

# Predicting Real Estate Funds: A Comparative Study of Machine Learning and Time Series Methods

Henrique Diniz<sup>1</sup>, Paulo Carneiro<sup>1</sup>, Fabrício A. Silva<sup>1</sup>

<sup>1</sup>Laboratório de Inteligência em Sistemas Pervasivos e Distribuídos (NESPeD-Lab)  
Universidade Federal de Viçosa (UFV) – Florestal – MG – Brazil

{henrique.p.diniz, paulo.h.carneiro, fabricio.asilva}@ufv.br

**Abstract.** *This work investigates different strategies for predicting the price variation of Real Estate Investment Funds (FIIs) using machine learning models, in comparison with a traditional time series method. We analyze the performance of the models in terms of fund categories (i.e., paper, brick, or hybrid), model settings (i.e., one model per fund or a general model), and forecast time window (i.e., 6 months or 1 month). An analysis was also carried out considering an enrichment of the data with characteristics of the properties belonging to the funds, which is a pioneering contribution of this work. The results reveal, among other conclusions, that machine learning models outperform the time series technique only for the medium term, and that the information on the properties belonging to the funds was important for improving forecasts.*

**Keywords:** Real Estate Investment Funds, Machine Learning, Time Series

## 1. Introdução

A crescente demanda por investimentos em fundos imobiliários tem chamado a atenção de investidores e acadêmicos, impulsionando a necessidade de uma análise mais aprofundada sobre esses ativos. Devido a sua popularização e recente ascensão, novos investidores são atraídos para o setor, visando ganhos maiores que os obtidos com renda fixa. Importante fator atrativo para esse tipo de investimento é o menor risco quando comparado com a compra de ações, resultando em maior interesse de pessoas físicas com menos capital disponível e mais conservadoras.

Os Fundos de Investimento Imobiliário (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. A maioria dos FIIs existentes no país segue a tendência internacional, focando em investimentos imobiliários para gerar renda de longo prazo a partir de imóveis comerciais. 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 [da Rocha Lima Jr 2011].

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.

O objetivo deste trabalho é analisar diferentes estratégias para a construção de modelos de previsão de fundos imobiliários. Serão respondidas perguntas relacionadas a quais técnicas utilizar, como treinar os modelos, se é relevante utilizar dados dos imóveis físicos, se técnicas de séries temporais funcionam bem e traçar caminhos iniciais para que outros pesquisadores possam contribuir e avançar na área de previsão de FIIs. Os resultados revelam, dentre outras conclusões, que modelos de aprendizado de máquina superam um método de séries temporais para uma perspectiva de previsão de 6 meses, e que o uso de características dos imóveis dos fundos ajuda a melhorar as previsões.

O restante do texto está organizado da seguinte forma: na Seção 2, será apresentada uma revisão da literatura relevante ao tema em questão. Em seguida, na Seção 3 serão descritos os procedimentos utilizados neste trabalho. Na Seção 4, os resultados da pesquisa serão apresentados e discutidos. O trabalho é concluído na Seção 5.

## **2. Trabalhos relacionados**

Poucos trabalhos de pesquisa foram realizados até o momento no contexto de FIIs. O estudo realizado por [Soares 2008] foi um dos primeiros na área, e foi feita uma análise preditiva para um único fundo imobiliário no ano de 2008. Apesar de pioneiro, os resultados estão desatualizados e nenhum avanço foi feito desde então. Os autores de [Zanandrea et al. 2018] criaram um modelo de regressão que considera a política de investimento, segmento, número de imóveis, características do gestor e taxas de administração, dentre outras variáveis. Como o viés comparativo temporal também é importante na escolha do fundo, [Manganotti 2014] verificou a ocorrência ou não de retorno anormal dos FIIs, no período de 2011 a 2013, fazendo comparação com o de imóveis comerciais que propiciam renda, entre o período de 2000 a 2013 no Brasil.

[Gertsekovich et al. 2019] consideram as condições de formação da carteira diversificada com abundância de FIIs de vários setores de propriedade residencial, hipoteca e mista. A respeito da formação de investimentos, determinaram o desenvolvimento de um modelo *Risk-Return*. Por outro lado, [Helfenstein et al. 2020] conduziram um estudo que avaliou sistematicamente fundos imobiliários nacionais e internacionais para investimentos, com o acompanhamento de ativos para alocação de recursos visando a proteção de patrimônio, através de fundos imobiliários.

