



**Universidade Federal de Pernambuco**

**Centro de Informática**

**Graduação em Engenharia da Computação**

Trabalho de Graduação

**Avaliação do método *Random Forest* para auxílio à seleção de fundos de investimento imobiliário listados na Bolsa de Valores**

**Aluno:** Victor Vianna dos Santos (vvs3@cin.ufpe.br)

**Orientador:** Adiel Teixeira de Almeida Filho (adielfilho@cin.ufpe.br)

**Área:** Ciência de Dados, Classificação, Mercado Financeiro, Random Forest, Fundos Imobiliários, Machine Learning

Setembro de 2023



**Universidade Federal de Pernambuco**

**Centro de Informática**

**Graduação em Engenharia da Computação**

Trabalho de Graduação

**Avaliação do método *Random Forest* para auxílio à seleção de fundos de investimento imobiliário listados na Bolsa de Valores**

Trabalho apresentado ao Programa de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

**Orientador:** Adiel Teixeira de Almeida Filho  
(adielfilho@cin.ufpe.br)

Setembro de 2023

Ficha de identificação da obra elaborada pelo autor,  
através do programa de geração automática do SIB/UFPE

Santos, Victor Vianna dos.

Avaliação do método Random Forest para auxílio à seleção de fundos de investimento imobiliário listados na Bolsa de Valores / Victor Vianna dos Santos. - Recife, 2023.

43 p.

Orientador(a): Adiel Teixeira de Almeida Filho

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Informática, Engenharia da Computação - Bacharelado, 2023.

1. Ciência de Dados. 2. Classificação. 3. Mercado Financeiro. 4. Random Forest. 5. Fundos Imobiliários. I. Almeida Filho, Adiel Teixeira de. (Orientação). II. Título.

000 CDD (22.ed.)

**VICTOR VIANNA DOS SANTOS**

**Avaliação do método *Random Forest* para auxílio à seleção de fundos de investimento imobiliário listados na Bolsa de Valores**

Trabalho apresentado ao Programa de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Aprovado em: 26/09/2023

**BANCA EXAMINADORA**

---

Prof. Adiel Teixeira de Almeida Filho (Orientador)

Universidade Federal de Pernambuco

---

Prof. Sérgio Ricardo de Melo Queiroz (Examinador Interno)

Universidade Federal de Pernambuco

## **Agradecimentos**

Acima de tudo, agradeço a Deus por guiar meus caminhos.

Agradeço a minha esposa, Mariana, que me acompanha e apoia em tudo que faço. A nossa filha, Teresa, pela benção que é em nossas vidas.

Agradeço aos meus pais, Regina e Sidnei, por serem minhas referências.

Aos professores que compõem o quadro docente do curso de graduação em Engenharia da Computação. Em especial, ao meu orientador, Adiel Teixeira de Almeida Filho, pela paciência e pela orientação generosa.

## Resumo

Fundos de investimento imobiliário são veículos de investimento cujo mercado vem crescendo exponencialmente, devido às vantagens e facilidades que a regulamentação proporciona, principalmente para pessoas físicas. O principal objetivo deste trabalho foi avaliar o método *Random Forest* como classificador para ordenar os FIIs analisados, mensalmente, de acordo com a probabilidade de haver uma variação positiva no preço dentro do próximo mês. Com os resultados das simulações, avaliamos as rentabilidades mensais dos ativos com maior probabilidade prevista, comparamos com índices de referência do mercado financeiro, como CDI e IFIX, e avaliamos a eficácia do método como ferramenta para auxílio a investidores na construção de uma carteira de investimentos. As simulações executadas demonstraram resultados promissores no período analisado, no qual a rentabilidade dos FIIs selecionados superou a dos *benchmarks*.

## **Abstract**

Real Estate Investment Funds (REIFs) are investment vehicles whose market has been growing exponentially, due to advantages and convenience provided by regulations, especially to individuals. This work's main objective is to evaluate the Random Forest method as a classifier to monthly rank the analyzed REIFs according to the probability of a positive price change during the next month of negotiations. In possession of the simulations' results, we evaluate the monthly returns of assets with the highest predicted probabilities, compare these to benchmarks, such as CDI and IFIX, and assess the effectiveness of the method to be used as a tool to aid investors in building an investment portfolio. The simulations that were executed showed promising results in the analyzed period, in which the selected REIFs' performance was superior to the benchmarks.

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	10
1.1 OBJETIVOS.....	10
<b>2. REVISÃO DE LITERATURA</b> .....	11
2.1 FUNDOS DE INVESTIMENTO IMOBILIÁRIO .....	13
2.2 IFIX .....	16
2.3 ANÁLISE DE FUNDOS IMOBILIÁRIOS .....	16
2.4 EXTRAÇÃO DE DADOS .....	17
2.5 RANDOM FOREST .....	17
<b>3. METODOLOGIA</b> .....	18
3.1 COLETA E PRÉ-PROCESSAMENTO DE DADOS .....	19
3.2 SELEÇÃO DE ATRIBUTOS .....	21
3.3 VARIÁVEL RESPOSTA .....	23
3.4 MODELO DE CLASSIFICAÇÃO .....	24
3.5 AVALIAÇÃO DO MODELO .....	25
3.6 PRÉ-SELEÇÃO E VALIDAÇÃO .....	26
3.6.1 PRÉ-SELEÇÃO.....	26
3.6.2 VALIDAÇÃO .....	26
<b>4. SIMULAÇÕES E RESULTADOS</b> .....	27
4.1 TREINAMENTO E TESTE.....	27
4.2 ANÁLISE DE ATRIBUTOS .....	28
4.3 PRÉ-SELEÇÃO .....	30
4.4 VALIDAÇÃO .....	31
<b>5. CONCLUSÕES E TRABALHOS FUTUROS</b> .....	39
<b>REFERÊNCIAS</b> .....	40

## LISTA DE FIGURAS

Figura 01 – Volume financeiro negociado em FIIs por ano .....	14
Figura 02 – Evolução do número de investidores em FIIs .....	15
Figura 03 – Fluxograma de atividades .....	18
Figura 04 – Exemplo de texto gerado pelo parser .....	20
Figura 05 – Exemplo de listas geradas após tratamento do texto .....	20
Figura 06 – Exemplo de DataFrame criado com os formatos corretos.....	20
Figura 07 – Janela expansível de treinamento .....	25
Figura 08 – Relação entre data do IPO e Precision Score atingido .....	28
Figura 09 – Média das Importâncias dos Atributos .....	28
Figura 10 – Importância acumulada média.....	30
Figura 11 – Gráficos Comparados Probabilidades x Fundos Imobiliários .....	32
Figura 12 – Comparativo gráfico de variações mensais .....	35
Figura 13 – Gráfico dos retornos mensais acumulados .....	36

## LISTA DE TABELAS

Tabela 01 – Descrições das informações extraídas do site Investing .....	19
Tabela 02 – Descrições dos atributos utilizados .....	21
Tabela 03 – Precision_Score atingido para cada FII .....	27
Tabela 04 - Média das Importâncias dos Atributos .....	29
Tabela 05 – Ranking por probabilidade de retorno mensal positivo .....	31
Tabela 06 – Tabela com variações mensais .....	33
Tabela 07 – Frequência de pré-seleção dos FIIs .....	35
Tabela 08 – Retorno mensal médio dos FIIs selecionados .....	37
Tabela 09 – Métricas comparativas para cada mês analisado.....	38
Tabela 10 – Retornos mensais de cada posição no ranking de probabilidade .....	38
Tabela 11 – Retorno médio e proporção de retornos positivos de acordo com o ranking.....	38

## 1. INTRODUÇÃO

O mercado de fundos de investimento imobiliário vem crescendo exponencialmente, chegando a multiplicar em mais de oito vezes o volume negociado e em mais de 16 vezes o número de investidores que aplicam nessa classe de ativos, entre os anos de 2017 e 2022, segundo dados da bolsa de valores (B3).

Uma maior disseminação de conteúdos financeiros nas redes sociais e a busca por investimentos mais rentáveis durante o ciclo de queda da taxa Selic entre os anos de 2017 e 2020 são exemplos de fatores que influenciaram esse movimento. Ao final desse ciclo, a taxa Selic chegou a atingir 2% a.a., o menor nível desde outubro de 1997, segundo histórico divulgado pelo Banco Central do Brasil. Já o IFIX, principal índice que mede o retorno médio dos fundos imobiliários negociados em bolsa, teve retornos anuais de aproximadamente 19,0%, 5,6% e 36,0% nos anos de 2017, 2018 e 2019, respectivamente.

Dado que os FIIs possuem características híbridas (SCOLESE et al, 2015), com componentes de renda fixa como contratos de aluguéis, LCIs, CRIs, e componentes de renda variável como a variação do preço dos imóveis e das cotas negociadas, é válido utilizar técnicas de inteligência artificial (IA) para auxiliar na escolha dos investimentos.

Apesar do potencial de uso da IA, ainda há poucos trabalhos que exploram esse tipo de ferramenta de previsão para FIIs, conforme afirma Diniz et al (2023). Um dos desafios de implementação é a seleção de dados históricos mais reduzido que os utilizados no mercado de ações. Como exemplo, o Ibovespa, principal indicador de desempenho das ações listadas, foi criado pela B3 em 1968. Já o IFIX foi criado em 2011 e muitos dos FIIs que o compõem surgiram nos últimos 5 anos.

### 1.1 OBJETIVOS

O principal objetivo deste trabalho foi justamente o de avaliar o método de Random Forest como classificador para ordenar os FIIs analisados, de acordo com a probabilidade de haver variação positiva dentro de um mês de negociação. Consequentemente, validar se o resultado do experimento pode ser útil no auxílio a investidores que estejam construindo uma carteira de

investimentos em FIIs. Utilizaremos séries históricas de dados quantitativos dos FIIs analisados e dos principais índices do mercado de investimentos como conjunto de dados de treinamento. Em seguida, armazenaremos as probabilidades previstas em cada execução para analisar os resultados obtidos.

Para atingir o principal objetivo exposto acima, foram definidos os seguintes objetivos específicos:

- implementar um código para coleta de dados históricos via técnicas de *web scraping*;
- comparar as rentabilidades dos FIIs selecionados pelo método com os principais índices de mercado, como CDI e IFIX;

No que se refere à estrutura deste trabalho, ele segue dividido da seguinte forma: seção 2, com a revisão da literatura sobre o tema; seção 3 descrevendo a metodologia utilizada para execução do trabalho; seção 4 consta do detalhamento dos resultados obtidos pela execução e, por fim, a seção 5 expondo as conclusões finais e as sugestões de trabalhos futuros.

## **2. REVISÃO DE LITERATURA**

Observamos que alguns autores desenvolveram trabalhos com objetivos similares, de auxiliar investidores na seleção de FIIs ou de previsão de valores.

Diniz et al (2023) trazem uma implementação dos métodos ARIMA, XGBoost e *Long Short Term Memory* (LSTM), a fim de comparar os resultados obtidos com técnicas de aprendizagem de máquina com um método tradicional de série temporal (ARIMA), concluindo que os métodos de aprendizagem tiveram um melhor desempenho no experimento realizado.

Utani et al (2022) propõem um sistema especialista com interface de linguagem natural e bases de conhecimento que alimentam um motor de inferência, a fim de executar uma análise qualitativa dos FIIs e retornando uma nota que representa o nível de convicção daquele diagnóstico. O experimento teve métricas de acurácia e de precisão baixas, mas sugeriu que uma recalibragem do motor e maior alimentação do sistema de regras poderiam gerar resultados mais precisos.

Pereira (2023) também busca implementar uma forma de auxiliar investidores na seleção de FIIs. Seu trabalho aplicou técnicas de ciência de dados para análise de fundos imobiliários de galpões logísticos, a fim de identificar padrões e definir regras para seleção dos melhores FIIs, que seriam os de menor risco, considerando as análises de dados quantitativos e qualitativos. Por fim, ele pontua como seria interessante aplicar um modelo de regressão linear múltipla para fazer a seleção e comparar os resultados.

Por sua vez, Waismann (2019) propôs uma ferramenta desenvolvida em *Microsoft Excel* para geração de carteiras de FIIs baseadas na Teoria de Markowitz. Ele conclui que os resultados obtidos não foram vantajosos se comparados ao IFIX.

Uma aplicação do método AHP-TOPSIS em uma carteira de FIIs foi usada por Lima et al (2021). O propósito desses autores foi o de elaborar um ranking com as melhores alternativas de acordo com as preferências do investidor na ponderação dos critérios utilizados, fornecendo os resultados para que o investidor tome a decisão final de investimento.

Dada a quantidade reduzida de trabalhos que utilizam métodos de aprendizagem de máquina para atingir esse objetivo, conforme apontado por Diniz et al (2013) e observado durante a pesquisa bibliográfica do presente trabalho, também buscamos por referências que focam no mercado de ações.

Abreu (2021) propõe a utilizar o método *Random Forest* para distinguir ativos presentes no IBOVESPA que teriam retorno positivo nos próximos 20 dias, utilizando a probabilidade prevista pelo modelo como fator de ordenação dos ativos. Seus achados foram interessantes ao concluir que o portfólio construído teve melhor retorno acumulado quando comparado a outras técnicas.

Silva et al (2023) também utilizam *Random Forest* para obter as probabilidades de retorno positivo a cada mês de avaliação. A alocação da carteira teórica para validação, por sua vez, era feita com base na estratégia de *Long & Short*, na qual o investidor compra ativos na perspectiva de que vão valorizar e vende aqueles ativos que esperam desvalorizar no futuro.

