

# Dados dos Fundos de Investimento Imobiliários (FIIs) Brasileiros Enriquecidos

Henrique P. B. Diniz<sup>1</sup>, Fabrício A. Silva<sup>1</sup>, Thais R. M. Braga Silva<sup>1</sup>

<sup>1</sup>Laboratório de Inteligência em Sistemas Pervasivos e Distribuídos (NESPED-LAB)  
Universidade Federal de Viçosa, Florestal, MG, Brasil

{henrique.p.diniz, fabricio.asilva, thais.braga}@ufv.br

**Abstract.** *This work presents a dataset on Real Estate Investment Funds (FIIs) in Brazil. The data were collected from various sources and enriched with detailed information about the properties included in the funds, enabling their use in predictive models, portfolio recommendation systems, and fundamental analysis, among other applications. The dataset is designed to benefit both finance researchers and investors, fostering advancements in analysis and decision-making within the sector.*

**Resumo.** *Este artigo apresenta um conjunto de dados sobre Fundos de Investimento Imobiliários (FIIs) no Brasil. Os dados foram coletados de diversas fontes e enriquecidos com informações detalhadas dos imóveis que compõem os fundos, possibilitando sua utilização em modelos preditivos, sistemas de recomendação de carteiras e análises fundamentalistas, entre outras aplicações. O conjunto de dados visa atender tanto pesquisadores em finanças quanto investidores, promovendo avanços na análise e tomada de decisões no setor.*

## 1. Introdução

Os investimentos são a espinha dorsal da economia global, agindo como motores que impulsionam o crescimento, a inovação e a estabilidade financeira. 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 e previsão de valores de ativos.

Os Fundos de Investimento Imobiliários (FIIs) são ativos que têm crescido bastante nos últimos anos. 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 [da Rocha Lima Jr 2011]. 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].

Uma grande dificuldade para quem começa a investir em FIIs é a escolha de quais fundos incluir em sua carteira de investimentos. Assim, constata-se uma escassez de estudos que realizem uma análise aprofundada sobre os FIIs, suas rentabilidades e a geração de modelos preditivos para estimar ganhos futuros. Um dos motivos para essa escassez é a falta de dados consolidados sobre os FIIs.

O objetivo deste trabalho é extrair, preparar, enriquecer e disponibilizar um conjunto de dados dos FIIs brasileiros, que possibilitará que pesquisadores e investidores possam realizar análises sobre esse tipo de investimento. Como os FIIs se diferenciam de outros fundos de investimento por estarem atrelados a imóveis, que são bens físicos, a extração feita neste trabalho tem um diferencial de enriquecer os dados dos FIIs com características da região em que os imóveis dos fundos se situam.

A criação da base de dados de FIIs seguiu o processo tradicional de ETL (extração, transformação e carregamento de dados). Primeiramente, os dados sobre os FIIs negociados no Brasil foram coletados, incluindo dividendos, preços de negociação e endereços dos imóveis. Em seguida, alguns índices relevantes para o mercado de FIIs foram também coletados, como inflação e taxa de juros. Esses dados foram então formatados e preparados, incluindo a transformação relativa aos desdobramentos que aconteceram com os FIIs. Por fim, os dados dos imóveis de cada fundo foram enriquecidos com características das regiões dos respectivos imóveis. Os dados foram então disponibilizados para acesso<sup>1</sup>.

O restante deste texto está organizado da seguinte forma. A Seção 2 discute os conjuntos de dados relacionados existentes. Em seguida, as Seções 3 e 4 descrevem os processos de extração e transformação dos dados, respectivamente. Algumas aplicações do conjunto de dados são listadas na Seção 5. Por fim, as conclusões estão na Seção 6.

## 2. Trabalhos Relacionados

Na literatura, existem dados disponíveis de outras aplicações financeiras, como da bolsa de valores [Cardoso et al. 2021]. Porém, poucos trabalhos relacionados a FIIs foram encontrados, como será descrito abaixo. A proposta de modelos para recomendar ou prever o comportamento de FIIs no Brasil e em outros países é ainda pouco explorada. Alguns trabalhos focam na recomendação de carteiras, como [Lauria 2020, SANTOS 2023, Rozário 2022, Utani et al. 2022, Barboza et al. 2022]. Porém, esses trabalhos não disponibilizam os seus dados para que o estudo seja replicado.