O aprendizado de máquina é considerado essencial para apoiar a decisão dos investidores, de acordo com [Brandão 2021]. Com o objetivo de maximizar os lucros, é apresentada uma ferramenta capaz de aprender e reconhecer padrões para apoiar a tomada de decisões de investimentos. A modelagem econométrica para predição do índice IFIX (Índice de Desempenho de Fundos Imobiliários) também merece destaque conforme [Carvalho 2019]. O autor tinha como objetivo estimar o comportamento futuro do IFIX, uma vez que é bastante atrativo a investidores pessoas físicas. Diversos índices econômicos, tais como IBOV, IPCA, IGPM, Câmbio, IFIX, NTN-B, IMA-B, IVG-R e MVG-R, foram minuciosamente examinados por [Alves 2021] quanto à sua influência sobre os preços dos FIIs. Como resultado, concluiu-se que a maior parte dos retornos desses fundos não pode ser explicada por fatores externos, mas sim por características

específicas de cada um deles.

De acordo com [Silva 2017], em sua pesquisa sobre os determinantes na escolha de um FII, foram utilizadas algumas características imobiliárias do fundo, como quantidade de imóveis, setor, locador e localização. Porém, o autor concluiu que tais variáveis são irrelevantes na escolha de um fundo imobiliário. Diferentemente, este trabalho atual propõe um enriquecimento mais detalhado sobre os imóveis dos fundos, a fim de analisar outros possíveis fatores que possam influenciar na variação de preço.

Vale destacar que grande parte dos trabalhos relacionados são da área das ciências econômicas, havendo ainda espaço para mais estudos relacionados ao uso de inteligência artificial nesta modalidade de investimento. Destaca-se a escassez de trabalhos que abordem a previsão dos preços de fundos imobiliários. Diferente dos estudos anteriores, o presente trabalho visa a construção e avaliação de diferentes modelos de predição de FIIs. Vale ressaltar que este é o primeiro estudo que se aprofunda nas características dos imóveis físicos dos fundos imobiliários para a construção dos modelos.

### 3. Metodologia

Nesta seção são apresentados 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 obtidos, incluindo a definição dos critérios utilizados para a avaliação do desempenho dos modelos.

#### 3.1. Coleta dos Dados

Foi estabelecido um protocolo de coleta de dados durante o período compreendido entre janeiro de 2021 e dezembro de 2022. Para realizar a coleta, foram empregadas técnicas de *web scraping* e a utilização de APIs, obtendo-se informações das seguintes fontes:

- **Funds Explorer:** Lista de Fundos Imobiliários listados na Bolsa de Valores<sup>1</sup>; Relação de dividendos por fundo pagos ao longo dos meses<sup>2</sup>;
- **InfoMoney:** Preço de fechamento por fundo ao longo dos meses<sup>3</sup>; Endereços de todos os imóveis dos fundos<sup>4</sup>; Valor do índice IFIX ao longo dos meses<sup>5</sup>.
- **Dados de Mercado:** Índices IGPM, SELIC e IPCA ao longo dos meses<sup>6</sup>;
- **Investing.com:** Desdobramentos de FIIs ocorridos ao longo do tempo<sup>7</sup>.

#### 3.2. 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.

---

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

<sup>2</sup><https://www.fundsexplorer.com.br/funds>

<sup>3</sup><https://fii-api.infomoney.com.br/api/v1/fii/cotacao/historico>

<sup>4</sup><https://fii-api.infomoney.com.br/api/v1/propertie/>

<sup>5</sup><https://www.infomoney.com.br/cotacoes/b3/indice/ifix/>

<sup>6</sup><https://www.dadosdemercado.com.br/economia/>

<sup>7</sup><https://br.investing.com/stock-split-calendar/>

Outro critério adotado no tratamento dos dados é a remoção de fundos imobiliários que não possuem pelo menos 20 registros de histórico, por não estarem ainda consolidados.