Costa (2022) implementa modelos de *Random Forest*, *Support Vector Regression (SVR)* e *K-Nearest Neighbor (KNN)* para pré-seleção de ativos e otimização de portfólio dentre dez ETFs (*Exchange Traded Funds*) de mercados internacionais. Concluíram que seus resultados

demonstram uma melhor performance em portfólios que utilizam as técnicas de aprendizagem de máquina, em relação ao retorno e ao risco.

Dois trabalhos desenvolvidos por alunos do Centro de Informática da UFPE também serviram como referência e inspiração para o presente estudo. Cavalcanti (2021) propõe a utilização de *Random Forest* para indicar momentos de compra e venda e alimentar uma estratégia de *Day Trading*. Por sua vez, Da Silva Junior (2021) implementa uma ferramenta para auxílio à tomada de decisão dos investidores. O objetivo da ferramenta é fornecer métricas das ações negociadas na bolsa de valores e notas de qualidade dadas pelo modelo de *Random Forest* utilizado como classificador.

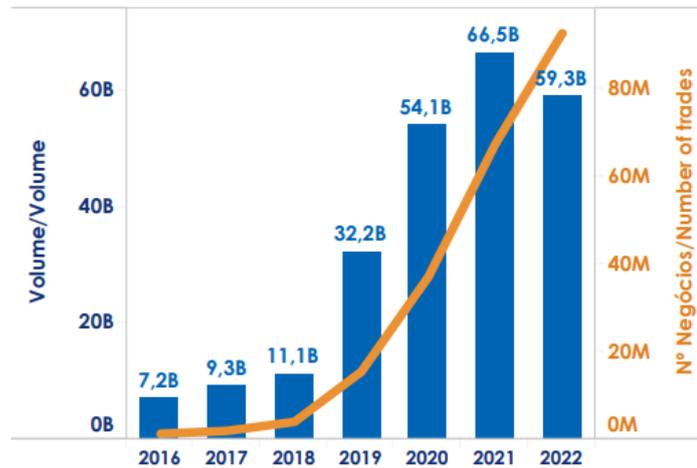
Ballings et al. (2015) recomendam o uso de *Random Forests* para prever a direção dos preços de ações, visto que o método teve os melhores resultados no estudo comparativo realizado. Escolhemos o método de Random Forest para prever a probabilidade de retornos positivos dos fundos imobiliários no presente trabalho, devido a essa comparação e diante das contribuições dos estudos elencados, pois se relacionam com o objetivo principal do presente trabalho – auxiliar investidores na seleção de ativos que tenham maior probabilidade de valorização.

## 2.1 FUNDOS DE INVESTIMENTO IMOBILIÁRIO

Os fundos de investimento imobiliário (somente fundos imobiliários ou FIIs), regulamentados pela Instrução CVM (Comissão de Valores Mobiliários) nº 472 de 31/10/2008, têm como objetivo investir recursos em empreendimentos imobiliários como empresariais, shopping centers, hospitais, hotéis, galpões logísticos e terrenos destinados ao agronegócio. Os recursos também podem ser alocados em títulos de dívida lastreados em ativos imobiliários, como CRIs e LCIs. Servem como uma alternativa para pessoas que desejam investir seu dinheiro no mercado imobiliário sem precisar comprar imóveis físicos. (Fonte: [https://www.b3.com.br/pt\\_br/produtos-e-servicos/negociacao/renda-variavel/fundos-de-investimento-imobiliario-fii.htm](https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/fundos-de-investimento-imobiliario-fii.htm))

Em 2022, o volume negociado em FIIs foi de R\$59,3 bilhões, segundo boletim da B3 divulgado em dezembro/2022. No fechamento do ano, o número de FIIs listados era 467, quase o triplo do número de 2017 que estava em 156.

Figura 01 – Volume financeiro negociado em FIIs por ano



Fonte: B3

Uma das principais vantagens de investimento por meio de FIIs em relação ao mercado imobiliário tradicional é a facilidade de acesso, já que o investimento em imóveis físicos exige quantias significativas de capital. Com os Fundos Imobiliários, não há valores mínimos de aplicação e é possível adquirir somente uma cota de um fundo imobiliário (VOGEL, 2022).

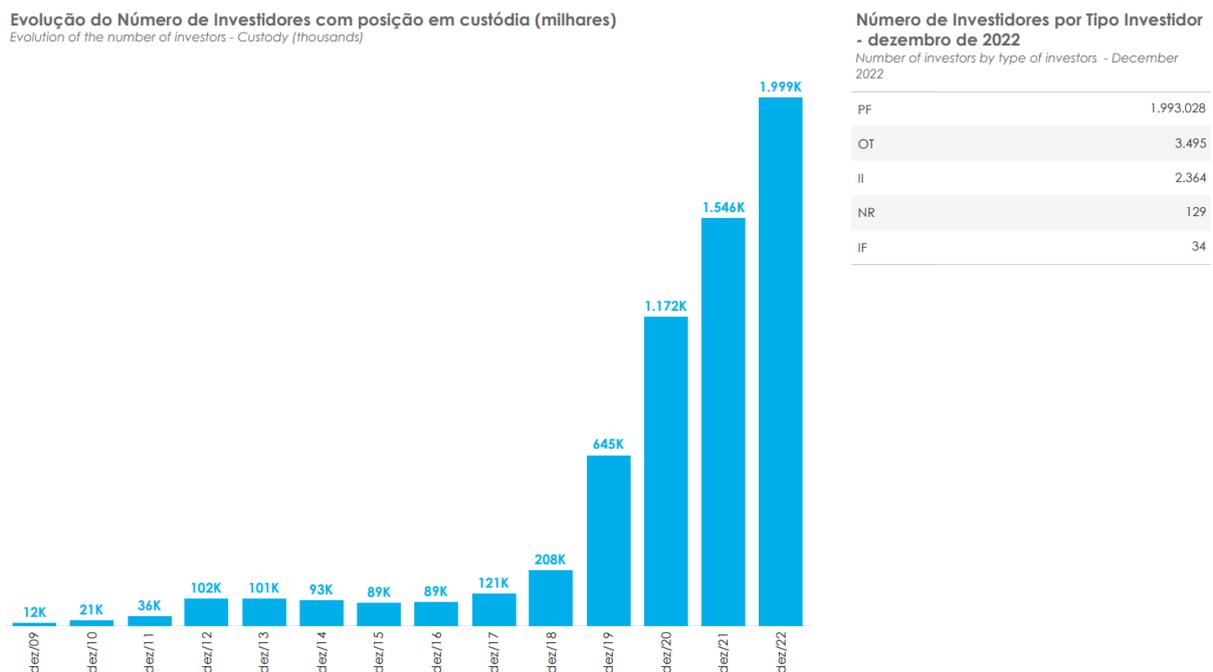
Outro benefício é a liquidez dessas aplicações. As cotas dos FIIs são negociadas em bolsa ou mercado de balcão organizado, permitindo ao investidor liquidar suas operações com mais rapidez, além da possibilidade de vendas parciais do capital investido, conforme apontado por Mugnaini et al (2008).

Mais uma vantagem de se investir no mercado imobiliário via FIIs está relacionada aos custos. As cotas negociadas via bolsa de valores podem ter custos de corretagem que se aproximam de 0,50% do valor transacionado. Vale ressaltar que diversas corretoras oferecem custos reduzidos ou até zerados como forma de atrair clientes. Por outro lado, a negociação com imóveis físicos pode trazer custos relevantes como ITBI, custos cartoriais, honorários advocatícios e comissões de corretores de imóveis, que somadas podem chegar a 6% do valor negociado (FRANKE, 2012).

Por fim, um grande incentivo aos investidores de FIIs é o benefício fiscal para pessoas físicas. Segundo o boletim da B3 em relação ao ano de 2022, dos 1.999.050 investidores em FIIs, 1.993.028 eram pessoas físicas. Havendo o cumprimento de critérios definidos pela Lei

11.033/04, os rendimentos distribuídos para estes serão isentos de imposto de renda. Alguns dos critérios são: que o fundo tenha pelo menos 50 cotistas; que as cotas sejam negociadas exclusivamente via bolsa de valores ou mercado de balcão organizado e que o cotista seja pessoa física detendo no máximo 10% do total de cotas emitidas pelo fundo. Caso um desses critérios não seja cumprido, o benefício fiscal é perdido e o cotista fica sujeito ao recolhimento de imposto de renda à alíquota de 20% dos rendimentos distribuídos (Scolese et al, 2015).

Figura 02 – Evolução do número de investidores em FIIs



Fonte: B3

Existem diversos tipos de risco no investimento em fundos imobiliários (VOGEL, 2022), alguns dos quais incluem:

- Risco de mercado: refere-se ao risco de que os preços das cotas dos fundos imobiliários fltuem devido a mudanças nas condições de mercado.
- Risco de crédito: relacionado à capacidade do emissor do título que está sendo financiado pelo fundo de pagar seus compromissos financeiros.
- Risco de liquidez: relacionado à possibilidade de não haver compradores suficientes no mercado quando o investidor deseja vender suas cotas.
- Risco de vacância: possibilidade de que os imóveis do fundo fiquem desocupados por períodos prolongados, o que reduziria a renda do fundo.

- Risco de inadimplência: relacionado à possibilidade de os inquilinos não pagarem o aluguel em dia ou até deixarem de pagar.
- Risco regulatório: relacionado a mudanças na legislação que possam afetar a capacidade do fundo de operar, gerar renda ou manter vantagens como isenção de imposto de renda para pessoas físicas.

## 2.2 IFIX

O IFIX é um índice de referência do mercado brasileiro de Fundos de Investimento Imobiliário (FII). Ele é calculado pela Bolsa de Valores de São Paulo (B3) desde dezembro de 2010 e reflete o desempenho das carteiras teóricas composta pelos FIIs mais negociados na B3, considerando tanto a valorização das cotas quanto a distribuição de rendimentos. O índice serve como uma referência para avaliação da performance média dos FIIs.

Fonte: ([https://www.b3.com.br/pt\\_br/market-data-e-indices/indices/indices-de-segmentos-e-setoriais/indice-fundos-de-investimentos-imobiliarios-ifix-estatisticas-historicas.htm](https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-de-segmentos-e-setoriais/indice-fundos-de-investimentos-imobiliarios-ifix-estatisticas-historicas.htm))

## 2.3 ANÁLISE DE FUNDOS IMOBILIÁRIOS

Segundo Kodama (2021), existem diversos indicadores que podem ser utilizados para analisar o desempenho de um fundo imobiliário. Alguns dos principais são:

- *Dividend Yield (DY)*: relação entre os rendimentos do fundo e o preço de suas cotas.
- Preço/Valor Patrimonial (P/VP): relação entre o preço de mercado da cota do fundo e o valor patrimonial da mesma cota.
- Retorno Total: é o retorno total que um investidor obteve em um determinado período, considerando tanto a valorização das cotas quanto os rendimentos.
- Vacância: percentual de imóveis do fundo que estão desocupados em um determinado momento.
- Vacância financeira: percentual do potencial de receita do fundo que não está sendo gerada em determinado momento.
- Inadimplência: percentual de inadimplência dos locatários do fundo.

É válido ressaltar que a análise de um fundo imobiliário deve considerar não apenas os indicadores financeiros, mas também as características dos imóveis que compõem a carteira do fundo e as perspectivas para o mercado imobiliário em geral.

## 2.4 EXTRAÇÃO DE DADOS

A extração de dados baseada em DOM (*Document Object Model*) é uma API utilizada para coletar informações de uma página na internet de acordo com a estrutura HTML ou XML. É necessário analisar o código da página para identificar a localização dos dados desejados.

Fonte: (<https://www.w3.org/TR/REC-DOM-Level-1/introduction.html>)

Ao contrário da extração de dados baseada em texto, a extração de dados baseada em DOM permite uma análise mais precisa e detalhada da página da web. É possível identificar a localização exata dos dados desejados e coletá-los de maneira mais eficiente. Além disso, a extração baseada em DOM é menos suscetível a erros devido a mudanças no *layout* ou na estrutura da página, pois ela é baseada na análise da estrutura da página (GUPTA et al, 2003).

Dioud et al (2019) descrevem algumas formas de abordar a extração de dados da *web*. Escolhemos o *Mimicry Approach*, ou abordagem por mimetismo, dada a especificidade das informações que precisávamos extrair e o fato de que todas as páginas teriam a mesma estrutura.

A biblioteca *Beautiful Soup 4* pode ser usada para realizar a extração de dados do DOM, baseada em HTML/XML (Zheng et al, 2015). Ela permite analisar e manipular a estrutura, facilitando a extração dos dados desejados. Através do uso de seletores e expressões regulares, é possível localizar elementos específicos da página e extrair seu conteúdo.

## 2.5 RANDOM FOREST

O método *Random Forest* consiste em construir um conjunto (*ensemble*) de árvores para melhorar o desempenho das árvores de decisão (DUDOIT et al., 2002). Inicia-se criando um subconjunto aleatório de variáveis candidatas e avaliando todas as possíveis divisões (BREIMAN, 2001). A melhor divisão é então usada para criar uma partição binária. Esse processo é repetido recursivamente em cada partição subsequente e se encerra quando o tamanho da partição for 1.

