

INSTITUTO FEDERAL DO ESPÍRITO SANTO
SISTEMAS DE INFORMAÇÃO

WELLINGTON BARBOSA JUNIOR

**APLICAÇÃO DE *MACHINE LEARNING* PARA AUXILIAR O PROCESSO
DECISÓRIO DE INVESTIMENTO EM FUNDOS IMOBILIÁRIOS**

Cachoeiro de Itapemirim

2022

WELLINGTON BARBOSA JUNIOR

**APLICAÇÃO DE *MACHINE LEARNING* PARA AUXILIAR O PROCESSO
DECISÓRIO DE INVESTIMENTO EM FUNDOS IMOBILIÁRIOS**

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Sistemas de Informação do Instituto Federal do Espírito Santo, Campus Cachoeiro de Itapemirim, como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Ricardo Maroquio Bernardo

Cachoeiro de Itapemirim

2022

(Biblioteca do Campus Cachoeiro de Itapemirim)

B238a Barbosa Júnior, Wellington .

Aplicação de machine learning para auxiliar o processo decisório de investimentos em fundos imobiliários / Wellington Barbosa Júnior. - 2022. 60 f. : il. ; 30 cm.

Orientador: Ricardo Maroquio Bernardo

TCC (Graduação) Instituto Federal do Espírito Santo, Campus Cachoeiro de Itapemirim, Sistemas de Informação, 2022.

1. Inteligência artificial. 2. Linguagem de programação (computadores). 3. Investimentos. I. Bernardo, Ricardo Maroquio. II. Título III. Instituto Federal do Espírito Santo.

CDD: 005.31

Bibliotecário/a: Jacqueline Machado Silva CRB-ES nº 640



**MINISTÉRIO DA EDUCAÇÃO
INSTITUTO FEDERAL DO ESPÍRITO SANTO
CAMPUS CACHOEIRO DE ITAPEMIRIM**



FOLHA DE APROVAÇÃO-TCC Nº 2 / 2023 - CAI (11.02.18)

Nº do Protocolo: 23151.000802/2023-96

Cachoeiro De Itapemirim-ES, 02 de março de 2023.

WELLINGTON BARBOSA JUNIOR

**APLICAÇÃO DE MACHINE LEARNING PARA AUXILIAR O PROCESSO DECISÓRIO DE
INVESTIMENTO EM FUNDOS IMOBILIÁRIOS**

Trabalho de Conclusão de Curso apresentado à Coordenadoria de Sistemas de Informação do Instituto Federal do Espírito Santo, como requisito parcial para obtenção de título de Bacharel em Sistemas de Informação.

Aprovado em 12 de dezembro de 2022

COMISSÃO EXAMINADORA

Prof. Dr. Ricardo Maroquio Bernardo
Instituto Federal do Espírito Santo
Orientador

Prof. Dr. Lucas Poubel Timm do Carmo
Instituto Federal do Espírito Santo

Prof. Dr. Rafael Vargas Mesquita dos Santos
Instituto Federal do Espírito Santo

(Assinado digitalmente em 03/03/2023 16:16)
LUCAS POUBEL TIMM DO CARMO
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO
CAI-CCSI (11.02.18.01.08.02.13)
Matrícula: 2417426

(Assinado digitalmente em 02/03/2023 21:35)
RAFAEL VARGAS MESQUITA DOS SANTOS
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO
CAI-CCSI (11.02.18.01.08.02.13)
Matrícula: 1544937

(Assinado digitalmente em 07/03/2023 23:14)
RICARDO MAROQUIO BERNARDO
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO
CAI-CCSI (11.02.18.01.08.02.13)
Matrícula: 2152606

Visualize o documento original em <https://sipac.ifes.edu.br/public/documentos/index.jsp> informando seu número: 2, ano: 2023, tipo: FOLHA DE APROVAÇÃO-TCC, data de emissão: 02/03/2023 e o código de verificação: 3e175e6867

DECLARAÇÃO DO AUTOR

Declaro, para fins de pesquisa acadêmica, didática e técnico-científica, que este Trabalho de Conclusão de Curso pode ser parcialmente utilizado, desde que se faça referência à fonte e ao autor.

Cachoeiro de Itapemirim, 12 de Dezembro de 2022.

Wellington Barbosa Junior

À minha esposa Marlene e filha Sabrina, que sempre com muita alegria me incentivaram e compreenderam os momentos de ausência.

AGRADECIMENTOS

Primeiramente a Deus por minha vida, família e amigos.

À minha querida esposa Marlene cuja presença sempre afetou positivamente a minha vida, em todos os aspectos.

Ao Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo - Cachoeiro de Itapemirim, pelo ambiente excelente que proporciona.

Ao Prof. Dr. Ricardo Maroquio Bernardo pela oportunidade e apoio na elaboração deste trabalho.

Aos membros da banca examinadora, Prof. Dr. Lucas Poubel Timm do Carmo e Prof. Dr. Rafael Vargas Mesquita dos Santos, pelo interesse e disponibilidade em avaliar o trabalho.

Com muita admiração e enorme respeito demonstro toda minha gratidão aos professores do Curso de Sistemas de Informação pela dedicação e competência profissional apresentadas.

A todos que direta ou indiretamente fizeram parte de minha formação, muito obrigado.

RESUMO

A decisão de investimento em ativos financeiros negociados em bolsa de valores é apontada como uma questão cheia de desafios devido à sua natureza volátil. Neste cenário, a aplicação de técnicas de aprendizado de máquina tem ganhado destaque, visto a possibilidade de se correlacionar atributos e grande quantidade de dados para modelar problemas complexos de forma a lidar com a volatilidade inerente desses ativos. O objetivo deste trabalho é verificar a possibilidade de aplicação de aprendizado de máquina a fundos imobiliários consonantes a estudos desenvolvidos para ações. Para o mercado de ações apurou-se uma série de estudos utilizando diversos métodos e abordagens de aprendizado de máquina, contudo, especificadamente para fundos imobiliários verifica-se uma escassez de referências. O modelo obtido conseguiu, predominantemente, superar quatro outras técnicas de negociação ao se avaliar o retorno anualizado em diferentes intervalos de tempo e de classificação de fundos, demonstrando assim a capacidade generalista do modelo.

Palavras-chave: aprendizado de máquina, fundos imobiliários, investimentos

ABSTRACT

The decision to invest in financial assets traded on the stock exchange is seen as an issue fraught with challenges due to its volatile nature. In this scenario, the application of machine learning techniques has gained prominence, given the possibility of correlating attributes and large amounts of data to model complex problems in order to deal with the inherent volatility of these assets. The objective of this work is to verify the possibility of applying machine learning to real estate funds in line with studies developed for stocks. For the stock market, a series of studies were found using different methods and machine learning approaches, however, specifically for real estate funds, there is a scarcity of references. The model obtained predominantly managed to outperform four other trading techniques when evaluating the annualized return at different time intervals and fund classification, thus demonstrating the generalist capacity of the model.

Keywords: machine learning, real estate funds, investments

LISTA DE FIGURAS

Figura 1 – Evolução do número de investidores na bolsa de valores com posição em custódia (mil).	13
Figura 2 – Relação entre inteligência artificial, aprendizado de máquina e aprendizado profundo em um diagrama de Veen.	19
Figura 3 – Hierarquia de aprendizado de máquina.	20
Figura 4 – Ilustração da operação de convolução com um <i>kernel</i> 3 x 3.	23
Figura 5 – Formato de uma amostra dos dados obtidos.	31
Figura 6 – Arranjo descritivo das imagens geradas a partir dos indicadores técnicos.	32
Figura 7 – Amostras das imagens geradas a partir dos indicadores técnicos.	33
Figura 8 – Estrutura do modelo elaborado.	35
Figura 9 – Matriz de confusão para os conjuntos de dados de treinamento, validação e teste.	39
Figura 10 – Amostra em que o modelo proposto obteve o melhor desempenho.	41
Figura 11 – Ocasão na qual a estratégia <i>Vortex</i> foi a melhor.	42
Figura 12 – Circunstância em que a estratégia de médias móveis foi superior as demais.	42
Figura 13 – Situação na qual a estratégia RSI foi vencedora.	42
Figura 14 – Distribuição dos resultado do modelo vs outras estratégias.	43
Figura 15 – Distribuição das operações do modelo vs outras estratégias.	44
Figura 16 – Fluxograma da metodologia de trabalho.	53

LISTA DE TABELAS

Tabela 1 – Configuração de hardware.	36
Tabela 2 – Especificação resumida da GPU.	36
Tabela 3 – Métricas da aplicação do modelo ao <i>dataset</i> de teste durante o treinamento.	40
Tabela 4 – Sumário do maior retorno anualizado de cada uma das estratégias para o conjunto dos fundos imobiliários analisados.	40
Tabela 5 – Sumário do maior retorno anualizado de cada uma das estratégias para o conjunto inicial de fundos imobiliários analisados entre 03/01/2022 e 31/08/2022.	45
Tabela 6 – Sumário do maior retorno anualizado de cada uma das estratégias para o conjunto teste dos fundos imobiliários com números de dias de negociação variados.	46
Tabela 7 – Comparação dos retornos anualizados do modelo e com as estratégias Comprar e Manter (B&H), Vortex, Médias Móveis Exponenciais (EMA) e Índice de Força Relativa (RSI).	55
Tabela 8 – Comparação dos retornos anualizados do modelo e com as estratégias Comprar e Manter (B&H), Vortex, Médias Móveis Exponenciais (EMA) e Índice de Força Relativa (RSI) para fundos imobiliários não utilizados no treinamento do modelo.	60

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Objetivos	14
1.2	Justificativa	14
1.3	Organização do Trabalho	16
1.4	Considerações sobre o capítulo	16
2	REFERENCIAL TEÓRICO	17
2.1	Fundo de Investimento Imobiliário	17
2.2	Inteligência artificial, aprendizado de máquina e aprendizado profundo	18
2.3	Linguagem de programação Python	23
2.3.1	Bibliotecas Python para Aprendizado de Máquina	24
2.4	Indicadores Técnicos	25
2.4.1	Índice de Força Relativa	25
2.4.2	Williams %R	26
2.4.3	Movimento Direcional	26
2.4.4	Média Móvel Convergente e Divergente	26
2.4.5	<i>Commodity Channel Index</i>	27
2.4.6	<i>Momentum</i>	27
2.4.7	<i>Trix</i>	27
2.4.8	Taxa de Mudança	28
2.4.9	Índice de Fluxo de Dinheiro	28
2.4.10	<i>Vortex</i>	29
2.4.11	Ciclo de Tendência de Schaff	29
2.5	Considerações sobre o capítulo	29
3	METODOLOGIA PROPOSTA	30
3.1	Coleta de dados	30
3.2	Geração das imagens	32
3.3	Desenvolvimento do modelo	33
3.4	Ambiente de desenvolvimento	35
3.4.1	Configurações de hardware	36
3.5	Considerações sobre o capítulo	37

4	RESULTADOS	38
4.1	GPU vs. CPU	38
4.2	Métricas de desempenho	38
4.3	Aplicação do modelo ao <i>dataset</i> de fundos imobiliários	40
4.4	Testes de generalização do modelo	44
4.5	Considerações sobre o capítulo	46
5	CONSIDERAÇÕES FINAIS	47
	REFERÊNCIAS	49
	APÊNDICE A – FLUXOGRAMA DA METODOLOGIA DE TRABALHO	53
	APÊNDICE B – ROTULAGEM DOS DADOS COM BASE NO INDICA-	
	DOR ROC	54
	APÊNDICE C – RETORNO ANUALIZADO DAS ESTRATÉGIAS . . .	55

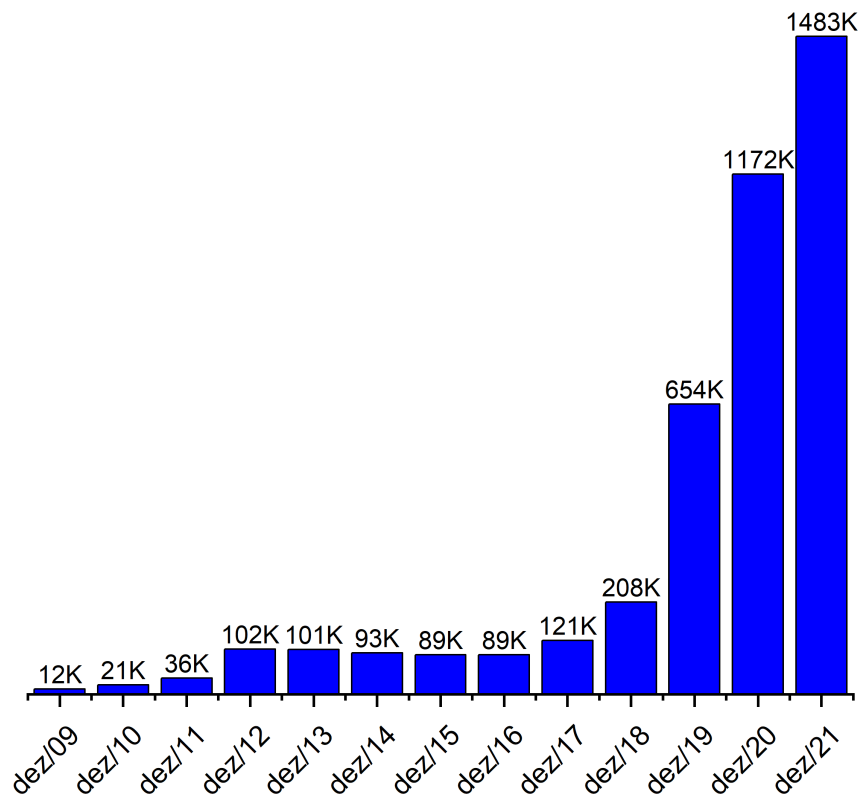
1 INTRODUÇÃO

Fundo de Investimento Imobiliário (FII) é uma comunhão de recursos captados por meio do sistema de distribuição de valores mobiliários e destinados à aplicação em empreendimentos imobiliários. Os fundos de investimento imobiliário são considerados condomínios fechados, ou seja, as suas cotas não podem ser resgatadas a pedido dos cotistas. O resgate, até então, só é permitido na ocasião de liquidação do fundo, deliberada por assembleia geral de cotistas ou ao término do seu prazo de duração, se houver, pois o fundo imobiliário também pode ter prazo de duração indeterminado, como é o caso da grande maioria dos fundos que compõem o mercado.

Assim, uma vez que o investidor só poderá se desfazer do investimento em um fundo imobiliário negociando suas cotas no mercado, o preço de compra/venda estará sujeito a diversas variáveis, não só aquelas referentes ao mercado imobiliário (vacância dos imóveis, rendimento, liquidez, contratos, volatilidade, etc.), mas também a outros tais como taxa de juros, inflação, expectativa e fundamentos da economia, etc. Santos (2020) cita o alto risco do investimento à natureza caótica altamente volátil do comportamento dos papéis, logo, os investimentos em ativos negociados em bolsa possuem alto risco. Para minimizar esse fato, o conhecimento do movimento dos preços no futuro pode vir a ser de grande importância.

Dada a quantidade e variedade de informações a serem consideradas no investimento ou desinvestimento em um fundo imobiliário, até mesmo para investidores com certa experiência no mercado, a tomada de decisão pode ser um exercício trabalhoso. De acordo com a bolsa de valores Brasil, Bolsa, Balcão - B3 (B3, 2021), somente no último ano (06/2020 a 06/2021), houve incremento de 42% no número de pessoas físicas na bolsa e, especificamente em fundos imobiliários, de 185% (12/2019 a 06/2021). Considerando que muitos desses entrantes não possuem experiência ou conhecimento em renda variável, a tomada de decisão pode se tornar uma tarefa árdua e complexa e, por fim, com resultados indesejados. A Figura 1 apresenta a evolução de investidores na bolsa de valores B3.