Além disso, também são removidos FIIs que possuem dados faltantes ou que apresentaram uma variação de 0% no preço da cota por três meses consecutivos. A disponibilidade desses dados é crucial para a análise, e a ausência ou inconsistência deles pode prejudicar a qualidade dos resultados obtidos. Por conseguinte, é essencial utilizar apenas dados completos e confiáveis para garantir a integridade das análises realizadas.

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 para prevenir a ocorrência de anomalias na variação do preço.

A seguir, procede-se com o enriquecimento dos dados de endereços por meio da obtenção das coordenadas geográficas dos imóveis. Para isso, utiliza-se uma expressão regular que procura por dois valores decimais separados por vírgula, imediatamente após o símbolo "@" na coluna que contém o link do Google Maps<sup>8</sup> com a localização do imóvel. Caso nenhum valor de coordenada geográfica seja obtido na etapa anterior, é realizada uma requisição para a URL do Google Places<sup>9</sup> com a inserção de todas as informações de endereço disponíveis, separadas por vírgula, após a última barra. Essa solicitação pode resultar na obtenção das coordenadas geográficas aproximadas do imóvel, que serão extraídas de acordo com o mesmo padrão mencionado anteriormente. Se ainda assim nenhuma coordenada tiver sido obtida, o campo ficará vazio.

Em seguida, tendo em vista que certos fundos imobiliários são formados por imóveis físicos, foi possível expandir as informações disponíveis através da ferramenta MoreData[Figueiredo et al. 2021], disponível publicamente. Foram incluídas informações sobre a quantidade de determinados tipos de estabelecimentos próximos a estas coordenadas, tais como shoppings tradicionais, prédios corporativos, lajes corporativas, galpões, agências bancárias, centros logísticos, faculdades, lojas de rua, terrenos, shoppings temáticos, hotéis e hospitais. Essa etapa visa enriquecer os dados de um fundo com informações sobre as regiões em que os seus imóveis físicos estão localizados.

Por fim, além dos atributos iniciais obtidos, outros foram derivados para complementar as informações e melhorar a precisão do modelo. Esses novos atributos representam as variações de outros já existentes, como o *CloseChange*, que é o atributo que se deseja prever. Adicionalmente, foram adicionados também atributos com um atraso temporal de 1, 3 e 6 meses. Estas novas colunas acrescentam informações relevantes para o modelo, o que pode resultar em melhorias na previsão.

### 3.3. Descrição dos dados

A seguir, serão apresentadas informações detalhadas sobre os atributos mais relevantes obtidos a partir da coleta e pré-processamento:

- *CloseChange*: indica a variação do preço de fechamento em relação ao intervalo de interesse e será utilizada como alvo nos modelos preditivos.

---

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

<sup>9</sup><https://www.google.com/maps/place/>

- *Close*: representa o preço de fechamento da cota do FII no mês.
- *Dividends*: representa o valor distribuído aos acionistas do FII no mês.
- *DividendsChange*: indica a variação do preço dos dividendos em relação ao intervalo de interesse e é um dos principais atrativos dos fundos imobiliários.
- *DividendYield*: indicador financeiro que mede a rentabilidade de um fundo imobiliário e é calculado dividindo o valor dos dividendos pelo preço do ativo.
- *DividendYieldChange*: indica a variação do *dividend yield* em relação ao intervalo de interesse e é útil para avaliar a tendência da rentabilidade do fundo.
- *Sector*: indica o setor de atuação do FII (Títulos e Valores Mobiliários, Lajes Corporativas, Logística, Híbrido, Outros, Shoppings, Residencial, Hotel ou Hospital).
- *SectorCloseChangeMean*: representa a variação do preço médio de fechamento de FIIs do mesmo setor.
- *SectorDividendsChangeMean*: indica a variação do preço médio dos dividendos de FIIs do mesmo setor.
- *Selic*: valor da SELIC (taxa básica de juros da economia) no mês.
- *IFIX*: valor do "Índice de Fundos de Investimentos Imobiliários" no mês.
- *IGPM*: valor do "Índice Geral de Preços do Mercado" no mês.
- *IPCA*: valor do "Índice Nacional de Preços ao Consumidor Amplo" no mês.

