPM	0,067918	0,072120
XPPR11	XGBoost: DC, DY, SDC, IPCA, IGPM	0,039858	0,049002
XPSF11	XGBoost: DC, IPCA, IGPM, IFIX	0,058233	0,069596

Tabela 30 – Melhores modelos para cada fundo de tijolo individualmente no médio prazo

Para a predição de 6 meses, o RMSE médio ficou em **0,066481**, enquanto o melhor RMSE obtido havia sido **0,082537** (estratégia E2 - XGBoost)

4.8.4 Discussão

Para o curto prazo, conforme evidenciado na tabela 31, o erro médio obtido conseguiu ser 31,87% menor nos FIIs da categoria papel, 30,95% menor na categoria Tijolo e 30,55% menor na categoria Híbrido.

Categoria	RMSE médio obtido	Melhor RMSE anterior
Papel	0,029668	0,043546
Tijolo	0,039998	0,057927
Híbrido	0,040233	0,055105

Tabela 31 – Comparativa dos resultados (Curto prazo)

Já para o médio prazo, conforme evidenciado na tabela 32, o erro médio obtido

conseguiu ser 25,11% menor nos FIIs da categoria papel, 19,50% menor na categoria Tijolo e 19,45% menor na categoria Híbrido.

Categoria	RMSE médio obtido	Melhor RMSE anterior
Papel	0,056380	0,075279
Tijolo	0,063718	0,079148
Híbrido	0,066481	0,082537

Tabela 32 – Comparativa dos resultados (Médio prazo)

Observa-se claramente que a abordagem de modelagem personalizada apresentou consistentemente uma redução significativa no erro médio em ambas as sazonalidades e em todas as categorias avaliadas. No entanto, é crucial salientar o contraponto do considerável custo computacional exigido para a criação dessa extensa quantidade de modelos.

4.9 Q9: É possível criar um classificador de modelos com erros aceitáveis?

Ao analisar os resultados apresentados pela seção anterior, foi possível identificar que alguns fundos imobiliários demonstraram um desempenho superior utilizando o modelo XGBoost, enquanto outros alcançaram melhores resultados com o ARIMA. Em uma análise mais detalhada, constatou-se que, no curto prazo, 58,64% dos melhores desempenhos foram alcançados pelo XGBoost, enquanto 41,36% foram atribuídos ao ARIMA. Já no médio prazo, 85,99% dos resultados favoráveis foram obtidos através do XGBoost, com apenas 14,01% utilizando o ARIMA. Esses números evidenciam que diferentes fundos exibem comportamentos distintos ao longo do tempo e que a escolha do algoritmo adequado pode variar conforme as características individuais de cada fundo. Portanto, esses resultados enfatizam a complexidade na construção de modelos preditivos para fundos imobiliários.

Dito isso, uma hipótese a ser investigada é a viabilidade de criar um modelo que possa determinar qual estratégia é mais adequada para cada fundo imobiliário, levando em consideração as características dos FIIs e o desempenho dos modelos em diferentes cenários. Portanto, a seguir é investigado se um classificador de modelos pode ser utilizado para escolher a melhor estratégia (ARIMA ou XGBoost) para cada fundo.

Para isso, serão avaliados os resultados comparando os erros nos seguintes cenários:

- FIIs treinados com ARIMA.
- FIIs treinados com XGBoost.
- FIIs treinados com o modelo M1.

- FIIs treinados com o modelo M2.
- FIIs treinados com o modelo ideal, que é um modelo hipotético perfeito usado apenas como base de comparação para o teto dos resultados.

4.9.0.1 Curto prazo

Na tabela 33 são apresentados os erros médios obtidos para cada estratégia, no período mensal, considerando os classificadores M1 e M2. Além disso, serão apresentados os resultados obtidos caso todos os fundos fossem treinados utilizando a metodologia ARIMA, o algoritmo XGBoost e, por fim, os resultados alcançados quando a exata melhor escolha possível é selecionada.