Figura 1 – Evolução do número de investidores na bolsa de valores com posição em custódia (mil).



Fonte: Adaptado de B3 (2021).

O aumento do número de investidores em fundos imobiliários pode ser explicado por um fator relevante, apontado por Mota (2013), que é permitir ao investidor pessoa física, "investidor comum", não qualificado, aplicar seus recursos em empreendimentos de grande porte, fato este que somente poderia ser possível se este disponibilizasse elevado volume financeiro para investir em ativos que necessitam de elevado volume financeiro para serem edificadas.

Diante deste cenário com diversos parâmetros, variáveis e quantidade de dados a serem considerados, convém a utilização de aprendizado de máquina (*machine learning*) para auxiliar a tomada de decisão de investimentos em fundos imobiliários. Logo, posto que fundos imobiliários são ativos de renda variável, estes estão submetidos, se não na mesma proporção, mas a fatores semelhantes aos quais as ações estão subordinadas.

1.1 OBJETIVOS

Este trabalho tem o objetivo de examinar a aplicabilidade da técnica de aprendizado de máquina a fundos imobiliários negociados em bolsa de valores. Para tanto, foram executadas as seguintes etapas:

- a. Realizar o levantamento dos dados adequados, ou seja, aqueles que detém relevante exposição a performance dos fundos imobiliários;
- b. Preparar os dados, ou seja, segregá-los em dados de treinamento, validação e teste;
- c. Consolidar a técnica de aprendizado de máquina a ser utilizada, com base na revisão bibliográfica realizada;
- d. Treinar o modelo;
- e. Validar o modelo;
- f. Avaliar os resultados obtidos com base em métricas estatísticas;
- g. Aprimorar o modelo.

1.2 JUSTIFICATIVA

Dado o recente crescimento de investidores na bolsa de valores, principalmente a partir do ano de 2019 e que muitos destes são provenientes estritamente do mercado de renda fixa. Mercado este geralmente menos arriscado onde investidor empresta dinheiro para um banco, o governo federal ou uma empresa e o dinheiro investido retornará com juros, segundo Perlin (2019).

Portanto, inúmeros investidores que sempre investiram em ativos com risco baixo, viram no mercado de renda variável possibilidades de diversificação da carteira de investimentos, de maior ganho de capital, de maior liquidez, etc.

Renda fixa é um tipo de investimento em que rendimentos reais, nominais ou indexados às taxas flutuantes são recebidos em intervalos de tempo regulares e definidos em documentos formais (NETO; SANTOS; MELLO, 2019).

De acordo com Scolese et al. (2015) a percepção das pessoas físicas é de que o investimento é seguro e apresenta baixa volatilidade em comparação com ativos de renda variável. Entretanto, mesmo que o fundo imobiliário esteja naturalmente lastreado a um ativo real, considerado um ativo seguro e com baixa volatilidade, verificou-se que, pela dinâmica inerente às variações de preços e riscos associados aos fundos imobiliários, há um caráter híbrido desses fundos no que tange ao mercado de renda fixa e variável.

Neste contexto, um modelo de inteligência artificial para auxílio no investimento em fundos imobiliários é interessante do ponto de vista do investidor para que possa melhor inferir sobre suas decisões.

Thakkar e Chaudhari (2021) ressalta que uma bolsa de valores representa uma abordagem coletiva de compra e venda de vários instrumentos públicos e/ou privados que apresentam volatilidade, ou seja, flutuações aleatórias ao longo do tempo e, portanto, a análise do comportamento do mercado é um trabalho desafiador.

A complexidade do problema também é destacada por Santos (2020), pois a volatilidade típica do mercado de ações torna a tarefa da previsão mais difícil. Não obstante, a previsão do movimento do preço de ações tem sido considerada há muito tempo como uma tarefa intrigante e desafiadora, tanto para estudos acadêmicos quanto para o mercado financeiro, apontam Huang e Liu (2019).

Yokoyama (2014) evidenciou que, além de variáveis macroeconômicas, as características de cada fundo (e.g. *market-to-book*, tamanho) podem ser responsáveis por explicar considerável parte da variância dos retornos obtidos.

Dentro do processo de tomada de decisão de compra ou venda, séries históricas dos preços das ações estão entre as informações mais importantes para os investidores. Entretanto, dado o caráter dinâmico dessas séries, a realização de previsões desse

tipo com alto grau de acerto é uma tarefa extremamente difícil, em virtude dos preços das ações serem afetados por diversos fatores sociais, políticos e macro-econômicos, como avalia Finkler (2017).

Deste modo, o desenvolvimento deste trabalho se justifica por almejar obter um modelo de classificação de pontos de negociação de fundos imobiliários utilizando uma técnica de inteligência artificial conhecida como aprendizado profundo, visando auxiliar a tomada de decisão, principalmente no que diz respeito à diminuição de risco da operação.

1.3 ORGANIZAÇÃO DO TRABALHO

No Capítulo 2 encontra-se a revisão da literatura, onde está descrita a fundamentação teórica base deste trabalho e a análise de alguns trabalhos relacionados. O Capítulo 3 trata dos métodos aplicados no desenvolvimento, enquanto no Capítulo 4 são abordados e discutidos os resultados. Por fim, no Capítulo 5 estão as conclusões e sugestões para trabalhos futuros.

1.4 CONSIDERAÇÕES SOBRE O CAPÍTULO

Neste Capítulo introdutório foram elencados os objetivos e apresentada a justificativa para desenvolvimento do tema proposto, além de resumidamente apontar como o trabalho está organizado.

No Capítulo 2 encontra-se a revisão da literatura, onde está descrita a fundamentação teórica base deste trabalho, tais como conceitos relacionados a inteligência artificial e indicadores técnicos.

2 REFERENCIAL TEÓRICO

No Capítulo anterior foram abordados os objetivos e justificativas para desenvolvimento do tema proposto, este tratará do referencial teórico base para o progresso do trabalho.

2.1 FUNDO DE INVESTIMENTO IMOBILIÁRIO

Conforme Zanadrea (2018), fundos imobiliários tornaram-se uma opção de investimento com certa popularidade nos últimos anos. Esta modalidade de investimento representa o acesso ao investimento com base imobiliária para qualquer pessoa.

Para a Comissão de Valores Mobiliários (CVM, 2019), fundo de investimento imobiliário é uma comunhão de recursos destinados à aplicação em empreendimentos imobiliários, tais como a construção e a aquisição de imóveis para posterior comercialização ou a aquisição de imóveis prontos para auferir renda de locação.

Os principais rendimentos obtidos pelos fundos de investimento imobiliário vêm do recebimento de aluguéis, dos juros dos títulos imobiliários investidos e do ganho de capital obtido na venda dos direitos reais sobre bens imóveis. A maioria dos fundos distribui os rendimentos mensalmente, em virtude da legislação tributária. Devido a essa característica de distribuição mensal de rendimentos, a CVM (2019) cita que:

Apesar de muitos FII distribuírem rendimentos mensalmente, não podemos considerar o investimento em FII como de renda fixa. Não apenas em função dos riscos do empreendimento (inadimplência de um inquilino, vacância de um imóvel etc.), mas também pelo fato de que o valor de suas cotas oscila.

Por meio dos fundos de investimento imobiliário, os investidores têm a oportunidade de aplicar em uma carteira diversificada de ativos financeiros ou em grandes empreendimentos imobiliários, que não seriam acessíveis individualmente (ANBIMA, 2014).

2.2 INTELIGÊNCIA ARTIFICIAL, APRENDIZADO DE MÁQUINA E APRENDIZADO PROFUNDO

Barr (1981) define inteligência artificial como a parte da ciência da computação que se preocupa em projetar sistemas computacionais inteligentes, ou seja, sistemas que exibem as características que associamos à inteligência e comportamento humanos - compreensão da linguagem, aprendizagem, raciocínio, resolução de problemas e assim por diante. Sendo assim, de forma concisa, a inteligência artificial pode ser descrita como o esforço para automatizar tarefas intelectuais normalmente realizadas por humanos (CHOLLET, 2021).

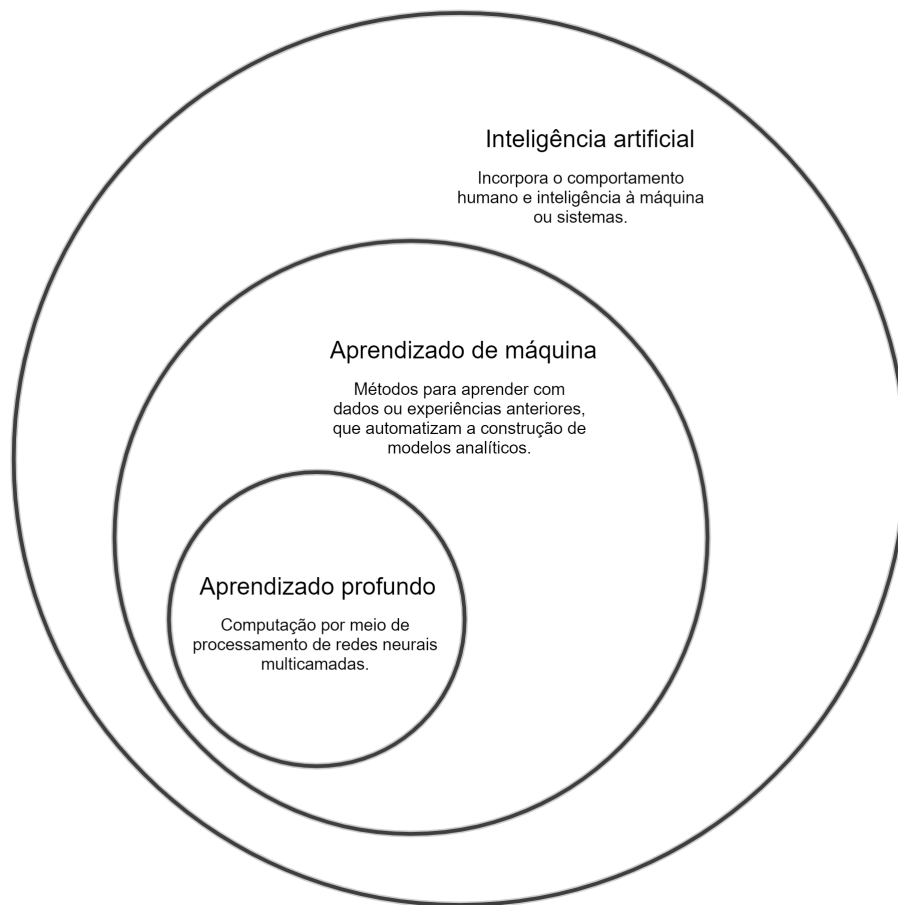
O aprendizado de máquina, por sua vez, é um campo de estudo que analisa o uso de algoritmos computacionais para transformar dados empíricos em modelos utilizáveis, conforme Edgar (2017). Por conseguinte, é geralmente considerado um subcampo da inteligência artificial e até mesmo um subcampo da ciência da computação em algumas perspectivas (SUBASI, 2020).

Para Das (2021), o uso das aplicações com aprendizado de máquina estão onipresentes na era digital atual e as metodologias vêm ganhando atenção devido à sua capacidade de resolver problemas complexos e fornecer um sistema de apoio à decisão a partir de grandes volumes de dados.

Enquanto, aprendizado profundo é uma subclasse de aprendizado de máquina que imita a função do cérebro humano para interpretar dados usando redes neurais de várias camadas, de acordo com Mao et al. (2019).

Por isso, as técnicas de aprendizado profundo podem desempenhar um papel fundamental na construção de sistemas inteligentes baseados em dados de acordo com as necessidades atuais, devido às suas excelentes capacidades de aprendizagem a partir de dados históricos (SARKER, 2021). A Figura 2 apresenta a relação entre inteligência artificial, aprendizado de máquina e aprendizado profundo.

Figura 2 – Relação entre inteligência artificial, aprendizado de máquina e aprendizado profundo em um diagrama de Venn.

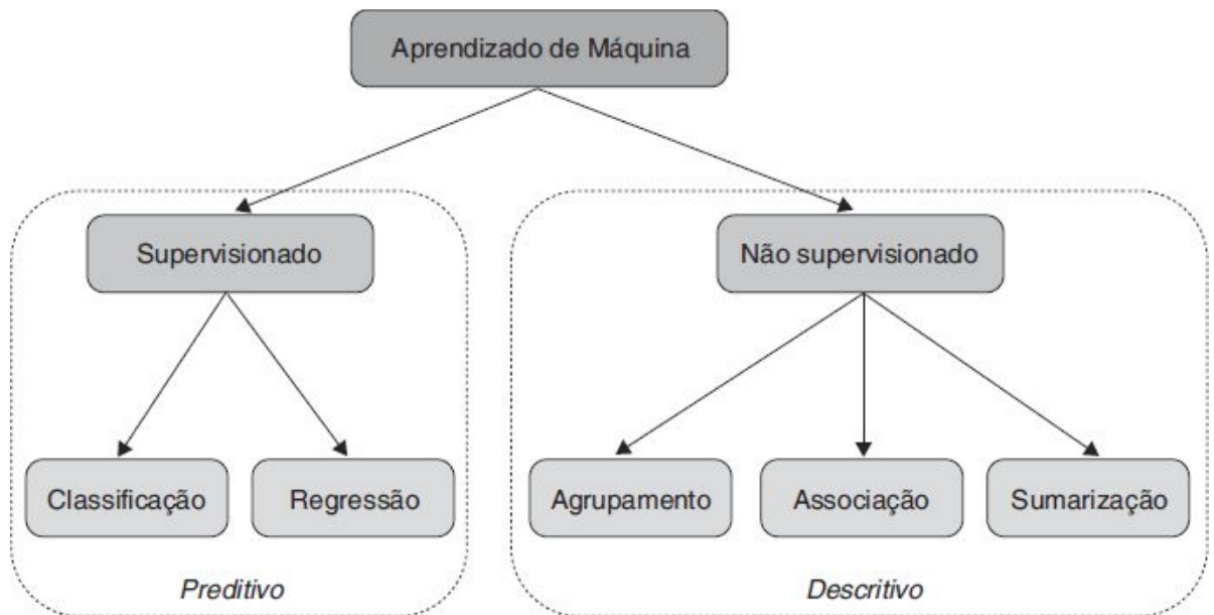


Fonte: Sarker (2021) e Patterson (2017).

Conforme Goodfellow (2016), algoritmos de aprendizagem de máquina podem ser categorizados como **não-supervisionados** ou **supervisionados**, de acordo com o tipo de experiência que eles podem ter durante o processo de aprendizagem.

Algoritmos de aprendizagem não-supervisionados experimentam um conjunto de dados (*dataset*) contendo muitas características, para, em seguida, aprender propriedades úteis da estrutura deste conjunto de dados. Os algoritmos de aprendizagem supervisionados, por outro lado, experimentam um *dataset* contendo exemplos característicos, mas cada exemplo também está associado a um rótulo ou alvo. Faceli (2011) apresenta uma hierarquia de aprendizado ilustrada na Figura 3.

Figura 3 – Hierarquia de aprendizado de máquina.



Fonte: Faceli (2011).

Tem-se, ainda, segundo Goodfellow (2016), que alguns algoritmos de aprendizado de máquina não experimentam apenas um *dataset* fixo. Por exemplo, algoritmos de aprendizagem **por reforço** interagem com um ambiente, então há um ciclo de *feedback* entre o sistema de aprendizagem e suas experiências.

Consoante Sarkar, Bali e Sharma (2018), métodos ou algoritmos de aprendizagem supervisionado incluem algoritmos de aprendizagem que coletam amostras de dados (conhecidas como dados de treinamento) e saídas associadas (conhecidas como rótulos ou respostas) com cada amostra de dados durante o processo de treinamento do modelo.