Além disso, outro fator relevante durante o processo de predição foi a categorização dos fundos, que podem ser classificados em três categorias:

**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.

### 3.4. Métodos

Para a geração dos modelos de previsão de preços dos ativos imobiliários, foram utilizados três algoritmos distintos: ARIMA, XGBoost e *Long Short Term Memory* (LSTM). O ARIMA foi escolhido por ser uma técnica muito conhecida para previsão de séries temporais. Apesar de existirem técnicas mais novas, 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. Por fim, o LSTM foi escolhido por ser uma técnica de redes neurais muito usada em problemas relacionados com investimentos.

Cada um desses algoritmos foi aplicado para gerar modelos com o objetivo de prever a variação do preço de fechamento dos fundos imobiliários com um horizonte temporal de 1 mês (curto prazo) e 6 meses (médio prazo). A combinação desses algoritmos e a utilização de diferentes horizontes temporais permitiram obter uma ampla gama de modelos que possibilitam analisar e prever o comportamento com diferentes níveis de precisão e abrangência temporal.

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

### 3.4.1. Estratégia 1: Um modelo por fundo

A **Estratégia 1** (E1) consiste no desenvolvimento de um modelo distinto para cada fundo imobiliário, avaliando todas as combinações possíveis de atributos para determinar qual modelo apresentou melhor desempenho para a maioria dos fundos.

Além disso, comparou-se a eficácia de técnicas de aprendizado de máquina (XG-boost), com um método tradicional de série temporal (ARIMA), em dois cenários distintos: curto e médio 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.

Por fim, foi avaliado se a previsão feita separadamente por categoria (Papel, Tijolo e Híbrido) apresentava resultados superiores àquela realizada de maneira geral.

### 3.4.2. Estratégia 2: Um modelo geral para todos os fundos

A **Estratégia 2** (E2) considera um único modelo que abrange todos os fundos. Nesse sentido, foram criados modelos com diferentes atributos, escolhidos empiricamente.

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. 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).

## 3.5. Treinamento e Validação

Para a realização do treinamento e validação dos modelos, adotou-se a metodologia de *Walk Forward*, ilustrada na Figura 1, 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. Para a previsão de 6 meses, o processo é o mesmo, mas com as janelas maiores.

Foram utilizados os métodos de erro absoluto, erro médio quadrático (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.

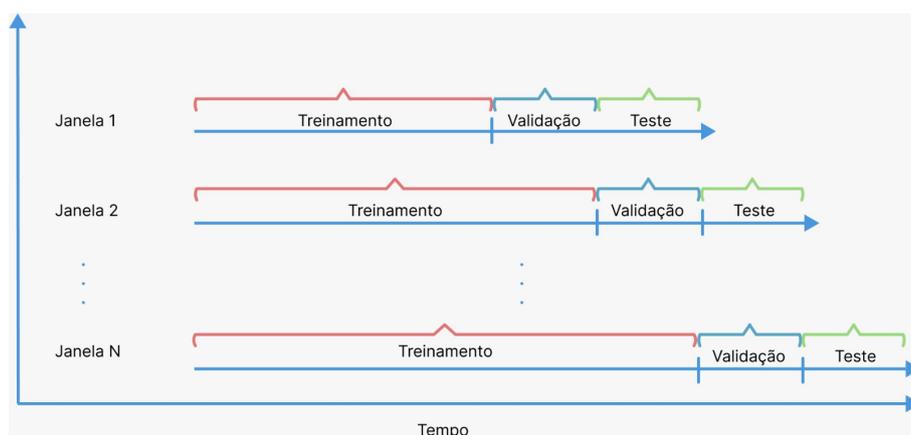


Figura 1. *Walk forward validation* [Kinoshita et al. 2022].

## 4. Resultados e Discussões

Nesta seção são apresentados os resultados visando responder a algumas perguntas. Os resultados apresentados são referentes aos melhores modelos encontrados com base em diferentes combinações de atributos.

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

Como pode ser visto na Tabela 1, a técnica de aprendizado de máquina que criava um modelo geral para todos os fundos (Estratégia 2) se saiu melhor para previsões de curto (1 mês) e médio prazo (6 meses), provavelmente por ser capaz de capturar relações mais complexas entre as variáveis e identificar padrões não lineares em dados históricos. Além disso, os modelos de aprendizado de máquina são capazes de lidar com grandes quantidades de dados de forma eficiente, permitindo que capturem nuances sutis e padrões que podem ser perdidos pelos métodos tradicionais de séries temporais.