Já com foco nas predições, a maioria dos trabalhos focam em fundos imobiliários americanos. O trabalho de [Freire 2021] aplicou três redes neurais com diferentes arquiteturas, usando dados abrangentes de 20 anos para prever preços dos fundos. Posteriormente, [Paulus et al. 2023] analisaram cerca de 4 milhões de *tweets*, postados ao longo de 10 anos, para prever retornos de fundos, descobrindo que o sentimento social é um indicador crucial para o mercado. Por fim, no mesmo ano, [Zhang et al. 2023] concentraram-se em otimizar a precisão da previsão de preços de fundos, experimentando com vários tamanhos de amostra e parâmetros de algoritmos de aprendizado de máquina. Como nos outros trabalhos, os dados utilizados nestes estudos não foram disponibilizados.

Apenas dois trabalhos com foco no Brasil foram encontrados. O trabalho de [Zanandrea et al. 2018] propõe desenvolvimento de um modelo para avaliar a rentabil-

<sup>1</sup><https://github.com/NESPEDUFV/repositorio-dados-fiiis>

idade 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. Em um cenário mais recente, [Barbosa Junior 2023] utilizou dados de 120 FIIs entre fevereiro de 2019 e dezembro de 2021 para desenvolver um modelo de aprendizado de máquina. Porém, nenhum desses trabalhos disponibilizou os dados utilizados.

Existem diversos conjuntos de dados públicos sobre FIIs disponíveis no Kaggle, que podem ser úteis em análises sobre o mercado financeiro. Alguns exemplos são:

- **Fundos Imobiliários Brasil - Brazilian Funds FIIs**<sup>2</sup>: Este dataset contém uma lista completa de todos os FIIs, semelhante ao conjunto de dados mais básico obtido neste trabalho.
- **Fundos de Investimentos Imobiliários - FII - StatusInvest**<sup>3</sup>: Fornece o valor de participação de cada fundo presente no IFIX em uma data específica. Neste trabalho, o IFIX é analisado de maneira mais abrangente, considerando sua variação ao longo do tempo.
- **Movimentações em Fundos Imobiliários**<sup>4</sup>: Apresenta dados detalhados sobre as aquisições trimestrais de cada FII. Esse conjunto de dados é mais adequado para quem deseja realizar análises periódicas com foco em ciclos trimestrais.

### 3. Extração dos Dados

A obtenção dos dados foi realizada por meio de técnicas de *web scraping* e APIs, provenientes de múltiplas fontes. Optou-se por selecionar um período recente de janeiro de 2021 a dezembro de 2022. No entanto, vale ressaltar que o código para a extração também está disponibilizado no repositório para que os dados possam ser atualizados.

#### 3.1. FIIs listados na bolsa de valores

A extração de dados de fundos de investimento foi conduzida através de requisições diretas à API do do site *Funds Explorer*<sup>5</sup>. Utilizou-se a URL `https://www.fundsexplorer.com.br/wp-json/funds/v1/get-ranking` para obter os dados necessários.

A resposta da API retorna um JSON contendo informações detalhadas sobre cada FII, semelhantes ao exemplo abaixo:

```

1  [
2    {
3      "ticker": "ALZR11",
4      "dividendo": "0.7108",
5      "yeld": "0.6500",
6      "media_yield_3m": "0.6600",
7      "soma_yield_3m": "1.9800",
8      "media_yield_6m": "0.6733",
9      "soma_yield_6m": "4.0400",
10     "media_yield_12m": "0.6958",

```

<sup>2</sup><https://www.kaggle.com/datasets/maverickjpa/fundos-imobiliarios-brasil-brazilian-f>

<sup>3</sup><https://www.kaggle.com/datasets/mateuscpinheiro/fiis-statusinvest>

<sup>4</sup><https://www.kaggle.com/datasets/franckepeixoto/movimentaes-em-fundos-imobilirios>