Segundo Cheung et al (2005) e Pesaran et al (1995), prever a direção da rentabilidade como positiva ou negativa é o suficiente para executar estratégias de investimentos lucrativas. Neste trabalho, buscamos prever a direção dos preços dos ativos ao invés dos preços nominais. Assim, tratamos um problema de classificação binária, onde a classe 1 representa ativos com

probabilidade de valorização acima de 60%, ou 0, caso contrário, conforme probabilidades previstas pelo modelo.

Segundo Fu et al (2018) há três principais estratégias de investimentos a serem construídas através de aprendizado de máquina: seleção de ativos para investir, gestão de portfólio e *timing*, esta tem o objetivo de determinar o momento mais apropriado para comprar ou vender ativos.

Neste trabalho, concentramos nossos esforços na seleção de FIIs para uma carteira de investimentos. Para atingir o objetivo principal de ordenar os ativos pela probabilidade de haver variação positiva na cota, utilizaremos atributos quantitativos dos FIIs e do mercado de investimentos para teste do modelo, buscando extrair a lista das probabilidades previstas para cada mês de negociação. A seção 3 entra em detalhes do fluxo de atividades necessários para desenvolvimento do trabalho.

### 3. METODOLOGIA

Este capítulo descreve a metodologia utilizada para a coleta, tratamento e armazenamento dos dados, a especificação dos atributos e do modelo de *Random Forest* utilizado para classificação dos FIIs conforme suas probabilidades mensais de retorno positivo.

Os códigos desse trabalho foram implementados em *Python 3.0* e executados no ambiente do *Google Colab*. As principais bibliotecas utilizadas foram *Scikit Learn*, *Pandas*, *NumPy*, *Matplotlib* e *Beautiful Soup 4*.

Figura 03 – Fluxograma de atividades



Fonte: Autor (2023)

### 3.1 COLETA E PRÉ-PROCESSAMENTO DE DADOS

Os dados históricos de negociação dos fundos imobiliários foram baixados diretamente do site Investing.com.br, o qual possibilita download de arquivos CSVs contendo as colunas indicadas na tabela 01. Para cada FII, o histórico foi baixado desde a respectiva data de início das negociações na Bolsa. Neste trabalho, foram considerados apenas os 111 FIIs que faziam parte do IFIX no momento da coleta (entre janeiro e março de 2023).

Tabela 01 – Descrições das informações extraídas do site Investing

Coluna	Descrição
Data	Data de referência
Último	Preço do último negócio realizado com o FII
Abertura	Preço do primeiro negócio realizado com o FII
Máxima	Maior cotação do FII atingida durante as negociações
Mínima	Menor cotação do FII atingida durante as negociações
Vol.	Volume de cotas negociado
Var%	Varição percentual do preço de fechamento de um dia para o outro

Fonte: Autor (2023)

Todos os dados eram lidos como *strings* pelo código. Para possibilitar a análise, tratamos a formatação das datas de cada linha para que estas fossem armazenadas como um *DateTimeIndex*, um método da biblioteca *Pandas* que converte colunas com datas em índices nos *DataFrames* criados. Os dados das outras colunas foram tratados e convertidos para *floats*.

Na análise de investimentos em fundos imobiliários, há outros valores históricos que podem ser relevantes para o modelo de predição, como os proventos pagos aos cotistas, necessários para cálculo do *Dividend Yield*, conforme descrito no capítulo 2.

Pela falta de fonte de informação pública que fornecesse o histórico de rendimentos num formato simples de baixar, no período de desenvolvimento do presente trabalho, foi necessário elaborar um código para extrair as informações de forma estruturada.

A fonte de informação escolhida foi o site *Funds Explorer*, pois identificamos que nas páginas de cada FII havia gráficos com os históricos mensais desses valores.

Fizemos um código em *Python* com auxílio da biblioteca *Beautiful Soup 4* para extrair os valores necessários, tratá-los e armazená-los em arquivos CSV, possibilitando a posterior

consolidação com os dados históricos de negociação. Descrevemos o funcionamento do código a seguir:

1. Acessar a página correspondente ao FII;
2. Coletar os elementos que contêm os dados de rendimentos através do *Mimicry Approach*;
3. Tratar as strings para armazenar em listas;
4. Formatar para que as datas fiquem no formato *Datetime* e os valores em *Float*;
5. Criar um *DataFrame* com *DateTimeIndex* e salvar em arquivos CSV, para cada FII.

Figura 04 – Exemplo de texto gerado pelo *parser*

```
//
(function() { var initChart = function() { var ctx = document.getElementById("chart-
2"); var chart = new Chart(ctx, { type: "line", data:
{"labels":["Dezembro/2017","Janeiro/2018","Fevereiro/2018","Março/2018","Abril/201
8","Maio/2018","Junho/2018","Julho/2018","Agosto/2018","Setembro/2018","Outubro/
2018","Novembro/2018","Dezembro/2018","Janeiro/2019","Fevereiro/2019","Março/2
019","Abril/2019","Maio/2019","Junho/2019","Julho/2019","Agosto/2019","Setembro/
2019","Outubro/2019","Novembro/2019","Dezembro/2019","Janeiro/2020","Fevereiro/
2020","Março/2020","Abril/2020","Maio/2020","Junho/2020","Julho/2020","Agosto/20
20","Setembro/2020","Outubro/2020","Novembro/2020","Dezembro/2020","Janeiro/20
21","Fevereiro/2021","Março/2021","Abril/2021","Maio/2021","Junho/2021","Julho/20
21","Agosto/2021","Setembro/2021","Outubro/2021","Novembro/2021","Dezembro/20
21","Janeiro/2022","Fevereiro/2022","Março/2022","Abril/2022","Junho/2022","Julho/
2022","Agosto/2022","Setembro/2022","Outubro/2022","Novembro/2022","Dezembro/
2022"],"datasets":[{"label":"VP","data":[95.105307,96.152696,96.198695,96.40097,96.
344896,96.359316,96.361801,95.754821,95.779681,95.683874,95.705101,95.859829,1
15.663442,109.855017,109.813298,109.754065,109.717641,109.646103,109.629277,1</pre></div><div data-bbox="115 600 284 617" data-label="Text"><p>Fonte: Autor (2023)</p></div><div data-bbox="236 639 757 656" data-label="Caption"><p>Figura 05 – Exemplo de listas geradas após tratamento do texto</p></div><div data-bbox="208 682 799 785" data-label="Text"><pre>[95.105307, 96.152696, 96.198695, 96.40097, 96.344896, 96.359316,
['Dezembro/2017', 'Janeiro/2018', 'Fevereiro/2018', 'Março/2018',
[0.0, 0.488042, 0.5, 0.67, 0.61, 0.62, 0.62, 0.62, 0.62, 0.62, 0.6
['Dezembro/2017', 'Janeiro/2018', 'Fevereiro/2018', 'Março/2018',</pre></div><div data-bbox="115 810 256 826" data-label="Text"><p>Fonte: Autor (2023)</p></div><div data-bbox="852 921 886 940" data-label="Page-Footer"><p>20</p></div>
```

Figura 06 – Exemplo de *DataFrame* criado com os formatos corretos

```

<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 63 entries, 2017-12-31 to 2023-02-28
Data columns (total 1 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   Rendimentos 63 non-null     float64
dtypes: float64(1)
memory usage: 1008.0 bytes

   Rendimentos
Date
2017-12-31    0.000000
2018-01-31    0.488042
2018-02-28    0.500000
2018-03-31    0.670000
2018-04-30    0.610000

```

Fonte: Autor (2023)

Em posse dos dados históricos de negociação e de rendimentos, incrementamos o código para armazenar todos os dados em CSVs consolidados para cada FII.

Como a frequência de negociação é diária (dias úteis) e a frequência dos rendimentos é mensal, utilizamos o método de *Forward Fill* para preencher as células vazias nestas colunas. Esse foi o método escolhido pois as análises dos indicadores para um determinado mês são feitas com base nos rendimentos do mês anterior.

### 3.2 SELEÇÃO DE ATRIBUTOS

Adicionamos colunas contendo as variáveis descritas a seguir, incrementando a base de treinamento. A relevância de cada atributo será verificada no experimento realizado, descrito no capítulo 4.

Tabela 02 – Descrições dos atributos utilizados

Atributo	Descrição
Último	Preço de fechamento
Abertura	Preço do primeiro negócio realizado com o FII
Máxima	Maior cotação do FII atingida durante as negociações
Mínima	Preço do último negócio realizado com o FII
Vol. BRL	Volume de negociação em Reais
Vol.	Volume de negociação em número de cotas
Var%	Varição percentual do preço de fechamento em um dia

Rendimentos	Valor dos rendimentos distribuídos naquele mês
DY_Mkt	Dividend Yield anualizado considerando a relação Rendimentos/Último
ultimo_1m_before	Preço de fechamento 21 dias antes
rend_1m_before	Valor dos rendimentos distribuídos no mês anterior
ret_1m	Retorno do último mês considerando os rendimentos e a variação da cota
tendencia_semana	Soma da coluna Target dos últimos 5 dias de negociação
media_semana	Razão entre média dos preços de fechamento da semana e o Último
media_trimestre	Razão entre média dos preços de fechamento do trimestre e o Último
media_ano	Razão entre média dos preços de fechamento do ano e o Último
media_ano_semana	Razão entre os valores media_ano e media_semana
media_ano_trimestre	Razão entre os valores media_ano e media_trimestre
abertura_vs_ultimo	Razão entre valores Abertura/Último
max_vs_ultimo	Razão entre valores Máxima/Último
min_vs_ultimo	Razão entre valores Mínima/Último
IFIX_close	Valor de fechamento do IFIX
IFIX_open	Valor de abertura do IFIX
IFIX_max	Maior valor do IFIX atingido durante o dia
IFIX_low	Menor valor do IFIX atingido durante o dia
var_IFIX	Variação do IFIX em um dia
CDI	Valor do CDI
var_CDI	Variação do CDI em um dia
IBOV	Valor do IBOV
var_IBOV	Variação do IBOV em um dia
IPCA	Valor do IPCA
var_IPCA	Variação do IPCA em um dia
IMAB	Valor do IMAB
var_IMAB	Variação do IMAB em um dia
IMAB5	Valor do IMAB5
var_IMAB5	Variação do IMAB5 em um dia
IMAB5+	Valor do IMAB5+
var_IMAB5+	Variação do IMAB5+ em um dia

Fonte: Autor (2023)

Além do IFIX, os dados históricos dos índices descritos a seguir foram incluídos na base de dados.

- IBOV: Principal indicador de desempenho das ações negociadas na B3, o índice é atualizado a cada quatro meses, composto por ativos listados que atendem aos critérios da metodologia de cálculo, como volume de negociação e valor de mercado. ([https://www.b3.com.br/pt\\_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm](https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm))
- CDI: Taxa de juros calculada baseada nos empréstimos realizados entre instituições financeiras (depósitos interbancários).

- IPCA: Índice Nacional de Preços ao Consumidor Amplo, calculado pelo IBGE com o objetivo de medir a inflação, também utilizado como índice de correção em ativos financeiros.
- IMAB: Índice que reflete a média da rentabilidade dos títulos públicos federais com taxas atreladas ao IPCA.
- IMAB5: Índice que reflete a média da rentabilidade dos títulos públicos federais com taxas atreladas ao IPCA e vencimento em até 5 anos.
- IMAB5+: Índice que reflete a média da rentabilidade dos títulos públicos federais com taxas atreladas ao IPCA e vencimento igual ou acima de 5 anos.

Fonte: [https://www.anbima.com.br/pt\\_br/informar/precos-e-indices/indices/ima.htm](https://www.anbima.com.br/pt_br/informar/precos-e-indices/indices/ima.htm)

Para aumentar a base de treinamento, incluímos também alguns atributos derivados dos dados iniciais, como razões entre algumas métricas e tendências. No capítulo 4, avaliaremos as importâncias de cada atributo e veremos que *tendencia\_semana* foi o de maior importância para o modelo.

### 3.3 VARIÁVEL RESPOSTA

O objetivo do trabalho é classificar os ativos avaliados servindo como uma pré-seleção para uma carteira de investimentos em fundos imobiliários. A variável resposta foi definida com base na variação do preço de fechamento da cota em um determinado período. Vamos classificar os FIIs de acordo com a probabilidade de haver uma variação positiva no preço de fechamento, após um período de 21 dias úteis, ou aproximadamente um mês de negociação. Quanto maior a probabilidade prevista para determinado FII, melhor será a colocação dele na pré-seleção.

Para tratar o caso como um problema de classificação, as probabilidades previstas foram discretizadas seguindo a equação descrita abaixo, onde  $x$  é a probabilidade de haver uma variação positiva no preço de fechamento da cota do FII, nos próximos 21 dias de negociação, prevista pelo modelo. Neste trabalho, escolhemos uma probabilidade prevista maior que 60% como o limite mínimo para considerar a classificação igual a 1.

$$f(x) = \begin{cases} 0, & x \leq 0.60 \\ 1, & x > 0.60 \end{cases}$$

Além da classificação como 0 ou 1, também armazenamos as probabilidades previstas pelo modelo em um arquivo, pois elas serão necessárias para ordenação dos ativos pré-selecionados.

