

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CAMPUS DE DOIS VIZINHOS
CURSO DE ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS

JOSÉ GREGÓRIO DO ROZÁRIO

**APRENDIZADO DE MÁQUINA APLICADO À SELEÇÃO DE
FUNDOS DE INVESTIMENTO IMOBILIÁRIOS**

TRABALHO DE CONCLUSÃO DE CURSO DE ESPECIALIZAÇÃO

DOIS VIZINHOS
2022

JOSÉ GREGÓRIO DO ROZÁRIO

**APRENDIZADO DE MÁQUINA APLICADO À SELEÇÃO DE
FUNDOS DE INVESTIMENTO IMOBILIÁRIOS**

**MACHINE LEARNING APPLIED ON REAL ESTATE INVESTMENT
FUNDS SORTING**

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Orientador: Prof. Me. Francisco Pereira Junior

DOIS VIZINHOS
2022



4.0 Internacional

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

JOSÉ GREGÓRIO DO ROZÁRIO

**APRENDIZADO DE MÁQUINA APLICADO À SELEÇÃO DE
FUNDOS DE INVESTIMENTO IMOBILIÁRIOS**

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Data de aprovação: 09/setembro/2022

Francisco Pereira Junior
Mestrado

Universidade Tecnológica Federal do Paraná - Câmpus Cornélio Procópio

Rosangela de Fátima Pereira Marquesone
Doutorado

Universidade Tecnológica Federal do Paraná - Câmpus Cornélio Procópio

Henrique Yoshikazu Shishido
Doutorado

Universidade Tecnológica Federal do Paraná - Câmpus Cornélio Procópio

DOIS VIZINHOS
2022

RESUMO

O presente texto visa apresentar uma estratégia de seleção automatizada de investimentos em fundos imobiliários, utilizando aprendizado de máquina, bancos de dados relacionais e parte de análise fundamentalista. O estudo tem por objetivo reduzir o leque de opções de investimentos para tornar a tarefa do investidor menos onerosa.

Palavras-chave: aprendizado de máquina; fundos de investimento imobiliário; mercado de ações.

ABSTRACT

This text aims to present a strategy for automatic real estate funds sorting, applying machine learning, relational databases and pieces of value investing. This research's main goal is to reduce the wide variety of investments options to ease investor's daily life.

Keywords: machine learning; real estates funds; stock market.

LISTA DE FIGURAS

Figura 1 – Indicadores fundamentalistas do MXRF11	14
Figura 2 – PDCA (esquerda) e CRISP-DM (direita)	19
Figura 3 – Modelo Entidade Relacionamento da Base de Dados	21
Figura 4 – Fluxo de Execuções (Elaborado pelo próprio autor, 2022)	22

LISTA DE TABELAS

Tabela 1 – Registro de FIIIs ao longo dos anos	15
Tabela 2 – Distribuição de FIIIs por faixa de valores	16
Tabela 3 – Intervalo de Índice Beta para Classificação de FIIIs	25
Tabela 4 – Intervalo de DY para Classificação de FIIIs	26
Tabela 5 – Acurácia de algoritmos de classificação por beta	28
Tabela 6 – Resultados de algoritmos de classificação por DY	29

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i>
B3	Brasil Bolsa Balcão
BDR	<i>Brazilian Depositary Receipts</i>
BOVESPA	Bolsa de Valores de São Paulo
CRISP	<i>Cross Industry Standard Process</i>
CVM	Comissão de Valores Mobiliários
ETF	<i>Exchange Traded Fund</i>
FII	Fundo de Investimento Imobiliário
IFIX	Índice de Fundos de Investimentos Imobiliários
JCP	Juros sobre Capital Próprio
MER	Modelo Entidade Relacionamento
PDCA	<i>Project, Do, Check, Act</i>

LISTA DE ALGORITMOS

Algoritmo 1 – Algoritmo Padrão de Atualização do Banco de Dados	23
Algoritmo 2 – Algoritmo Para Inserção de Novo FII	23

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Fundos de Investimento Imobiliários	11
1.2	Problema de Pesquisa	11
1.3	Justificativa	11
1.4	Objetivos	11
1.4.1	Objetivo Geral	12
1.4.2	Objetivos Específicos	12
1.5	Organização do Trabalho	12
2	REVISÃO DE LITERATURA	13
2.1	Mercado de Ações e a Bolsa de Valores Brasileira	13
2.2	Análise de Mercado	13
2.3	Fundos de Investimento Imobiliários	14
2.4	A Escolha de Fundos para Investir	16
2.4.1	Beta	16
2.4.2	Dividendos	17
2.5	Trabalhos relacionados	17
2.6	Utilização de algoritmos de aprendizado de máquina supervisionado	18
3	DESENVOLVIMENTO	19
3.1	Metodologia	19
3.2	Materiais	19
3.3	Métodos	20
3.4	Recuperação dos dados	21
3.4.1	Atualizações diárias	21
3.4.2	Atualizações a cada duas horas	22
3.4.3	Algoritmo de atualização dos dados	22
3.5	Armazenamento de Dados	23
3.6	Interpretação/Modelagem	23
3.6.1	Vetores de características	23
3.6.2	Caracterização de datas	24
3.6.3	Intervalo de tempo dos conjuntos de treinamento/teste e validação	24
3.6.4	Seleção de Fundos de Investimentos Imobiliários para modelagem de aprendizado de máquina	25
3.7	Definição de intervalos de beta para classificação	25
3.8	Definição de intervalos de DY para classificação	26
3.9	Treinamento e teste dos modelos	26
3.10	Validação dos modelos	27
4	RESULTADOS	28
4.1	Resultados de classificação por beta	28
4.2	Resultados de classificação por DY	29
5	CONCLUSÃO	30
5.1	Considerações Finais	30

5.2	Limitações	30
5.3	Trabalhos Futuros	30
	REFERÊNCIAS	32

1 INTRODUÇÃO

1.1 Fundos de Investimento Imobiliários

A história dos Fundos de Investimento Imobiliários (FIIs) remonta a 1993, com a promulgação da Lei 8668/93, no Brasil, seguindo tendências do mercado Norte Americano, conforme indicado por (BARONI; BASTOS, 2018). Ainda segundo (BARONI; BASTOS, 2018), a tendência de investimento nesse tipo de fundo se dá pela obrigatoriedade de pagamentos mensais de dividendos, algo facultativo para outros ativos da bolsa.

Uma característica que atrai *stakeholders*¹ para essa categoria de investimento é a obrigatoriedade de pagamento mensal de proventos (BARONI; BASTOS, 2018). Além dessa obrigatoriedade, a lei que regulamenta estes fundos exige um piso para esses proventos, que está atrelado aos rendimentos mensais do fundo. Em resumo: investidores buscam esse tipo de ativo como uma fonte de renda passiva, com pagamentos periódicos, aproximadamente estáveis.

Vale ressaltar que FIIs são tratados como ações na bolsa de valores. Por esse motivo, fundos imobiliários são negociados em pregões da bolsa, no caso do Brasil, na B3.

Com o intuito de apresentar ao leitor ferramentas mínimas para discorrer pelo presente texto, essa seção visa explicar o mercado de ações, a bolsa de valores brasileira e o subgrupo de ações intitulado Fundos de Investimento Imobiliários (FIIs).

1.2 Problema de Pesquisa

É difícil escolher onde investir. Em linhas gerais, dentre as centenas de opções do mercado de ações, é complexo definir quanto investir e onde, ou seja, que ações negociar. Se restringirmos a categoria de ações a FIIs, mesmo assim, o leque ainda é grande. Há cerca de 320 FIIs atualmente negociados na B3.

Não é objetivo desse trabalho indicar vantagens e desvantagens de FIIs perante outros tipos de investimentos, como imóveis em si.

Esse estudo visa tornar a escolha de negociação mais restrita, deixando o investidor livre para utilizar seu tempo em outras atividades;

1.3 Justificativa

Existem inúmeras variáveis para se levar em conta quando se deseja investir. Índices, conceitos, notícias, entre outros tópicos.