Modelo	RMSE	MAE
E1 - ARIMA	0,056331	<b>0,039462</b>
E1 - XGBoost	0,064434	0,045758
E2 - XGBoost	<b>0,054430</b>	0,041151
E2 - LSTM	0,062336	0,045382

(a) Horizonte temporal de 1 Mês

Modelo	RMSE	MAE
E1 - ARIMA	0,150348	0,104505
E1 - XGBoost	0,133602	0,080815
E2 - XGBoost	<b>0,082978</b>	<b>0,063845</b>
E2 - LSTM	0,090898	0,068835

(b) Horizonte temporal de 6 Meses

Tabela 1. Melhores modelos de cada estratégia comparados ao ARIMA

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

A separação dos fundos em diferentes categorias não apresentou melhorias significativas nos resultados, como pode ser visto nas Tabelas 2 para os fundos de papel, 3 para os híbridos e 4 para os de tijolo, em comparação com a média geral na Tabela 5. Em termos de previsão de curto prazo (1 mês), o ARIMA apresentou melhor desempenho apenas para os fundos de tijolo, enquanto que para os fundos híbridos e de papel, a Estratégia 2 utilizando XGBoost obteve melhores resultados. Entretanto, para previsões de

médio prazo (6 meses), a Estratégia 2 se mostrou superior para todos os tipos de fundos. Portanto, observa-se que a categorização dos fundos não é um fator determinante para a melhoria dos resultados na previsão da variação temporal dos FIIs.

Modelo	RMSE	MAE
E1 - ARIMA	0,045126	0,032468
E1 - XGBoost	0,052190	0,036700
E2 - XGBoost	<b>0,043549</b>	<b>0,029513</b>
E2 - LSTM	0,061486	0,046391

(a) Horizonte temporal de 1 Mês

Modelo	RMSE	MAE
E1 - ARIMA	0,118805	0,090949
E1 - XGBoost	0,080742	0,059437
E2 - XGBoost	0,075478	<b>0,051876</b>
E2 - LSTM	<b>0,075279</b>	0,054272

(b) Horizonte temporal de 6 Meses

**Tabela 2. Melhores modelos de cada estratégia para categoria Papel**

Modelo	RMSE	MAE
E1 - ARIMA	0,062200	<b>0,043555</b>
E1 - XGBoost	0,071832	0,050244
E2 - XGBoost	<b>0,057927</b>	0,044485
E2 - LSTM	0,080482	0,056590

(a) Horizonte temporal de 1 Mês

Modelo	RMSE	MAE
E1 - ARIMA	0,156162	0,108825
E1 - XGBoost	0,107010	0,079168
E2 - XGBoost	<b>0,079148</b>	<b>0,062825</b>
E2 - LSTM	0,084323	0,066216

(b) Horizonte temporal de 6 Meses

**Tabela 3. Melhores modelos de cada estratégia para categoria Híbrido**

Modelo	RMSE	MAE
E1 - ARIMA	<b>0,055105</b>	<b>0,041208</b>
E1 - XGBoost	0,068239	0,049951
E2 - XGBoost	0,065325	0,047277
E2 - LSTM	0,069525	0,051446

(a) Horizonte temporal de 1 Mês

Modelo	RMSE	MAE
E1 - ARIMA	0,171944	0,128760
E1 - XGBoost	0,162330	0,090787
E2 - XGBoost	<b>0,082537</b>	<b>0,063694</b>
E2 - LSTM	0,091898	0,071003

(b) Horizonte temporal de 6 Meses

**Tabela 4. Melhores modelos de cada estratégia para categoria Tijolo**

O desempenho dos modelos não apresentou variações significativas ao considerar as diferentes categorias de fundos imobiliários. Essa constatação é relevante, uma vez que a ausência de diferenças significativas entre os erros por categoria indica que o modelo possui uma capacidade consistente de fazer previsões precisas, independentemente da categoria do fundo imobiliário em questão.