### 3.4 MODELO DE CLASSIFICAÇÃO

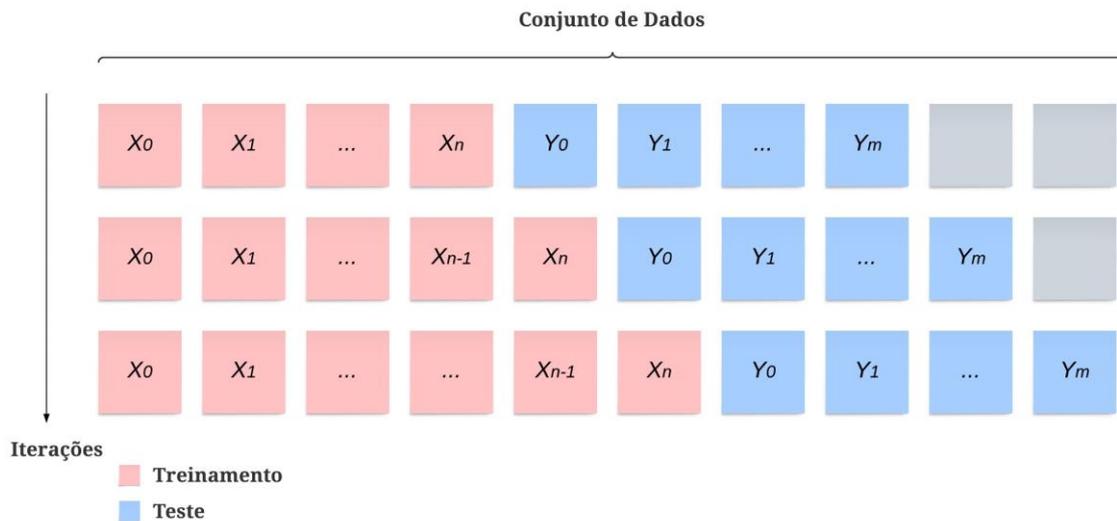
O método de *Machine Learning* utilizado neste trabalho foi o *Random Forest*, com auxílio da biblioteca *Scikit Learn 1.2.2*. Segundo Ballings et al., 2015, *Random Forest* foi o *ensemble method* com os melhores resultados para previsão da direção de preços de ações no mercado financeiro. Visto que nosso objetivo é pré-selecionar ativos que tenham maior probabilidade de variação positiva, decidimos trabalhar com esse método.

Como apontado por Cacalvanti (2021) mantivemos a parametrização padrão, alterando apenas o número de estimadores para 500, um número alto como recomendado por Breiman (2001), e a profundidade máxima para 3. O modelo foi executado dez vezes em cada FII e os resultados analisados representam as médias das execuções.

O treinamento do modelo de *Random Forest* utilizado neste trabalho foi efetuado em janelas expansíveis da base de dados. Iniciamos com os 500 dias de negociação iniciais sendo utilizados como a base de treinamento. A base para teste é composta pelos 21 dias subsequentes. Após o teste, incluímos esses 21 dias que foram testados na base de treinamento para testar outros 21 dias à frente e assim por diante, até cobrir toda a base. A cada iteração, os valores de teste e as respectivas predições do modelo são concatenados num único *DataFrame* para análise posterior.

Antes de executar essa etapa, filtramos a lista de FIIs para trabalhar somente com aqueles que começaram a ser negociados na Bolsa de Valores antes de 05/01/2019. Após o filtro, ficamos com 40 dos 98 FIIs cujos dados foram coletados.

Figura 07 – Janela expansível de treinamento



Fonte: Autor (2023)

### 3.5 AVALIAÇÃO DO MODELO

Num problema de classificação binária, há dois tipos de classe: positiva ou negativa. O modelo atribui uma dessas classes para cada instância da base de dados, resultando em quatro tipos de resultado:

- *True Positive* (TP): o modelo prevê uma classe positiva corretamente;
- *True Negative* (TN): o modelo prevê uma classe negativa corretamente;
- *False Positive* (FP): o modelo prevê uma classe negativa como positiva;
- *False Negative* (FN): o modelo prevê uma classe positiva como negativa.

A partir das quantidades de cada resultado citado acima, podemos calcular as métricas a seguir, conforme a documentação disponível em [scikit-learn.org](https://scikit-learn.org):

- $Precision = TP / (TP + FP)$   
*Precision* pode ser interpretado como a habilidade do classificador de não classificar classes negativas como positivas.
- $Recall = TP / (TP + FN)$   
*Recall* pode ser interpretado como a habilidade do classificador de encontrar todas as amostras positivas.

- $F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$

O *F1 score* pode ser interpretado como a média harmônica de *Precision* e *Recall*.

Escolhemos *Precision* como a métrica de avaliação do modelo, pois é mais importante que haja um maior número de classes positivas corretas quando avaliamos um modelo para suporte à decisão de investimentos. Queremos que o máximo de decisões positivas estejam corretas. Em contrapartida, podemos ter um baixo valor de *Recall*, o que representa um alto número de falsos negativos. Isso não é impeditivo para uma boa avaliação do modelo, pois limitar as perdas é mais importante que maximizar os ganhos na nossa proposta.

### 3.6 PRÉ-SELEÇÃO E VALIDAÇÃO

Após as etapas de treinamento e testes, temos as principais métricas resultantes armazenadas e prontas para serem utilizadas na pré-seleção de ativos. Posteriormente, validaremos a eficácia da pré-seleção comparando as variações dos preços de fechamento dos ativos pré-selecionados com as variações dos principais índices utilizados como *benchmark* no mercado de investimentos.

#### 3.6.1 PRÉ-SELEÇÃO

Durante a execução dos testes da seção 3.4, armazenamos todas as probabilidades previstas para que sejam utilizadas como critério de pré-seleção dos ativos avaliados. Quanto maior a probabilidade prevista pelo modelo, melhor será a classificação do ativo como possível investimento para os próximos 21 dias úteis, aproximadamente 1 mês.

O período escolhido para essa etapa foi o ano de 2022. Separamos as probabilidades de cada FII avaliado no primeiro dia de negociação de cada mês desse ano. Em seguida, colocamos em ordem decrescente de probabilidade e descartamos os que não estiverem entre os 10 primeiros colocados.

#### 3.6.2 VALIDAÇÃO

Dentre os 10 FIIs separados em cada mês, descartamos aqueles que não tiverem uma probabilidade prevista acima de 60%, visto que este foi o limite mínimo que consideramos para o modelo classificar como 1. Após esse filtro, para cada FII no conjunto restante, a variação do preço da cota entre o primeiro e o último dia de negociação de cada mês é comparada com as variações dos índices utilizados como *benchmark* no mesmo período.

Nesta validação, os valores dos FIIs e dos índices são normalizados pelos valores iniciais de cada histórico para melhor visualização nos gráficos gerados.

#### 4. SIMULAÇÕES E RESULTADOS

Neste capítulo, a implementação da metodologia definida no capítulo 3 é detalhada, assim como os resultados do modelo de classificação e das regras de pré-seleção.

##### 4.1 TREINAMENTO E TESTE

Como descrito na seção 3.3, selecionamos apenas os FIIs cuja data inicial de negociação fosse menor que 05/01/2019. O treinamento e o teste foram executados dez vezes para cada FII, desde a primeira data disponível no histórico até a data mais recente de cada base. 40 FIIs foram submetidos a essa etapa, devido ao filtro da data inicial. Os resultados obtidos representam a média das dez execuções do classificador para cada um e estão expostos na tabela 03 por ordem decrescente.

Tabela 03 – *Precision\_Score* atingido para cada FII

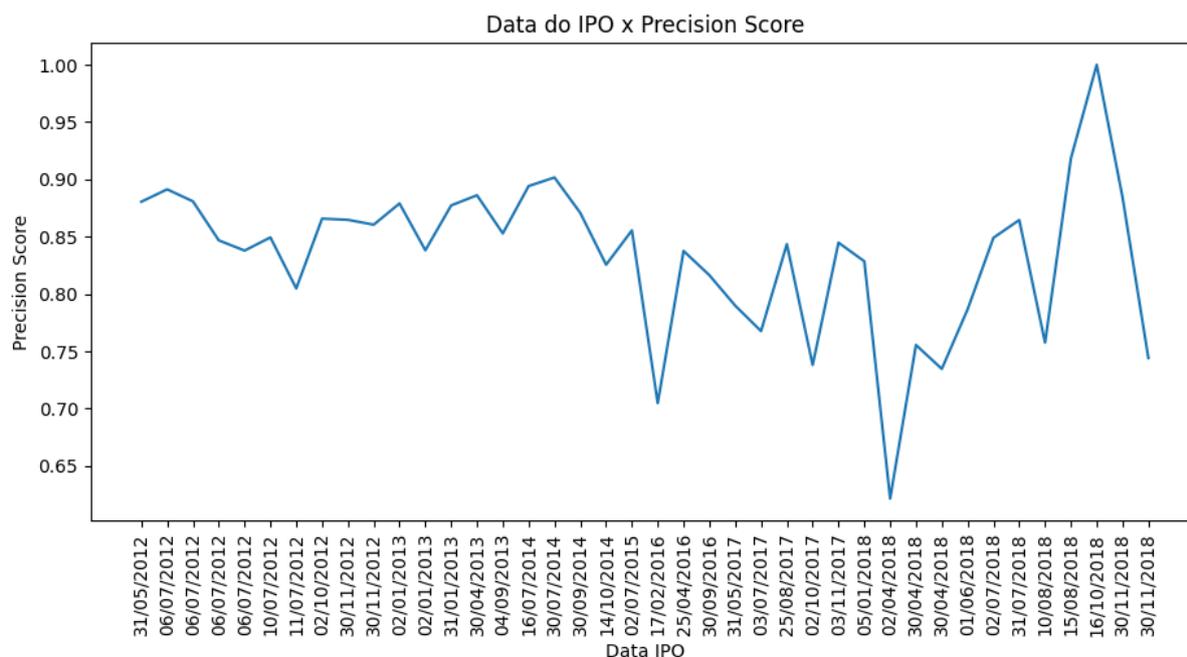
Ticker	Precision	Ticker	Precision	Ticker	Precision	Ticker	Precision
XPIN11	100.00%	HGCR11	87.73%	HGBS11	84.68%	GGRC11	78.99%
RECR11	91.84%	CPTS11	87.08%	VISC11	84.48%	RBRR11	78.67%
BCFF11	90.17%	BRCR11	86.58%	TGAR11	84.35%	OUJP11	76.77%
NSLU11	89.42%	RBVA11	86.47%	SDIL11	83.82%	VTLT11	75.76%
KNRI11	89.13%	HGRU11	86.46%	HGLG11	83.79%	HFOF11	75.56%
BPFF11	88.62%	KNCR11	86.05%	CARE11	83.76%	KNHY11	74.42%
VGIR11	88.44%	BCRI11	85.56%	ALZR11	82.86%	RBRF11	73.81%
VRTA11	88.08%	MFII11	85.29%	PORD11	82.56%	MGFF11	73.47%
MXRF11	88.05%	HGRE11	84.94%	KNIP11	81.64%	BCIA11	70.48%
BBPO11	87.90%	XPLG11	84.91%	JSRE11	80.49%	IRDM11	62.16%

Fonte: Autor (2023)

O valor médio foi de 83,38% e a mediana foi de 84,79%, sendo esta nos FIIs XPLG11 e HGBS11.

Na figura 08, podemos ver a relação entre as datas iniciais de negociação dos FIIs e o *Precision Score* atingido pelo modelo. O gráfico foi elaborado para verificarmos se havia algum padrão de acordo com a quantidade de dados históricos disponíveis no treinamento, porém este não foi o caso.

Figura 08 – Relação entre data do IPO e *Precision Score* atingido

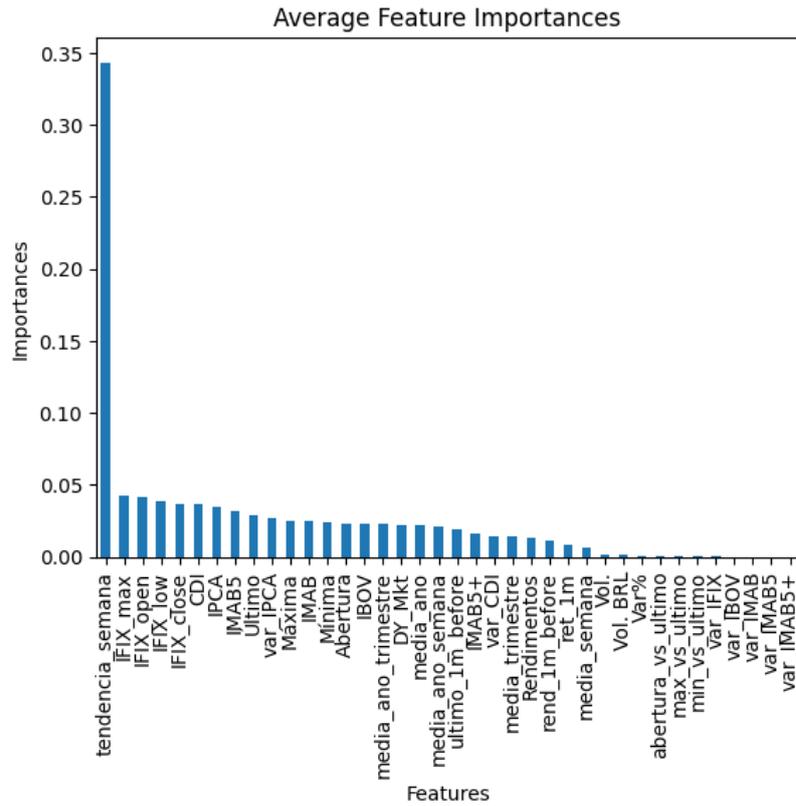


Fonte: Autor (2023)

#### 4.2 ANÁLISE DE ATRIBUTOS

Na seção 3.2 foram listadas as *features* utilizadas no treinamento, com base em outros trabalhos, nos indicadores de análise de FIIs e nas relações entre algumas *features*. Para validar a relevância de cada uma, após a execução do modelo para um FII, armazenamos as importâncias de cada atributo (*feature importances*) com o auxílio de uma propriedade fornecida pelo módulo *RandomForestClassifier* da biblioteca *Scikit Learn*. Na figura 09, podemos observar a média de importância para cada *feature*.