A análise fundamentalista, muito defendida por ícones como Warren Buffett e Luiz Barsi Filho, possui ferramentas interessantes para o estudo de ações, porém essa análise se limita a valores objetivos, sem levar em consideração o uso de ferramentas contemporâneas, como *machine learning*.

1.4 Objetivos

Os principais objetivos do trabalho são apresentados a seguir.

¹ Stakeholders em mercado de ações são acionistas

1.4.1 Objetivo Geral

Criar um banco de dados, com informações válidas e condizentes com cotações de FIIs, bem como manter esses dados atualizados. É importante ressaltar que nem todos os dados alimentados no banco de dados são utilizados para classificação de FIIs, nesse texto. Parte desses dados não utilizados ficam disponíveis para posteriores estudos.

Classificar FIIs para investimentos, através de classificação de risco e de retorno através de dividendos.

1.4.2 Objetivos Específicos

1. Montar uma base de dados com valores estruturados sobre FIIs;
2. Manter essa base de dados periodicamente atualizada;
3. Utilizar *machine learning* para classificar FIIs segundo critérios de risco e pagamento de proventos;
4. Criar uma máquina virtual para armazenar o banco de dados criado no trabalho, bem como para executar os *scripts* que mantém esse banco atualizado;

1.5 Organização do Trabalho

O trabalho está organizado de maneira a pormenorizadamente explicar sobre a pesquisa, apresentá-la em detalhes e concluir o que foi alcançado.

Cada capítulo visa esclarecer o caminho trilhado, a saber:

- [Capítulo 2](#): indica as fontes geradoras de dados que embasam tanto a pesquisa quanto fatos;
- [Capítulo 3](#): aponta a maneira como o trabalho foi desenvolvido, indicando algoritmos e como esses foram aplicados;
- [Capítulo 4](#): resume os resultados alcançados no [Capítulo 3](#);
- [Capítulo 5](#): aponta as considerações finais acerca dos métodos aplicados e resultados obtidos;

2 REVISÃO DE LITERATURA

Esta seção visa dar o embasamento teórico do estudo, indicando as fontes utilizadas que guiaram a pesquisa e o presente texto.

2.1 Mercado de Ações e a Bolsa de Valores Brasileira

Tendo como objetivo fomentar a troca de bens e serviços, não apenas daquilo que já existe mas também o que poderá trazer lucros aos investidores, criou-se em meados do século XVI o princípio do mercado de ações, também conhecido como *stock market* (BIANCA, 2019). Tem-se relatos de que a primeira bolsa de valores iniciou suas atividades em 1531, na Bélgica. Entretanto, o primeiro relato de ações comercializadas nesse formato remonta a 1602. Também há relatos de mercadores que frequentavam a casa do Sr. Van der Burse, em meados de 1487, também na Bélgica, para comercializar metais, moedas e letras de câmbio. Possivelmente daí venha o nome de Bolsa de Valores (WIKIPÉDIA, 2022).

Inglaterra e Estados Unidos, com o intuito de proteger seu capital e acompanhar a tendência de mercado, instituíram suas bolsas de valores em 1571 e 1792, respectivamente (BIANCA, 2019). Nesse mercado podia-se investir em empresas desbravadoras, como foi o caso da Companhia Holandesa das Índias Orientais, ou ainda comprar frações (ações) de empresas de mineração, têxteis ou de transporte. O comércio desses títulos começou a ocorrer em pregões, isso é, seções de comércio dentro da bolsa de valores (WIKIPÉDIA, 2022).

No Brasil essa tendência levou mais tempo para ser implantada, e muito mais tempo para ser absorvida por investidores pequenos. A primeira bolsa de valores brasileira iniciou suas atividades em 1843, no Rio de Janeiro, graças às políticas financeiras do então Ministro da Fazenda, Alves Branco (BIANCA, 2019).

O Brasil chegou a contar com 27 bolsas operando simultaneamente, na década de 1960. Mas, os investidores pequenos, também conhecidos como investidores pessoa física, tiveram mais espaço na Bolsa de Valores brasileira somente no início da década de 1990 (REIS; TOSETTO, 2017).

Com o intuito de unificar as operações das bolsas, no ano 2000, houve um acordo entre as duas maiores bolsas de valores brasileiras, à época, para direcionar os títulos negociados por cada uma: a Bolsa de Valores de São Paulo (BOVESPA) ficaria com títulos privados, enquanto a Bolsa de Valores do Rio de Janeiro (BVRJ) negociaria títulos públicos (WIKIPÉDIA, 2022).

2.2 Análise de Mercado

A análise técnica de mercados, também conhecida como análise gráfica, consiste numa ferramenta indicativa de probabilidade de tendências de preços, que possibilitam ao investidor antever momentos de compra e venda de ativos (SANTANA, 1997). Amplamente utilizada por investidores do *day-trade*, foca no curto prazo, visto que parte da premissa: comportamentos ocorridos por um dado ativo, no passado, podem ocorrer novamente e indicam tendências de compra ou venda.

Em contrapartida, a análise fundamentalista apresenta o estudo da situação financeira, contábil e as perspectivas de uma empresa, empregando análise de muitas variáveis envolvidas com o ativo (MONEY, 2021). Como exemplo, pode-se citar a análise feita pelo sítio eletrônico Fundamentus (Figura 1), em que indicadores como receita trimestral, lucro por ação, *dividend yield*, e muitos outros estão disponíveis para investidores consultarem. De maneira muito

simplista a análise fundamentalista visa separar ações baratas de ações caras, sendo que o preço da ação não é o único fator relevante para essa conclusão (MEIRELLES, 2012).

Figura 1 – Indicadores fundamentalistas do MXRF11

FII	MXRF11	Cotação	9,92
Nome	MAXI RENDA FUNDO DE INVESTIMENTO IMOBILIARIO - FII	Data últ cot	11/08/2022
Mandato	Renda	Min 52 sem	8,54
Segmento	Híbrido	Max 52 sem	9,92
Gestão	Ativa	Vol \$ méd (2m)	6.004.090
Valor de mercado	2.242.310.000	Nro. Cotas	226.039.663
Relatório	31/05/2022 ↓	Últ Info Trimestral	31/03/2022
Oscilações		Indicadores	
Dia	0,10%	FFO Yield	8,62%
Mês	0,81%	Div. Yield	10,2%
30 dias	2,06%	P/VP	0,98
12 meses	10,08%	Resultado	
2022	6,50%	Últimos 12 meses	
2021	5,29%	Receita	219.038.000
2020	-14,57%	Venda de ativos	51.501.600
2019	54,82%	FFO	193.201.000
2018	16,63%	Rend. Distribuído	230.560.000
		Últimos 3 meses	
		Receita	59.996.300
		Venda de ativos	10.850.700
		FFO	54.020.300
		Rend. Distribuído	47.468.300

Fonte: Adaptado de MXRF11

Ícones do cenário de investimos como Luiz Barsi Filho e Warren Buffett se destacam pelo uso da Análise Fundamentalista, que prega como *slogan* a frase: "Comprar e Manter" (*buy and hold*). Diferente de estratégias gráficas, a abordagem fortemente defendida por Buffett e Barsi Filho se baseia no uso de estatísticas e foco em pontos chave de investimentos, ora encarados de maneira individual, ora encarados de maneira coletiva (MONEY, 2021). Essa análise não conta apenas com simpatizantes de peso, mas também com resultados expressivos, que justificam seu uso (BOMFIM, 2022).

Aqueles que se interessarem por investimentos na bolsa de valores podem iniciar sua trajetória acessando o sítio eletrônico da B3 (Brasil, Bolsa, Balcão). Lá se tem uma visão geral de ativos relacionados na Bovespa, bem como informações detalhadas de cada ativo, sendo eles ações ordinárias e preferenciais de empresas, fundos de investimento imobiliário, pacotes de ações (conhecidos como ETFs) e espelhamentos de ações estrangeiras (conhecidos como BDRs) (B3, 2022).