Estratégia	ARIMA	XGBoost	M1	M2	Modelo ideal
S1 - Mensal	0,049891	0,044908	0,043303	0,045422	0,035328
S2 - Mensal	0,049891	0,044908	0,045173	0,047464	0,035328

Tabela 33 – Erros médios obtidos por cada estratégia no curto prazo

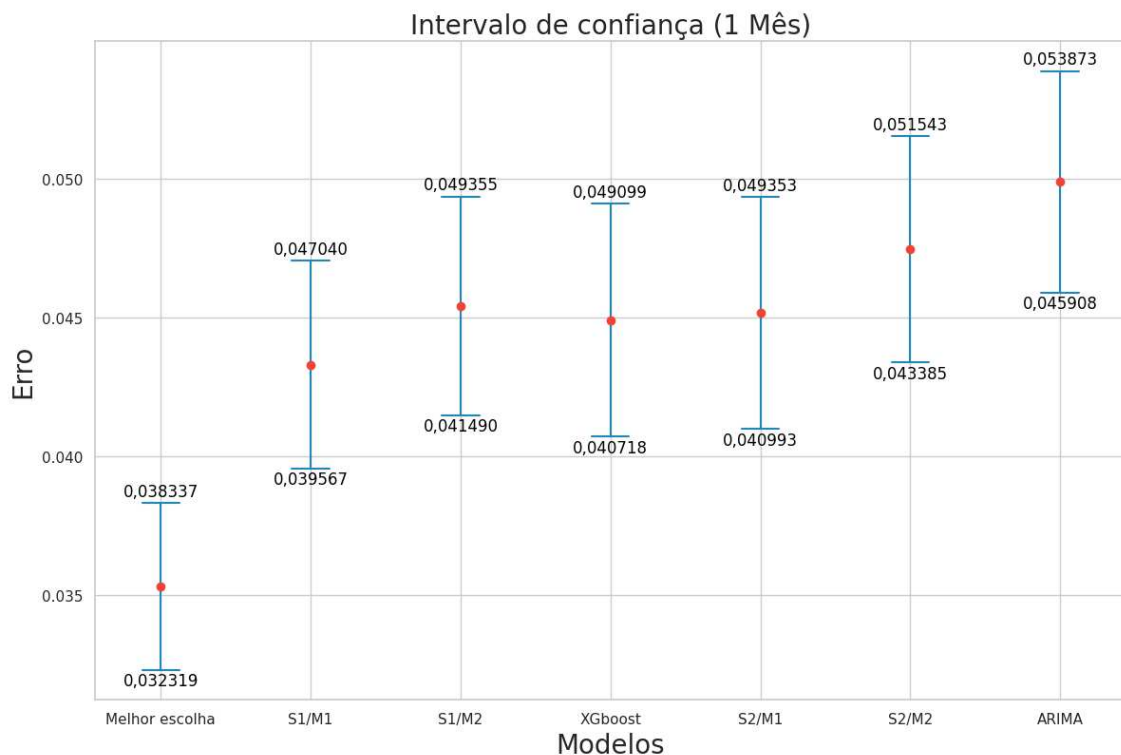


Figura 18 – intervalo de confiança dos erros de cada estratégia para o curto prazo

O classificador M1 ao utilizar a estratégia S1, apesar de haver sobreposição do intervalo de confiança com as demais estratégias, como visto na figura 18, demonstrou a menor dispersão e a menor mediana em comparação com os demais. Esses resultados corroboram a evidência de que essa estratégia se destaca como a opção mais promissora e eficaz.

4.9.0.2 Médio prazo

Como mencionado anteriormente, os resultados médios obtidos para cada estratégia serão apresentados novamente, desta vez considerando o período semestral.