Figura 09 – Média das Importâncias dos Atributos



Fonte: Autor (2023)

Tabela 04 - Média das Importâncias dos Atributos

Feature	Média
tendencia_semana	34.33%
IFIX_max	4.25%
IFIX_open	4.15%
IFIX_low	3.90%
IFIX_close	3.70%
CDI	3.65%
IPCA	3.45%
IMAB5	3.23%
Último	2.90%

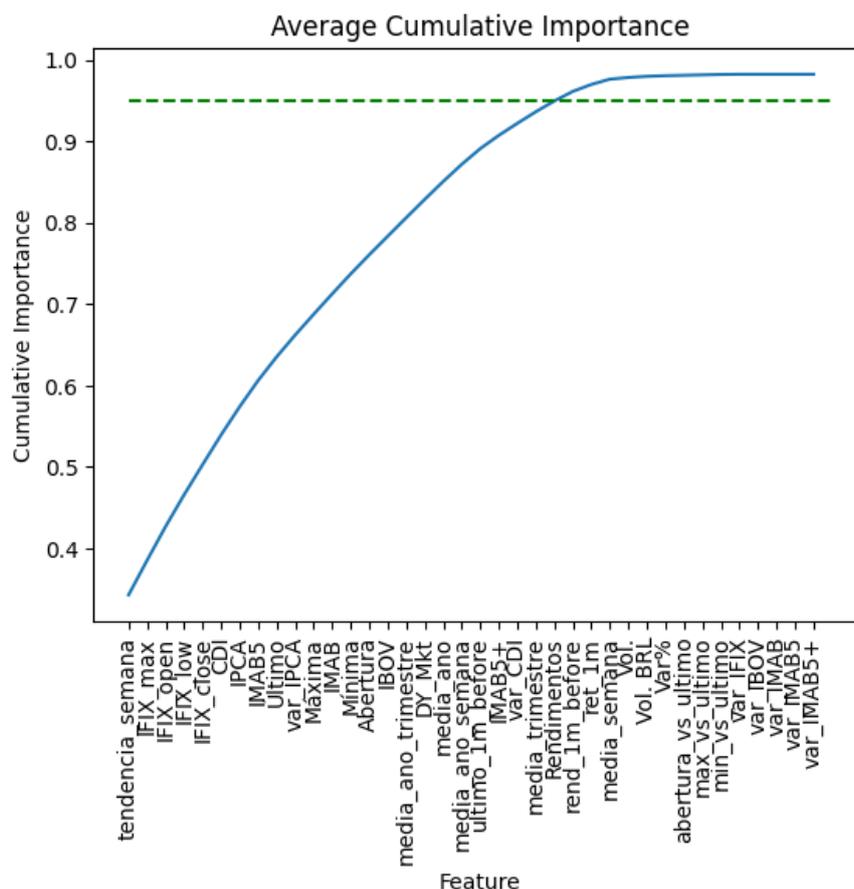
Feature	Média
var_IPCA	2.68%
Máxima	2.55%
IMAB	2.50%
Mínima	2.45%
Abertura	2.35%
IBOV	2.28%
media_ano_trimestre	2.28%
DY_Mkt	2.25%
media_ano	2.20%

Feature	Média
media_ano_semana	2.10%
ultimo_1m_before	1.93%
IMAB5+	1.63%
var_CDI	1.48%
media_trimestre	1.43%
Rendimentos	1.33%
rend_1m_before	1.18%
ret_1m	0.85%
media_semana	0.65%

Feature	Média
Vol.	0.20%
Vol. BRL	0.15%
Var%	0.08%
abertura_vs_ultimo	0.05%
max_vs_ultimo	0.05%
min_vs_ultimo	0.05%
var_IFIX	0.03%
var_IBOV	0.00%
var_IMAB	0.00%
var_IMAB5	0.00%
var_IMAB5+	0.00%

Fonte: Autor (2023)

Figura 10 – Importância acumulada média



Fonte: Autor (2023)

Outra forma de visualizar as importâncias é através do gráfico cumulativo, na figura 10. Vemos que a *Cumulative Importance* atinge 94,95% no atributo de Rendimentos. Ou seja, 14 atributos que aparecem após este limiar representam menos de 5% da importância cumulativa média. Vale ressaltar que o atributo *ret\_1m*, que representa o retorno total do mês anterior ao observado, está nessa parcela menos significativa. Isso vai contra a intuição de muitos investidores que tendem a supervalorizar essa medida, apontado pelas finanças comportamentais como a heurística da disponibilidade, a qual se refere sobre a influência de fatos recentes na tomada de decisão.

Dito isso, o atributo *tendencia\_semana* que indica a medida em que o ativo teve de valorização ou desvalorização na última semana observada, teve uma importância média altíssima, em 34,33%.

#### 4.3 PRÉ-SELEÇÃO

Em posse de todas as probabilidades de que haja uma variação positiva na cota do FII após 21 dias de negociação previstas pelo modelo, montamos rankings de acordo com essa probabilidade para cada início de mês ao longo do ano de 2022. O objetivo é utilizar essa pré-seleção para auxiliar na construção de uma carteira de investimentos em FIIs. No máximo dez fundos com as maiores probabilidades são pré-selecionados. Para que permaneçam na pré-seleção, a probabilidade deve ser maior que 60%, mesmo limite que foi escolhido para o modelo classificar uma predição como 1. A seguir, temos os rankings de cada mês.

Tabela 05 – Ranking por probabilidade de retorno mensal positivo

jan/22		fev/22		mar/22		abr/22		mai/22		jun/22	
Ticker	Prob										
HGRU11	89.96%	VRTA11	72.84%	VGIR11	92.68%	VGIR11	92.67%	VGIR11	91.05%	KNCR11	80.79%
MFII11	87.50%	KNCR11	69.90%	VRTA11	77.84%	MXRF11	79.83%	MXRF11	79.66%	KNHY11	74.53%
HFOF11	85.05%	BBPO11	69.25%	HFOF11	60.76%	HGCR11	79.51%	KNCR11	78.96%	OUJP11	72.08%
NSLU11	78.58%	CPTS11	67.68%	MXRF11	60.72%	KNCR11	76.39%	GGRC11	77.10%	CARE11	71.22%
KNCR11	78.32%	HGCR11	66.71%	CARE11	60.37%	VRTA11	75.78%	MFII11	76.38%	JSRE11	69.92%
ALZR11	78.13%	RBRR11	66.18%	RBRR11	59.71%	OUJP11	72.06%	HGCR11	70.43%	PORD11	64.49%
JSRE11	77.71%	TGAR11	65.68%	SDIL11	59.12%	VTLT11	70.91%	OUJP11	66.66%	RECR11	63.57%
RECR11	75.45%	GGRC11	65.34%	KNCR11	56.45%	KNRI11	70.46%	KNRI11	66.34%	KNIP11	61.45%
BPFF11	74.70%	OUJP11	64.79%	VISC11	54.30%	BCRI11	65.28%	ALZR11	65.38%	HGCR11	61.28%
HGCR11	74.65%	BCRI11	64.44%	HGCR11	51.54%	SDIL11	65.10%	SDIL11	64.93%	HGLG11	59.94%

jul/22		ago/22		set/22		out/22		nov/22		dez/22	
Ticker	Prob										
KNCR11	80.63%	KNCR11	80.84%	HGRU11	87.56%	HGRU11	83.04%	VTLT11	86.54%	VGIR11	54.35%
VGIR11	73.89%	VGIR11	74.85%	ALZR11	84.48%	NSLU11	78.82%	HFOF11	84.87%	KNHY11	50.88%
JSRE11	73.29%	HGRU11	70.90%	KNCR11	82.56%	HFOF11	77.33%	GGRC11	81.97%	BBPO11	49.40%
RECR11	73.06%	XPLG11	70.09%	VGIR11	81.43%	VTLT11	76.74%	BBPO11	79.81%	VTLT11	47.61%
HGCR11	69.97%	TGAR11	69.13%	HGCR11	75.15%	TGAR11	76.04%	MXRF11	72.58%	MXRF11	43.00%
BCRI11	68.24%	HGLG11	68.48%	BPFF11	74.62%	OUJP11	72.58%	MFII11	71.98%	OUJP11	42.71%
OUJP11	67.95%	PORD11	67.23%	JSRE11	71.53%	HGCR11	69.34%	VRTA11	61.93%	KNIP11	40.38%
PORD11	66.44%	BPFF11	65.09%	HGLG11	71.38%	IRDM11	68.08%	VISC11	60.69%	BCIA11	38.93%
HGRU11	59.60%	IRDM11	63.83%	NSLU11	69.28%	MXRF11	66.74%	BPFF11	56.94%	VISC11	38.87%
HGLG11	58.79%	BCIA11	63.75%	HGBS11	67.69%	BPFF11	65.54%	HGRU11	56.38%	RECR11	37.03%

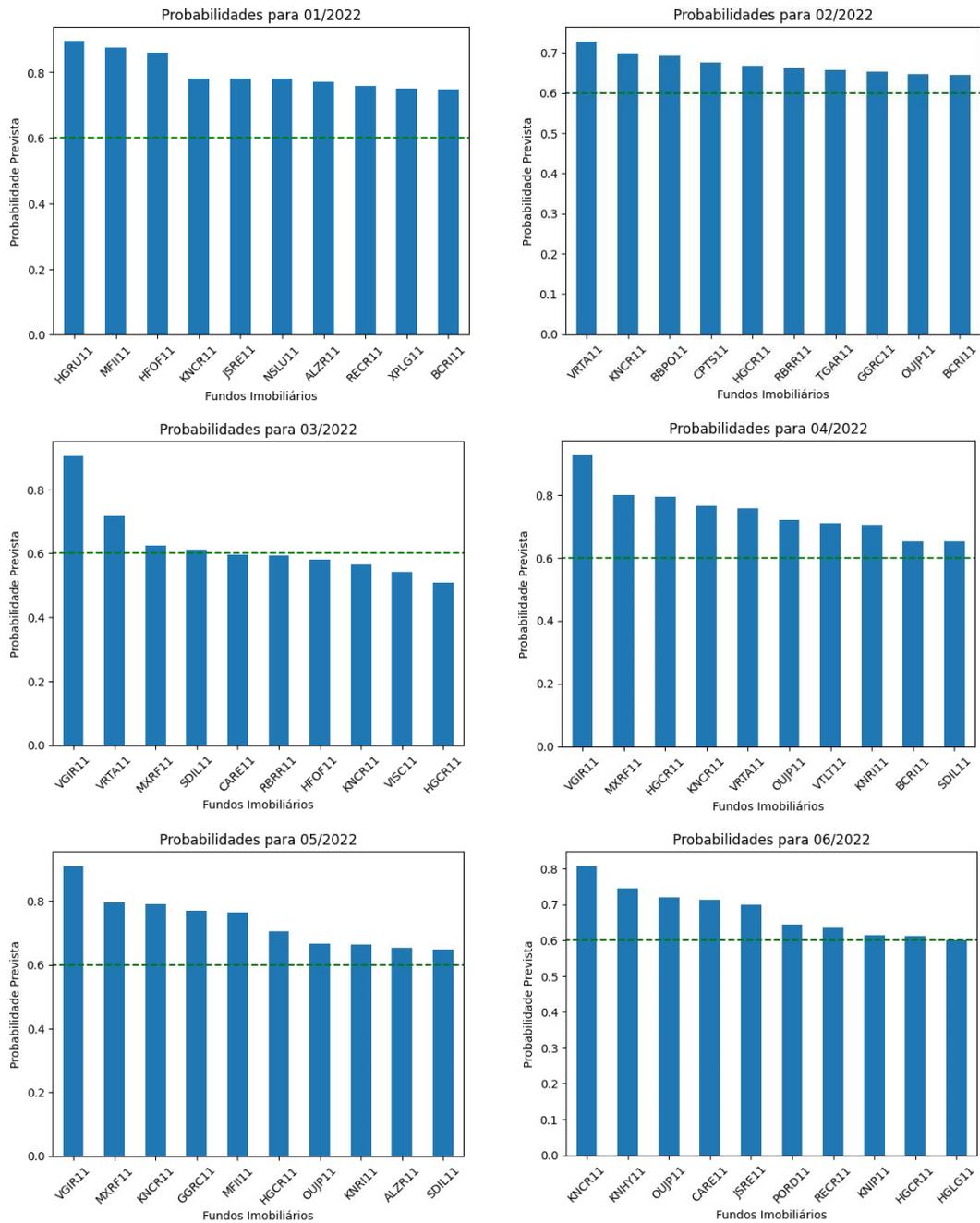
Fonte: Autor (2023)