2.3 Fundos de Investimento Imobiliários

Criados em 1993 através da lei 8668/93, os FIIs são fundos cujo foco de aplicação está ligado a empreendimentos e/ou ativos do setor imobiliário e são negociados da mesma maneira de ações regulares, BDRs e ETFs (BARONI; BASTOS, 2018) - na bolsa de valores. Como quaisquer ações do mercado, os FIIs também sofrem oscilações, ocasionadas por fatores internos, tais quais índice SELIC, taxa de inadimplência, vacância de imóveis, taxa de endividamento da população, ou por fatores externos, tais como Risco Brasil, variação cambial, guerras, entre outros (BARONI; BASTOS, 2018).

Conforme pode ser observado na [Tabela 1](#), entre 1995 e 2010 foram registrados 58 FIs na Comissão de Valores Mobiliários (CVM). Já entre 2011 e 2021, últimos 10 anos, 230 fundos foram registrados. Destes, 51 só em 2020. Além disso, segundo ([ROCHA, 2022](#)), a procura por esse tipo de investimento teve um aumento em se tratando da quantidade de investidores, com um crescimento de 660% entre os anos de 2018 e 2022.

Tabela 1 – Registro de FIs ao longo dos anos.

Ano de Registro	Quantidade
1995	3
1996	3
1997	2
1998	1
1999	2
2000	2
2001	1
2002	1
2003	1
2004	4
2006	4
2007	5
2008	7
2009	7
2010	15
2011	22
2012	24
2013	14
2014	9
2015	5
2016	18
2017	25
2018	31
2019	5
2020	51
2021	26

Fonte: [fiis.com.br](https://www.fiis.com.br) (2021)

Além de serem muito mais estáveis do que outros ativos operados em bolsa, contam com uma característica que atrai muitos investidores - a obrigatoriedade de se pagar dividendos, mensalmente, com valor mínimo estipulado por lei. Isso, atrelado ao fato de que muitos FIs são acessíveis até para o investidor menos abastado, é muito convidativo para quem está iniciando no mercado de ações, e para aqueles que desejam rendimentos mensais passivos. A [Tabela 2](#) mostra que há FIs para diferentes poderes de compra, como o caso daqueles abaixo de 10 reais por cota.

Ainda, por se tratar de um tipo especial de ação, foi criado o Índice de Fundos Imobiliários (IFIX) que é um índice calculado pela média ponderada de 106 FIs. Esse índice é utilizado como referência para acionistas e apresenta uma visão simplificada do segmento

Tabela 2 – Distribuição de FIIs por faixa de valores

Acima de 100 reais	Entre 10 e 100 reais	Abaixo de 10 reais
63	161	27

Fonte: fiis.com.br (2021)

de fundos imobiliários perante outros investimentos (B3, 2021). Tais fundos são escolhidos conforme o grau de relevância dentro do conjunto de fundos imobiliários ativos (B3, 2021).

Porém, nem tudo é vantagem: diferente de ações regulares, que tem imposto de renda retido apenas quando o montante de ações vendidas, no período de um mês ultrapassar R\$ 20.000,00, fundos de investimento imobiliários são taxados em 20% sobre qualquer lucro líquido obtido, independente dos lucros de dividendos ou JCP. Isso indica que o investidor deve pensar muito bem ao comprar um FII, pois o momento da venda pode ser delicado, caso haja lucro líquido, ou seja, o valor de venda superar o valor de compra (UOL, 2022).

Ressalta-se também que fundos imobiliários não podem nem devem ser tratados como renda fixa, mesmo que o pagamento de proventos seja regular. O conceito vem da diferença entre as rendas variável e fixa: na primeira não é possível provisionar o valor do retorno de investimento, no momento em que esse é feito. Na segunda é possível calcular a rentabilidade (BARONI; BASTOS, 2018).

2.4 A Escolha de Fundos para Investir

Pode-se dizer que o ato de escolher onde investir é a segunda tarefa mais difícil para o investidor: a primeira é a escolha de iniciar a investir. Levando isso em consideração, e sabendo a vasta gama de investimentos no mercado, é natural para os investidores pensar em criar estratégias de investimentos e/ou ferramentas que facilite a escolha de um investimento.

Neste presente estudo foram utilizados dois indicadores, para classificação de FIIIs:

2.4.1 Beta

Esse índice leva em conta um critério balizador (referência) para medir o quão similar um ativo (ação, fundo de investimento, etc) se comporta com relação a essa baliza (PÓVOA, 2012). O Beta é um excelente indicador de risco de investimentos. Em outras palavras, utilizando-se um critério estável, com baixo risco, tem-se o seguinte cenário:

- Investimentos com **beta alto**, relativos ao critério estável, levam o investidor a correr um alto risco;
- Investimentos com **beta baixo**, relativos ao critério estável, indicam baixo risco ao investidor.

Uma característica muito peculiar é a possível existência de beta negativo, que indica que o ativo tem um comportamento avesso ao critério balizador:

- Quando os retornos monetários da referência sobem, os retornos do ativo descem;
- Quando os retornos monetários da referência descem, os retornos do ativo sobem;

Em se tratando de FIIIs, o critério de referência é o Índice de Fundos Imobiliários (IFIX). O cálculo do beta, para um FII, é conseguindo através da seguinte equação:

$$\text{beta} = \text{covariância}(\text{FII}, \text{IFIX}) / \text{variância}(\text{IFIX})$$

2.4.2 Dividendos

Outro critério classificador é o pagamento de dividendos. Dividendo é uma palavra derivada do termo dividir. No mercado de ações, dividendo é a distribuição de parte dos lucros recebidos por empresas listadas em bolsa (REIS; TOSETTO, 2017).

Conforme (BARONI; BASTOS, 2018) e (REIS; TOSETTO, 2017) um dos grandes atrativos de FIs é a periodicidade e piso de pagamento de dividendos. Essa constância de pagamentos, somada a uma estratégia de investimentos, garante recebimentos mensais, que em longo prazo refletem em renda passiva e aposentadoria. Como mencionado, FIs são obrigados por lei a pagar mensalmente uma quota mínima de sua receita, dividida igualmente entre seus investidores.

Para fins de classificação, utiliza-se a informação *Dividend Yield* (DY), que é a média de retornos dos dividendos pagos pelo fundo, dividida pela média do retorno do fundo, num período de 12 meses (REIS; TOSETTO, 2017). Para quaisquer classes de investidores que prezem retornos através de dividendos, a informação do DY é imprescindível. Altos DYs não significam necessariamente bons investimentos pois, dificilmente um fundo conseguirá manter esse retorno por muito tempo (REIS; TOSETTO, 2017).

O uso do DY para classificação de FIs visa munir investidores que foquem em renda passiva a longo prazo. Os intervalos de tempo utilizados para classificação de Beta e DY são explanados mais detalhadamente na [Seção 3.2](#).

2.5 Trabalhos relacionados

Em se tratando de aquisição de dados financeiros, o trabalho relacionado em (CANTO, 2021) é o mais próximo. A referência utilizada para captação de cotações diárias é a mesma, o portal investing.com, tal qual o exemplo: (INVESTING, 2021).

A maneira como os dados são captados difere: utiliza-se para essa biblioteca o método RESTful API, que nada mais é do que requisições do tipo GET, PUT e POST.

Essa abordagem é muito mais elegante do que o *web scrapping*. Porém, sua limitação reside no fato de o servidor do portal investing.com limitar requisições consecutivas, em tempo, ou seja, entre uma requisição e outra, é necessário aguardar no mínimo dois segundos (BARTOLOME, 2022), tempo esse testado pelo autor.

Tal limitação não seria problema, visto que para atualizar os Fundos Imobiliários brasileiros seriam necessários em torno de quinhentos segundos, ou aproximadamente oito minutos e trinta segundos, desconsiderando-se possíveis latências de conexão ou lentidões de processamento. Porém, isso pode ser um limitante caso a taxa de amostragem das cotações seja reduzida ao menor valor possível, que é de cinco minutos. Se tal taxa for desejada, utilizar a biblioteca investpy não é uma opção. Por esse motivo, além é claro de não depender de outra API, optou-se por manter o *web scrapping*.