Estratégia	ARIMA	XGBoost	M1	M2	Melhor escolha
S1 - Semestral	0.117670	0.058204	0.059216	0.058982	0.055464
S2 - Semestral	0.117670	0.058204	0.058805	0.059027	0.055464

Tabela 34 – Erros médios obtidos por cada estratégia no médio prazo

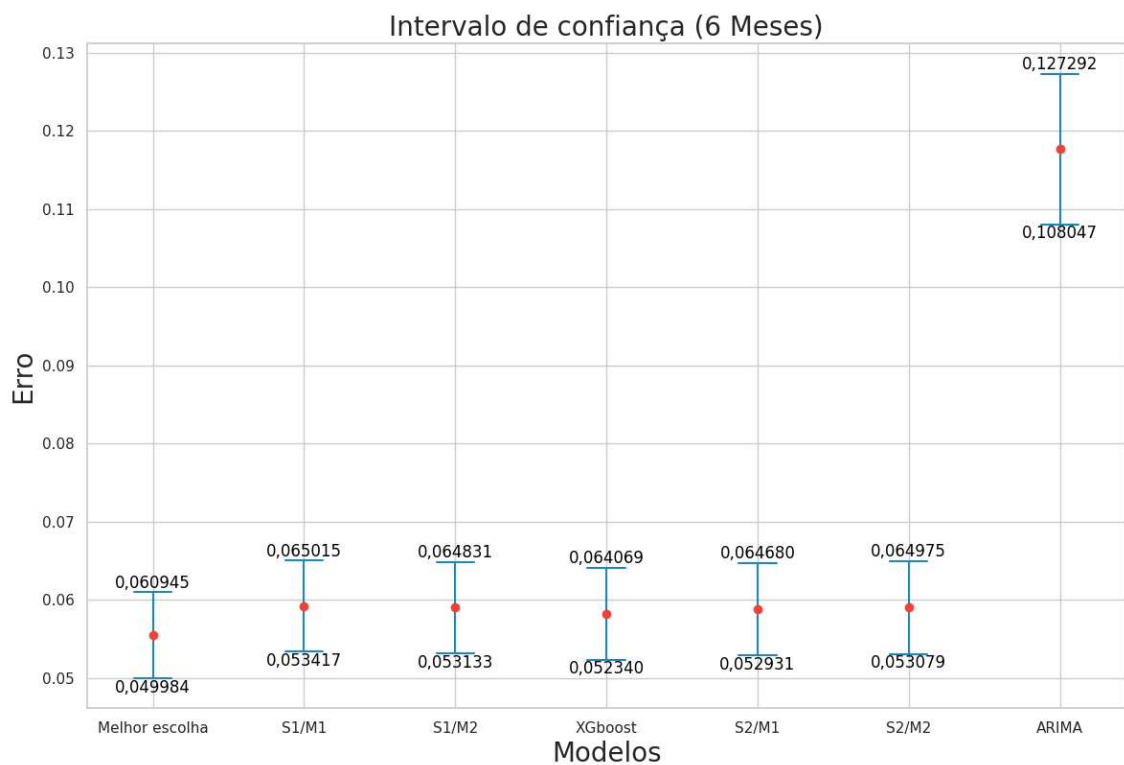


Figura 19 – intervalo de confiança dos erros de cada estratégia para o médio prazo

Como pode ser observado pela tabela 34 e pela figura 19, para o médio prazo a discrepância entre a melhor escolha e os demais classificadores foi relativamente pequena que torna-se insignificante tentar chegar a uma conclusão. Isso indica que, nesse contexto, as estratégias e modelos utilizados possuem um desempenho bastante semelhante, dificultando a identificação de uma opção superior clara.

5 Conclusões e Trabalhos Futuros

Este trabalho investigou estratégias para a previsão de valorização de Fundos de Investimento Imobiliário (FIIs) utilizando modelos de aprendizado de máquina em contraste com métodos tradicionais de séries temporais. A metodologia adotada incluiu a coleta e análise de um conjunto robusto de dados, abrangendo cotações diárias, dividendos, detalhes dos imóveis, entre outros atributos significativos. O estudo explorou diversas estratégias de modelagem, incluindo a construção de modelos específicos por fundo e um modelo geral para todos os fundos.

Um aspecto inovador do estudo foi o enriquecimento dos dados com informações adicionais sobre os imóveis, buscando entender como características específicas influenciam a valorização dos FIIs. Isso incluiu análises geoespaciais e a avaliação do impacto de fatores econômicos externos.