Basicamente tenta-se modelar a relação entre as entradas e suas saídas correspondentes dos dados de treinamento para prever respostas de saída para novas entradas de dados com base no conhecimento que adquirido anteriormente referente às relações e mapeamentos entre as entradas e seus resultados. Estes métodos de aprendizagem são de duas classes principais, de acordo com os tipos de tarefas que se dispõem a resolver, a saber, **classificação** e **regressão**.

Métodos de aprendizagem não-supervisionados são denominados desta forma, segundo Sarkar, Bali e Sharma (2018), porque o modelo ou algoritmo tenta aprender estruturas latentes inerentes, padrões e relações nos dados fornecidos sem qualquer ajuda ou supervisão, tal como notas ou resultados rotulados.

Deste modo, a aprendizagem não-supervisionada está mais preocupada em tentar extrair *insights* ou informações significativas a partir dos dados, em vez de tentar prever algum resultado com base em treinamento supervisionado previamente realizado. Esta aprendizagem pode ser categorizada em quatro grandes áreas, conforme as tarefas a serem realizadas: **agrupamento, redução dimensional, detecção de anomalia e mineração de regras de associação.**

Os métodos de aprendizagem por reforço são um pouco diferentes dos convencionais métodos supervisionados e não-supervisionados, conforme Sarkar, Bali e Sharma (2018). Neste contexto, existe um agente a ser treinado por um determinado período de tempo para interagir com um ambiente específico e ao longo do tempo seu desempenho é melhorado no que diz respeito aos tipos de ações que este agente efetua no ambiente.

Normalmente, o agente começa com um conjunto de estratégias ou políticas para interagir com o ambiente. Ao observar o ambiente, o agente efetua uma determinada ação baseada em uma regra ou política, observando o estado atual do ambiente. Com base na ação, o agente recebe uma recompensa, se for benéfica ou uma penalidade se for prejudicial. Então, atualiza suas políticas e estratégias, se necessário, e esse processo iterativo continua até que o agente aprenda o suficiente sobre seu ambiente para obter as recompensas desejadas.

Aprendizagem profunda é definida por Ketkar e Moolayil (2021) como um subcampo dentro do aprendizado de máquina que lida com os algoritmos que se assemelham a uma versão simplificada do cérebro humano e que resolve uma vasta categoria de problemas de inteligência de máquina moderna. Do mesmo modo, Chollet (2021) delimita o aprendizado profundo como:

... um subcampo específico do aprendizado de máquina: uma nova abordagem sobre o aprendizado de representações a partir de dados que

ênfatiza o aprendizado de camadas sucessivas de representações cada vez mais significativas. O "profundo" em "aprendizado profundo" não é uma referência a qualquer tipo de compreensão mais profunda alcançada pela abordagem; em vez disso, significa essa ideia de camadas sucessivas de representações.

Redes neurais recorrentes, de acordo com Campesato (2020) é um tipo de arquitetura que foi desenvolvido durante a década de 1980 e são adequadas para conjuntos que contêm dados sequenciais, bem como para tarefas de processamento de linguagem natural, como modelagem de linguagem, geração de texto ou autocompletar de frases, enquanto a rede LSTM é um tipo particular de rede neural recorrente com uma ampla gama de utilizações, como classificação de documentos, análise de séries temporais, reconhecimento de fala e voz (NABIPOUR et al., 2020), além de reconhecimento de escrita (CAMPESATO, 2020).

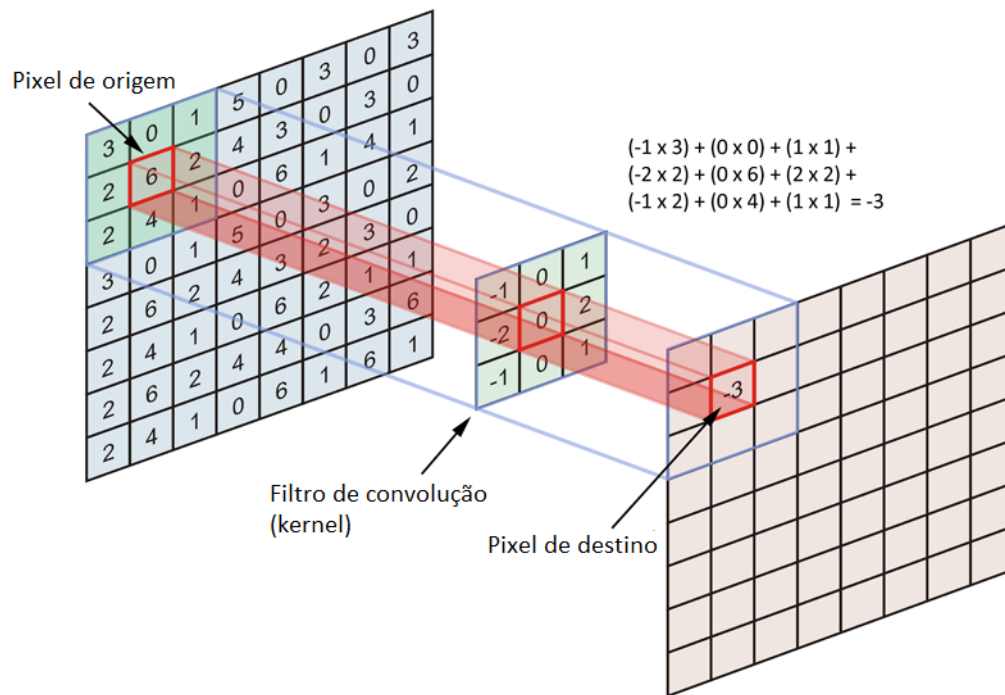
As redes neurais convolucionais (CNN) são enquadradas como classificadores nas técnicas de aprendizado profundo, sendo utilizadas principalmente para analisar imagens.

Uma rede neural convolucional é um algoritmo de aprendizado profundo que prossegue com a aplicação de filtros simples aplicados repetidamente para obter um mapa de características, que é uma extração implícita de características significativas de uma imagem (SRIVASTAVA et al., 2021).

Sewak, Karim e Pujari (2018) também denominam o filtro aplicado na operação de convolução como *kernel* ou janela de convolução. A convolução de forma simples é a soma da multiplicação de matrizes elemento a elemento entre o *kernel* e a vizinhança que o *kernel* cobre da imagem de entrada. A Figura 4 demonstra o processo de convolução¹.

¹ Adaptado de <https://dev.to/sandeepbalachandran/machine-learning-convolution-with-color-images-2p41>

Figura 4 – Ilustração da operação de convolução com um *kernel* 3 x 3.



Fonte: Adaptado de Lima, Rubik e Morais (2020).

Assim, conforme explicam Xiang, Zhang e Seeling (2020) as camadas convolucionais são muito consistentes em extrair características as quais os filtros são capazes de encontrar nas imagens.

2.3 LINGUAGEM DE PROGRAMAÇÃO PYTHON

No contexto da ciência de dados, a linguagem de programação Python tem se mostrado a mais apropriada para tal finalidade, pois, de acordo com Gaurav (2018), desde a coleta até a visualização dos dados a linguagem é extremamente útil por dispor de diferentes bibliotecas para manipulação, análise, testes e apresentação gráfica dos resultados.

Desta forma, podemos dizer que Python atua como uma linguagem de programação chave para a ciência de dados. Podemos concluir que o Python é uma linguagem flexível e fornece ferramentas adequadas para executar técnicas de ciências de dados (GAURAV, 2018).

De acordo com McKinney (2019), Python surgiu em 1991 e tornou-se uma das lin-

guagens de programação interpretadas mais populares, principalmente nos últimos anos, passando de uma linguagem de computação científica inovadora para uma das linguagens mais importantes em ciência de dados, aprendizado de máquina e desenvolvimento em geral, tanto no ambiente acadêmico quanto no mercado.

2.3.1 Bibliotecas Python para Aprendizado de Máquina

De modo a efetuar estudos de ciência de dados e aprendizado de máquina, o emprego de algumas bibliotecas é indispensável para implementação do código de maneira produtiva e eficiente. Para McKinney (2019) as bibliotecas Python essenciais para análise de dados são: *NumPy*, *pandas*, *matplotlib*, *SciPy* e *scikit-learn*.

Johansson (2018) explana sobre cada uma dessas bibliotecas:

- a. *NumPy* fornece estruturas de dados para representar uma rica variedade de matrizes, métodos e funções para operar em tais matrizes, assim, proporciona o *back-end* numérico para quase todas as bibliotecas científicas ou técnicas para a linguagem.
- b. A biblioteca *Pandas* fornece principalmente estruturas de dados e métodos de representação e manipulação. As duas principais estruturas de dados são os objetos *Series* e *DataFrame*, que são usados para representar séries e dados tabulares, respectivamente.
- c. *Matplotlib* é uma biblioteca Python para produção de gráficos 2D e 3D de qualidade, com suporte para uma variedade de formatos de saída diferentes.
- d. A biblioteca *SciPy* é construída sobre *NumPy*, e oferece as estruturas de dados de *array* básicas e suas operações fundamentais. Os módulos no *SciPy* proveem manipulação de alto nível para operações de álgebra linear, otimização, interpolação, integração e muito mais.
- e. *Scikit-learn* contém uma coleção abrangente de algoritmos relacionados ao aprendizado de máquina, incluindo regressão, classificação, redução de dimensionalidade e agrupamento.

TensorFlow² é uma plataforma de aprendizado de máquina de código aberto, gratuita e baseada em Python, desenvolvida principalmente pelo Google®. Muito parecido com o NumPy, o objetivo principal do TensorFlow é permitir que engenheiros e pesquisadores manipulem expressões matemáticas sobre tensores numéricos (CHOLLET, 2021).

Keras é uma biblioteca Python de alto nível e compacta para aprendizado que pode ser executada no TensorFlow. Ela permite que os desenvolvedores se concentrem nos principais conceitos de aprendizado profundo, como a criação de camadas para redes neurais, ao mesmo tempo em que trata dos detalhes essenciais de tensores, suas formas e seus detalhes matemáticos (MANASWI, 2018).

2.4 INDICADORES TÉCNICOS

Indicadores técnicos representam linhas de oscilação ou tendência, construídas com base em médias móveis ou em fórmulas matemáticas ou estatísticas, com o objetivo de antecipar, confirmar ou descrever tendências (CVM, 2019). Nesta seção encontram-se os indicadores técnicos utilizados neste trabalho.

2.4.1 Índice de Força Relativa

Henriques (2016) descreve que o Índice de Força Relativa (RSI, do inglês *Relative Strength Index*) permite identificar o enfraquecimento de uma tendência, a quebra de um suporte ou de uma resistência antes que estes ocorram, pois compara a variação média de preços dos períodos de avanços com a variação média dos períodos de declínios (LEMOS, 2015). A Equação 1 refere-se ao cálculo do índice.

$$RSI = 100 - \frac{100}{1 + RS} \quad (1)$$

onde *RS* é a razão entre a média de fechamentos de alta e a média de fechamentos de baixa em determinado período de tempo.

² <https://www.tensorflow.org/>

2.4.2 Williams %R

Este indicador de momento, de acordo com Lemos (2015), representa o nível de fechamento relativo à maior máxima do período, servindo para identificar níveis de sobrecompra e sobrevenda. O cálculo pode ser feito a partir da Equação 2:

$$\%R = \frac{M_n - F_i}{M_n - m_n} \times 100 \quad (2)$$

sendo, que i e n correspondem ao dia em questão e ao período estipulado, respectivamente, e M_n é o preço máximo no período, F_i o fechamento do dia e m_n o preço mínimo do período.

2.4.3 Movimento Direcional

De acordo com Lemos (2015) *Average Directional Index* (ADX) mede a força da tendência sem levar em conta sua direção. O movimento é considerado positivo quando a máxima corrente subtraída da máxima anterior é maior do que a mínima anterior subtraída da mínima corrente, ao contrário, o movimento direcional é caracterizado como negativo quando a mínima anterior subtraída da mínima corrente é maior do que a máxima atual subtraída da máxima anterior (LEMOS, 2015).

Da conceituação deste indicador, desdobram-se duas outras características a serem consideradas, o Indicador Direcional Positivo (DI+) e Indicador Direcional Negativo (DI-), que complementam o ADX ao definir a direção da tendência.

2.4.4 Média Móvel Convergente e Divergente

Média Móvel Convergente e Divergente ou *Moving Average Convergence/Divergence* (MACD) é um dos indicadores de impulsão mais simples e eficazes disponíveis, segundo Lemos (2015), que complementa ao explicitar que o MACD transforma dois indicadores seguidores de tendência, duas médias móveis, em um oscilador de *momentum*, subtraindo a média móvel maior da menor média móvel.

Semelhante ao indicador anterior, do conceito obtém-se mais duas características que serão utilizadas, a saber, a linha de sinal do MACD, que corresponde a média móvel

exponencial de nove dias dos preços de fechamento, e o histograma de MACD, que representa a diferença entre este e sua linha de sinal.

2.4.5 *Commodity Channel Index*

Este indicador denominado *Commodity Channel Index* (CCI) mede os desvios de preço de um título em relação a uma média móvel e serve para detectar o início e o fim de uma tendência de mercado, de acordo com Lemos (2015). A Equação 3, baseada em Colby (2002), mostra como determinar o indicador:

$$CCI = \frac{M - A}{0.015 \times \sigma \times M} \quad (3)$$

onde M é o típico para um período, A é a média de M para o período em análise e σ é o desvio padrão.

2.4.6 *Momentum*

Lemos (2015) traduz o *Momentum* como uma espécie de medida de velocidade da tendência em determinado mercado ou ativo. Este indicador pode ser definido como uma diferença ou uma razão, com o método da razão sendo preferível para manter a comparabilidade ao longo do tempo, conforme Colby (2002). A Equação 4 demonstra como calcular o indicador *momentum*.

$$MOM = \frac{C}{C_n} \times 100 \quad (4)$$

sendo C o preço de fechamento mais recente e C_n o preço de fechamento há n períodos.

2.4.7 *Trix*

Conforme Colby (2002), *Trix* é um oscilador de momento de preço. Lemos (2015) ressalta que é um indicador de impulsão que mostra a variação percentual de uma média exponencial de um preço de fechamento, projetado para filtrar os movimentos de preços insignificantes graças à sua suavização tripla.

2.4.8 Taxa de Mudança

Kirkpatrick e Dahlquist (2015) estabelecem que a Taxa de Mudança (ROC, do inglês *Rate of Change*) é o mais simples de todos os osciladores, correspondendo a uma medida do valor que o preço de um título mudou em um determinado número n de períodos anteriores. A taxa de variação pode ser utilizada para identificar os períodos em que a variação percentual se aproxima de um nível que antecipa um ponto de virada no passado (LEMOS, 2015). A Equação 5 demonstra como calcular a taxa de mudança.

$$ROC = \frac{C - C_n}{C_n} \times 100 \quad (5)$$

onde C é preço de fechamento na atual data e C_n é o preço de fechamento de n períodos atrás.

2.4.9 Índice de Fluxo de Dinheiro

Também denominado Money Flow Index (MFI), é um indicador utilizado para verificar a força de uma tendência através das entradas e saídas de dinheiro de um ativo. Para isso, utiliza preço e volume na quantificação das pressões de compra e venda, como explicam Vijaykumar e Kuma (2020). Assim sendo, Lemos (2015) demonstra como calcular fluxo de dinheiro no dia i como apresentado na Equação 6.

$$MF_i = \frac{M_i + m_i + F_i}{3} \times V_i \quad (6)$$

sendo M_i a máxima cotação do dia, m_i a mínima, F_i o preço de fechamento e V_i o volume negociado.

