

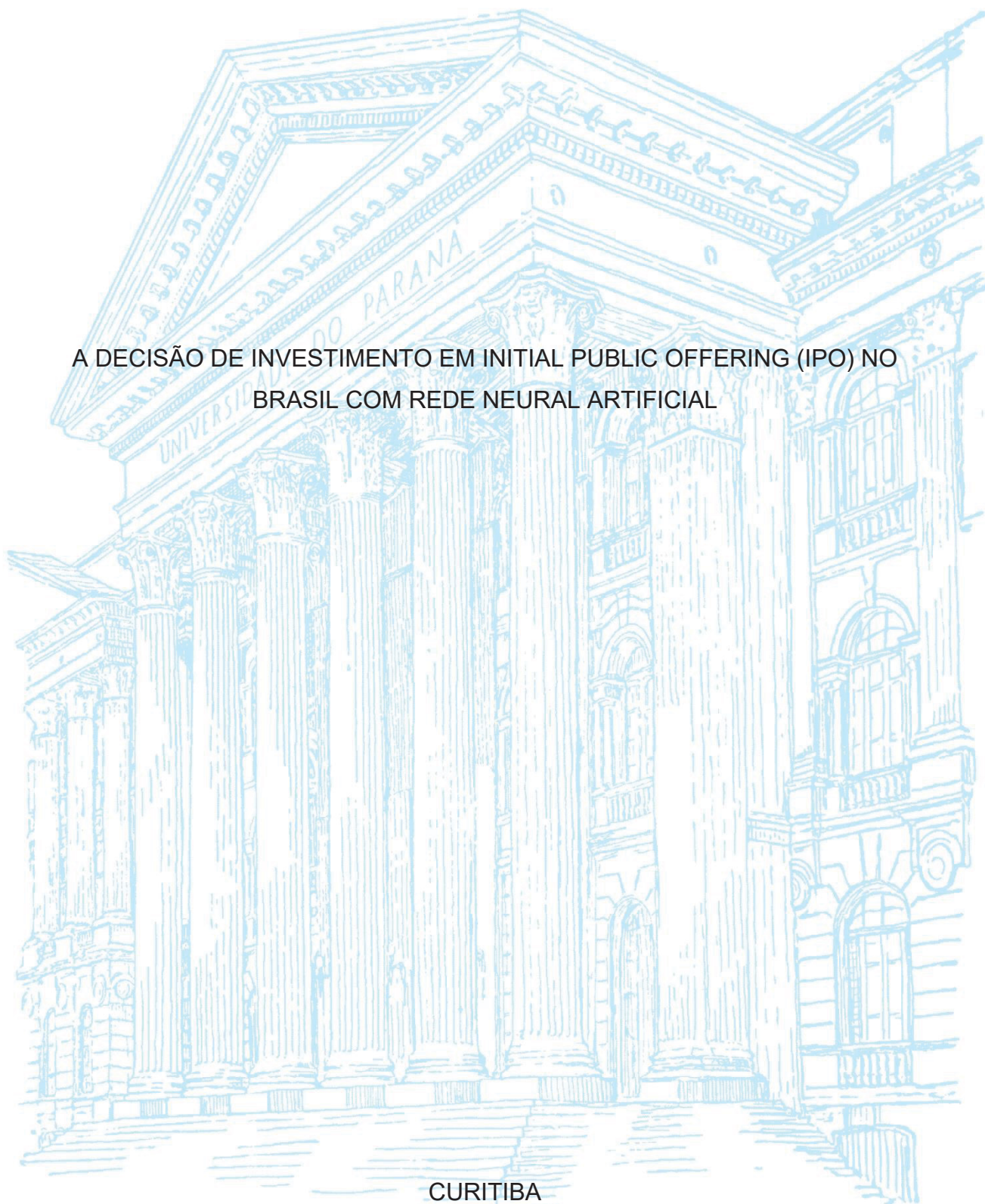
UNIVERSIDADE FEDERAL DO PARANÁ

JULIANY HELEN DAS GRAÇAS PINTO

A DECISÃO DE INVESTIMENTO EM INITIAL PUBLIC OFFERING (IPO) NO
BRASIL COM REDE NEURAL ARTIFICIAL

CURITIBA

2023



JULIANY HELEN DAS GRAÇAS PINTO

A DECISÃO DE INVESTIMENTO EM INITIAL PUBLIC OFFERING (IPO) NO
BRASIL COM REDE NEURAL ARTIFICIAL

Dissertação apresentada ao curso de Pós-Graduação em Gestão de Organizações, Liderança e Decisão (PPGOLD), Setor de Sociais Aplicadas, Universidade Federal do Paraná, como requisito à obtenção do título de Mestra em Gestão.