Os resultados demonstraram que, em muitos casos, os modelos de aprendizado de máquina superaram as técnicas tradicionais de análise de séries temporais, especialmente na previsão de tendências de médio prazo. Isso sugere um potencial significativo para a aplicação dessas tecnologias no setor financeiro.

A abordagem de desenvolver modelos personalizados para cada FII mostrou-se particularmente eficaz, evidenciando a importância de considerar as especificidades de cada fundo. Este resultado destaca o valor de abordagens de modelagem mais granulares e adaptadas.

Embora promissores, os resultados também revelaram desafios, incluindo a necessidade de extensos conjuntos de dados históricos e a sensibilidade dos modelos a anomalias do mercado. A pesquisa destacou a importância de técnicas robustas de tratamento de dados e a avaliação contínua da performance dos modelos.

Para futuras investigações, recomenda-se a exploração de modelos mais complexos, como redes neurais profundas e algoritmos de aprendizado por reforço, que podem capturar dinâmicas de mercado mais sutis e oferecer previsões ainda mais precisas. Outra direção importante para pesquisas subsequentes é a inclusão de outras variáveis macroeconômicas, para avaliar seu impacto sobre a valorização dos FIIs. Isso pode ajudar a entender melhor como fatores externos influenciam o mercado.

Uma extensão interessante seria a criação de um simulador financeiro que utilizasse as previsões dos modelos desenvolvidos neste estudo. Ao invés de simplesmente visualizar o desempenho dos modelos em termos estatísticos, como precisão e erro médio, este simulador permitiria aos investidores simular cenários de investimento com base nas previsões, considerando custos, taxas e outras despesas associadas. Isso proporcionaria

uma compreensão mais tangível e prática do valor das previsões geradas pelos modelos, ajudando os investidores a tomar decisões mais informadas e realistas.

Outra direção promissora é explorar uma abordagem de *stack* de classificadores para aprimorar a precisão das previsões. Nesta abordagem, os resultados dos modelos individuais podem ser usados como atributos de entrada para um meta-classificador, combinando suas previsões para gerar uma previsão final mais precisa. Esta técnica de “empilhamento” tem sido amplamente utilizada em problemas de aprendizado de máquina para melhorar a precisão das previsões, especialmente quando os modelos base têm vieses diferentes ou são especializados em diferentes aspectos do problema. Avaliar sua eficácia neste contexto pode fornecer insights valiosos sobre a otimização do desempenho dos modelos desenvolvidos neste estudo.

Este estudo abre novas perspectivas para a análise e previsão da valorização de FIIs, demonstrando o potencial do aprendizado de máquina nesse contexto. A pesquisa não só forneceu insights valiosos sobre a aplicabilidade de técnicas avançadas em finanças, mas também estabeleceu uma base sólida para futuras investigações que podem expandir ainda mais o conhecimento na área, beneficiando tanto acadêmicos quanto profissionais do mercado.

Referências

ANBIMA. **Princípios de Investimento. Material de Estudos 4 da Certificação CPA-10.** 2018. Disponível em: <<https://www.anbima.com.br/data/files/74/40/1D/33/466A4810EA926748882BA2A8/CPA-10-Cap4.pdf>>.

ANTÔNIO, R. M.; LIMA, F. G.; JUNIOR, T. P. Stock recommendations and investment portfolio formation: A study in the brazilian market. **Contaduría y administración**, Elsevier, v. 60, n. 4, p. 874–892, 2015.

B3. **Fundos de Investimento Imobiliário (FII).** 2021. Disponível em: <https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/fundos-de-investimento-imobiliario-fii.htm>.

BARBOZA, J. P. et al. Alocação de fundos de investimento imobiliários e estratégias de negociação: Carteiras elaboradas em algoritmos de reinforcement learning. **Encontro de Gestão e Negócios**, 2022.

BRANDÃO, I. V. **Framework de apoio à tomada de decisão no mercado de ações baseado em aprendizado por reforço profundo.** Dissertação (Mestrado), Universidade de Brasília, 2021.