Assim, se no dia i , o preço médio for maior que o do dia anterior, então o fluxo de dinheiro é positivo (PMF), caso contrário o fluxo de dinheiro será negativo (NMF). Então o índice de fluxo de dinheiro é determinado pela Equação 7.

$$MFI = 100 - \frac{100}{\left(1 + \frac{\sum PMF}{\sum NMF}\right)} \quad (7)$$

2.4.10 *Vortex*

O indicador *Vortex* (BOTES; SIEPMAN, 2010) é um indicador composto por duas linhas, sendo uma linha de tendência de alta (*VI+*) e uma linha de tendência de baixa (*VI-*). Desta forma, capturam movimento ascendente e descendente nas tendências de preços.

2.4.11 Ciclo de Tendência de Schaff

O Ciclo de Tendência de Schaff (STC, do inglês Schaff *Trend Cycle*) é um indicador de ciclo que é usado para analisar os ciclos de mercado na análise técnica. De acordo com Huang, Huang e Ni (2019), este indicador combina um indicador estocástico lento com MACD e é frequentemente usado para prever taxas de câmbio.

2.5 CONSIDERAÇÕES SOBRE O CAPÍTULO

O Capítulo 2 tratou da fundamentação teórica base. Foi abordado o conceito de fundo imobiliário, realizada revisão bibliográfica da aplicação da inteligência artificial em vários contextos e apresentada a série de indicadores técnicos utilizados, além dos motivos da escolha da linguagem programação Python para desenvolver o trabalho. Detalhamento da metodologia aplicada encontra-se a seguir, no Capítulo 3.

3 METODOLOGIA PROPOSTA

O Capítulo 2 tratou do referencial teórico, onde foram apresentados os indicadores técnicos utilizados neste trabalho, bem como os conceitos de fundo imobiliário e de técnicas de inteligência artificial.

Este trabalho está inserido no cenário da ciência de dados. Corrobora para isso Molin (2021), que explana que a ciência de dados é frequentemente descrita como um campo interdisciplinar onde a interseção de habilidades de programação, *know-how* estatístico e conhecimento de domínio se cruzam.

... um novo campo multidisciplinar surgiu: ciência de dados, um campo que combina estatística, análise de dados, aprendizado de máquina, matemática, ciência da computação, ciência da informação e seus métodos relacionados para compreender e analisar fenômenos da vida real (THACH, 2021).

Deste modo, para desenvolvimento de aplicações nesta área de ciência de dados, a linguagem de programação Python mostra-se adequada pelas vantagens já abordadas no capítulo anterior e a versão utilizada foi a 3.7.13. O detalhamento da metodologia de trabalho encontra-se em um fluxograma presente no Apêndice A.

3.1 COLETA DE DADOS

Os dados brutos foram obtidos do site de finanças Yahoo Finance¹ por meio da biblioteca Python yfinance². Esses dados consistem das séries históricas de cotações que trazem o histórico de preços dos títulos negociados com dados de cotação de abertura, maiores alta e baixa, a cotação de fechamento, fechamento ajustado, desdobramentos e desdobramentos/grupamento de cotas.

Uma vez obtidos os dados brutos, que originalmente possuem sete características (*features*), faz-se necessário selecionar os dados de interesse, que são os preços de

¹ <https://finance.yahoo.com/>

² <https://pypi.org/project/yfinance/>

abertura, máximo, mínimo e de fechamento diário, além do volume negociado dos ativos elencados para o estudo. Posteriormente, foi realizado o ajuste relativo aos desdobramentos e/ou grupamentos de cotas que por ventura cada fundo selecionado tenha sido submetido.

Com a finalidade de efetuar o tratamento dos dados anteriormente proposto e também para armazenar o resultado fez-se uso da plataforma Kaggle³. Foram selecionados 120 fundos imobiliários de diferentes classificações, segmentos, níveis de liquidez, patrimônio líquido e número de cotistas. A Figura 5 apresenta uma amostra do *dataframe* dos dados de um dos fundos imobiliários selecionados.

Figura 5 – Formato de uma amostra dos dados obtidos.

	Open	High	Low	Close	Volume
Date					
2014-08-06	115.750000	117.000000	115.110001	117.000000	19141
2014-08-07	117.000000	118.000000	115.059998	115.199997	27491
2014-08-08	115.010002	116.330002	115.010002	115.290001	19303
2014-08-11	115.309998	119.500000	115.309998	119.500000	24466
2014-08-12	118.349998	118.699997	117.150002	117.900002	14718
...
2021-12-23	133.440002	133.690002	133.149994	133.520004	17926
2021-12-27	133.520004	135.000000	133.300003	134.750000	25418
2021-12-28	134.100006	140.000000	134.100006	139.350006	41219
2021-12-29	138.500000	142.899994	138.009995	141.270004	29689
2021-12-30	141.270004	142.600006	141.000000	141.889999	16002

Fonte: Autor.

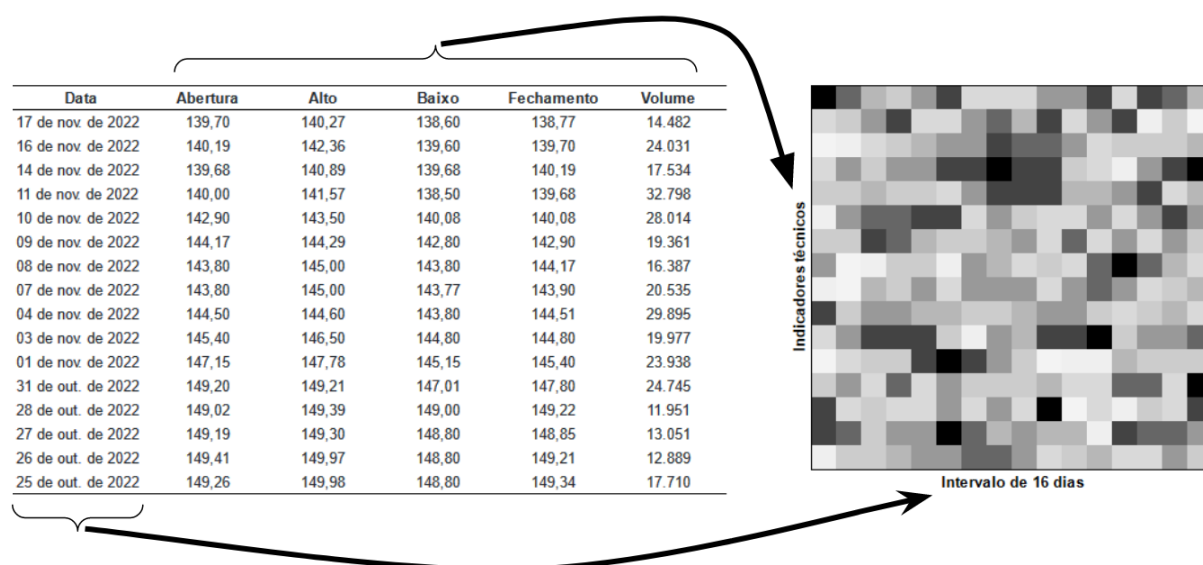
A contribuição da quantidade de dados de cada um dos fundos para o total é variada, em virtude do respectivo tamanho da base dados de negociação diária disponível para cada um destes, contudo, a data final para todos é 30/12/2021.

³ <https://www.kaggle.com/>

3.2 GERAÇÃO DAS IMAGENS

A partir dos indicadores técnicos listados no Capítulo 2, foram geradas imagens 16 x 16 para cada um dos dias de negociação dos fundos imobiliários selecionados. No eixo x incide a série temporal dos 16 indicadores presentes no eixo y, conforme exposto na Figura 6. O cálculo dos indicadores foi realizado com o auxílio da biblioteca Python Pandas Technical Analysis (Pandas TA)⁴.

Figura 6 – Arranjo descritivo das imagens geradas a partir dos indicadores técnicos.



Fonte: Autor.

As imagens foram rotuladas como **manter**, **comprar** e **vender** com base no ROC (Apêndice B), consonante aos sinais de negociação descritos por Lemos (2015) e Kirkpatrick e Dahlquist (2015). Optou-se pelo indicador ROC para rotular as imagens, pois este, conforme Pring (1992), ajuda a explicar alguns movimentos cíclicos nos mercados, muitas vezes dando aviso prévio de uma reversão na tendência predominante.

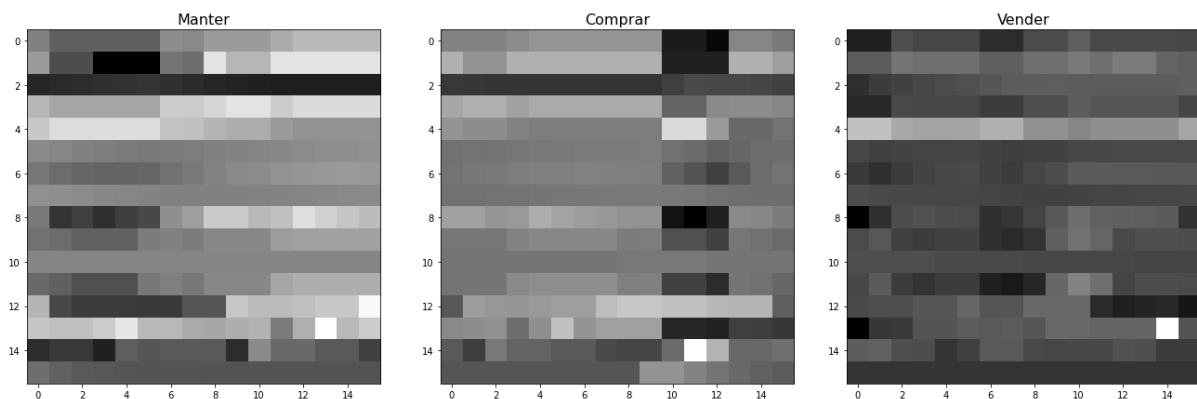
Sezer e Ozbayoglu (2018) ressaltam que a ordem dos indicadores para a confecção das imagens é importante, pois diferentes arranjos resultarão em diferentes formações de imagem. Assim, recomendam que para obter uma imagem de representação consistente e significativa, os indicadores sejam agrupados em duas categorias, a

⁴ <https://github.com/twopirllc/pandas-ta>

saber, tendência e oscilador.

A Figura 7 apresenta uma amostra das imagens geradas com três características distintas. Cada linha das imagens correspondem, sequencialmente, aos indicadores RSI, Williams %R, ADX, DI+, DI-, MACD, histograma MACD, linha de sinal MACD, CCI, MOM, Trix, ROC, MFI, VI+, VI- e STC, todos categorizados como osciladores.

Figura 7 – Amostras das imagens geradas a partir dos indicadores técnicos.



Fonte: Autor.

No total foram obtidas 126.997 imagens rotuladas ao final desta etapa. Posteriormente, este *dataset* foi submetido ao modelo de rede convolucional proposto.

3.3 DESENVOLVIMENTO DO MODELO

O modelo proposto foi desenvolvido utilizando Keras com base na arquitetura proposta por Sezer e Ozbayoglu (2018).

Keras é uma API de aprendizado profundo para Python, desenvolvida com base no TensorFlow, que fornece uma maneira conveniente de definir e treinar qualquer tipo de modelo de aprendizado profundo. Keras foi inicialmente desenvolvido para pesquisa, com o objetivo de permitir a experimentação rápida de aprendizagem profunda (CHOLLET, 2021).

Sumariamente, o modelo é constituído (em sequência) por uma camada de entrada (*input*), duas camadas convolucionais bidimensionais (Conv2D), uma camada de achatamento (*Flatten*) e duas camadas densas (*Dense*). Após as camadas convolucionais

encontram-se camadas de normalização (*BatchNormalization*), como também depois da primeira camada densa.

A camada *flatten* tem a função de transformar a bidimensionalidade da imagem convolucionada em uma *array* unidimensional antes de alimentar as camadas *dense* (totalmente conectadas). Por fim, as camadas densas devem resultar no mesmo número de nós que os rótulos de classificação.

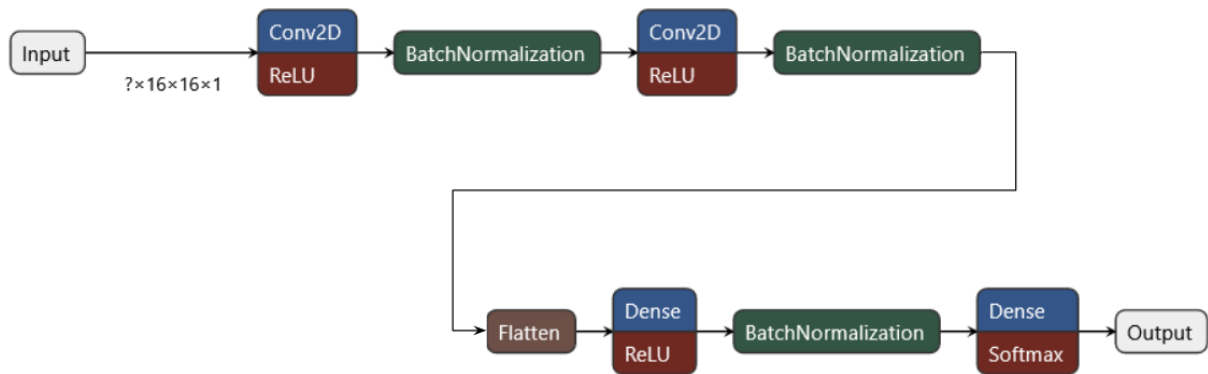
BatchNormalization ou normalização em lote é aplicado ao modelo para normalizar as contribuições para uma camada de cada mini-lote processado pela rede, sendo, dessa forma, uma técnica que auxilia a coordenar a atualização das várias camadas no modelo.

O treinamento de redes neurais profundas é complicado pelo fato de que a distribuição das entradas de cada camada muda durante o treinamento, à medida que os parâmetros das camadas anteriores mudam. Isso retarda o treinamento, exigindo taxas de aprendizado mais baixas e inicialização de parâmetros cuidadosa, tornando notoriamente difícil treinar modelos com não-linearidades saturantes (GOODFELLOW, 2016).

Santurkar et al. (2018) explica que a normalização em lote impacta o treinamento da rede de uma maneira fundamental, fazendo com que o cenário do problema de otimização correspondente seja significativamente mais suave. Isso garante, em particular, que os gradientes sejam mais preditivos e, portanto, permite o uso de uma faixa maior de taxas de aprendizado e uma convergência de rede mais rápida.

Para avaliar o desempenho do modelo foram escolhidas algumas estratégias de negociação baseadas nos indicadores *Vortex*, RSI, além do cruzamento de Médias Móveis Exponenciais (EMA, do inglês Exponential Moving Average) (ROCKEFELLER, 2019). A Figura 8 apresenta a estrutura do modelo elaborado.

Figura 8 – Estrutura do modelo elaborado.



Fonte: Autor.

3.4 AMBIENTE DE DESENVOLVIMENTO

Este trabalho fez uso de dois ambientes desenvolvimento em nuvem, sendo estes Kaggle e Google Colaboratory, cujo primeiro foi utilizado para tratamento dos dados brutos e o segundo para treinamento e testes do modelo.

Kaggle é uma plataforma para competições de modelagem e análise em que os participantes competem para produzir o melhor modelo preditivo para um determinado conjunto de dados (POLAK; COOK, 2021) e foi empregada para extração, preparação e, principalmente, para armazenamento dos conjuntos de dados.

O ecossistema Google Colaboratory, mais comumente denominado como “Google Colab” ou simplesmente “Colab”, é um produto do Google Research⁵ que permite que qualquer pessoa escreva e execute arbitrariamente código Python por meio de um navegador.