Modelo	RMSE	MAE
E1 - ARIMA	<b>0,054144</b>	<b>0,039077</b>
E1 - XGBoost	0,064087	0,045632
E2 - XGBoost	0,055600	0,040425
E2 - LSTM	0,070498	0,051476

(a) Horizonte temporal de 1 Mês

Modelo	RMSE	MAE
E1 - ARIMA	0,148970	0,109511
E1 - XGBoost	0,116694	0,076464
E2 - XGBoost	<b>0,079054</b>	<b>0,059465</b>
E2 - LSTM	0,251500	0,063830

(b) Horizonte temporal de 6 Meses

**Tabela 5. Média dos melhores modelos comparados as modelos gerais**

### 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 2** que utilizam as informações dos imóveis físicos e os que não utilizaram. Essa comparação não faz sentido para a Estratégia 1, pois quando um modelo é criado por fundo, todos os atributos de imóveis serão idênticos para todas as entradas.

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, com o seu equivalente no conjunto C2 (em que os dados dos imóveis são adicionados). 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.

Desta maneira, para uma melhor visualização dos resultados, os modelos foram agrupados de dois em dois. Os modelos de cor azul representam aqueles que não utilizaram as informações dos imóveis físicos, já os modelos de cor laranja as utilizaram, como pode ser visualizado na Figura 2 para a previsão de 1 mês e Figura 3 para de 6 meses.

Com a inclusão dos atributos dos imóveis é possível visualizar uma pequena melhora em alguns modelos, como pode ser melhor observado na Tabela 6. Nessa perspectiva, tendo em vista a predição com horizonte temporal de 1 mês, foi possível observar uma melhora em apenas 2 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. Além disso, em relação aos modelos que apresentaram os melhores resultados, apenas um não utilizou os dados de imóveis, como destacado em **negrito** na Tabela 6.