BRAV, A.; LEHAVY, R. An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics. **The Journal of Finance**, Wiley Online Library, v. 58, n. 5, p. 1933–1967, 2003.

CARVALHO, A. L. d. P. **Modelagem econométrica para forecasting do IFIX.** Tese (Doutorado), Fundação Getúlio Vargas, 2019.

COSTA, K. V. S. d. et al. **Redes Neurais Artificiais para Análise e Previsão de Fenômenos Econômicos: Avaliando o Grau de Monetização da Economia Brasileira.** [S.l.], 2023.

DIMSON, E.; MARSH, P. An analysis of brokers' and analysts' unpublished forecasts of uk stock returns. **The Journal of Finance**, Wiley Online Library, v. 39, n. 5, p. 1257–1292, 1984.

FIGUEIREDO, L. J. A. S. et al. Moredata: A geospatial data enrichment framework. In: . New York, NY, USA: Association for Computing Machinery, 2021. (SIGSPATIAL '21), p. 419–422. ISBN 9781450386647. Disponível em: <<https://doi.org/10.1145/3474717.3484210>>.

FREIRE, D. V. **Forecasting U.S. Reit Index Prices with Artificial Neural Networks.** Dissertação (Mestrado), 2021. AAI29335585.

FREITAS, D. d. C.; REIS, D. L. Educação financeira por meio de uma plataforma interativa de investimento: a aplicabilidade da inteligência artificial (ia): Financial education through an interactive investment platform: the applicability of artificial intelligence (ia). **Brazilian Journal of Development**, v. 8, n. 11, p. 75525–75545, Nov. 2022. Disponível em: <<https://ojs.brazilianjournals.com.br/ojs/index.php/BRJD/article/view/54733>>.

GABRIEL, F. S.; POST, K. C. d. S. R.; ROGERS, P. Clustering real estate investment trusts: Brazil versus united states. **Journal of Management Research**, Macrothink Institute, 2015.

GERTSEKOVICH, D. et al. Return on investment in reit real estate funds. In: IOP PUBLISHING. **IOP Conference Series: Materials Science and Engineering**. [S.l.], 2019. v. 667, n. 1, p. 012025.

HELFENSTEIN, R.

Um modelo de rede neural de classificação para predição da tendência do preço bitcoin, 2022.

JR, J. da R. L. Fundos imobiliários têm futuro no brasil? **Real Estate Mercado**, Universidade de São Paulo (USP), 2011.

JUNIOR, W. B. Trabalho de Conclusão de Curso, **Aplicação de Machine Learning para Auxiliar o Processo Decisório de Investimento em Fundos Imobiliários**. Cachoeiro de Itapemirim: [s.n.], 2023. Disponível em: <<https://repositorio.ifes.edu.br/handle/123456789/3205>>.

KIM, Y.; JEONG, S. R.; GHANI, I. Text opinion mining to analyze news for stock market prediction. **Int. J. Advance. Soft Comput. Appl**, v. 6, n. 1, p. 2074–8523, 2014.

KINOSHITA, J. K.; MENEGHETTI, D. D. R.; BIANCHI, R. A. d. C. Aprendizado por reforço profundo com redes recorrentes aplicado à negociação do minicontrato futuro de dólar. **Anais do I Brazilian Workshop on Artificial Intelligence in Finance (BWAIF 2022)**, Sociedade Brasileira de Computação, 2022.

LAURIA, R. d. S. **Aplicação de sistemas de recomendação em finanças**. Tese (Doutorado), 2020.

LEITE, R. S. M. M. et al. Predição da cotação real/bitcoin usando a rede neural long short term memory. **Brazilian Journal of Development**, v. 9, n. 6, p. 21160–21177, 2023. Disponível em: <<https://ojs.brazilianjournals.com.br/ojs/index.php/BRJD/article/view/57659>>.

LINS, R. N. F.

Previsão de fundos de investimentos com o uso de machine learning, Londrina, 2020. 74 f. Trabalho de Conclusão de Curso (Graduação).