De acordo com Paper (2021), Google Colab é especialmente adequado para aprendizado profundo, análise de dados, pesquisa e educação em aprendizado de máquina. Em resumo, a plataforma é um serviço de notebook semelhante ao Jupyter⁶. Ele é hospedado em nuvem e não requer configuração alguma para utilização, além de fornecer acesso gratuito a recursos de computação poderosos, como Unidades de Processamento Gráfico (GPUs) e Unidades de Processamento de Tensor (TPUs).

⁵ <https://research.google/>

⁶ <https://jupyter.org/>

3.4.1 Configurações de hardware

Conforme mencionado anteriormente, o desenvolvimento do modelo foi realizado no ambiente Google Colab. A seguir, a Tabela 1 apresenta a especificação de hardware fornecida pela plataforma, enquanto a Tabela 2 trata da GPU utilizada.

Tabela 1 – Configuração de hardware.

	Ambiente de execução	
	Sem GPU	Com GPU
Arquitetura	x86_64	
Processador	Intel® Xeon® CPU @ 2.20GHz	
CPU(s)	2	
Thread(s) por core	2	
Família da CPU	6	
Modelo	79	
Cache L1	64kB	
Cache L2	256kB	
Cache L3	56320kB	
RAM	13GB	
Disco	100,7GB	78,2GB

Fonte: Autor.

Tabela 2 – Especificação resumida da GPU.

Fabricante	NVIDIA®
Modelo	Tesla T4
Memória	16GB
Versão do Driver	460.32.03
Versão CUDA	11.2

Fonte: Autor.

Inicialmente, com a configuração de hardware disponibilizada pelo Google Colab, percebeu-se a importância da GPU para estudos e elaboração de modelos *deep*

learning, pois o treinamento do modelo e a avaliação dos resultados demandam tempo significativamente menores quando comparados com o uso exclusivo da CPU.

O aumento de performance pode ser explicado em virtude da arquitetura da CPU em relação a GPU, conforme descreve Zocca (2017), pois enquanto a CPU é composta por alguns núcleos que podem lidar com algumas *threads* por vez, a GPU é composta por centenas de núcleos que podem manipular milhares de *threads* ao mesmo tempo. Não obstante, uma GPU é altamente paralelizável em comparação a uma CPU, que é formalmente uma unidade serial.

3.5 CONSIDERAÇÕES SOBRE O CAPÍTULO

Este capítulo abordou a metodologia aplicada para desenvolvimento do trabalho, apresentando as principais ferramentas, métodos e *hardware* utilizados. No Capítulo 4, em seguida, são apresentados e discutidos os resultados.

4 RESULTADOS

Uma vez aplicadas as ferramentas e a metodologia tratada no Capítulo 3, foram obtidos os resultados que serão apresentados a seguir, descritos textualmente, de forma tabular e gráfica.

4.1 GPU VS. CPU

Após consolidar o modelo da rede, seu treinamento exclusivamente efetuado por CPU demandou em média 55min ou, aproximadamente, 02min20s por época. Quando o treinamento do modelo foi realizado com auxílio de GPU, o tempo para conclusão permaneceu em torno de 1min ou 2s por época, isto é, 1,8% do tempo total gasto pelo processo com CPU.

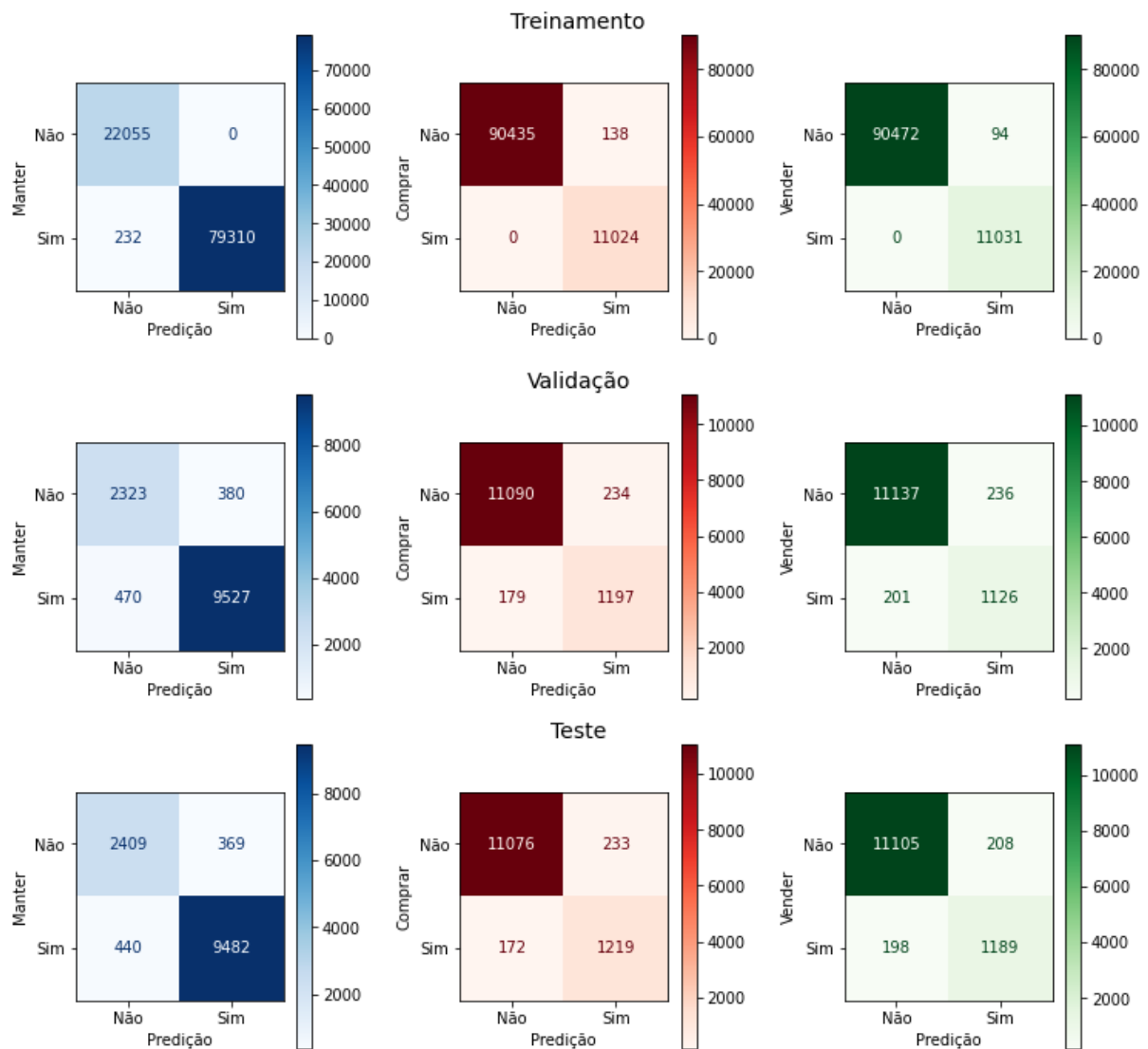
4.2 MÉTRICAS DE DESEMPENHO

Geron (2019) cita que, na maioria dos casos, avaliar um classificador é significativamente mais complicado do que avaliar um regressor. Ainda, a acurácia para os classificadores geralmente não é a medida preferencial de desempenho, especialmente quando se estiver lidando com conjuntos de dados assimétricos, como é a situação dos dados utilizados neste trabalho.

Entretanto, a matriz de confusão é uma das formas mais populares de avaliar um modelo de classificação e, embora não seja uma métrica em si, a representação matricial pode ser usada para definir uma variedade de métricas (SARKAR; BALI; SHARMA, 2018), sendo uma maneira melhor de se avaliar o desempenho de um classificador (GERON, 2019).

Subasi (2020) explana que o desempenho dos modelos classificadores pode ser resumido por meio de uma matriz de confusão. Deste modo, a Figura 9 exibe a performance de classificação em uma perspectiva matricial.

Figura 9 – Matriz de confusão para os conjuntos de dados de treinamento, validação e teste.



Fonte: Autor.

Ademais, conforme definido por Geron (2019), a métrica de acurácia determinada para o *dataset* de teste foi de 0,94. As métricas de precisão, recall e f1-score encontram-se na Tabela 3.

Tabela 3 – Métricas da aplicação do modelo ao *dataset* de teste durante o treinamento.

	precisão	recall	f1-score
Manter	0.96	0.96	0.96
Comprar	0.84	0.88	0.86
Vender	0.85	0.86	0.85

Fonte: Autor.

4.3 APLICAÇÃO DO MODELO AO *DATASET* DE FUNDOS IMOBILIÁRIOS

O período médio de negociação dos fundos presentes no *dataset* foi de 1058 dias. Os resultados apresentados encontram-se descontados de imposto de renda, visto que ganhos de capital auferidos na alienação de cotas de fundos imobiliários tem alíquota de 20% (BRASIL, 1993).

Ao simular a aplicação do modelo a cada um dos 120 fundos imobiliários, este obteve o maior retorno anualizado em 87% das ocasiões, ou seja, para 104 fundos. A Tabela 4 apresenta o sumário dos resultados, enquanto a Tabela 7, presente no Apêndice C, dispõe do comparativo detalhado dos resultados obtidos.

Os resultados obtidos demonstram que o modelo proposto, em termos de retorno anualizado do investimento, superou as estratégias de *benchmark* *Vortex*, EMA e RSI. A estratégia base comprar e manter (*Buy and Hold*) não obteve o melhor desempenho em situação alguma e o modelo resultou no pior desempenho em duas oportunidades.

Tabela 4 – Sumário do maior retorno anualizado de cada uma das estratégias para o conjunto dos fundos imobiliários analisados.

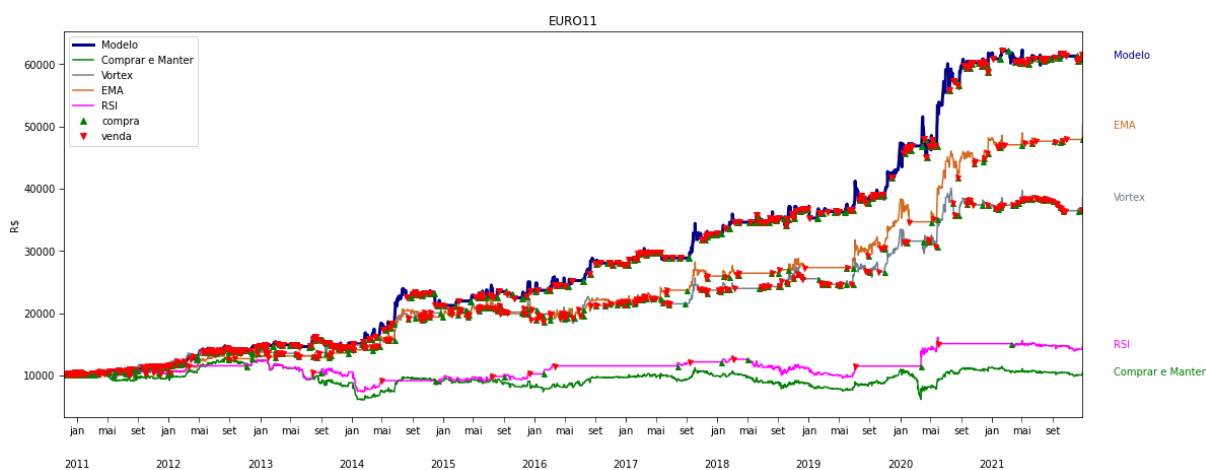
Estratégia	Fundos	Percentual
Modelo	104	87%
Vortex	10	8%
Médias Móveis Exponenciais	5	4%
Índice de Força Relativa	1	1%

Fonte: Autor.

Para o *dataset* (120 fundos) as estratégias do modelo, *Buy and Hold*, *Vortex*, EMA e RSI tiveram retorno anualizado médio de 19,65%, -1,93%, 13,11%, 9,79% e 0,87%, respectivamente. Uma vez que os dados de negociação dos fundos correspondem a intervalos (número de dias) de negociação diferentes, é válida a informação do retorno médio ponderado, que foram de 19,43%, -1,69%, 13,21%, 11,97% e 1,08% para a negociação proposta, respectivamente, pelo modelo, *Buy and Hold*, *Vortex*, EMA e RSI, respectivamente. Diante disso, verifica-se que o resultado médio do modelo foi 47,08% superior à segunda melhor estratégia, isto é, *Vortex*.

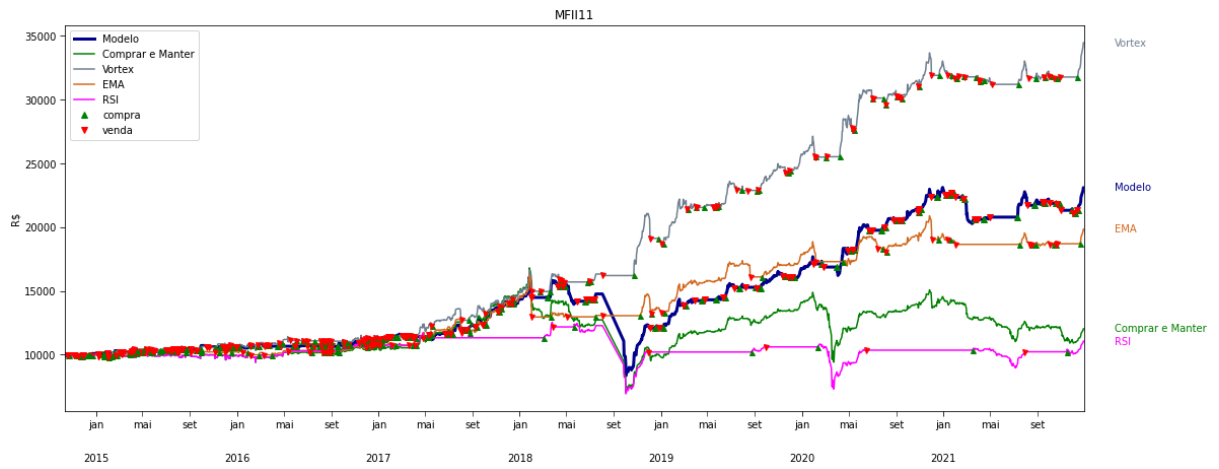
As Figuras 10, 11, 12 e 13 apresentam amostras de desempenho do modelo frente a quatro estratégias de negociação. Na primeira, o modelo obteve melhor performance; na segunda, a estratégia *Vortex* foi a vencedora; na terceira, a negociação com base em médias exponenciais se sobressaiu; e na quarta, a estratégia baseada no indicador técnico RSI foi melhor.

Figura 10 – Amostra em que o modelo proposto obteve o melhor desempenho.



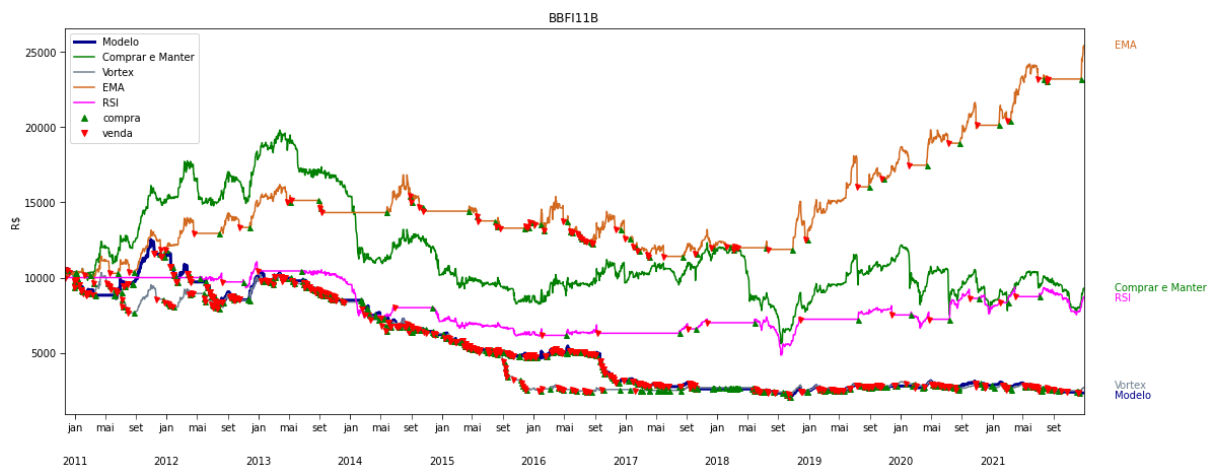
Fonte: Autor.