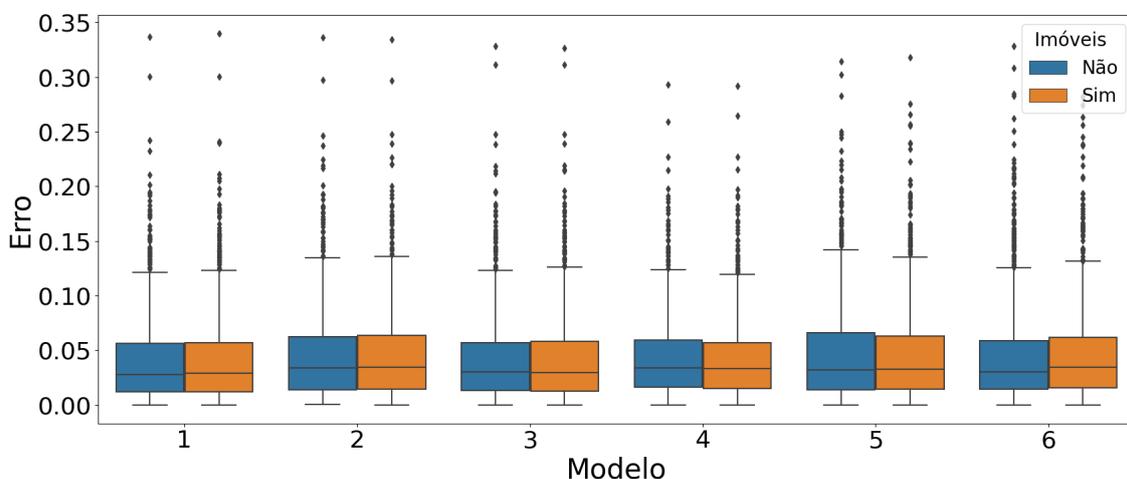


Figura 2. informação dos imóveis x sem informações dos imóveis para 1 mês

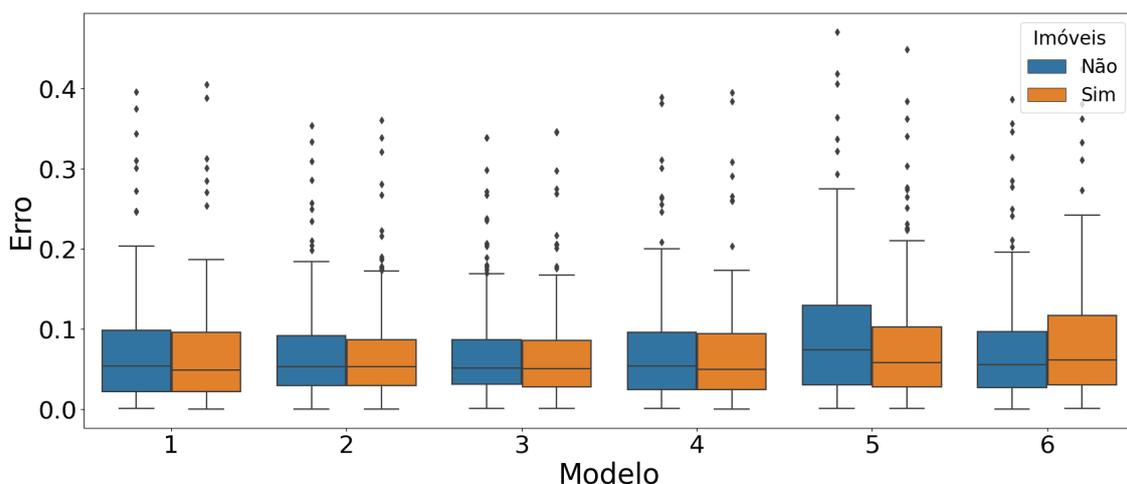


Figura 3. informação dos imóveis x sem informações dos imóveis para 6 meses

Meses		1		6	
Modelo	Imóveis	RMSE	MAE	RMSE	MAE
XGBoost	Não	0.056973	<b>0.040585</b>	0.091063	0.068062
	Sim	0.057352	0.040953	0.087407	0.065073
XGBoost	Não	0.060335	0.044585	0.086372	0.066843
	Sim	0.060673	0.044900	0.084629	0.065122
XGBoost	Não	0.057324	0.041399	0.084545	0.065044
	Sim	0.057750	0.041735	<b>0.082978</b>	<b>0.063845</b>
XGBoost	Não	0.055232	0.042214	0.089387	0.067649
	Sim	<b>0.054430</b>	0.041151	0.086922	0.065262
LSTM	Não	0.066055	0.047423	0.122220	0.092475
	Sim	0.062908	0.045415	0.101670	0.076491
LSTM	Não	0.066116	0.046098	0.090898	0.068835
	Sim	0.062336	0.045382	0.099929	0.077442

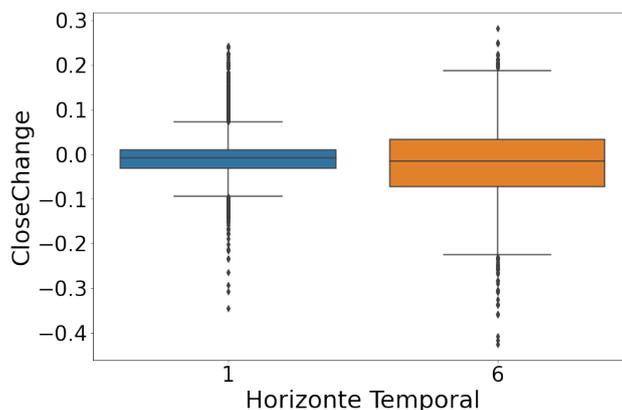
Tabela 6. Comparação dos modelos com e sem o uso das características dos imóveis.

Pode-se constatar que as informações sobre os imóveis físicos dos fundos não demonstraram uma melhoria significativa nos resultados. Essa observação indica a necessidade de investigações futuras que abordem a hipótese em questão por meio de testes de hipótese, a fim de avaliar de forma mais precisa qualquer impacto positivo que possa existir. Esta abordagem estatística será essencial para fornecer uma contribuição sólida e validada quanto à relação entre as informações imobiliárias e os resultados dos fundos, uma vez que os resultados do presente estudo não forneceram evidências convincentes de tal associação.

#### 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)?

Modelos com horizonte temporal de 1 mês têm significativamente métricas de erro menores que modelos de 6 meses, como mostra a Tabela 5 para a **Estratégia 1** e a Tabela 6 para

a **Estratégia 2**. Nessa perspectiva, pode-se destacar como obstáculos para os modelos de 6 meses, tanto uma menor quantidade de dados para treinamento, quanto uma maior variabilidade dos valores a serem preditos, como pode-se observar na Figura 4, que apresenta a variabilidade nos preços muito maior para o horizonte de 6 meses (barra laranja) que de 1 mês (barra azul).