O trabalho descrito em (REIS, 2021), em que a classificação de ativos, não apenas de FIs, leva em conta análise fundamentalista e indicadores que vão além de beta e DY, indica trabalhos nesse campo, porém sem aplicação de técnicas computacionais para isso.

Em se tratando de *machine learning* aplicado à predição de ativos o texto (LI, 2022) aborda o uso da técnica *Long Short-term Memory*, ou LSTM, porém não aplica outros algoritmos tais quais KNN, MLP e outros, abordados no presente texto.

2.6 Utilização de algoritmos de aprendizado de máquina supervisionado

A escolha de aprendizado supervisionado de máquina, em específico os algoritmos aplicados no estudo, deve-se à aplicabilidade dessa técnica em abordagens classificatórias.

Conforme (REZENDE, 2014), utilizar rótulos facilita a visualização de dados, já que ajuda aquele que recebe a informação a classificá-la. Classes tensionam o usuário à entender conceitos como "risco alto", "risco baixo" (Subseção 2.4.1), "elevado pagamento de proventos" ou "pagamento de proventos moderado" (Subseção 2.4.2).

As técnicas aplicadas e indicadas no texto são:

- KNN: k vizinhos mais próximos;
- *Random Forest*: floresta aleatória, em que múltiplas árvores de decisão são construídas, com aplicação da técnica de inteligência de enxame;
- *Decision Tree*: árvore de decisões, em que apenas um árvore de decisões por modelo, ao invés de múltiplas, como o caso da floresta aleatória;
- *Bagging*: agregação de *bootstrap*, ou método de ensacamento, no qual técnicas de classificação e regressão estatística são aplicadas. Reduz variações e ajustes excessivos;
- *Boosting*: algoritmo de impulsionamento, também com aplicação da técnica de inteligência de enxame, com redução de variações e polarizações;
- *Voting*: algoritmo de votação, em que uma média dos resultados por mais de um método é realizada, para que um resultado mais estável seja alcançado;
- MLP: perceptron (rede neural) multi camadas;

3 DESENVOLVIMENTO

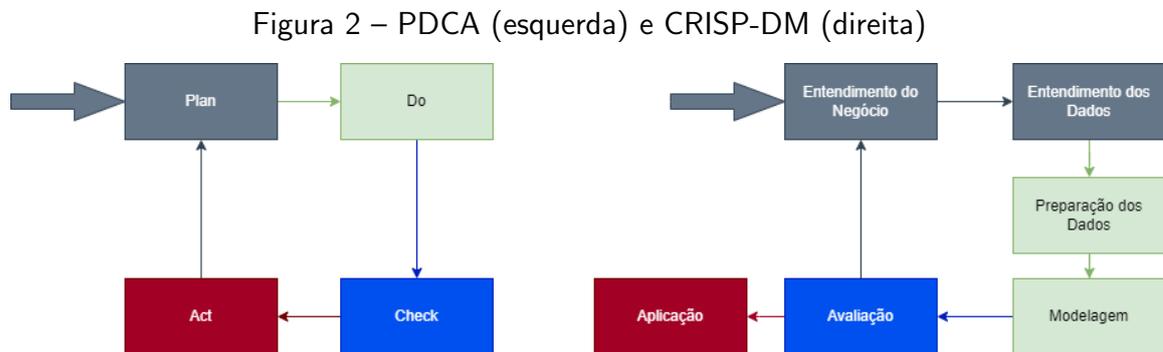
A presente seção visa explicar a maneira como este estudo foi conduzido.

3.1 Metodologia

A metodologia aplicada neste estudo baseia-se em CRISP-DM (*Cross Industry Standard Process for Data Mining*, ou Processo Padrão Inter-Indústria para Mineração de Dados (HOTZ, 2021)) para estruturar o estudo e apresentar os seus resultados. Focada em ciência de dados, com forte apelo à padronização entre indústrias e retroalimentação de dados/ensaios, o uso da CRISP-DM ajudou consideravelmente no desenvolvimento deste projeto, principalmente pelo quesito de realimentações de algoritmos e repetições desses, em diferentes cenários.

Os algoritmos utilizados pelo estudo encontram-se em sítio eletrônico público ¹.

O entendimento do negócio e dos dados estudados, coleta, limpeza e estruturação dos dados, modelagem da solução, avaliação e aplicação da solução são detalhados na Seção 3.3. De acordo com (KARAIVANOV, 2021) pode-se ver uma semelhança entre CRISP/DM e *Plan, Do, Check, Act* (PDCA, ou Planejar, Atuar, Verificar, Agir), o que tornou sua escolha natural para o autor, ambientado com a segunda devido a seu forte uso em gerenciamento de projetos. Esta comparação pode ser observada na Figura 2.



Fonte: Adaptado de (KARAIVANOV, 2021)

3.2 Materiais

O banco de dados relacional utilizado para o presente estudo é o PostgreSQL, versão 14. A linguagem de programação utilizada foi Python 3, subversão 9, juntamente com o ambiente integrado Microsoft VSCode, versão 1.63.0.

A escolha de um banco de dados relacional deve-se aos seguintes fatores:

- o volume de dados do presente trabalho é pequeno, o que não justifica o uso de um banco de dados para grandes volumes;
- afinidade do autor com bancos de dados relacionais, em comparação a NoSQL.

As bibliotecas utilizadas, conforme a categoria, foram:

¹ <https://github.com/GregorioRozario/cccdados>

- Visualização em cores, no terminal
colorama, versão 0.4.4
termcolor, versão 1.1.0
- Visualização de gráficos:
matplotlib, versão 3.5.0
- Manipulações e operações de *dataframes*:
numpy, versão 1.21.4
pandas, versão 1.3.4
- Conexão ao banco de dados:
psycpg, versão 2.9.2
- Funcionalidades de aprendizado de máquina:
scikit_learn, versão 1.0.1
- Visualização de mapa de calor para correlação de variáveis de um *dataframe*:
seaborn, versão 0.11.2
- Aquisição de dados através de protocolo HTTP/HTTPS para *web scrapping*:
beautifulsoup4, versão 4.10.0
requests, versão 2.26.0
Unidecode, versão 1.3.2

Para tornar o estudo mais produtivo, duas outras bibliotecas foram criadas pelo autor. Elas foram desenvolvidas para facilitar atividades gerais e atividades inerentes a FIs, respectivamente **rozlib** e **fiislib**. Ambas estão na versão 0.1.

O conteúdo dessas bibliotecas, a saber:

- rozlib: manipulação de dados em banco de dados PostgreSQL, através de métodos de conveniência. Manipulação de algoritmos de *machine learning*, também utilizando métodos de conveniência;
- fiislib: manipulação de dados oriundos da internet, através de *web scrapping*, bem como ordenação e estruturação de dados para armazenamento em banco de dados;

Em se tratando de *hardware* utilizou-se um *notebook* com processador Intel Core i5, 12GB de memória RAM e SSD com 500GB de espaço, para as escritas e testes de códigos fonte.

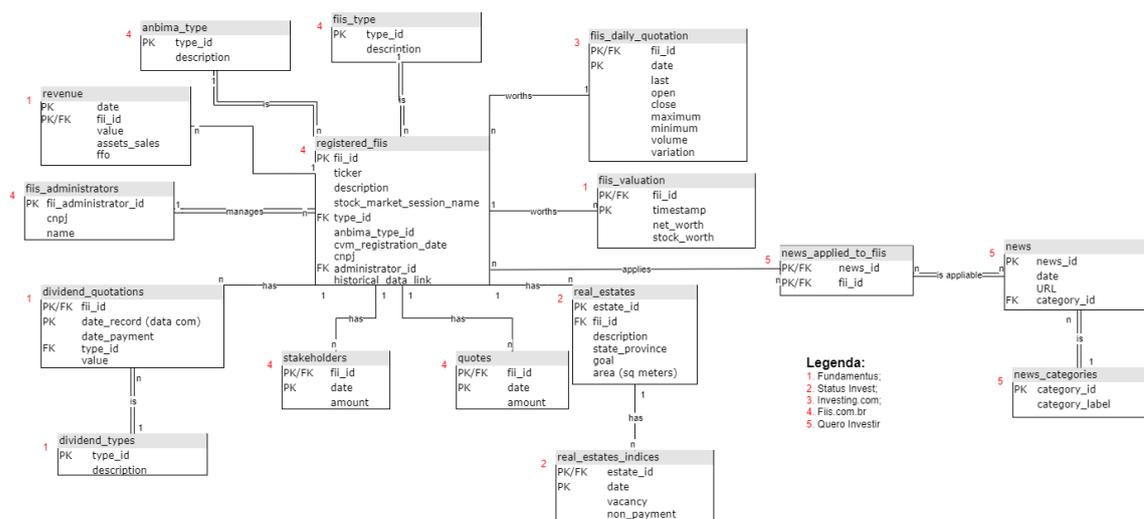
Para armazenamento do banco de dados e execução dos códigos fonte testados utilizou-se um Raspberry Pi Model 3, com 1GB de RAM, e SD *Card* de 32GB de espaço.