Figura 11 – Ocasião na qual a estratégia *Vortex* foi a melhor.



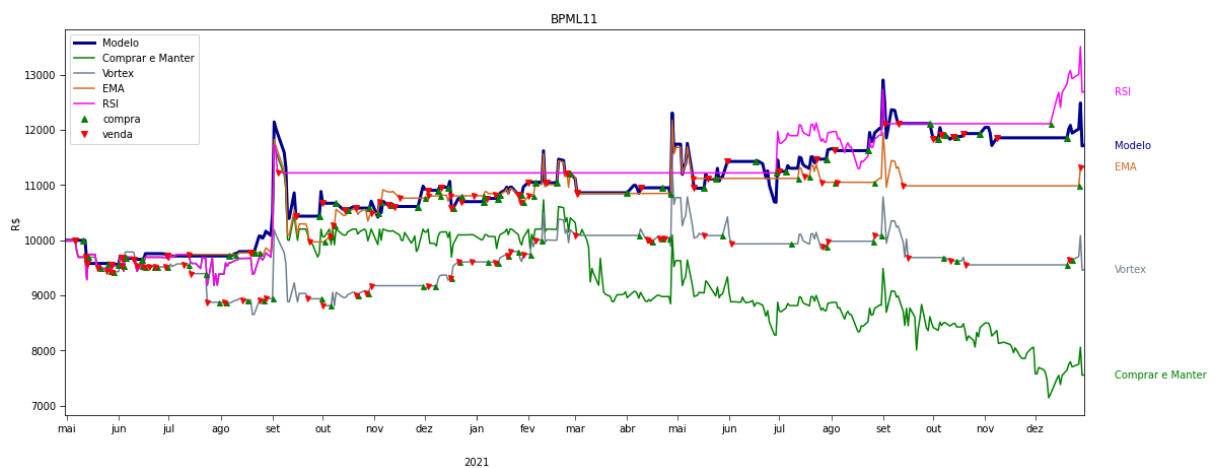
Fonte: Autor.

Figura 12 – Circunstância em que a estratégia de médias móveis foi superior as demais.



Fonte: Autor.

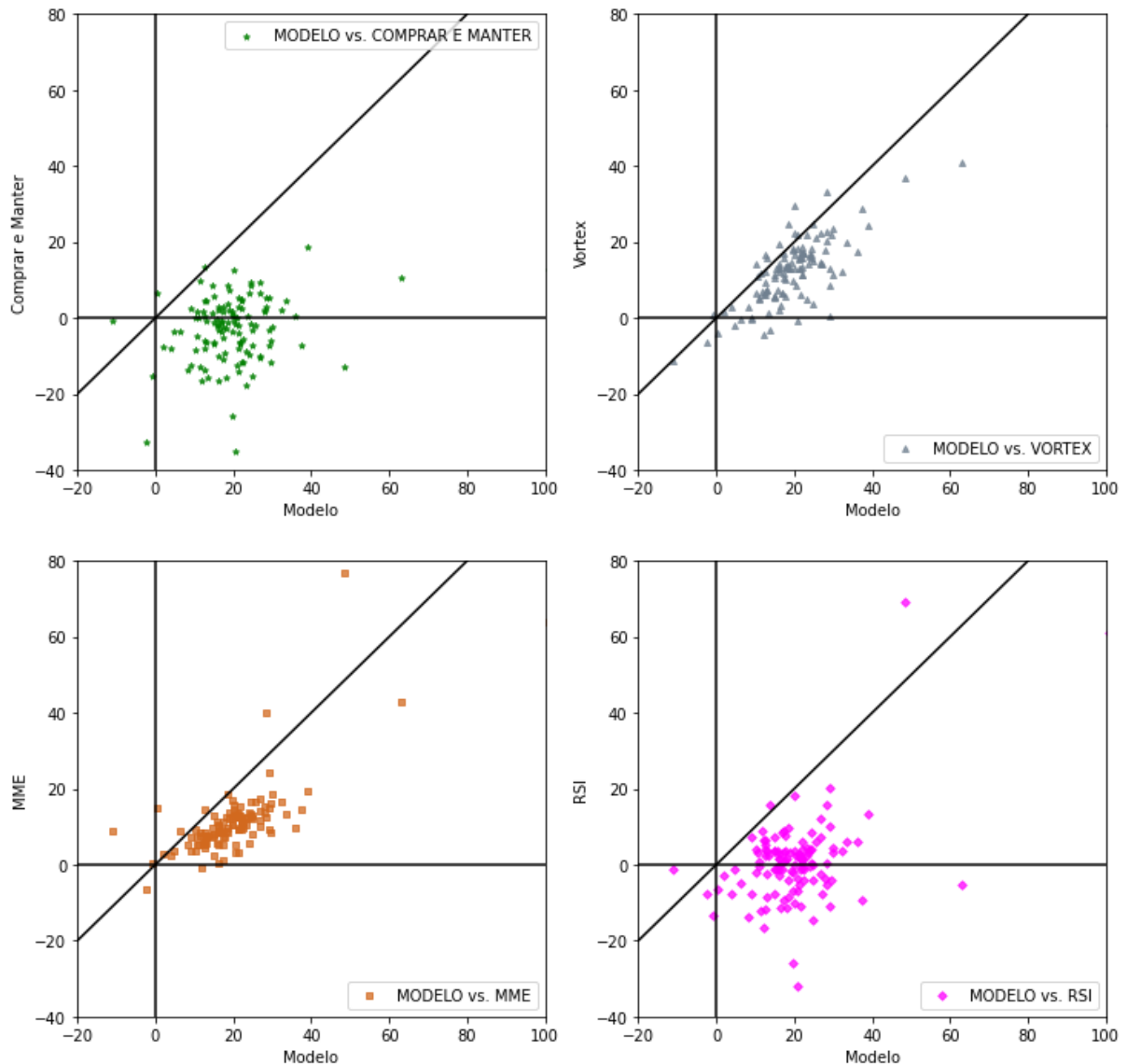
Figura 13 – Situação na qual a estratégia RSI foi vencedora.



Fonte: Autor.

Para melhor entendimento do modelo perante as demais estratégias de comparação, tem-se os gráficos presentes na Figura 14, cuja representação possibilita visualizar como se deu a distribuição dos resultados entre os pares. É possível visualizar que, predominantemente, o modelo obteve melhores resultados.

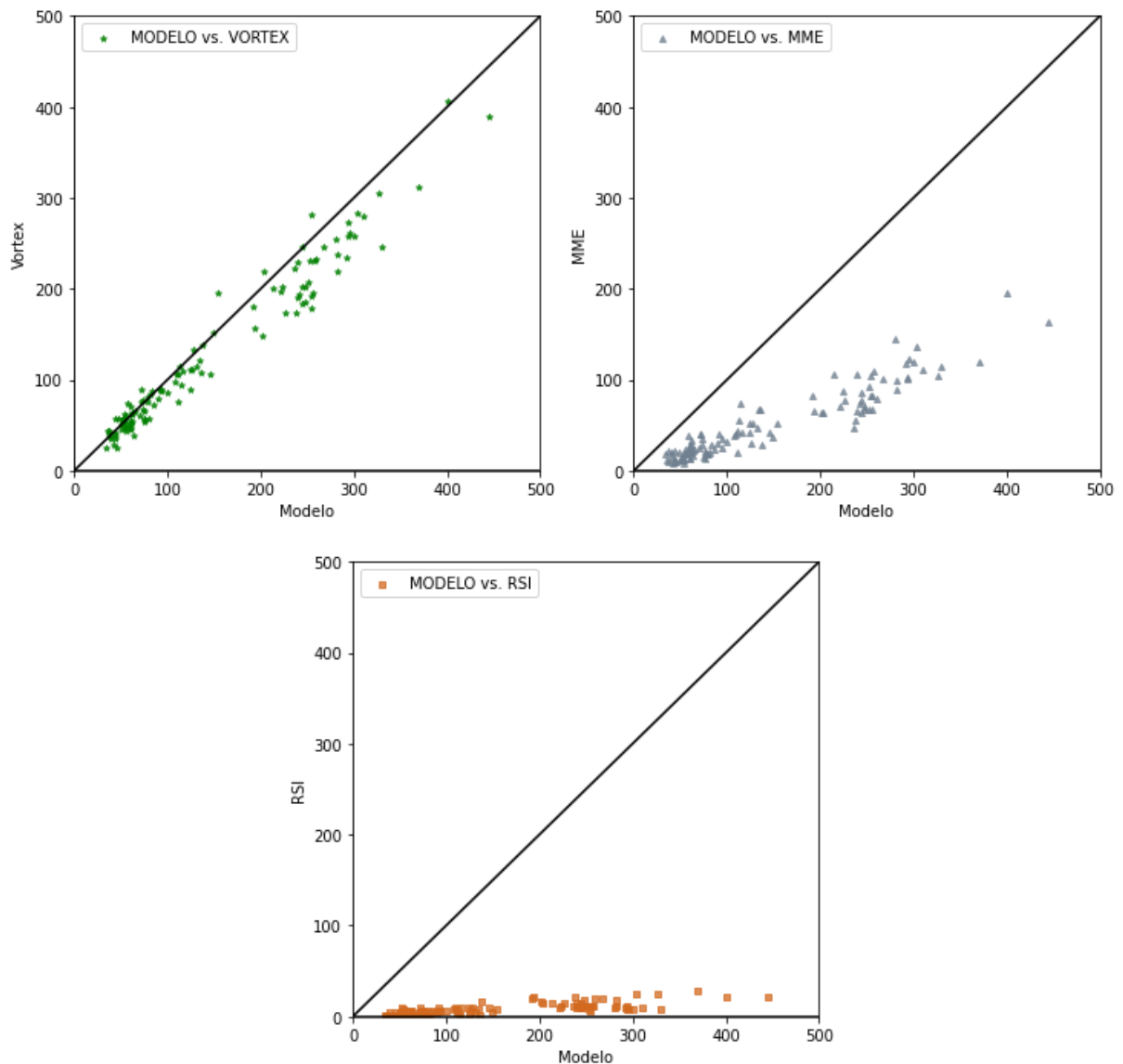
Figura 14 – Distribuição dos resultado do modelo vs outras estratégias.



Fonte: Autor.

De forma semelhante o número de operações (compra ou venda) também foi apurado e as distribuições encontram-se na Figura 15, onde é possível observar que a estratégia baseada no indicador *Vortex*, em média, se aproximou bastante do número de operações proposto pelo modelo.

Figura 15 – Distribuição das operações do modelo vs outras estratégias.



Fonte: Autor.

4.4 TESTES DE GENERALIZAÇÃO DO MODELO

Um ponto importante para qualquer modelo de classificação é sua capacidade de generalizar as soluções. Então, decorrente dos resultados até então obtidos e com o intuito de avaliar se o modelo proposto traz consigo um caráter generalista, foram realizadas duas avaliações.

A primeira foi realizada com os 120 fundos imobiliários utilizados na elaboração do modelo, porém, com dados em um intervalo de tempo futuro (03/01/2022 à 31/08/2022) totalizando. A segunda foi realizada com um conjunto de 17 fundos imobiliários distintos

do conjunto de dados inicial, além dos intervalos de tempo de negociação variados (1175 dias em média).

No contexto da primeira análise, os retornos médios anualizados foram de 26,39%, 12,67%, 25,54%, 19,96% e 7,83%, respectivamente, para as operações com o modelo proposto, *Buy and Hold*, *Vortex*, EMA e RSI. O modelo não disponibilizou do pior resultado para quaisquer dos fundos.

Na maioria das ocasiões, o modelo alcançou o melhor desempenho, conforme Tabela 5, porém, em percentual menor quando comparado ao *dataset* utilizado inicialmente que dispunha de uma média de 1058 dias de negociação por fundo. Neste cenário o modelo obteve o melhor retorno em 48% das ocasiões.

Tabela 5 – Sumário do maior retorno anualizado de cada uma das estratégias para o conjunto inicial de fundos imobiliários analisados entre 03/01/2022 e 31/08/2022.

Estratégia	Fundos	Percentual
Modelo	57	48%
Vortex	36	30%
Médias Móveis Exponenciais	12	10%
Comprar e Manter	10	8%
Índice de Força Relativa	5	4%

Fonte: Autor.

Referente à segunda avaliação, obtiveram-se os retornos anualizados médios de 17,96%, -10,57%, 9,19%, 9,79% e -1,67%, e ponderados pelos dias de negociação de 18,75%, -14,86%, 6,54%, 9,71% e -3,46% para o modelo e as demais estratégias de negociação *Buy and Hold*, *Vortex*, EMA e RSI, respectivamente.

Na Tabela 6, é possível perceber que o modelo também atingiu o melhor retorno anualizado para a grande maioria dos fundos imobiliários que integravam este conjunto. Nesta circunstância o modelo conseguiu atingir o melhor retorno em 82% dos casos.

Tabela 6 – Sumário do maior retorno anualizado de cada uma das estratégias para o conjunto teste dos fundos imobiliários com números de dias de negociação variados.

Estratégia	Fundos	Percentual
Modelo	14	82%
Vortex	2	12%
Médias Móveis Exponenciais	1	6%

Fonte: Autor.

Na Tabela 8, constante do Apêndice C, estão detalhados os resultados dos 17 fundos imobiliários utilizados no teste.

As avaliações efetuadas anteriormente refletiram uma diferença entre os percentuais de fundos em que o modelo conseguiu atingir o melhor resultado. Esta diferença pode ser explicada em virtude dos intervalos de tempo de avaliação. Enquanto o primeiro foi de 120 dias (Tabela 5), o segundo foi, em média, de 1175 dias (Tabela 6).

Em vista disso, pode-se inferir que a exatidão das operações sugeridas pelo modelo se sobressaem ao longo do tempo, garantindo que grande parcela dos fundos tenham seu retorno otimizado. O resultado obtido para os conjuntos de teste corrobora a premissa de que o modelo tem caráter generalista. Em ambas avaliações o modelo foi superior às demais estratégias de comparação.

4.5 CONSIDERAÇÕES SOBRE O CAPÍTULO

Neste capítulo foram apresentados e discutidos os resultados obtidos com o modelo de rede resultante do presente trabalho. Foram descritos os retornos anualizados obtidos pelo modelo, bem como as comparações com os retornos obtidos a partir de estratégias de negociação baseadas em três indicadores técnicos. O caráter generalista do modelo também foi examinado a partir de duas avaliações. No Capítulo 5, a seguir, são expostas as considerações finais sobre o trabalho.

5 CONSIDERAÇÕES FINAIS

Neste trabalho foi utilizada uma rede convolucional profunda baseada no estudo de Sezer e Ozbayoglu (2018), que resultou em um modelo para auxiliar operações de compra e venda de cotas de fundos imobiliários negociadas em bolsa de valores.

A partir de uma série temporal de vários indicadores técnicos, foram geradas imagens rotuladas pelo indicador técnico ROC. Deste modo, pode-se inferir que o modelo conseguiu identificar o comportamento dos sinais de compra e venda deste indicador nas imagens geradas para treinamento.