Orientador: Prof. Dr. José Roberto Frega.

CURITIBA

2023

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIAS SOCIAIS APLICADAS

Pinto, Juliany Helen das Graças

A decisão de investimento em *Initial Public Offering* (IPO) no Brasil com rede neural artificial / Juliany Helen das Graças Pinto. – Curitiba, 2023.

1 recurso on-line : PDF.

Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Ciências Sociais Aplicadas, Programa de Pós-Graduação em Gestão de Organizações, Liderança e Decisão. Orientador: Prof. Dr. José Roberto Frega.

1. Investimentos. 2. Oferta pública (Finanças). 3. Redes neurais (Computação). I. Frega, José Roberto. II. Universidade Federal do Paraná. Programa de Pós-Graduação em Gestão de Organizações, Liderança e Decisão. III. Título.

Bibliotecária: Maria Lidiane Herculano Graciosa CRB-9/2008

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação GESTÃO DE ORGANIZAÇÕES, LIDERANÇA E DECISÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **JULIANY HELEN DAS GRAÇAS PINTO** intitulada: **A decisão de investimento em Initial Public Offering (IPO) no Brasil com Rede Neural Artificial**, sob orientação do Prof. Dr. JOSÉ ROBERTO FREGA, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestra está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 07 de Julho de 2023.

Assinatura Eletrônica
10/07/2023 15:13:44.0
JOSÉ ROBERTO FREGA
Presidente da Banca Examinadora

Assinatura Eletrônica
10/07/2023 17:48:55.0
ADALTO ACIR ALTHAUS JUNIOR
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
10/07/2023 12:00:56.0
VAGNER ALVES ARANTES
Avaliador Externo (40001016)

Assinatura Eletrônica
10/07/2023 09:45:17.0
PEDRO GUILHERME RIBEIRO PICCOLI
Avaliador Externo (PONTIFÍCIA UNIVERSIDADE CATÓLICA DO
PARANÁ)

AGRADECIMENTOS

Nesta jornada desafiadora de aprendizado, sou profundamente grata a todos que generosamente contribuíram para a minha trajetória.

Aos meus pais, Edson e Mary, sou grata por me mostrarem as letras e incentivarem meus sonhos com amor. Ao meu irmão, Juninho, agradeço por ser meu amigo de vida, compartilhando momentos de alegria e enfrentando desafios familiares lado a lado. Meu sincero agradecimento também se estende ao meu companheiro, Levy, cujo apoio e encorajamento foram fundamentais para não desistir de projeto algum, desde o primeiro dia que nos conhecemos.

Sou grata também às instituições de ensino que fizeram parte da minha formação, e a todos os funcionários que as compõem, especialmente à UFPR e ao PPGOLD, que proporcionaram um ambiente de aprendizado enriquecedor e ferramentas que abriram portas inimagináveis. No mestrado, em especial, agradeço aos profs. Gustavo Loch, Adalto Althaus Junior e Rodrigo Soares pelas valiosas contribuições desde os estágios iniciais da pesquisa. Ao prof. José Frega, minha gratidão por me orientar e apoiar minhas escolhas desde o primeiro momento. E, também, ao prof. Pedro G. Ribeiro Piccoli, meu agradecimento pelos valiosos comentários, que não apenas auxiliaram na melhoria desta dissertação, mas também abriram oportunidades para trabalhos futuros.

Além disso, sou grata à prof.^a Jane Mendes, cujos ensinamentos sobre redação acadêmica foram de extrema importância para o desenvolvimento do trabalho. Também desejo agradecer ao prof. Guilherme Frederico, que gentilmente compartilhou seu conhecimento sobre como conduzir uma revisão sistemática da literatura, assim como ao prof. Hideo Araki, minha atual chefia imediata, pelas conversas inspiradoras e trocas de experiências no campo da Inteligência Artificial.