3.3 Métodos

O desenho e população do banco de dados foram as tarefas mais onerosas do trabalho. Pré-processamento, identificação de padrões, classificação e agrupamento foram tarefas relativamente breves se comparadas à montagem do banco. O Modelo Entidade-Relacionamento (MER) encontra-se na [Figura 3](#).

Nota-se no MER da [Figura 3](#) a presença da tabela principal, intitulada “registered_fiis” e as demais, orbitando ao redor. É possível notar também uma semelhança a um *data warehouse*.

Figura 3 – Modelo Entidade Relacionamento da Base de Dados



Fonte: Autoria própria.

A origem dos dados de cada tabela está indicada no canto superior esquerdo, através de um número, em vermelho.

Os dados inseridos nas tabelas do banco de dados são oriundos de cinco origens (fontes externas):

- <https://fiis.com.br/>: lista de FIIs, dados gerais sobre cada fundo, como nome registrado, CNPJ, tipo, entre outros dados;
- <https://www.fundamentus.com.br/>: dados de pagamentos de proventos e receita (lucro líquido);
- <https://br.investing.com/>: dados diários de cotações dos índices dos FIIs e notícias;
- <https://statusinvest.com.br/>: dados de imóveis pertencentes aos FIIs e índices de vacância e inadimplência desses imóveis;
- <https://www.euqueroinvestir.com/>: notícias.

A escolha de páginas da internet para recuperação de dados permitiu a recuperação de informações dinâmicas, constantemente atualizadas, concomitantes com o uso da técnica *web scrapping*. Infelizmente, apenas uma fonte não teria todas as informações desejadas, por isso a necessidade de se buscar dados em mais de um sítio eletrônico.

A Figura 4 é um guia para detalhar as etapas de criação do projeto, bem como o fluxo de execuções de *scripts* e algoritmos.

A primeira parte do *pipeline* já foi detalhada no Capítulo 2, em que bolsa de valores, FIIs e estratégias de investimento são embasados.

3.4 Recuperação dos dados

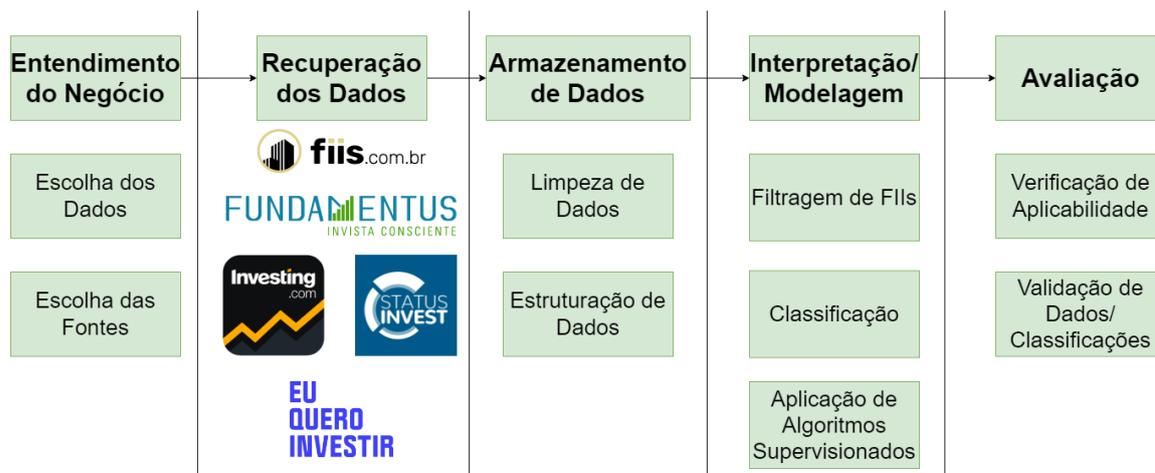
Para melhor explicar como os dados são inseridos no banco de dados, três subseções foram criadas, a seguir.

3.4.1 Atualizações diárias

Diariamente, e apenas em dias úteis (quando ocorrem os pregões), *scripts* executam tarefas de:

- Registro de novos FIIs (caso algum tenha sido recentemente registrado na CVM);

Figura 4 – Fluxo de Execuções (Elaborado pelo próprio autor, 2022)



Fonte: Autoria própria.

- Inserção de cotações de FIIs (somente aqueles que participaram do pregão);
- Informações de pagamentos de dividendos e JCP;
- Informações de quantidade de cotas;
- Informações de quantidade de cotistas (*stakeholders*).

A escolha pelo ciclo diário para essas atualizações se dá pelo fato de que tais dados apenas tem variações significativas uma vez ao dia. Atualizações com ciclos menores enquadram-se melhor para investimentos similares a *day trade*, que necessitam de múltiplas intervenções rápidas por parte do investidor. Fundos de investimento imobiliários são ativos mais inertes e demandam intervenções em menor quantidade.

3.4.2 Atualizações a cada duas horas

A cada duas horas ocorre a execução de *scripts* que:

- Registra novos FIIs (caso algum tenha sido recentemente registrado na CVM);
- Atualiza informações de imóveis pertencentes ao FII, além de informações de vacância e inadimplência desses imóveis;
- Atualiza informações de valor de mercado e valor patrimonial.

Para esses dados percebeu-se que um ciclo mais rápido de atualizações era necessário. Vale lembrar que esses dados não dependem de dias úteis para serem modificados, portanto essas atualizações são feitas inclusive em feriados e finais de semana.

3.4.3 Algoritmo de atualização dos dados

Independentemente do período de atualização dos dados referentes aos FIIs (diário ou a cada duas horas), os algoritmos são semelhantes em sua essência. Observa-se a estrutura em [Algoritmo 1](#).

O algoritmo de inserção de novos FIIs, [Algoritmo 2](#), difere um pouco desse padrão. Caso algum novo FII seja cadastrado na CVM, esse aparecerá na lista contida em (FIIS.COM.BR, 2021). A comparação da lista contida no banco de dados com a lista contida na fonte recém citada indicará divergência, gatilho necessário para a inserção do novo FII no banco de dados.

Algoritmo 1: Algoritmo Padrão de Atualização do Banco de Dados

```

recuperar dados de fiis;
while dados de fiis do
  recuperar dados da internet
  if dados atuais  $\neq$  dados no banco then
    | inserir novos dados no banco
  end
end

```

Algoritmo 2: Algoritmo Para Inserção de Novo FII

```

recuperar lista de fiis do banco;
recuperar lista de fiis da internet;
if lista de fiis do banco  $\neq$  lista de fiis da internet then
  | inserir novo(s) fii(s) no banco
end

```

3.5 Armazenamento de Dados

Essa etapa evita dados incompletos, redundantes, imprecisos ou desnecessários, bem como garante que as informações sejam armazenadas de maneira simétrica, uma característica de bancos de dados relacionais.