Tomar como base o estudo de Sezer e Ozbayoglu (2018) e aplicá-lo a fundos imobiliários mostrou-se uma estratégia interessante, pois, conforme demonstrado nos resultados, o modelo superou a estratégia base *Buy and Hold*, além de outras estratégias de negociação baseadas em outros três indicadores técnicos, sendo estes *Vortex*, EMA e RSI.

O caráter de generalização do modelo também foi avaliado. Ao submetê-lo a *datasets* distintos de fundos imobiliários, bem como a intervalos de tempo não utilizados nas etapas de treinamento e validação, o modelo obteve a melhor performance em ambas as situações, conseguindo em grande parte das vezes atingir o melhor retorno anualizado do investimento.

Como sugestão de trabalhos futuros, tem-se:

- a. Avaliar a utilização de outros indicadores técnicos para formação das imagens, bem como de outras propostas de rotulagem de modo a classificá-las como pontos de compra, manutenção ou venda;
- b. Considerar o uso de indicadores técnicos a partir de uma abordagem discreta (binária), pois conforme observou Nabipour et al. (2020) existe melhora significativa no desempenho dos modelos quando estes usam dados binários em vez de contínuos;

- c. Incrementar o *dataset* com dados de outros ativos de renda variável (ex.: ações) de modo a promover maior caráter generalista ao modelo;
- d. Considerar na avaliação do retorno anualizado do investimento, além do ganho de capital proveniente das operações de compra e venda, também o retorno com o rendimento (*dividend yield*) dos fundos.

REFERÊNCIAS

- ANBIMA. *Fundos de Investimento Imobiliário*. Rio de Janeiro: Associação Brasileira das Entidades dos Mercados Financeiros e de Capitais, 2014. ISBN 978-85-86500-69-5.
- B3. *Perfil Pessoa Física: B3*. 2021. Disponível em: <https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/perfil-pessoas-fisicas/perfil-pessoa-fisica/>. Acesso em: 20 out. 2021.
- BARR, A. *The Handbook of artificial intelligence*. Stanford, Calif: HeurisTech Press, 1981. ISBN 0-86576-089-6.
- BOTES, E.; SIEPMAN, D. The vortex indicator. *Stocks & Commodities*, Seattle, v. 28:1, 2010.
- BRASIL. Lei nº 8.668, de 25 de junho de 1993. *Diário Oficial [da] República Federativa do Brasil*, Brasília, DF, 1993. ISSN 1677-7042. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/l8668.htm>.
- CAMPESATO, O. *Artificial intelligence, machine learning, and deep learning*. Dullers, Virginia: Mercury Learning and Information, 2020. ISBN 978-1-68392-467-8.
- CHOLLET, F. *Deep Learning with Python, Second Edition*. [S.l.]: Manning, 2021. ISBN 9781617296864.
- COLBY, R. W. *The encyclopedia of technical market indicators, second edition*. 2. ed. [S.l.]: McGraw-Hill Education, 2002.
- CVM, C. d. V. M. *Mercado de valores mobiliários brasileiro*. Rio de Janeiro: Comissão de Valores Mobiliários, 2019. ISBN 978-85-67896-04-5.
- DAS, H. *Applied intelligent decision making in machine learning*. Boca Raton: CRC Press, Taylor & Francis Group, 2021. ISBN 978-0-367-50336-9.
- EDGAR, T. *Research methods for cyber security*. Cambridge, MA: Syngress, 2017. ISBN 978-0-12-805349-2.
- FACELI, K. *Inteligência artificial : uma abordagem de aprendizado de máquina*. Rio de Janeiro: Grupo Gen - LTC, 2011. ISBN 978-85-216-1880-5.
- FINKLER, A. C. *Aprendizagem de máquina aplicada à previsão dos movimentos do Ibovespa*. Dissertação (Mestrado) — Universidade Federal do Paraná - Programa de Pós-Graduação em Matemática, Curitiba, 2017.
- GAURAV, R. S. Python as a key for data science. *International Journal of Computer Sciences and Engineering, IJCSE*, Indore, INDIA, v. 6, p. 325–328, 4 2018. ISSN 2347-2693. Disponível em: <https://www.ijcseonline.org/full_paper_view.php?paper_id=1894>.
- GERON, A. *Hands-on machine learning with scikit-learn, keras, and TensorFlow*. 2. ed. Sebastopol, CA: O'Reilly Media, 2019.

GOODFELLOW, I. *Deep learning*. Cambridge, Massachusetts: The MIT Press, 2016. ISBN 978-0262035613.

HENRIQUES, S. A. A. *Análise técnica a índices bolsistas*. Dissertação (Mestrado) — Instituto Politécnico de Lisboa - Instituto Superior de Contabilidade e Administração de Lisboa, Lisboa, 2016. Disponível em: <<http://hdl.handle.net/10400.21/8437>>.

HUANG, C.-S.; LIU, Y.-S. Machine learning on stock price movement forecast: The sample of the taiwan stock exchange. *International Journal of Economics and Financial Issues*, EconJournals, v. 9, n. 2, p. 189–201, fev. 2019. Disponível em: <<https://doi.org/10.32479/ijefi.7560>>.

HUANG, J.-Z.; HUANG, W.; NI, J. Predicting bitcoin returns using high-dimensional technical indicators. *The Journal of Finance and Data Science*, Elsevier BV, v. 5, n. 3, p. 140–155, set. 2019. Disponível em: <<https://doi.org/10.1016/j.jfds.2018.10.001>>.

JOHANSSON, R. *Numerical Python: Scientific Computing and Data Science Applications with Numpy, SciPy and Matplotlib*. Apress, 2018. ISBN 9781484242469. Disponível em: <<https://books.google.com.br/books?id=AzKBDwAAQBAJ>>.

KETKAR, N.; MOOLAYIL, J. *Deep Learning with Python*. Apress, 2021. Disponível em: <<https://doi.org/10.1007/978-1-4842-5364-9>>.

KIRKPATRICK, C. D.; DAHLQUIST, J. R. *Technical analysis*. 3. ed. Upper Saddle River, NJ: Financial Times Prentice Hall, 2015.

LEMOES, F. *Análise técnica dos mercados financeiros: um guia completo e definitivo dos métodos de negociação de ativos*. 1. ed. São Paulo: Saraiva Educação, 2015. ISBN 978-8502637825.

LIMA, M.; RUBIK, E.; MORAIS, R. *Introdução ao reconhecimento de imagens*. FACE - Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas (UNB), 2020. Disponível em: <<https://lamfo-unb.github.io/2020/12/05/Captcha-Break/>>.

MANASWI, N. *Deep learning with applications using Python: chatbots and face, object, and speech recognition with TensorFlow and Keras*. Berkeley, CA: Apress, 2018. ISBN 978-1-4842-3516-4.

MAO, S. et al. Opportunities and challenges of artificial intelligence for green manufacturing in the process industry. *Engineering*, v. 5, n. 6, p. 995–1002, 2019. ISSN 2095-8099. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2095809919300074>>.

MCKINNEY, W. *Python para análise de dados: Tratamento de dados com Pandas, NumPy e IPython*. Novatec Editora, 2019. ISBN 9788575227510. Disponível em: <<https://books.google.com.br/books?id=4hmWDwAAQBAJ>>.

MOLIN, S. *Hands-On Data Analysis with Pandas - Second Edition*. City: Packt Publishing, 2021. ISBN 9781800563452.

MOTA, R. R. *A evolução do mercado de fundos de investimento imobiliários no Brasil no período de 1994 a março/2013 e a utilização dessa alternativa de investimento para o investidor pessoa física no Brasil*. Dissertação (Mestrado) — Universidade de Brasília, Brasília, 2013. Disponível em: <<https://repositorio.unb.br/handle/10482/15165>>.

NABIPOUR, M. et al. Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. *IEEE Access*, v. 8, p. 150199–150212, 2020.

NETO, J.; SANTOS, J. de S.; MELLO, E. *O mercado de renda fixa no Brasil: conceitos, precificação e risco*. [S.l.]: Saint Paul, 2019. ISBN 9788580041507.

PAPER, D. *State-of-the-art deep learning models in TensorFlow*. 1. ed. Berlin, Germany: APress, 2021.

PATTERSON, J. *Deep learning: a practitioner's approach*. Sebastopol, CA: O'Reilly, 2017. ISBN 978-1-491-91425-0.

PERLIN, M. S. *Poupando e Investindo em Renda Fixa*. [S.l.]: Publicação independente, 2019. ISBN 978-85-92243-5-86.

POLAK, J.; COOK, D. A study on student performance, engagement, and experience with kaggle inclass data challenges. *Journal of Statistics and Data Science Education*, Taylor & Francis, v. 29, n. 1, p. 63–70, 2021. Disponível em: <<https://doi.org/10.1080/10691898.2021.1892554>>.

PRING, M. J. Rate of change. *Stocks & Commodities*, Seattle, v. 10:8, 1992.

ROCKEFELLER, B. *Technical analysis for dummies*. 4. ed. [S.l.]: Standards Information Network, 2019.

SANTOS, G. C. *Algoritmos de machine learning para previsão de ações da B3*. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, Programa de Pós-Graduação em Engenharia Elétrica, 2020. Disponível em: <<https://doi.org/10.14393/ufu.di.2020.640>>.

SANTURKAR, S. et al. How does batch normalization help optimization? arXiv, 2018. Disponível em: <<https://arxiv.org/abs/1805.11604>>.

SARKAR, D.; BALI, R.; SHARMA, T. *Practical Machine Learning with Python*. Apress, 2018. Disponível em: <<https://doi.org/10.1007/978-1-4842-3207-1>>.

SARKER, I. H. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, Springer Science and Business Media LLC, v. 2, n. 6, ago. 2021. Disponível em: <<https://doi.org/10.1007/s42979-021-00815-1>>.

SCOLESE, D. et al. Análise de estilo de fundos imobiliários no brasil. *Revista de Contabilidade e Organizações*, Universidade de São Paulo, Agencia USP de Gestão da Informação Acadêmica (AGUIA), v. 9, n. 23, p. 24, abr. 2015. Disponível em: <<https://doi.org/10.11606/rco.v9i23.83452>>.

SEWAK, M.; KARIM, R.; PUJARI, P. *Practical Convolutional Neural Networks*. Birmingham, England: Packt Publishing, 2018.

SEZER, O. B.; OZBAYOGLU, A. M. Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. *Applied Soft Computing*, Elsevier BV, v. 70, p. 525–538, set. 2018. Disponível em: <<https://doi.org/10.1016/j.asoc.2018.04.024>>.

SRIVASTAVA, S. et al. (Ed.). *Concepts and Real-Time Applications of Deep Learning*. Springer International Publishing, 2021. Disponível em: <<https://doi.org/10.1007/978-3-030-76167-7>>.

SUBASI, A. *Practical machine learning for data analysis using python*. London: Academic Press, 2020. ISBN 978-0-12-821379-7.

THACH, N. *Data science for financial econometrics*. Cham: Springer, 2021. ISBN 978-3-030-48853-6.

THAKKAR, A.; CHAUDHARI, K. A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions. *Expert Systems with Applications*, Elsevier BV, v. 177, p. 114800, set. 2021. Disponível em: <<https://doi.org/10.1016/j.eswa.2021.114800>>.

VIJAYKUMAR, K.; KUMA, B. S. A study on money flow indices using technical analysis with reference to selected securities broking company. *Mukt Shabd Journal*, Pune, IX: VIII, 2020.

XIANG, Z.; ZHANG, R.; SEELING, P. Machine learning for object detection. In: *Computing in Communication Networks*. Elsevier, 2020. p. 325–338. Disponível em: <<https://doi.org/10.1016/b978-0-12-820488-7.00034-7>>.

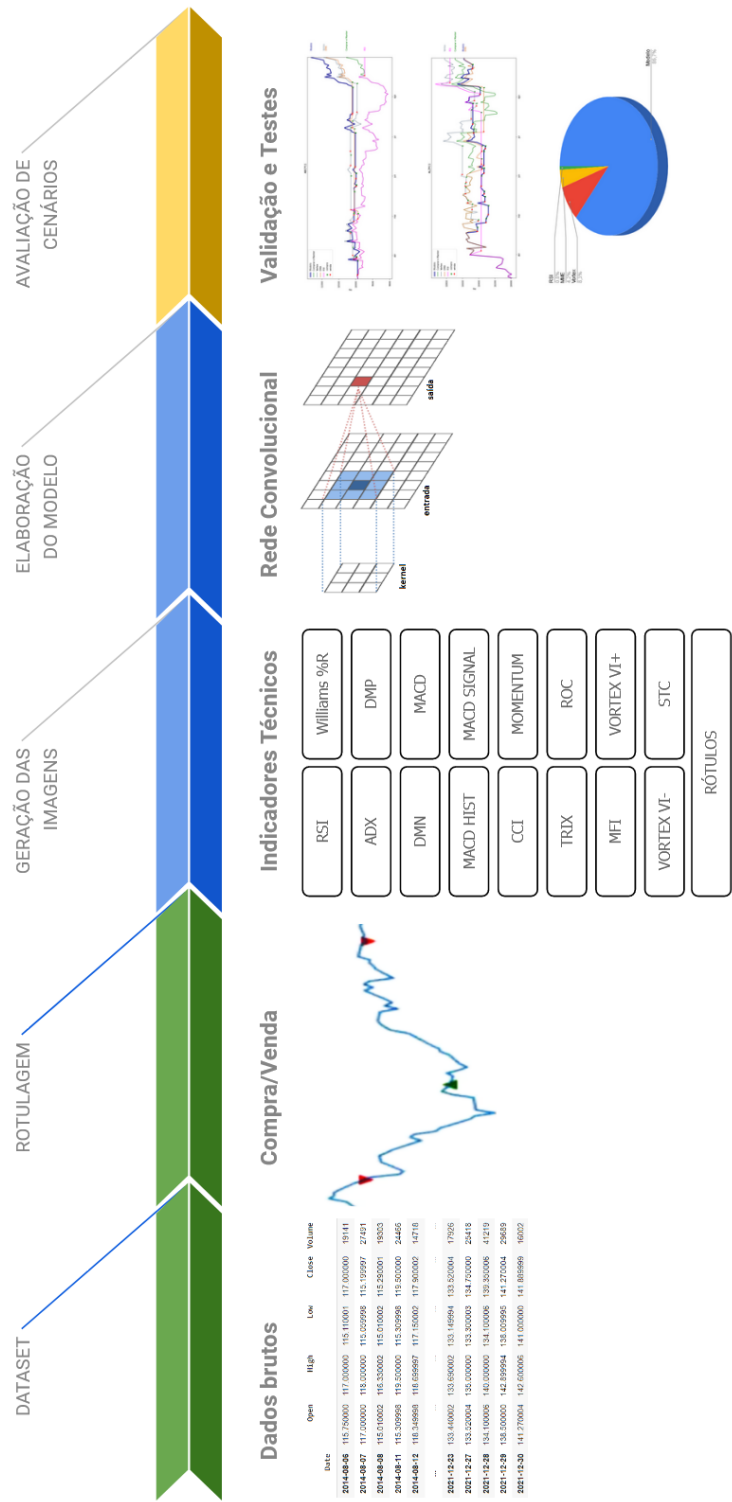
YOKOYAMA, K. Y. *Os Fundos de Investimento Imobiliário Listados em Bolsa e as Variáveis que Influenciam seus Retornos*. Dissertação (Mestrado) — Universidade Federal Do Espírito Santo, Programa de Pós-Graduação em Ciências Contábeis, Vitória, 2014. Disponível em: <<http://repositorio.ufes.br/handle/10/1307>>.