<sup>5</sup><https://www.fundsexplorer.com.br/ranking>

```

11     "soma_yield_12m": "8.3500",
12     "variacao_cotacao_mes": "0.7372",
13     "rentabilidade": "-0.2217",
14     "rentabilidade_mes": "1.3920",
15     "cotacao_fechamento": "109.3200",
16     "soma_yield_ano_corrente": "4.7200",
17     "ano": "2024",
18     "vpa_yield": "0.6924",
19     "vpa": "105.7800",
20     "vpa_change": "-1.0385",
21     "pl": "1293278937.9800",
22     "setor": "Misto",
23     "valor": "108.48",
24     "liquidezmediadiaria": "1899874.64",
25     "patrimonio": "1293278937.98",
26     "pvp": "1.0256",
27     "ativos": "14",
28     "numero_cotista": "151678"
29 }
30 ]

```

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 de Ativos'.

### 3.2. Dividendos pagos ao longo do tempo

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

Para exemplificar o processo, considera-se o fundo **ALZR11**. Primeiramente, uma requisição HTTP foi realizada, adicionando o ticker do FII ao final da URL da API do Funds Explorer, conforme ilustrado a seguir: <https://www.fundsexplorer.com.br/wp-json/funds/v1/fund/ALZR11/incomes>. A resposta obtida em formato JSON foi então convertida para um *DataFrame* utilizando a biblioteca *pandas*.

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

```

1 [
2   {
3     "ticker": "ALZR11",
4     "created_at": "2024-07-22 19:07:54",
5     "updated_at": "2024-07-22 19:07:54",
6     "valor": "0.7108",
7     "tipo": "Rendimento",
8     "data_pagamento": "2024-07-25",
9     "data_base": "2024-07-18",
10    "yeld": "0.6500",
11    "rentabilidade": "-0.2217",
12    "cotacao_fechamento": "109.3200",
13    "variacao_cotacao_mes": "0.7372",

```

```
14     "mes": "7",
15     "ano": "2024"
16   }
17 ]
```

Para assegurar a consistência nos cálculos subsequentes de atributos derivados e prevenir erros de divisão por zero, valores nulos de dividendos foram substituídos por uma quantia simbólica de 1 centavo (R\$ 0,01).

### 3.3. 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:

```
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. 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.

### 3.4. Endereços dos imóveis dos FIIs

A relação de endereços dos imóveis por fundo foi 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://fii-api.infomoney.com.br/api/v1/property/ALZR11`. A resposta da API será semelhante ao 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*, *Latitude*, recebendo *NaN* como valor padrão e *Longitude*, também recebendo *NaN* como valor padrão.

### 3.5. Valor do índice IFIX ao longo do tempo

O índice IFIX consolida os valores de FIIs em um indicador que serve como termômetro para o setor. Os valores do IFIX ao longo do tempo também foram obtidos pelo InfoMoney. 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   }

```

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 JSON abaixo:

```

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.

### 3.6. 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 Mercado<sup>6</sup>. Para cada índice ( IGP-M, Selic e IPCA), foi repetido o processo abaixo.

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.

Ao final é gerada uma tabela mais condensada e estruturada, com a data (mês e ano) e o valor do índice, facilitando análises e visualizações temporais mais eficientes.

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

### 3.7. Desdobramentos de FIIs ocorridos ao longo do tempo

É importante destacar que um desdobramento 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.

A obtenção dos desdobramentos é realizada por meio de uma requisição POST à API do *Investing.com*, através da URL <https://br.investing.com/stock-split-calendar/Service/getCalendarFilteredData>. Os dados enviados na requisição devem seguir o seguinte exemplo:

```
1 {  
2   "country[]": "32",  
3   "dateFrom": "2019-01-01",  
4   "dateTo": "2024-01-01",  
5   "currentTab": "custom",  
6   "submitFilters": "1",  
7   "limit_from": "0"  
8 }
```

A resposta da requisição é processada para identificar cada bloco de desdobramento. Em seguida, expressões regulares são usadas para extrair a data, o ticker e a proporção do desdobramento no formato N:N.

Os dados extraídos são armazenados em um *DataFrame* do pandas.

### 3.8. 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, foi feito um enriquecimento dos dados com características dos imóveis de cada fundo. Esse é um diferencial deste conjunto de dados, já que adiciona atributos relevantes para as futuras aplicações.

Inicialmente, procede-se com o enriquecimento dos dados de endereços por meio da obtenção das coordenadas geográficas dos imóveis. Para isso, utiliza-se a expressão regular “(?:=@)[-]\*[\\d.]\*,-[\\d.]\*” para procurar por dois valores decimais separados por vírgula, imediatamente após o símbolo “@” no link do Google Maps<sup>7</sup> com a localização do imóvel, obtido na coleta dos endereços como descrito na seção 3.4.

Caso nenhum valor de coordenada geográfica seja obtido na etapa anterior, é realizada uma requisição para o Google Places<sup>8</sup> 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 Selenium<sup>9</sup> foi utilizada para tentar fazer a coleta individual das coordenadas faltantes. Para isso, a automação acessou cada uma das URLs dos

<sup>7</sup><https://www.google.com/maps>

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

<sup>9</sup><https://www.selenium.dev/>



endereços, aguardou o carregamento da página e após isso, aplicou a mesma expressão regular na URL carregada para a obtenção das coordenadas.

Em seguida, 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 alguns tipos de 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 e foram gerados atributos para cada informação, contendo a quantidade e a média das notas de avaliação.

### **3.8.1. Escolha dos tipos de estabelecimentos**

A escolha dos tipos de estabelecimentos para o enriquecimento dos dados foi realizada considerando aspectos que podem influenciar diretamente a percepção de valor dos imóveis, tanto para inquilinos quanto para investidores. Estabelecimentos próximos podem adicionar conveniência e qualidade de vida, impactando positivamente na atratividade e valorização das propriedades.

Foram incluídas categorias de serviços essenciais, como restaurantes, farmácias e academias, que são amplamente procurados por moradores devido à conveniência que oferecem no dia a dia. Além disso, escolas e faculdades foram consideradas por sua influência no mercado de aluguel, especialmente em áreas com grande concentração de estudantes, enquanto supermercados foram incluídos por serem fundamentais para atender às necessidades básicas dos moradores.

A presença de postos policiais próximos foi um critério relevante, pois pode aumentar a sensação de segurança na região, enquanto a proximidade de estações de metrô contribui para a mobilidade e acessibilidade, características muito valorizadas em grandes centros urbanos.

Por fim, shoppings e hotéis foram selecionados por sua capacidade de atrair fluxo de pessoas e gerar desenvolvimento econômico local, enquanto hospitais foram considerados por seu papel fundamental no atendimento à saúde, um fator decisivo para muitos residentes.

Este enfoque arbitrário foi adotado para fornecer uma base inicial de análise. No entanto, reconhece-se que a escolha pode não abranger todos os aspectos relevantes para todas as situações. Portanto, a inclusão de outras categorias de estabelecimentos poderá ser considerada em futuras análises, conforme a disponibilidade de dados e a evolução das necessidades identificadas.

## 4. Transformação e 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.

O primeiro passo no tratamento dos dados de FIIs 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.

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 para prevenir a ocorrência de anomalias na variação do preço.

Na sequência, os seguintes atributos foram derivados para complementar:

- *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 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 anteriores.
- *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 nas análises. 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 excluídos. 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. Para cada fundo, foram coletados e/ou gerados 33 atributos no total.

## 5. Aplicações

O conjunto de dados descrito neste trabalho tem utilidade tanto para pesquisadores quanto para investidores. As principais aplicações vislumbradas são:

- Pesquisas para modelos preditivos: pesquisadores da área de IA para o mercado financeiro podem utilizar o conjunto de dados para validar modelos preditivos de fundos imobiliários. Os autores deste trabalho inclusive já utilizaram os dados em um artigo [Diniz et al. 2023];
- Pesquisas para recomendações de carteiras de investimento: sistemas de recomendação de carteiras devem diversificar os ativos para reduzir os riscos dos investidores. A organização de carteiras de investimentos é uma tarefa complexa que também tem sido pesquisada. Com os dados descritos aqui, será possível incluir os ativos FIIs nas recomendações, o que ainda não é feito;
- Análise fundamentalista: investidores poderão utilizar os dados, em conjunto com uma ferramenta de *Business Intelligence* (BI), para auxiliar nas análises fundamentalistas dos FIIs. Os dados enriquecidos com as características dos imóveis de cada FII são essenciais para esse aspecto;
- Ferramentas para automatização de investimentos: os dados também poderão ser utilizados para treinar robôs investidores que realizam transações de forma automatizada, incluindo os FIIs em suas listas de ativos.

## 6. Conclusão

Este trabalho apresenta um conjunto de dados<sup>10</sup> de fundos de investimento imobiliário (FIIs) do Brasil, enriquecidos com informações sobre os imóveis físicos que são gerenciados pelos respectivos fundos. Os dados foram coletados de diversas fontes e organizados de forma tabular, contendo 33 atributos para cada fundo. Esse conjunto pode beneficiar tanto pesquisadores da área de IA para o mercado financeiro quanto investidores. Como trabalhos futuros, pretende-se criar um processo automatizado para atualizar periodicamente os dados, que variam ao longo do tempo.

## Referências

B3 (2021). Fundos de investimento imobiliário (fii).

<sup>10</sup>[https://github.com/NESPEDUFV/repositorio\\_dados\\_fiis](https://github.com/NESPEDUFV/repositorio_dados_fiis)

- Barbosa Junior, W. (2023). Aplicação de machine learning para auxiliar o processo decisório de investimento em fundos imobiliários.
- Barboza, J. P., Santos, G. C., Garruti, D. V. T., and Barboza, F. L. d. M. (2022). Alocação de fundos de investimento imobiliários e estratégias de negociação: Carteiras elaboradas em algoritmos de reinforcement learning. *Encontro de Gestão e Negócios*.
- Cardoso, F., Malska, J., Ramiro, P., Lucca, G., Borges, E., Mattos, V., and Berri, R. (2021). Bovdb: A data set of stock quotes for machine learning on all companies from b3 between 1995 and 2020. In *Anais do III Dataset Showcase Workshop*, pages 21–32, Porto Alegre, RS, Brasil. SBC.
- da Rocha Lima Jr, J. (2011). Fundos imobiliários têm futuro no brasil? *Real Estate Mercado*.
- Diniz, H., Carneiro, P., and Silva, F. (2023). Predicting real estate funds: A comparative study of machine learning and time series methods. In *Anais do II Brazilian Workshop on Artificial Intelligence in Finance*, pages 49–60, Porto Alegre, RS, Brasil. SBC.
- 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.
- Freire, D. V. (2021). Forecasting u.s. reit index prices with artificial neural networks. Master's thesis. AAI29335585.
- Lauria, R. d. S. (2020). *Aplicação de sistemas de recomendação em finanças*. PhD thesis.
- Paulus, N., Lautenschlaeger, L., and Schäfers, W. (2023). Social media and real estate: Do twitter users predict reit performance?
- Rozário, J. G. d. (2022). Aprendizado de máquina aplicado à seleção de fundos de investimento imobiliários. Trabalho de conclusão de curso de especialização, Universidade Tecnológica Federal do Paraná, Campus de Dois Vizinhos. Curso de Especialização em Ciência de Dados.
- SANTOS, V. V. d. (2023). Avaliação do método random forest para auxílio à seleção de fundos de investimento imobiliário listados na bolsa de valores.
- Steffen, M. A. (2015). Rentabilidade dos fundos de investimento imobiliários: uma análise da influência da bolsa de valores, dos juros e de fatores específicos do mercado. Master's thesis, Universidade de São Paulo.
- Utani, L. A., Silva, G. R. d., and Faria, E. B. P. (2022). Sistema especialista para apoio a decisão de investidores na aquisição de fundos imobiliários. *Revista Eletrônica Anima Terra*, VII(15):105–118. Mogi das Cruzes-SP.
- Zanandrea, V. et al. (2018). Avaliação de fundos de investimento imobiliário no brasil. Master's thesis, Universidade Federal de Santa Maria.
- Zhang, W., Li, B., and Liew, A. e. a. (2023). Predicting the returns of the us real estate investment trust market: evidence from the group method of data handling neural network. *Financ Innov*, 9:98.