A etapa de Limpeza de Dados garante a não inserção de dados redundantes ou imprecisos no banco de dados, tais como datas em formatos incoerentes, ou valores monetários incompatíveis com a moeda vigente, na qual os FIIs são cotados, no caso o Real Brasileiro. De maneira mais detalhada: cotações de FIIs, dividendos, número de acionistas e demais dados são procurados na internet, nas fontes listadas no [Seção 3.3](#). No momento em que esses valores serão confrontados com o banco de dados, verifica-se se esses já se encontram armazenados. Em caso negativo, a estruturação encarrega-se do armazenamento dos dados.

A etapa de Estruturação de dados insere cada informação referente à cotações, dividendos, inadimplência e afins em seus devidos locais, em suas devidas tabelas. É nessa etapa que toma-se o cuidado de inserir datas no devido formato *date*, dados numéricos respeitando essa convenção e *strings* também seguindo suas restrições.

3.6 Interpretação/Modelagem

O *pipeline* utilizado em ciência de dados versa sobre a seleção daqueles dados que têm real valor no treinamento, teste e validação dos modelos de aprendizado de máquina. O presente estudo utilizou como ponto de partida dados de cotação diários dos FIIs, para montagem de vetores de características. Diferentes combinações de características formaram diferentes vetores, para fins de testes e melhores resultados nos modelos finais.

3.6.1 Vetores de características

Entendem-se por vetores de características informações distintas, não repetitivas, que identificam um elemento de estudo, nesse caso um FII. Em se tratando de valores diários de cotação, as informações disponíveis para formação dos vetores de características são:

- Data (trabalhada para gerar uma referência ao início do ano, explicada a seguir);
- Cotação de abertura do pregão;

- Cotação de fechamento do pregão;
- Cotação máxima atingida pelo FII, no pregão;
- Cotação mínima atingida pelo FII, no pregão;
- Volume de negociações do FII, no pregão;
- Variação da cotação do FII, referenciada ao pregão anterior.

3.6.2 Caracterização de datas

Visto que a informação de data é dado importante do vetor de características, utilizado no treinamento/teste/validação dos modelos de *machine learning*, e levando em conta que épocas do ano podem ser consideradas semelhantes para anos distintos, foi necessário substituir as datas com a informação do ano. Por exemplo 21/03/2021, pode ser interpretada sem a informação do ano, 21/03, visto que a importância da data reside em sua referência desde o começo de seu respectivo ano, não no ano em si. Essa abordagem é conhecida como *Year To Date*, ou YTD.

É possível que estudos futuros cheguem à conclusão que, por exemplo, anos pares, eleitorais no Brasil, tenham comportamentos diferenciados de anos ímpares, principalmente para datas próximas ao período eleitoral.

3.6.3 Intervalo de tempo dos conjuntos de treinamento/teste e validação

Diversos testes foram realizados no banco de dados para se obter resultados melhores, e mais fiéis à realidade. É necessário levar em conta o que versa a Análise Fundamentalista: ganhos na bolsa de valores ocorrem a longo prazo ([MONEY, 2021](#)).

Como exemplo é possível utilizar o FII mais popular da Bovespa: o MXRF11, que em agosto de 2021 contava com 416044 cotistas (Fonte: banco de dados da pesquisa). Observa-se uma queda acentuada em seu beta conforme o intervalo de tempo para esse cálculo aumenta, ou seja, o prazo aumenta. O que isso significa? Que o risco diminui drasticamente conforme o prazo do investimento aumenta.

O beta para o MXRF11, calculado entre os dias 01/01/2018 e 01/01/2020 é de 0.00458. O que isso quer dizer? Que a cada R\$ 1,00 investido no MXRF11, há chance de se ganhar ou perder R\$ 0,00458, dentro do período de dois anos citado. Por si só esse risco já é muitíssimo baixo, mas como dito anteriormente: os riscos dentro do conjunto de FIIs são muito baixos, se comparados a riscos de outros ativos, como cripto ativos.

Ainda, o MXRF11 é um FII com data de lançamento antiga (20/09/2011), o que o torna bem aceito na comunidade de investimentos e o faz alcançar grandes volumes de negociações diárias, mantendo sua cotação estável. Agora, caso o beta seja calculado, para o MXRF11, levando em conta as datas 01/01/2015 e 01/01/2020, ou seja, cinco anos, o valor do beta cai para 0.00364. Isso embasa a teoria fundamentalista: ganhos em longo prazo minimizam perdas ([REIS, 2021](#)).

Em se tratando de escolha de intervalos para os conjuntos de treinamento/teste e validação, diversas combinações foram testadas, também com o intuito de maximizar os resultados, mantendo-os fiéis à realidade.

Um grande cuidado que se deve ter ao escolher os conjuntos de treinamento/teste e de validação é que o segundo deve ser maior ou igual ao primeiro, isso é, os vetores de características do conjunto de validação devem ter dimensões no mínimo iguais aos vetores do conjunto de treinamento/teste, ou seja, a quantidade de elementos do conjunto de validação deve ter no mínimo a mesma quantidade de elementos do conjunto de treinamento/testes.

Também é mandatório que as informações presentes nos vetores de características de ambos os conjuntos sejam idênticas, conforme as possibilidades contidas em [Subseção 3.6.1](#).

Isso se faz necessário porque os modelos treinados esperam conjuntos de tamanhos iguais para que novas previsões sejam realizadas. Os dados excedentes do conjunto de validação são descartados, iniciando o descarte a partir dos elementos mais antigos desse conjunto, para garantir dimensões iguais entre ambos.

3.6.4 Seleção de Fundos de Investimentos Imobiliários para modelagem de aprendizado de máquina

O banco de dados conta com mais de 300 FII's cadastrados. Isso não significa que todos os fundos ali cadastrados puderam ser utilizados para treinamento, teste e validação de modelos de *machine learning*. O presente estudo utilizou os seguintes critérios para seleção de FII's, para modelagens:

- FII's com informações de dividendos ao menos nos últimos 60 dias;
- FII's com informações de quantidade de investidores;
- FII's com informações de cotações diárias;
- FII's com informações de valores patrimoniais e de mercado.

As explicações para o uso de cada um dos critérios:

- FII's sem informações de dividendos nos últimos 60 dias estão inativos, já que a lei os obriga a pagar dividendos mensalmente em caso de atividade normal;
- FII's sem informações de investidores podem estar inativos ou mesmo não possuírem investidores naquele instante. Um FII que não pode ser comercializado é um ativo inalcançável;
- Sem informações de cotações diárias não é possível montar um vetor de características com volume mínimo de dados;
- FII's sem informações de valores patrimoniais e/ou de mercado são ativos com informações defasadas ou inexistentes. Isso causa, em investidores experientes, aversão à compra. O mercado de ações presa por transparência de informações.

3.7 Definição de intervalos de beta para classificação

Como dito anteriormente, espera-se que FII's tenham risco reduzido, se comparados a outros investimentos. Levando isso em conta e sabendo que algoritmos de classificação de conjuntos esperam divisões que garantam certas quantias de elementos para cada classe, nos conjuntos de treinamento e teste, ensaios foram realizados para se chegar nos intervalos de classificação de beta, conforme [Tabela 3](#).

Tabela 3 – Intervalo de Índice Beta para Classificação de FII's

Classe	Intervalo
Beta elevado	$\text{beta} \geq 0.2$
Beta intermediário	$0.1 \leq \text{beta} < 0.2$
Beta baixo	$\text{beta} < 0.1$

Fonte: Banco de Dados desse texto (2022)

Esses valores foram estipulados levando em conta a distribuição normal dos índices beta dos FII's utilizados para criação dos vetores de características. Para fins comparativos

criou-se outro conjunto de intervalos de beta, através de percentis, para se fazer classificações por esse indicador. Os intervalos utilizados foram subdivididos em 3 intervalos distintos, com quantidades idênticas de amostras para cada intervalo.

3.8 Definição de intervalos de DY para classificação

Como conceituado anteriormente, o cálculo do *Dividend Yield* (DY) é feito levando em conta os 12 meses anteriores à data de referência. Sendo assim, o DY tem intervalo fixo, devido a seu conceito, porém a data de referência pode ser utilizada como variável. O presente estudo utiliza a mesma data de encerramento do cálculo do beta como data de referência para cálculo do DY. Dito isso, o passo seguinte foi definir os intervalos de classificação de DY, tais quais os intervalos de beta.