ZANADREA, V. *Avaliação de fundos de investimento imobiliário no Brasil*. Dissertação (Mestrado) — Universidade Federal de Santa Maria, Programa de Pós-Graduação em Engenharia de Produção, Santa Maria, 2018. Disponível em: <<http://repositorio.ufsm.br/handle/1/14888>>.

ZOCCA, V. *Python deep learning: next generation techniques to revolutionize computer vision, AI, speech and data analysis*. Birmingham: Packt Publishing, 2017. ISBN 978-1-78646-445-3.

APÊNDICE A – FLUXOGRAMA DA METODOLOGIA DE TRABALHO

Figura 16 – Fluxograma da metodologia de trabalho.



Fonte: Autor.

APÊNDICE B – ROTULAGEM DOS DADOS COM BASE NO INDICADOR ROC

```

1 # Entrada:  dataframe com dados diários com as cotações de
2 #           abertura, máxima, mínima, fechamento e volume
3 #           negociado.
4 # Saída:    array com rótulos diários correspondentes a
5 #           comprar(1), vender(2) ou manter(0).
6
7 def rotulagem_roc(ohlcv):
8     roc = ta.roc(ohlcv["close"])
9     length = sum(np.isnan(roc))
10    label = [np.NaN] * length
11    label.append(0)
12    for index, roc_value in enumerate(roc[length+1:]):
13        if roc_value > 0 and roc.iloc[length+index-1] < 0:
14            label.append(1)
15        elif roc_value < 0 and roc.iloc[length+index-1] > 0:
16            label.append(2)
17        else:
18            label.append(0)
19    return label

```

De forma semelhante outros rótulos podem ser obtidos a partir de indicadores distintos. Para tanto, altera-se a linha 8 para definir o novo indicador (biblioteca Pandas TA) e implementa-se a lógica de indicação de sinais de compra e venda na estrutura de repetição da linha 12.

APÊNDICE C – RETORNO ANUALIZADO DAS ESTRATÉGIAS

Tabela 7 – Comparação dos retornos anualizados do modelo e com as estratégias Comprar e Manter (B&H), Vortex, Médias Móveis Exponenciais (EMA) e Índice de Força Relativa (RSI).

Ticker	Data inicial	Dias	Retorno anualizado das estratégias (%)				
			Modelo	B&H	Vortex	EMA	RSI
ABCP11	24/11/2010	2752	30,42	4,56	19,71	13,38	6,04
ALZR11	15/03/2018	938	27,33	2,86	20,25	12,39	0,39
ARRI11	24/01/2020	479	38,80	-3,58	29,48	16,89	-10,15
BARI11	12/08/2020	342	12,04	1,49	5,30	7,24	0,76
BBFI11B	23/11/2010	2755	-12,45	-0,67	-11,29	8,91	-1,25
BBPO11	13/10/2014	1792	14,54	-2,01	12,31	11,23	-1,04
BBRC11	13/10/2014	1792	18,09	2,25	14,74	13,70	-1,59
BCIA11	14/12/2017	1001	43,44	-7,06	28,73	14,46	-9,38
BCRI11	14/12/2017	1001	11,83	0,52	9,34	7,16	2,52
BNFS11	15/10/2014	1790	21,34	4,41	15,81	11,28	1,42
BPFF11	13/10/2014	1792	21,23	-0,82	16,74	13,00	-0,16
BPML11	30/04/2020	414	10,16	-15,67	-3,29	7,86	15,63
BRCO11	30/04/2020	414	12,15	-6,08	15,91	5,56	-8,47
BRCR11	13/10/2014	1792	19,21	-6,45	15,84	9,90	1,69
BTCR11	02/04/2019	681	11,61	-8,20	6,06	5,27	5,24
BTLG11	13/10/2014	1753	28,66	1,56	22,75	15,45	-5,35
CBOP11	14/10/2014	1791	16,21	-3,00	14,22	8,30	1,41
CEOC11	13/10/2014	1792	20,90	0,74	17,73	10,64	-7,03
CPFF11	30/04/2020	414	3,06	-7,81	1,86	2,87	-2,96
CTXT11	16/10/2014	1789	9,25	-12,83	11,95	8,42	-12,30
CVBI11	30/04/2020	414	15,60	13,54	16,82	14,53	6,50
CXCE11B	13/10/2014	1789	25,45	-6,62	18,60	12,89	0,40

Continua na próxima página.

Tabela 7 – Continuação.

Ticker	Data inicial	Dias	Retorno anualizado das estratégias (%)				
			Modelo	B&H	Vortex	EMA	RSI
CXRI11	12/12/2014	1683	27,61	-2,37	23,68	18,70	4,28
EDFO11B	13/10/2014	1790	29,79	0,47	17,57	9,72	6,04
EURO11	09/11/2010	2764	18,10	0,62	13,19	15,95	3,83
FCFL11	13/10/2014	1792	24,41	8,50	18,11	16,75	3,84
FIGS11	13/10/2014	1792	22,50	-3,80	17,06	13,51	-1,41
FIIB11	14/10/2014	1791	17,43	5,60	13,07	12,94	-1,30
FIIP11B	13/10/2014	1792	22,57	2,34	13,59	12,54	-0,84
FIVN11	15/10/2015	1202	27,03	-13,03	36,99	76,76	69,08
FLMA11	13/10/2014	1792	23,34	3,62	24,95	18,80	-8,49
FPAB11	16/10/2014	1789	13,83	-6,36	7,44	7,48	-2,47
FVPQ11	09/06/2015	1182	26,55	-9,48	16,52	5,79	-0,55
GGRC11	14/12/2017	1001	16,94	-2,00	13,56	9,98	-0,21
GSFI11	10/07/2019	614	56,23	10,58	40,92	42,93	-5,22
GTWR11	30/08/2019	577	27,39	-6,04	22,03	16,18	-3,89
HABT11	11/10/2019	548	28,19	2,00	18,02	13,77	-3,56
HCTR11	17/09/2019	566	37,53	2,25	12,21	16,59	3,71
HFOF11	10/05/2018	901	17,15	-5,24	15,96	11,49	-9,29
HGBS11	21/05/2012	2385	19,09	-0,06	18,28	13,86	-3,50
HGCR11	13/10/2014	1792	15,68	1,91	12,88	9,07	2,77
HGFF11	04/11/2019	532	23,71	-12,48	22,06	12,44	-6,91
HGLG11	24/05/2011	2625	22,26	4,86	17,23	10,49	4,23
HGRE11	16/03/2010	2927	22,17	1,69	18,25	12,50	-1,95
HGRU11	04/10/2018	798	22,50	5,14	15,35	12,75	2,80
HLOG11	30/04/2020	414	2,97	-15,35	0,85	0,44	-13,27
HGPO11	12/08/2019	591	20,29	12,80	22,14	14,40	18,05
HSML11	23/10/2019	540	29,90	-10,22	22,38	17,42	-2,31
HUSC11	05/11/2018	765	49,00	5,19	0,35	24,13	20,42

Continua na próxima página.

Tabela 7 – Continuação.

Ticker	Data inicial	Dias	Retorno anualizado das estratégias (%)				
			Modelo	B&H	Vortex	EMA	RSI
IBFF11	30/04/2020	414	17,60	-13,95	6,07	3,31	-10,79
IRDM11	12/08/2020	342	29,37	-3,15	8,73	14,82	-10,89
JSRE11	13/10/2014	1792	14,24	-1,17	5,61	8,28	-1,26
KFOF11	08/05/2019	658	25,85	-10,58	14,19	11,76	-7,77
KINP11	12/12/2019	506	21,05	5,41	33,07	40,03	15,68
KNCR11	13/10/2014	1792	11,57	-0,13	4,82	5,30	-0,24
KNHY11	19/10/2018	789	11,62	1,51	6,14	5,72	6,56
KNIP11	14/12/2017	1001	10,22	0,05	6,45	3,81	4,23
KNRI11	13/10/2014	1792	20,01	2,71	13,02	12,56	2,14
KNRE11	14/12/2017	1001	17,32	-35,23	-0,91	3,11	-31,97
LGCP11	30/04/2020	414	14,41	-8,43	14,31	9,45	-2,15
LUGG11	30/04/2020	414	6,23	-8,22	2,78	2,34	-7,71
LVBI11	03/06/2020	391	18,32	-9,33	10,18	7,16	7,85
MALL11	26/02/2019	705	23,62	-3,11	24,61	12,78	-3,94
MBRF11	13/10/2014	1792	7,35	-3,72	-0,26	8,70	-4,87
MCCI11	30/04/2020	414	12,26	9,64	6,74	7,84	8,82
MFII11	13/10/2014	1755	12,86	2,78	19,54	10,42	1,54
MGFF11	12/06/2018	878	21,50	-7,88	15,80	9,44	-5,45
MXRF11	13/10/2014	1792	17,77	2,80	13,37	8,53	3,61
NEWL11	30/04/2020	414	4,83	-3,51	-2,08	3,52	-1,41
NVHO11	27/06/2018	868	89,91	12,55	51,65	63,90	61,19
ONEF11	13/10/2014	1792	25,58	6,52	16,14	13,73	3,96
OUFF11	20/09/2019	563	22,44	-11,78	11,58	9,28	6,17
OUJP11	14/12/2017	1001	20,43	-2,12	12,97	9,14	2,53
PATC11	11/06/2019	634	18,27	-15,88	13,77	12,51	-11,29
PLCR11	30/04/2020	414	23,51	9,17	14,16	14,34	12,12
PORD11	14/10/2014	1766	12,08	-0,79	7,20	7,17	3,64

Continua na próxima página.

Tabela 7 – Continuação.

Ticker	Data inicial	Dias	Retorno anualizado das estratégias (%)				
			Modelo	B&H	Vortex	EMA	RSI
PQDP11	13/10/2014	1792	-8,91	6,65	-3,82	14,78	-6,35
QAGR11	20/02/2020	460	-2,60	-32,53	-6,35	-6,53	-7,66
RBCO11	02/09/2020	327	11,33	-25,86	9,82	10,40	-25,86
RBED11	11/07/2019	613	18,99	-10,93	10,71	5,41	0,97
RBFF11	02/09/2020	327	16,34	-16,69	3,47	5,12	-11,46
RBRD11	13/10/2014	1792	21,37	-7,32	14,98	11,95	0,00
RBRF11	14/12/2017	1001	7,73	-3,51	11,64	9,43	3,16
RBRP11	01/11/2018	735	21,59	-2,37	17,24	1,30	8,80
RBRR11	13/07/2018	857	13,90	1,21	5,79	7,08	7,26
RBRY11	31/07/2019	584	10,94	-4,72	11,14	5,32	3,17
RBVA11	11/07/2019	613	22,47	-11,51	17,03	11,46	-1,08
RCRB11	30/04/2020	414	14,56	-10,08	1,92	0,34	-2,88
RECT11	10/07/2019	615	30,16	-11,67	11,40	8,53	3,22
RFOF11	11/05/2020	408	23,67	-1,38	16,01	12,69	0,52
RNGO11	13/10/2014	1792	13,75	-6,23	10,58	7,67	3,58
RVBI11	30/04/2020	414	21,04	-1,69	8,39	5,34	2,36
SADI11	12/08/2020	342	37,79	18,57	24,16	19,49	13,23
SARE11	30/04/2020	414	25,66	-15,12	3,90	7,93	-14,58
SCPF11	25/10/2016	1269	22,99	-5,72	17,16	9,76	-4,22
SDIL11	13/10/2014	1792	25,89	0,57	14,03	12,72	3,37
SHPH11	13/10/2014	1792	13,95	4,70	1,20	5,52	2,29
SPTW11	13/10/2014	1792	21,88	-6,23	11,30	15,52	1,17
TEPP11	14/07/2020	363	10,86	-12,43	-0,15	3,62	-7,65
TRXF11	30/04/2020	414	11,30	4,51	12,39	8,65	3,94
VCJR11	30/04/2020	414	13,03	2,35	0,20	7,16	7,39
VGIP11	01/06/2020	393	15,93	1,87	5,75	7,44	1,15
VGIR11	30/04/2020	414	15,72	8,36	13,12	8,65	9,61

Continua na próxima página.

Tabela 7 – Continuação.

Ticker	Data inicial	Dias	Retorno anualizado das estratégias (%)				
			Modelo	B&H	Vortex	EMA	RSI
VIFI11	18/08/2020	338	5,67	-16,58	-4,43	-0,67	-16,58
VILG11	28/05/2019	644	25,54	-1,85	21,18	16,70	6,24
VINO11	30/04/2020	414	18,61	7,51	13,09	11,98	3,06
VISC11	09/01/2018	984	19,50	-1,25	14,02	9,81	-4,04
VLOL11	13/10/2014	1792	23,40	6,40	14,72	10,24	7,07
VRTA11	13/10/2014	1792	17,15	-0,66	6,57	8,67	-1,41
VSHO11	01/04/2019	682	24,29	-8,82	22,11	10,41	4,53
VTLT11	10/08/2018	804	17,86	-1,20	6,64	8,58	8,42
VVPR11	30/04/2020	414	15,24	-0,47	7,22	5,47	3,57
XPCI11	30/04/2020	414	16,31	8,63	11,23	9,71	3,72
XPCM11	13/10/2014	1792	24,51	-17,74	5,10	12,27	2,53
XPHT11	30/04/2020	414	19,75	9,27	16,80	13,43	8,45
XPIN11	12/08/2020	342	11,91	-13,85	2,75	5,17	-13,85
XPLG11	12/08/2020	342	12,39	-13,87	6,86	4,47	-11,58
XPML11	12/08/2020	342	6,68	-6,74	5,46	2,53	-7,54
XPPR11	30/04/2020	414	36,01	-7,11	12,98	9,12	10,25
XPSF11	30/04/2020	414	18,06	-2,84	13,60	9,04	1,78

Fonte: Autor.

Tabela 8 – Comparação dos retornos anualizados do modelo e com as estratégias Comprar e Manter (B&H), Vortex, Médias Móveis Exponenciais (EMA) e Índice de Força Relativa (RSI) para fundos imobiliários não utilizados no treinamento do modelo.

Ticker	Data inicial	Dias	Retorno anualizado das estratégias (%)				
			Modelo	B&H	Vortex	EMA	RSI
BMLC11	13/10/2014	1669	19,31	0,75	14,15	10,99	2,66
CARE11	13/10/2014	1669	45,03	-36,57	-25,70	42,22	27,38
CRFF11	28/11/2019	684	21,33	-9,13	15,89	9,35	4,43
CXTL11	14/10/2014	1952	26,82	-7,15	8,70	4,19	-0,38
HOSI11	06/05/2020	579	16,15	-7,34	13,76	5,47	-1,35
HPDP11	30/04/2020	582	25,00	2,24	16,68	4,95	3,52
JPPA11	30/04/2020	582	12,08	3,67	9,98	4,77	8,01
NAVT11	22/12/2021	174	16,31	9,48	39,04	18,96	6,61
NEWU11	02/09/2020	495	22,15	-15,66	7,75	9,16	-15,66
PLRI11	14/10/2014	1935	18,55	-16,61	12,97	9,90	-12,51
PRSV11	13/10/2014	1960	16,26	-23,37	11,93	-1,04	-15,10
RBIR11	13/05/2020	574	13,28	-6,07	0,80	14,44	-0,66
RBVO11	13/10/2014	1960	12,90	-24,72	3,82	9,21	-9,61
RNDP11	13/10/2014	1960	-12,73	-13,39	-9,04	-3,13	-7,17
RECR11	19/08/2021	259	4,80	-8,74	4,63	2,69	-3,28
VOTS11	25/09/2018	974	16,02	-1,69	12,68	12,08	3,07
XTED11	13/10/2014	1960	32,07	-25,39	18,17	12,24	-18,33

Fonte: Autor.