**Figura 4. Boxplot da variação do "CloseChange"**

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

Em geral, os modelos LSTM apresentaram resultados piores que os modelos XGBoost para horizontes temporais de 1 e 6 meses, como é possível visualizar nas Tabelas 5 e 6. Em comparação ao ARIMA, os modelos apresentam piores resultados no curto prazo e melhores no médio prazo. Em contraponto, por se tratar de uma rede neural, pode-se em trabalhos futuros apresentar melhores resultados com uma base de dados maior em termos de período, e com uma melhor configuração das camadas da rede.

Pode-se depreender que, com os dados disponíveis, os modelos LSTM não conseguem apresentar os melhores resultados em nenhum dos cenários abrangidos. Porém, mais estudos devem ser realizados para que essa conclusão seja realmente aceita.

#### **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?**

Ao analisar a Tabela 1(a), observa-se que utilizar um modelo individual para cada fundo (E1) foi pior do que utilizar um modelo geral (E2) para a perspectiva de previsão de curto e médio prazo.

### **5. Conclusões e Trabalhos Futuros**

Este trabalho investigou diferentes estratégias para a previsão de fundos imobiliários. Com base nos resultados obtidos, pode-se concluir que diferentes técnicas de previsão têm desempenhos diferentes em diferentes horizontes de tempo.

Uma contribuição importante e pioneira deste trabalho foi avaliar o impacto da inclusão das informações dos imóveis físicos na previsão dos preços dos fundos imobiliários. No entanto, sugere-se que trabalhos futuros realizem um teste de hipóteses para

avaliar o impacto real desse incremento nos modelos treinados, além de um enriquecimento mais detalhado dos imóveis pertencentes ao fundo e melhorias nos modelos de redes neurais.

## **Agradecimentos**

Os autores agradecem à Fapemig e ao CNPq pelo suporte financeiro.

## **Referências**

- Alves, F. N. (2021). Fatores dinâmicos aplicados ao mercado de fundos imobiliários bvmf e cetip. Master's thesis, Universidade de Brasília.
- Brandão, I. V. (2021). Framework de apoio à tomada de decisão no mercado de ações baseado em aprendizado por reforço profundo. Master's thesis, Universidade de Brasília.
- Carvalho, A. L. d. P. (2019). *Modelagem econométrica para forecasting do IFIX*. PhD thesis, Fundação Getúlio Vargas.
- da Rocha Lima Jr, J. (2011). Fundos imobiliários têm futuro no brasil? *Real Estate Mercado*.
- Figueiredo, L. J. A. S., dos Santos, G. B., Souza, R. P. P. M., Silva, F. A., and Silva, T. R. M. B. (2021). Moredata: A geospatial data enrichment framework. SIGSPATIAL '21, page 419–422, New York, NY, USA. Association for Computing Machinery.
- Gertsekovich, D., Gorbachevskaya, L., Grigorova, L., and Peshkov, V. (2019). Return on investment in reit real estate funds. In *IOP Conference Series: Materials Science and Engineering*, volume 667, page 012025. IOP Publishing.
- Helfenstein, L. G. et al. (2020). A análise sistemática de fundos imobiliários nacionais e internacionais para investimento. Master's thesis, Universidade Estadual do Oeste do Paraná.
- Kinoshita, J. K., Meneghetti, D. D. R., and Bianchi, R. A. d. C. (2022). 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)*.
- Manganotti, K. H. (2014). *Estudo sobre a ocorrência ou não de retorno anormal nos Fundos Imobiliários de Renda e a comparação com o retorno de imóveis físicos que propiciam renda*. PhD thesis, Universidade de São Paulo.
- Silva, A. E. M. d. (2017). Determinantes de performance dos fundos de investimentos imobiliários brasileiros (fiis). Mestrado em Contabilidade e Finanças, Fucape Business School.
- Soares, F. (2008). Predição de séries financeiras utilizando wavelets e redes neurais: um modelo para os fundos de investimentos imobiliários. *Programa de Pós-Grad. em Sistemas e Processos Industriais - Mestrado*.
- Zanandrea, V. et al. (2018). Avaliação de fundos de investimento imobiliário no brasil. Master's thesis, Universidade Federal de Santa Maria.