Além de levar em conta os retornos desejáveis, aceitos pela comunidade investidora, também atentou-se à distribuição mínima de elementos por classe, para que treinamento e teste pudessem ocorrer. Treinamentos de algoritmos e testes no banco de dados levaram em conta os intervalos contidos na [Tabela 4](#).

Tabela 4 – Intervalo de DY para Classificação de FIIs

Classe	Intervalo
DY elevado	$DY \geq 1.0\%$
DY intermediário	$0.6\% \leq DY < 1.0\%$
DY baixo	$DY < 0.6\%$

Fonte: Banco de Dados desse texto (2022)

Assim como ocorreu com o indicador beta, criou-se outro conjunto de intervalos de DY, através de percentis, para se fazer classificações por esse indicador. Foram utilizados 3 intervalos, com quantidades idênticas de amostras para cada.

Seria possível utilizar o valor anual de DY, para os fins classificatórios indicados, porém por conta de simplificação o autor preferiu normalizar esses valores para os valores mensais.

3.9 Treinamento e teste dos modelos

Como informado no início dessa seção, utilizou-se a biblioteca scikit-learn, versão 1.0.1 para uso de algoritmos de aprendizado de máquina. Os algoritmos utilizados para modelagem de aprendizado de máquina foram:

- Boosting;
- Bagging;
- KNN (k vizinhos mais próximos);
- MLP (perceptron multi camadas);
- Decision Tree (árvore de decisões);
- Random Forest;

- Voting.

A motivação para utilização de múltiplos métodos se deve à necessidade de fazer *benchmarking* entre esses. A seção de resultados traz esse comparativo. Seguindo o fluxo do *pipeline* de aprendizado de máquina, utilizou-se a própria biblioteca scikit learn para dividir o conjunto de dados entre treinamento e testes, treiná-lo e testá-lo.

3.10 Validação dos modelos

Ferramentas de aprendizado supervisionado, tais quais a biblioteca scikit learn, são muito eficientes para treinar e testar os modelos de *machine learning*, utilizando um conjunto inicial de dados para tal efeito. Porém, há um próximo passo necessário, inserido no *pipeline* de processamento de dados, que é a validação dos modelos treinados.

Essa validação deve ocorrer com um novo conjunto de dados, pertencente ao mesmo banco de dados, porém preferencialmente com elementos disjuntos ao primeiro conjunto, utilizado no treinamento/teste dos modelos de aprendizado de máquina. A necessidade de ter um conjunto preferencialmente disjunto se deve ao fato de a validação ser um processo muito semelhante ao teste, do *pipeline*: caso haja dados compartilhados entre os conjuntos de testes e de validação, os resultados serão sobrepostos, levando a falsos resultados válidos.

A validação, como será expresso no capítulo [Capítulo 4](#), foi feita de duas maneiras: levando em conta apenas o algoritmo com melhor resultado na fase de testes e todos os algoritmos testados.

4 RESULTADOS

O banco de dados gerado pelo estudo, como objetivo secundário, abre possibilidades para trabalhos futuros. Em se tratando dos objetivos primários, aplicação de aprendizado supervisionado de máquina para classificação de FIIs, os resultados são detalhados nesse capítulo.

Vale ressaltar que o autor aplicou dois conjuntos de classes para os resultados, tanto de beta quanto de DY. Um conjunto utiliza estratégia estatística, com limiares por percentis. O outro conjunto utiliza limiares definidos pelo autor, após diversos ensaios realizados, optando pelos melhores resultados.

4.1 Resultados de classificação por beta

Os resultados de classificação de FIIs segundo o grau de risco de investimento, beta, foram acima do esperado para os conjuntos de testes indicados na [Seção 3.7 \(Conjunto 1\)](#), porém abaixo do esperado para os conjuntos divididos por percentis ([Conjunto 2](#)).

Diversos intervalos de datas foram testados no algoritmo de classificação, para se conseguir melhores resultados. Utilizando-se datas entre 31/10/2020 e 31/10/2021 para construção do conjunto de treinamento/teste e datas entre 26/04/2021 e 26/05/2022 para o conjunto de validação, chegou-se aos resultados contidos na [Tabela 5](#).

Tabela 5 – Acurácia de algoritmos de classificação por beta

Método	Acurácia (Conjunto 1)	Acurácia (Conjunto 2)
Decision Tree	85,00%	57,70%
Voting	85,00%	60,00%
Random Forest	85,00%	62,50%
KNN	82,50%	62,50%
Bagging	82,50%	60,00%
Boosting	75,00%	52,50%
MLP	70,00%	35,00%

Fonte: Testes de precisão de algoritmos de *machine learning*, para previsão de beta (2022)

Validações feitas apenas com o método mais eficaz, nesse caso, árvore de decisões, sobre o **Conjunto 1**, levaram à acurácia de 80,81%, e sobre o **Conjunto 2** 47,47%. Validações feitas com todos os métodos testados, sobre o **Conjunto 1**, levaram à acurácia final de 73,45%, e de 51,37% para o **Conjunto 2**.

Para um investidor procurando resultados como "comprar" ou "vender", os dados resultantes da abordagem detalhada na [Seção 3.2](#) indicam que aproximadamente 3/4 das previsões realizadas sobre o **Conjunto 1** indicaram previsões que se concretizaram. Sendo assim, o investidor que busca uma faixa de indicador de risco (beta) terá aproximadamente 75% de chances de adquirir ou vender um fundo de investimento imobiliário com nível de risco conforme desejado.

4.2 Resultados de classificação por DY

Os resultados de classificação de FII's segundo o retorno mensal, DY, foram abaixo do esperado, tanto para o conjunto de intervalos descritos na [Seção 3.8](#), quanto pelo conjunto de intervalos com percentis. Utilizando-se as mesmas datas para treinamento/teste e validação da seção anterior, obtiveram-se os resultados contidos na [Tabela 6](#)

Tabela 6 – Resultados de algoritmos de classificação por DY

Método	Acurácia (Conjunto 1)	Acurácia (Conjunto 2)
Bagging	65,00%	42,50%
Decision Tree	60,00%	50,00%
Voting	52,50%	45,00%
Boosting	52,50%	42,50%
Random Forest	52,50%	32,50%
MLP	50,00%	35,00%
KNN	40,00%	32,50%

Fonte: Resultados obtidos pelo estudo (2022)

Validações feitas apenas com o método mais eficaz, nesse caso, *random tree*, levaram à acurácia de 45,45%, para o **Conjunto 1** e de 33,33% para o **Conjunto 2**. Validações feitas com todos os métodos testados, sobre o **Conjunto 1**, levaram à acurácia final de 51,08% e de 34,34% para o **Conjunto 2**.

Como indicado, os resultados para a classificação por DY ainda ficaram muito aquém do esperado. Estima-se que utilizar outras características no vetor de cada FII possa acarretar numa melhora dos resultados, ou seja, modelos mais próximos da realidade.

Para o DY, um investidor esperando informações como "comprar" ou "vender", a assertividade seria de aproximadamente 50%, sobre o **Conjunto 1**. Sendo assim, as previsões de retorno de proventos para FII's teriam um grau de precisão de 1 para 2. Cabe ao investidor decidir quais são os intervalos de DY que se enquadram melhor com sua estratégia de investimentos.

Em palavras, os resultados obtidos indicam que, tomando como exemplo a classificação de investimentos por beta, para o **Conjunto 1**, para cada 10 investimentos em diferentes FII's, há probabilidade de se investir em 8,5 FII's que estarão no intervalo de beta previsto pelos modelos de *machine learning* aplicados pelo texto. A mesma ideia é válida para a classificação por DY.

5 CONCLUSÃO

A presente seção vem ratificar o que foi exposto no corpo do trabalho.

5.1 Considerações Finais

Parte dos resultados finais do estudo foram alcançados a contento, com elevados níveis de acurácia em testes e validações. A utilização de um conjunto de intervalos para o índice beta, conforme critério apontado pelo autor, gerou os melhores resultados, com aproximadamente 75% de precisão. A utilização de percentis, sobretudo para fins comparativos, indicou que o segundo critério não foi tão efetivo, com pouco mais de 50% de precisão.