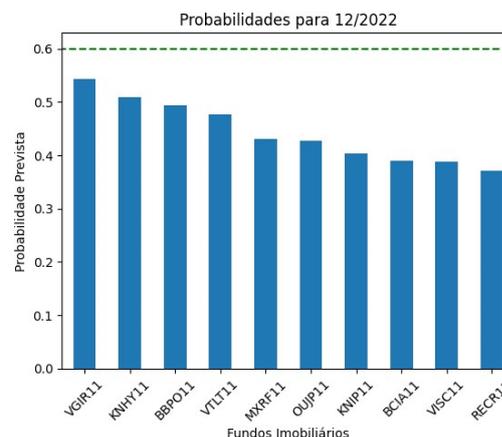
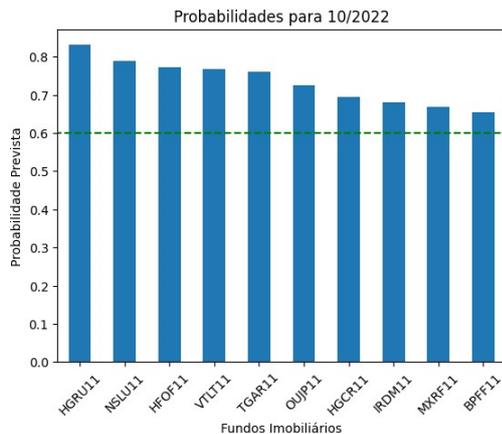
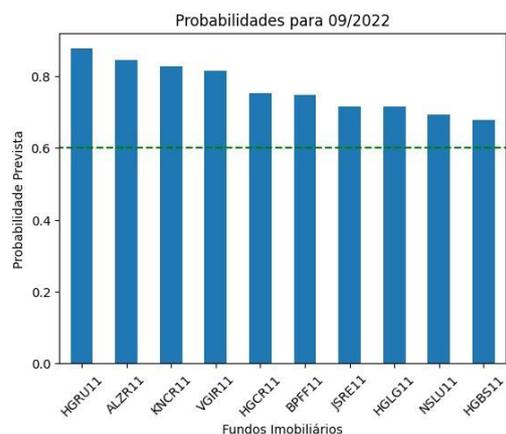
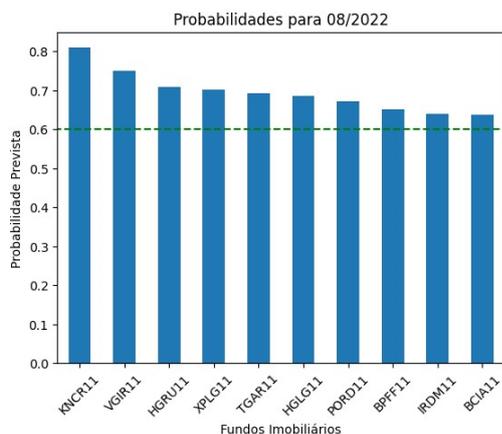
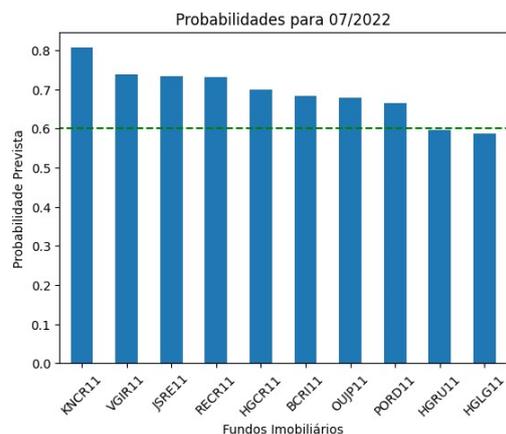
#### 4.4 VALIDAÇÃO

Para validação dos ativos pré-selecionados em cada mês de 2022, comparamos a variação dos preços de fechamento das cotas dos FIIs com as variações dos *benchmarks* IFIX, CDI,

IBOV, IPCA, IMAB, IMAB5, IMAB5+. Os resultados podem ser visualizados nas imagens e tabelas a seguir.

Figura 11 – Gráficos Comparados Probabilidades x Fundos Imobiliários





Fonte: Autor (2023)

Tabela 06 – Tabela com variações mensais

jan/22			fev/22			mar/22		
Ticker	Variação	Var. % CDI	Ticker	Variação	Var. % CDI	Ticker	Variação	Var. % CDI
IBOV	7.91%	1134.65%	BCRI11	1.10%	163.23%	CARE11	43.70%	5095.31%
RECR11	3.58%	512.88%	IMAB5	0.92%	136.31%	IBOV	4.19%	488.49%
HGCR11	2.05%	294.53%	OUIP11	0.88%	130.15%	IMAB5+	3.66%	427.06%
ALZR11	1.50%	214.86%	IPCA	0.70%	103.54%	VGIR11	3.28%	381.84%
MFII11	1.07%	154.02%	VRTA11	0.69%	101.83%	IMAB	3.08%	358.55%
HGRU11	0.73%	105.39%	CDI	0.68%	100.00%	IMAB5	2.52%	293.51%
CDI	0.70%	100.00%	IMAB	0.26%	38.94%	IFIX_close	1.59%	184.86%
IPCA	0.58%	83.05%	IBOV	-0.08%	-11.28%	MXRF11	1.41%	163.82%

IMAB5	0.09%	13.18%
<b>JSRE11</b>	<b>-0.35%</b>	<b>-50.59%</b>
IFIX_close	-0.44%	-62.83%
<b>KNCR11</b>	<b>-0.66%</b>	<b>-94.05%</b>
IMAB	-0.67%	-95.56%
<b>BPFF11</b>	<b>-1.01%</b>	<b>-145.49%</b>
IMAB5+	-1.45%	-207.80%
<b>HFOF11</b>	<b>-2.88%</b>	<b>-413.02%</b>
<b>NSLU11</b>	<b>-3.86%</b>	<b>-553.96%</b>

IMAB5+	-0.41%	-61.26%
<b>KNCR11</b>	<b>-0.90%</b>	<b>-132.38%</b>
IFIX_close	-1.28%	-188.81%
<b>CPTS11</b>	<b>-1.68%</b>	<b>-248.10%</b>
<b>RBRR11</b>	<b>-1.96%</b>	<b>-289.88%</b>
<b>BBPO11</b>	<b>-2.01%</b>	<b>-297.57%</b>
<b>HGCR11</b>	<b>-2.68%</b>	<b>-396.24%</b>
<b>TGAR11</b>	<b>-5.77%</b>	<b>-852.55%</b>
<b>GGRC11</b>	<b>-10.31%</b>	<b>-1524.13%</b>

IPCA	1.17%	136.75%
CDI	0.86%	100.00%
<b>VRTA11</b>	<b>0.41%</b>	<b>48.04%</b>
<b>HFOF11</b>	<b>-0.14%</b>	<b>-16.85%</b>

abr/22		
Ticker	Varição	Var. % CDI
<b>MXRF11</b>	<b>6.93%</b>	<b>877.18%</b>
<b>OUJP11</b>	<b>4.38%</b>	<b>553.68%</b>
<b>BCRI11</b>	<b>2.33%</b>	<b>294.67%</b>
<b>VGIR11</b>	<b>1.79%</b>	<b>226.69%</b>
<b>KNCR11</b>	<b>1.47%</b>	<b>186.15%</b>
IPCA	1.29%	163.29%
IMAB5	1.20%	151.81%
CDI	0.79%	100.00%
IFIX_close	0.51%	64.88%
IMAB	0.11%	13.85%
<b>KNRI11</b>	<b>-0.52%</b>	<b>-66.43%</b>
<b>HGCR11</b>	<b>-0.57%</b>	<b>-71.70%</b>
IMAB5+	-1.01%	-128.18%
<b>VRTA11</b>	<b>-1.16%</b>	<b>-146.45%</b>
<b>SDIL11</b>	<b>-1.19%</b>	<b>-150.46%</b>
<b>VTLT11</b>	<b>-1.34%</b>	<b>-169.71%</b>
IBOV	-11.26%	-1425.43%

mai/22		
Ticker	Varição	Var. % CDI
IBOV	4.42%	459.39%
<b>MXRF11</b>	<b>3.16%</b>	<b>328.08%</b>
<b>HGCR11</b>	<b>2.51%</b>	<b>260.95%</b>
<b>SDIL11</b>	<b>2.13%</b>	<b>221.69%</b>
<b>KNCR11</b>	<b>2.01%</b>	<b>208.87%</b>
IMAB5+	1.38%	143.40%
IMAB	1.07%	111.34%
CDI	0.96%	100.00%
IPCA	0.85%	88.30%
IFIX_close	0.80%	82.82%
IMAB5	0.78%	81.12%
<b>OUJP11</b>	<b>0.65%</b>	<b>67.19%</b>
<b>GGRC11</b>	<b>0.34%</b>	<b>35.36%</b>
<b>ALZR11</b>	<b>-0.64%</b>	<b>-66.36%</b>
<b>KNRI11</b>	<b>-0.82%</b>	<b>-85.34%</b>
<b>VGIR11</b>	<b>-1.46%</b>	<b>-151.46%</b>
<b>MFI11</b>	<b>-3.56%</b>	<b>-369.96%</b>

jun/22		
Ticker	Varição	Var. % CDI
<b>PORD11</b>	<b>3.86%</b>	<b>404.58%</b>
<b>OUJP11</b>	<b>2.68%</b>	<b>281.01%</b>
<b>KNCR11</b>	<b>1.00%</b>	<b>105.38%</b>
CDI	0.95%	100.00%
<b>HGCR11</b>	<b>0.89%</b>	<b>93.48%</b>
<b>RECR11</b>	<b>0.74%</b>	<b>77.36%</b>
<b>KNHY11</b>	<b>0.60%</b>	<b>62.91%</b>
IPCA	0.56%	58.85%
IMAB5	0.18%	18.67%
IMAB	-0.28%	-28.86%
<b>KNIP11</b>	<b>-0.49%</b>	<b>-51.57%</b>
IMAB5+	-0.76%	-79.44%
IFIX_close	-0.97%	-102.03%
<b>JSRE11</b>	<b>-2.27%</b>	<b>-238.23%</b>
<b>CARE11</b>	<b>-11.30%</b>	<b>-1185.32%</b>
IBOV	-11.51%	-1207.32%

jul/22		
Ticker	Varição	Var. % CDI
IBOV	4.26%	431.87%
<b>PORD11</b>	<b>3.51%</b>	<b>355.84%</b>
<b>VGIR11</b>	<b>3.12%</b>	<b>316.50%</b>
<b>KNCR11</b>	<b>2.05%</b>	<b>208.11%</b>
<b>OUJP11</b>	<b>1.06%</b>	<b>107.39%</b>
CDI	0.99%	100.00%
IFIX_close	0.79%	79.89%
<b>BCRI11</b>	<b>0.33%</b>	<b>33.03%</b>
IMAB5	0.20%	20.11%
IPCA	-0.01%	-0.70%
<b>HGCR11</b>	<b>-0.50%</b>	<b>-50.97%</b>
IMAB	-0.65%	-65.89%
IMAB5+	-1.56%	-158.78%
<b>RECR11</b>	<b>-1.85%</b>	<b>-187.59%</b>
<b>JSRE11</b>	<b>-1.87%</b>	<b>-190.17%</b>

ago/22		
Ticker	Varição	Var. % CDI
<b>XPLG11</b>	<b>12.19%</b>	<b>1102.74%</b>
<b>BCIA11</b>	<b>10.21%</b>	<b>923.83%</b>
<b>BPFF11</b>	<b>8.64%</b>	<b>781.80%</b>
IBOV	7.14%	645.70%
<b>HGRU11</b>	<b>6.30%</b>	<b>569.79%</b>
IFIX_close	6.17%	557.88%
<b>TGAR11</b>	<b>5.06%</b>	<b>457.51%</b>
<b>HGLG11</b>	<b>3.72%</b>	<b>336.06%</b>
IMAB5+	2.47%	223.27%
<b>KNCR11</b>	<b>1.79%</b>	<b>162.05%</b>
<b>VGIR11</b>	<b>1.62%</b>	<b>146.62%</b>
CDI	1.11%	100.00%
IMAB	1.06%	95.65%
<b>IRDM11</b>	<b>-0.04%</b>	<b>-3.55%</b>
IMAB5	-0.07%	-6.69%
IPCA	-0.52%	-47.06%
<b>PORD11</b>	<b>-1.50%</b>	<b>-135.48%</b>

set/22		
Ticker	Varição	Var. % CDI
<b>NSLU11</b>	<b>5.42%</b>	<b>531.37%</b>
<b>BPFF11</b>	<b>3.41%</b>	<b>334.20%</b>
<b>HGRU11</b>	<b>2.54%</b>	<b>248.53%</b>
<b>HGBS11</b>	<b>1.67%</b>	<b>163.71%</b>
<b>JSRE11</b>	<b>1.33%</b>	<b>130.10%</b>
IMAB5+	1.27%	124.65%
<b>HGCR11</b>	<b>1.17%</b>	<b>114.54%</b>
CDI	1.02%	100.00%
<b>KNCR11</b>	<b>0.82%</b>	<b>80.06%</b>
IMAB	0.80%	78.55%
IFIX_close	0.74%	72.04%
<b>VGIR11</b>	<b>0.60%</b>	<b>58.73%</b>
IMAB5	0.26%	25.18%
<b>HGLG11</b>	<b>0.24%</b>	<b>23.05%</b>
IPCA	-0.30%	-29.32%
IBOV	-0.33%	-32.70%
<b>ALZR11</b>	<b>-0.65%</b>	<b>-63.93%</b>