MACHADO, E. J.; ASSIS, C. A. S. de; PEREIRA, A. C. M. Modelagem, implementação e avaliação de estratégias de negociação baseadas em algoritmos de aprendizado de máquina para o mercado financeiro. **Revista Brasileira de Computação Aplicada**, v. 12, n. 1, p. 16–31, jan. 2020. Disponível em: <<https://seer.upf.br/index.php/rbca/article/view/9106>>.

MARTINEZ, L. C.; PEREIRA, A. M. Beating the stock market with a deep reinforcement learning day trading system. **2020 International Joint Conference on Neural Networks (IJCNN)**, p. 1–8, 2020. Disponível em: <<https://api.semanticscholar.org/CorpusID:221665598>>.

- MESQUITA, C. M.; OLIVEIRA, R.; PEREIRA, A. C. M. Utilização de uma rede neural lstm e testes da razão da variância para previsões em séries de ativos da bovespa. **Journal of Financial Forecasting**, 2024.
- MESQUITA, C. M.; OLIVEIRA, R. Arantes de; PEREIRA, A. C. M. Combining an lstm neural network with the variance ratio test for time series prediction and operation on the brazilian stock market. In: **2020 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2020. p. 1–8.
- MOSHIRIAN, F.; NG, D.; WU, E. The value of stock analysts' recommendations: Evidence from emerging markets. **International Review of Financial Analysis**, Elsevier, v. 18, n. 1-2, p. 74–83, 2009.
- PAULUS, N.; LAUTENSCHLAEGER, L.; SCHÄFERS, W. **Social Media and Real Estate: Do Twitter users predict REIT performance?** Working Paper, 2023. Disponível em: <https://ideas.repec.org/p/arz/wpaper/eres2023_200.html>.
- PORTALDOINVESTIDOR. **Fundos de investimento**. 2021. Disponível em: <https://www.investidor.gov.br/menu/Menu_Investidor/fundos_investimentos/introducao>.
- REZENDE, S. O. Predição do movimento do Índice ibovespa a partir de notícias. 2021.
- ROZÁRIO, J. G. d. **Aprendizado de Máquina Aplicado à Seleção de Fundos de Investimento Imobiliários**. Dissertação (Trabalho de Conclusão de Curso de Especialização) — Universidade Tecnológica Federal do Paraná, Campus de Dois Vizinhos, 2022. Curso de Especialização em Ciência de Dados.
- SANTOS, F. R. d. S. **Predição de preço de ações utilizando aprendizado de máquina**. Monografia (Monografia) — Universidade Federal de Ouro Preto, 2022.
- SANTOS, R. N. **FIIS como alternativa para rentabilizar capital ao invés do tradicional investimento brasileiro da aquisição direta de imóveis**. Trabalho de Conclusão de Curso, Curitiba, Brasil, 2019.
- SANTOS, V. V. d. **Avaliação do método Random Forest para auxílio à seleção de fundos de investimento imobiliário listados na Bolsa de Valores**, 2023.
- STEFFEN, M. A. **Rentabilidade dos fundos de investimento imobiliários: uma análise da influência da Bolsa de Valores, dos juros e de fatores específicos do mercado**. Dissertação (Mestrado) — Universidade de São Paulo, 2015.
- UTANI, L. A.; SILVA, G. R. d.; FARIA, E. B. P. Sistema especialista para apoio a decisão de investidores na aquisição de fundos imobiliários. **Revista Eletrônica Anima Terra**, VII, n. 15, p. 105–118, 2º semestre 2022. ISSN 2526-1940. Mogi das Cruzes-SP. Disponível em: <[URL_DO_ARTIGO](#)>.
- ZANANDREA, V. et al. **Avaliação de fundos de investimento imobiliário no Brasil**. Dissertação (Mestrado), Universidade Federal de Santa Maria, 2018.
- ZHANG, W.; LI, B.; LIEW, A. e. a. Predicting the returns of the us real estate investment trust market: evidence from the group method of data handling neural network. **Financ Innov**, v. 9, p. 98, 2023. Disponível em: <<https://doi.org/10.1186/s40854-023-00486-2>>.