Em se tratando de DY, os resultados alcançados, tanto com intervalos definidos pelo autor, quanto com percentis, mostram que ainda há um caminho a ser trilhado, procurando uma melhor técnica de classificação, talvez alterando intervalo de datas ou inserindo/removendo dados do vetor de características, ou até mesmo utilizando outra técnica de aprendizado de máquina, tal qual *data mining* ou *auto encoders*.

5.2 Limitações

Uma parcela pequena do conjunto de dados foi utilizada nos algoritmos apresentados no texto. Mesmo limitando-se o volume de dados, em se tratando de tabelas e colunas dessas, a manipulação de dados foi extensa.

Além de se manipular dados em volume elevado, os processamentos foram repetidos exaustivamente, para se chegar à conclusão de quais combinações de **vetores de características** gerariam dados com alto grau de precisão, mantendo a fidelidade desses.

Além dessa variável, os **intervalos de datas** também possibilitaram novas combinações, que obrigou o autor a realizar mais testes e novas validações, tendo como base a variável **vetor de características**.

Somando esses fatores, não seria possível expor nesse texto todas as combinações realizadas, de **intervalos de datas** e **vetores de características**, ou ainda utilizar mais informações, contidos no banco de dados, para incrementar esses vetores. Listar as possíveis combinações de dados, contidos no banco, também seria parte introdutória de trabalhos futuros.

5.3 Trabalhos Futuros

O presente texto deixa um caminho a trilhar em diversos aspectos. Alguns deles são apontados:

- Utilização de *data warehouse* para recuperação de dados de maneira padronizada, conforme essa abordagem, para aumentar a velocidade de recuperação de dados;
- Utilização de outros rótulos/grandezas, tais quais, receita crescente/decrecente, inadimplência/vacância baixa/alta, para classificação de FIIIs, além de beta e DY;
- Utilização de linguagem natural para análise de notícias vinculadas a FIIIs, com posterior análise de cotações/rendimentos de ativos;
- Utilização de lógica nebulosa para classificação contínua de FIIIs, não apenas discreta, tais quais uma escala de pagamento de dividendos, de 0% a 100%, em que fundos classificados próximos a 0% seriam FIIIs menos aderentes ao DY pretendido pelo investidor, ao passo

que fundos classificados próximos a 100% teriam a característica mais próxima ao que o investidor busca;

Referências

- B3. **Índice de Fundos de Investimentos Imobiliários (IFIX B3)**. 2021. <https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-de-segmentos-e-setoriais/indice-fundos-de-investimentos-imobiliarios-ifix-composicao-da-carteira.htm>. Accessed: 2021-12-08. Citado na página 16.
- B3. **Ações**. 2022. <https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/empresas-listadas.htm>. Accessed: 2022-08-11. Citado na página 14.
- BARONI, M.; BASTOS, D. **Guia Suno de Fundos Imobiliários**. Paulínia: Editora Vivalendo, 2018. Citado 4 vezes nas páginas 11, 14, 16 e 17.
- BARTOLOME, A. **how to resolve ConnectionError: ERR#0015: error 429?** 2022. <<https://github.com/alvarobartt/investpy/issues/530>>. Accessed: 2022-06-21. Citado na página 17.
- BIANCA, A. **Conheça a origem da Bolsa de Valores**. 2019. <<https://gorila.com.br/blog/a-origem-da-bolsa-de-valores/#:~:text=Primeira%20Bolsa-,Criada%20em%201531%2C%20a%20Bolsa%20da%20Antu%C3%A9ria%2C%20na%20B%C3%A9lgica%2C,1602%2C%20na%20Bolsa%20de%20Amsterd%C3%A3.>> Accessed: 2022-06-02. Citado na página 13.
- BOMFIM, R. **Luiz Barsi, o megainvestidor que fez fortuna com dividendos**. 2022. <<https://valor.globo.com/financas/noticia/2022/04/25/luiz-barsi-o-megainvestidor-que-fez-fortuna-com-dividendos.ghtml>>. Accessed: 2022-08-11. Citado na página 14.
- CANTO, A. B. del. **Welcome to investpy's documentation!** 2021. <<https://investpy.readthedocs.io/>>. Accessed: 2022-06-20. Citado na página 17.
- FIIS.COM.BR. **Lista de FIIs**. 2021. <<https://fiis.com.br/lista-de-fundos-imobiliarios/>>. Accessed: 2021-07-30. Citado 3 vezes nas páginas 15, 16 e 22.
- HOTZ, N. **What is CRISP DM?** 2021. <<https://www.datascience-pm.com/crisp-dm-2>>. Accessed: 2022-01-08. Citado na página 19.
- INVESTING. **Detalhes MXRF11**. 2021. <<https://br.investing.com/equities/fii-maxi-ren>>. Accessed: 2021-06-20. Citado na página 17.
- KARAIVANOV, D. **O que é o Ciclo Plan-Do-Check-Act?** 2021. <<https://kanbanize.com/pt/gestao-lean/melhoria/o-que-e-o-ciclo-pdca>>. Accessed: 2022-01-08. Citado na página 19.
- LI, A. K. Y. **Predicting Stock Prices Using Machine Learning**. 2022. <<https://neptune.ai/blog/predicting-stock-prices-using-machine-learning>>. Accessed: 2022-09-12. Citado na página 17.
- MEIRELLES, R. M. P. **Análise Técnica e a Fundamentalista**. 20 2. Monografia (Especialização) — Universidade Federal de Juiz de Fora, Minas Gerais, 2012. Citado na página 14.

MONEY, I. **Análise Fundamentalista de Ações: como identificar empresas sólidas e rentáveis a longo prazo**. 2021. <<https://www.infomoney.com.br/guias/analise-fundamentalista/>>. Accessed: 2022-02-15. Citado 3 vezes nas páginas 13, 14 e 24.

PÓVOA, A. **Valuation: como precificar ações**. Rio de Janeiro: Editora Campus, 2012. Citado na página 16.

REIS, T. **Análise fundamentalista: o que é? Guia completo de análise de ações**. 2021. <<https://www.suno.com.br/guias/analise-fundamentalista/>>. Accessed: 2022-08-22. Citado 2 vezes nas páginas 17 e 24.

REIS, T.; TOSETTO, J. **Guia Suno Dividendos**. Paulínia: Editora Vivalendo, 2017. Citado 2 vezes nas páginas 13 e 17.

REZENDE, D. D. de. **Um Modelo de Classificação Supervisionada com Rotulagem Automática para Reconhecimento de Áudio de Web Rádios**. 84 2. Monografia (Mestrado) — Universidade Federal de Ouro Preto, Minas Gerais, 2014. Citado na página 18.

ROCHA, D. **Número de brasileiros investindo em FIs cresce 660% em 3 anos**. 2022. <<https://investidor.estadao.com.br/investimentos/fundos-imobiliarios-investidores-brasil-alta>>. Accessed: 2022-05-27. Citado na página 15.

SANTANA, G. F. **O Poder de Previsão da Análise Técnica**. 60 1. Monografia (Mestrado) — Fundação Getúlio Vargas, São Paulo, 1997. Citado na página 13.

UOL, C. para o. **Imposto de Renda 2022: Saiba como declarar fundos imobiliários**. 2022. <<https://economia.uol.com.br/imposto-de-renda/noticias/redacao/2022/05/07/imposto-de-renda-2022-saiba-como-declarar-fundos-imobiliarios.htm#:~:text=Ganho%20com%20venda%20de%20cotas,Bolsa%20paga%2020%25%20de%20IR&text=O%20imposto%20%C3%A9%20de%2020,da%20Bolsa%20e%20de%20corretagem.>>. Accessed: 2022-06-02. Citado na página 16.

WIKIPÉDIA. **Bolsa de valores**. 2022. <https://pt.wikipedia.org/wiki/Bolsa_de_valores>. Accessed: 2022-06-02. Citado na página 13.