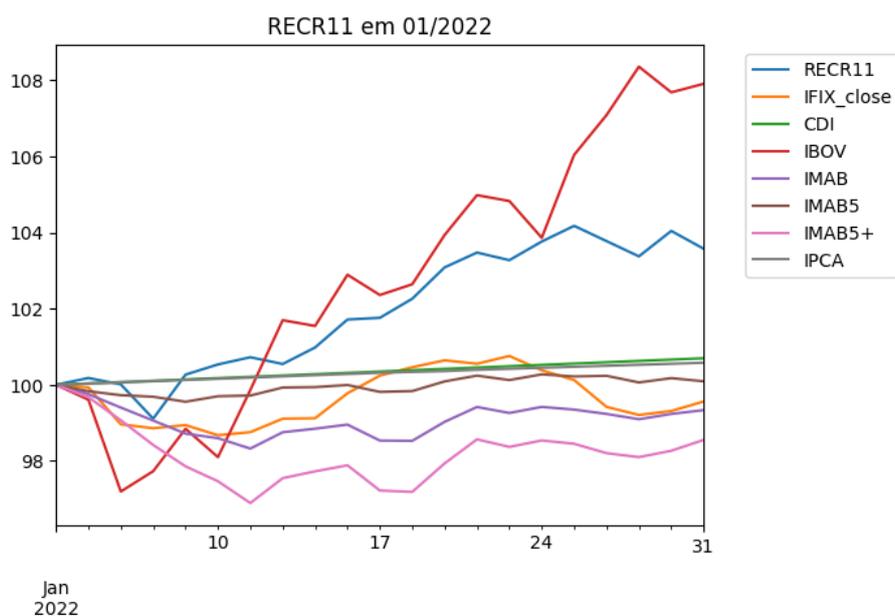
out/22		
Ticker	Varição	Var. % CDI
<b>OUJP11</b>	<b>3.87%</b>	<b>399.57%</b>
<b>VTLT11</b>	<b>2.96%</b>	<b>304.98%</b>
IMAB5	1.81%	186.41%
IMAB	1.05%	107.96%

nov/22		
Ticker	Varição	Var. % CDI
CDI	0.97%	100.00%
IPCA	0.46%	47.52%
IMAB5	-0.32%	-32.65%
IMAB	-0.79%	-81.04%

CDI	0.97%	100.00%	<b>MXRF11</b>	<b>-0.98%</b>	<b>-101.04%</b>
<b>HFOF11</b>	<b>0.67%</b>	<b>68.88%</b>	IMAB5+	-1.18%	-121.58%
<b>MXRF11</b>	<b>0.49%</b>	<b>50.32%</b>	<b>MFII11</b>	<b>-1.85%</b>	<b>-191.05%</b>
IMAB5+	0.40%	41.36%	<b>VTLT11</b>	<b>-2.50%</b>	<b>-257.89%</b>
<b>TGAR11</b>	<b>0.38%</b>	<b>39.45%</b>	<b>BBPO11</b>	<b>-3.80%</b>	<b>-391.71%</b>
IFIX_close	0.29%	29.57%	IBOV	-3.80%	-391.94%
IPCA	0.17%	17.52%	IFIX_close	-4.08%	-420.82%
IBOV	-0.08%	-8.65%	<b>VRTA11</b>	<b>-4.33%</b>	<b>-447.10%</b>
<b>BPFF11</b>	<b>-0.12%</b>	<b>-11.89%</b>	<b>GGRC11</b>	<b>-4.37%</b>	<b>-450.42%</b>
<b>HGCR11</b>	<b>-0.53%</b>	<b>-54.53%</b>	<b>VISC11</b>	<b>-7.69%</b>	<b>-793.09%</b>
<b>HGRU11</b>	<b>-0.63%</b>	<b>-64.62%</b>	<b>HFOF11</b>	<b>-11.11%</b>	<b>-1146.05%</b>
<b>IRD11</b>	<b>-2.64%</b>	<b>-271.86%</b>			
<b>NSLU11</b>	<b>-6.83%</b>	<b>-705.06%</b>			

Fonte: Autor (2023)

Figura 12 – Comparativo gráfico de variações mensais



Fonte: Autor (2023)

Dos 40 FIIs trabalhados, 33 foram pré-selecionados pelo menos uma vez. Os mais frequentes foram HGCR11 e KNCR11, pré-selecionados oito vezes cada, e os seguintes fundos foram pré-selecionados apenas uma vez: BCIA11, CPTS11, HGBS11, KNHY11, KNIP11, RBRR11, VISC11, XPLG11.

Tabela 07 – Frequência de pré-seleção dos FIIs

Ticker	Frequência	Ticker	Frequência	Ticker	Frequência
HGCR11	8	BCRI11	3	IRD11	2
KNCR11	8	GGRC11	3	KNR11	2
OUJP11	6	MFII11	3	SDIL11	2
VGIR11	6	NSLU11	3	BCIA11	1

MXRF11	5	PORD11	3	CPTS11	1
BPFF11	4	RECR11	3	HGBS11	1
HFOF11	4	TGAR11	3	KNHY11	1
HGRU11	4	VTLT11	3	KNIP11	1
JSRE11	4	BBPO11	2	RBRR11	1
VRTA11	4	CARE11	2	VISC11	1
ALZR11	3	HGLG11	2	XPLG11	1

Fonte: Autor (2023)

No período analisado, 100 pré-seleções de FIIs tiveram a classificação 1, ou seja, probabilidade acima de 60% para variação positiva naquele mês. Dos 100, 56 tiveram rentabilidade positiva, resultando em 56% de assertividade. A rentabilidade mensal média entre os 100 FIIs foi de 0,70%. Para o mês de dezembro de 2022, nenhum FII obteve probabilidade acima de 60% na média das dez execuções. A rentabilidade média dos FIIs ficou acima do CDI em 4 meses (março, abril, agosto e setembro) e ficou acima do IFIX em 5 meses (janeiro, março, abril, junho e setembro). A rentabilidade média foi positiva nos meses de janeiro, março, abril, maio, julho, agosto e setembro.

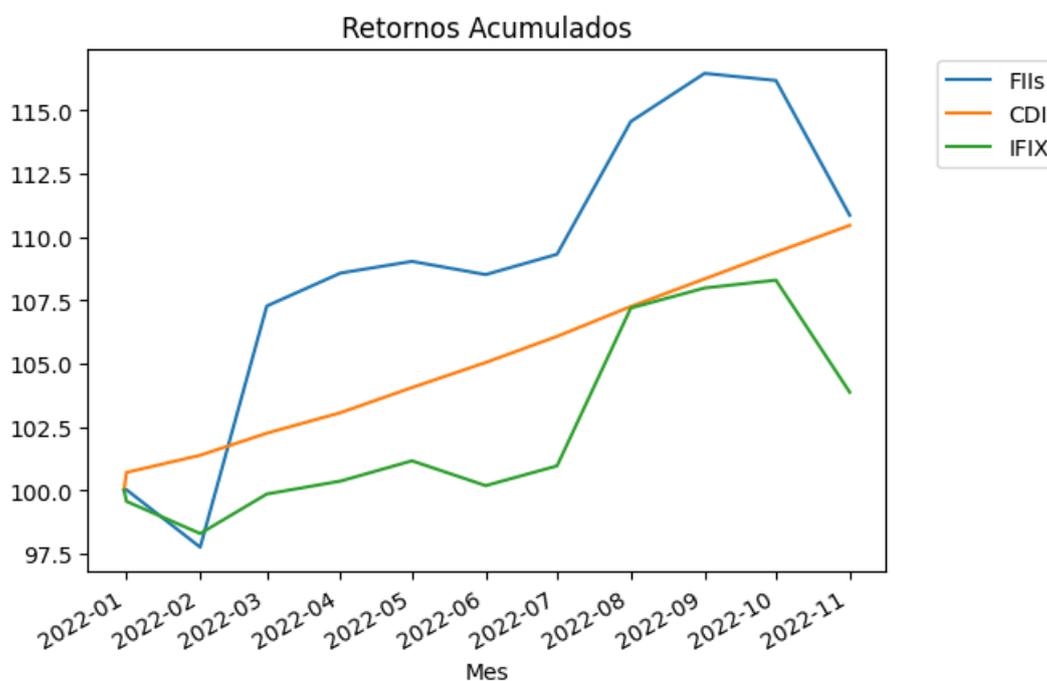
No acumulado dos 11 meses, considerando uma carteira de FIIs composta pelos FIIs pré-selecionados em cada mês, distribuídos igualmente, teríamos os seguintes resultados:

Rentabilidade acumulada da carteira de FIIs: 10,85%, ou 103,82% do CDI e 6,98 pontos percentuais acima do IFIX.

Rentabilidade do CDI no período: 10,45%.

Rentabilidade do IFIX no período: 3,87%.

Figura 13 – Gráfico dos retornos mensais acumulados



Fonte: Autor (2023)

Tabela 08 – Retorno mensal médio dos FIIs selecionados

Mês	Retorno Médio	Ret. Médio em %CDI	Retorno Médio - IFIX
jan/22	0.02%	2.46%	0.46%
fev/22	-2.26%	-334.56%	-0.99%
mar/22	9.73%	1134.43%	8.14%
abr/22	1.21%	153.36%	0.70%
mai/22	0.43%	44.90%	-0.36%
jun/22	-0.48%	-50.04%	0.50%
jul/22	0.73%	74.02%	-0.06%
ago/22	4.80%	434.13%	-1.37%
set/22	1.65%	162.04%	0.92%
out/22	-0.24%	-24.48%	-0.52%
nov/22	-4.58%	-472.29%	-0.50%

Fonte: Autor (2023)

O intuito era apenas uma pré-seleção, mas, no período analisado, uma simples carteira distribuída uniformemente entre os FIIs superou ligeiramente o CDI, mas foi bastante superior ao retorno do IFIX, mesmo que estes não fossem os objetivos da ferramenta.

Podemos observar que, no mês de março/2022, a carteira de FIIs teve uma rentabilidade bem acima da média, sendo responsável pela maior parte do *alpha* – termo que representa a rentabilidade que excedeu o *benchmark*, IFIX – gerado no período dos 11 meses. Também é válido ressaltar que este foi o mês com menos FIIs pré-selecionados – apenas 5 tiveram

probabilidade acima de 60%. Dentre esses 5 FIIs, o CARE11 e o VGIR11 superaram o IFIX, mostrando que o algoritmo foi capaz de fazer uma boa seleção dentre o conjunto total.

Tabela 09 – Métricas comparativas para cada mês analisado

Mês	Qtd FIIs	Qtd FIIs > CDI	Qtd FIIs > IFIX	Qtd FIIs Positivos	% > CDI	% > IFIX	% Positivos
jan/22	10	5	6	5	50.00%	60.00%	50.00%
fev/22	10	3	4	3	30.00%	40.00%	30.00%
mar/22	5	3	2	4	60.00%	40.00%	80.00%
abr/22	10	5	5	5	50.00%	50.00%	50.00%
mai/22	10	4	4	6	40.00%	40.00%	60.00%
jun/22	9	3	7	6	33.33%	77.78%	66.67%
jul/22	8	4	4	5	50.00%	50.00%	62.50%
ago/22	10	8	4	8	80.00%	40.00%	80.00%
set/22	10	6	7	9	60.00%	70.00%	90.00%
out/22	10	2	5	5	20.00%	50.00%	50.00%
nov/22	8	0	4	0	0.00%	50.00%	0.00%

Fonte: Autor (2023)

Tabela 10 – Retornos mensais de cada posição no ranking de probabilidade

Rank Prob.	jan/22	fev/22	mar/22	abr/22	mai/22	jun/22	jul/22	ago/22	set/22	out/22	nov/22
1	0.73%	0.69%	3.28%	1.79%	-1.46%	1.00%	2.05%	1.79%	2.54%	-0.63%	-2.50%
2	1.07%	-0.90%	0.41%	6.93%	3.16%	0.60%	3.12%	1.62%	-0.65%	-6.83%	-11.11%
3	-2.88%	-2.01%	-0.14%	-0.57%	2.01%	2.68%	-1.87%	6.30%	0.82%	0.67%	-4.37%
4	-3.86%	-1.68%	1.41%	1.47%	0.34%	-11.30%	-1.85%	12.19%	0.60%	2.96%	-3.80%
5	-0.66%	-2.68%	43.70%	-1.16%	-3.56%	-2.27%	-0.50%	5.06%	1.17%	0.38%	-0.98%
6	1.50%	-1.96%	n/a	4.38%	2.51%	3.86%	0.33%	3.72%	3.41%	3.87%	-1.85%
7	-0.35%	-5.77%	n/a	-1.34%	0.65%	0.74%	1.06%	-1.50%	1.33%	-0.53%	-4.33%
8	3.58%	-10.31%	n/a	-0.52%	-0.82%	-0.49%	3.51%	8.64%	0.24%	-2.64%	-7.69%
9	-1.01%	0.88%	n/a	2.33%	-0.64%	0.89%	n/a	-0.04%	5.42%	0.49%	n/a
10	2.05%	1.10%	n/a	-1.19%	2.13%	n/a	n/a	10.21%	1.67%	-0.12%	n/a

Fonte: Autor (2023)

Tabela 11 – Retorno médio e proporção de retornos positivos de acordo com o ranking

Rank Prob.	Retorno Médio	Qtd Positivos	% Positivos
1	0.84%	8	72.73%
2	-0.23%	7	63.64%
3	0.06%	5	45.45%
4	-0.32%	6	54.55%
5	3.50%	4	36.36%
6	1.98%	8	80.00%
7	-1.01%	4	40.00%
8	-0.65%	4	40.00%
9	1.04%	5	62.50%
10	2.27%	5	71.43%

Fonte: Autor (2023)

Após a análise dos desempenhos obtidos e das métricas observadas nas simulações, também podemos concluir que uma melhor colocação no ranking de probabilidade não implicou num melhor retorno médio. Selecionar os FIIs com maior retorno não era um dos nossos objetivos e esse resultado reforça o potencial de melhoria numa carteira de investimentos, se o método for utilizado em conjunto com outras técnicas de análise.

A seguir, discorreremos sobre as considerações finais do presente trabalho.

## 5. CONCLUSÕES E TRABALHOS FUTUROS

O objetivo do trabalho foi avaliar o método de *Random Forest* para classificar os FIIs como forma de auxiliar investidores na seleção de ativos. Foi necessário implementar um código para coletar e tratar os dados históricos, analisar as informações obtidas, selecionar os atributos a serem considerados, implementar o modelo de RF e validar os resultados obtidos.

Mesmo com séries históricas reduzidas, quando comparadas ao mercado de ações, consideramos que os resultados obtidos foram satisfatórios. Como observado na seção 4.4, da validação, uma simples distribuição de alocação uniforme entre os FIIs pré-selecionados superou o CDI e o IFIX no período analisado. Vemos isso como um indicador de que o objetivo foi atingido e que tem um bom potencial futuro, pois as classificações poderiam ser utilizadas apenas como um filtro inicial a ser complementado pela análise qualitativa, considerando atributos como cenário macroeconômico, histórico do gestor, situações atípicas em contratos do FII, entre outras. Adicionalmente, em períodos de taxa Selic alta como o que analisamos, superar a rentabilidade do CDI com ativos mais voláteis não é tarefa trivial.

O objetivo específico de implementação do código para coleta e tratamento dos dados históricos também foi um desafio que tomou boa parte do tempo de trabalho, até estar pronto e cumprir seu papel. Esse código final pode auxiliar na economia desse valioso tempo em futuros trabalhos que precisem coletar esses dados.

Sugestões para serem exploradas em trabalhos futuros do mesmo tema:

- Considerar tributação de ganho de capital e custos de operações em caso de rebalanceamento;

- Complementar a ferramenta com uma análise qualitativa de relatórios de *research* dos bancos/corretoras/casas de análise. Por exemplo, a ferramenta proposta por Da Silva Junior, 2021;
- Implementar outros métodos de aprendizagem de máquina como Redes Neurais Artificiais (RNA), KNN, SVM;
- Incluir mais indicadores dos FIIs como *vacância*, *cap rate*, número de ativos, localização dos ativos físicos, concentração por emissor/inquilino;
- Análise de probabilidade de desvalorização para auxílio na decisão de venda de ativos;
- Incluir medidas de índice de Sharpe e de volatilidade para avaliar a carteira formada pelos FIIs.

## REFERÊNCIAS

GUPTA, Suhit et al. **DOM-based content extraction of HTML documents**. In: Proceedings of the 12th international conference on World Wide Web. 2003. p. 207-214.

DIOUF, Rabiyaou et al. **Web scraping: state-of-the-art and areas of application**. In: 2019 IEEE International Conference on Big Data (Big Data). IEEE, 2019. p. 6040-6042.

ZHENG, Chunmei; HE, Guomei; PENG, Zuojie. **A Study of Web Information Extraction Technology Based on Beautiful Soup**. J. Comput., v. 10, n. 6, p. 381-387, 2015.

BALLINGS, Michel et al. Evaluating multiple classifiers for stock price direction prediction. **Expert systems with Applications**, v. 42, n. 20, p. 7046-7056, 2015.

VOGEL, Eduardo Telli. **Análise da classe de fundos imobiliários brasileiros e estudo comparativo com ativos atrelados ao CDI**. 2022.

KODAMA, Leonardo Souza. **Estudo da rentabilidade de fundos imobiliários com utilização da média móvel**. 2021.

SCOLESE, Daniel et al. Análise de estilo de fundos imobiliários no Brasil. **Revista de Contabilidade e Organizações**, v. 9, n. 23, p. 24-35, 2015

MAIA, Paulo Magno Silva; DE SOUZA, Rafael Morais. Análise do mercado de fundos de investimento imobiliário negociados na Bolsa. **Revista de Contabilidade do Mestrado em Ciências Contábeis da UERJ**, v. 20, n. 2, p. 18-36, 2015.

LEUNG, Mark T.; DAOUK, Hazem; CHEN, An-Sing. Forecasting stock indices: a comparison of classification and level estimation models. **International Journal of forecasting**, v. 16, n. 2, p. 173-190, 2000.

PESARAN, M. Hashem; TIMMERMANN, Allan. Predictability of stock returns: Robustness and economic significance. **The Journal of Finance**, v. 50, n. 4, p. 1201-1228, 1995.

CHEUNG, Yin-Wong; CHINN, Menzie D.; PASCUAL, Antonio Garcia. Empirical exchange rate models of the nineties: Are any fit to survive? **Journal of international money and finance**, v. 24, n. 7, p. 1150-1175, 2005.

FU, XingYu et al. **A machine learning framework for stock selection**. arXiv preprint arXiv:1806.01743, 2018.

CAVALCANTE, Rodolfo C. et al. Computational intelligence and financial markets: A survey and future directions. **Expert Systems with Applications**, v. 55, p. 194-211, 2016.

SAITO, Takaya; REHMSMEIER, Marc. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. **PloS one**, v. 10, n. 3, p. e0118432, 2015.

DUDOIT, Sandrine; FRIDLAND, Jane; SPEED, Terence P. Comparison of discrimination methods for the classification of tumors using gene expression data. **Journal of the American statistical association**, v. 97, n. 457, p. 77-87, 2002.

UTANI, LEONARDO AKIRA; DA SILVA, GABRIEL RODRIGUES; FARIA, EDQUEL BUENO PRADO. **SISTEMA ESPECIALISTA PARA APOIO A DECISÃO DE INVESTIDORES NA AQUISIÇÃO DE FUNDOS IMOBILIÁRIOS**, 2022.

DINIZ, Henrique; CARNEIRO, Paulo; SILVA, Fabrício A. 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**. SBC, 2023. p. 49-60.

BARBOSA JUNIOR, Wellington. **Aplicação de Machine Learning para Auxiliar o Processo Decisório de Investimento em Fundos Imobiliários**. 2023.

MUGNAINI, Alexandre; VIEIRA DA SILVA, Wesley; SOUZA, Alceu; MAIA DEL CORSO, Jansen. Análise da eficiência e mercado e performance de fundos de investimentos imobiliários negociados na Bovespa. **Revista Capital Científico do Setor de Ciências Sociais Aplicadas Vol. 6 n° 1 Jan/Dez**, 2008. p. 65-86.

FRANKE, Gabriel de Azevedo. **Fundo de Investimento imobiliário como alternativa de investimento no mercado financeiro – Um estudo comparativo**. Monografia (Trabalho de conclusão - Bacharel em Administração) - Departamento de Ciências Administrativas da Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

Zanandrea, Vinícius. **Avaliação de fundos de investimento imobiliário no Brasil**. Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia de Produção. Universidade Federal de Santa Maria. 2018.

PEREIRA, Victor Borges. **A aplicação da ciência de dados para análise de fundos imobiliários com foco na carteira de galpões logísticos no brasil**. 2023.

WAISMANN, Hugo Casado. **Aplicação do modelo de Markowitz em fundos de investimento imobiliário e análise de desempenho mensal entre janeiro de 2018 e junho de 2019**. 2019. 78 f. Trabalho de Conclusão de Curso (Graduação em Engenharia de Produção) - Universidade Federal Fluminense, Niterói, 2019.

ABREU, André Fidelis Figueiredo de. **Aplicação de machine learning na pré-seleção de ativos para portfólios de investimento**. FGV EESP - MPE: Dissertações, Mestrado Profissional em Economia. 2021. <https://hdl.handle.net/10438/31361>

Lima, Flavio De Souza & da Hora, Henrique & Barcellos, Renato. (2021). **GESTÃO DE PORTFÓLIO DE FUNDOS IMOBILIÁRIOS: UTILIZANDO O MÉTODO AHP-TOPSIS**. 10.29327/154013.24-19.

SILVA, José Erasmo; DE LIMA MONTEBELLO, Maria Imaculada; DOS SANTOS SILVA, Jorge Luiz. INVESTMENTS IN TIMES OF UNCERTAINTY: FORMATION OF PORTFOLIOS USING RANDOM FOREST. **Revista Gestão em Análise**, v. 12, n. 1, p. 121-137, 2023.

COSTA, Thiago Raymon Cruz Cacique da. **Otimização de portfólio com pré-seleção de ativos usando Machine Learning: uma aplicação no contexto dos mercados emergentes**. 2023. 45 f. Dissertação (Mestrado em Economia) — Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, Brasília, 2022.

CAVALCANTI, Adrion de Albuquerque Filho. **Avaliação do método de Random Forest para indicar o melhor momento de Compra e Venda de ações no mercado financeiro**. Trabalho de Graduação em Engenharia da Computação do Centro de Informática, Universidade Federal de Pernambuco, 2021.

DA SILVA JUNIOR, Adilson Angelo. **FERRAMENTA DE AUXÍLIO À TOMADA DE DECISÃO PARA INVESTIMENTOS EM AÇÕES NA BOLSA DE VALORES**. Trabalho de Graduação em Ciência da Computação do Centro de Informática, Universidade Federal de Pernambuco, 2021.

<https://www.b3.com.br/data/files/18/14/1F/30/C9885810F534EB48AC094EA8/Boletim%20FII%20-%2012M22.pdf>

[https://www.b3.com.br/pt\\_br/market-data-e-indices/indices/indices-de-segmentos-e-setoriais/indice-fundos-de-investimentos-imobiliarios-ifix-estatisticas-historicas.htm](https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-de-segmentos-e-setoriais/indice-fundos-de-investimentos-imobiliarios-ifix-estatisticas-historicas.htm)

## APÊNDICE A – COMPOSIÇÃO DO IFIX ENTRE JANEIRO E MARÇO DE 2023

Código	Participação								
KNIP11	6.91%	MCCI11	1.38%	GGRC11	0.90%	GALG11	0.46%	BPFF11	0.26%
KNCR11	5.38%	BRCO11	1.31%	GTWR11	0.87%	RECT11	0.42%	BLMG11	0.24%
HGLG11	3.60%	VRTA11	1.28%	MALL11	0.87%	MFH11	0.41%	SNFF11	0.23%
KNR11	3.13%	HGRE11	1.26%	RBRF11	0.85%	RCRB11	0.40%	NSLU11	0.21%
IRDM11	3.11%	VCJR11	1.25%	BTCI11	0.84%	SNCI11	0.39%	VSLH11	0.21%
XPLG11	2.39%	VILG11	1.21%	PVBI11	0.83%	BARI11	0.35%	WHGR11	0.21%
CPTS11	2.37%	HSML11	1.19%	ALZR11	0.79%	KISU11	0.34%	QAGR11	0.20%
MXRF11	2.22%	KNSC11	1.09%	RZAK11	0.73%	KFOF11	0.33%	LGCP11	0.19%
RECR11	2.07%	RBRR11	1.09%	VINO11	0.69%	PORD11	0.31%	HGFF11	0.19%
HGRU11	2.04%	URPR11	1.08%	XPCI11	0.67%	CPFF11	0.30%	HSAF11	0.19%
XPML11	1.92%	TRXF11	1.07%	HABT11	0.67%	RZAT11	0.30%	XPPR11	0.18%
BTLG11	1.90%	LVBI11	1.05%	VGHF11	0.63%	AIEC11	0.29%	CACR11	0.18%
VISC11	1.87%	DEVA11	1.03%	SDIL11	0.55%	BCIA11	0.29%	VTLT11	0.18%
HGBS11	1.72%	RBVA11	1.01%	SARE11	0.53%	OUJP11	0.29%	RBFF11	0.18%
KNHY11	1.71%	BBPO11	1.25%	RBRP11	0.52%	PATL11	0.29%	TORD11	0.17%
HCTR11	1.63%	JSRE11	1.24%	BCRI11	0.52%	XPSF11	0.29%	PLCR11	0.17%
TGAR11	1.53%	VGIR11	0.97%	RBRY11	0.51%	MCHF11	0.28%	VIUR11	0.17%
HGCR11	1.50%	VGIP11	0.94%	BTAL11	0.49%	BTRA11	0.27%	BLMR11	0.15%
BCFF11	1.50%	CVBI11	0.94%	XPIN11	0.48%	SADI11	0.27%	NCHB11	0.15%
HFOF11	1.44%	HSLG11	0.92%	MGFF11	0.47%	AFHI11	0.27%	MORE11	0.14%
BRCR11	1.44%	RZTR11	0.92%	RBRL11	0.46%	TEPP11	0.27%	RBHG11	0.14%
								ARRI11	0.13%
								MORC11	0.12%
								HOF11	0.09%
								RELG11	0.09%
								CARE11	0.07%
								IBCR11	0.07%

Fonte: B3 (2023)