Aos meus colegas, agradeço por compreenderem minha dedicação aos estudos, em especial à Kelly e à Elaine, que demonstraram entusiasmo com minha carreira acadêmica desde a graduação em Administração, e ao Anderson, por dedicar-se à leitura deste trabalho desde o projeto piloto.

Por fim, quero expressar meu profundo agradecimento a todos que, de forma indireta, contribuíram financeiramente para o meu percurso acadêmico na rede pública. Cada um desempenhou um papel valioso e meu coração transborda de gratidão!

“Faça o seu melhor, na condição que você tem, enquanto você não tem condições melhores para fazer melhor ainda!” (Mário Sérgio Cortella).

RESUMO

Este estudo tem o **objetivo** geral de desenvolver um modelo supervisionado de Rede Neural Artificial (RNA) com capacidade preditiva para auxiliar investidores na construção da decisão de compra de ações de Ofertas Públicas Iniciais (IPOs) no curto prazo no Brasil, integrando a tecnologia de Inteligência Artificial (IA) ao problema de investimento em IPO. Para alcançar esse objetivo, o **método** adotado envolve a coleta e preparação dos dados socioeconômico-financeiros disponíveis no momento de publicação dos prospectos definitivos dos IPOs, seguido pela configuração dos elementos constitutivos da RNA, configuração dos hiper parâmetros, escolha apropriada da função de ativação e avaliação do desempenho dos modelos por meio de métricas como acurácia, precisão, recall e F1. Os **resultados** revelaram que combinações de variáveis, envolvendo indicadores econômicos, características das empresas emissoras e dados financeiros, apresentaram um desempenho satisfatório na classificação das ações de IPOs. A **contribuição** deste estudo reside na demonstração da viabilidade e utilidade das RNAs supervisionadas como ferramentas de suporte à tomada de decisões financeiras no contexto específico das ações de IPOs no Brasil. No entanto, ressalta-se que a utilização das RNAs deve ser considerada como um recurso informacional complementar, sendo necessário ponderar as informações fornecidas pelo modelo junto a outras análises e estratégias de investimento, levando em conta a imprevisibilidade do mercado financeiro. Conclui-se que a aplicação das RNAs supervisionadas no contexto das ações de IPOs no Brasil é viável e promissora, permitindo a identificação de ações com maior potencial de retorno financeiro.

Palavras-chave: Decisão de Investimento. Oferta Pública Inicial. IPO. Rede Neural Artificial. RNA.

ABSTRACT

This study **aims** to develop a supervised Artificial Neural Network (ANN) model with predictive capability to assist investors in making buying decisions for Initial Public Offerings (IPOs) in the short term in Brazil, integrating Artificial Intelligence (AI) technology into the IPO investment problem. To achieve this objective, the adopted **method** involves collecting and preparing the socioeconomic and financial data available in the definitive prospectuses of IPOs, followed by configuring the constitutive elements of the ANN, setting the hyperparameters, choosing an appropriate activation function, and evaluating the models' performance using metrics such as precision, recall, and F1 score. The **results** revealed that combinations of variables involving economic indicators, issuer-specific characteristics, and financial data demonstrated satisfactory performance in classifying IPO stocks. The **contribution** of this study lies in demonstrating the feasibility and usefulness of supervised ANNs as tools to support financial decision-making in the specific context of IPOs in Brazil. However, it is important to note that the use of ANNs should be considered as a complementary informational resource, and the information provided by the model should be carefully considered in conjunction with other analyses and investment strategies, taking into account the unpredictability of the financial market. In conclusion, the application of supervised ANNs in the context of IPOs in Brazil is viable and promising, enabling the identification of stocks with higher potential for financial returns.

Keywords: Investment Decision Making. Initial Public Offering. IPO. Artificial Neural Network. ANN.

LISTA DE FIGURAS

FIGURA 1 - Organização da dissertação	8
FIGURA 2 - Estrutura do Referencial Teórico a partir do processo decisório	9
FIGURA 3 - Tipo de receita no Ciclo de Vida das empresas.....	11
FIGURA 4 - Resumo de temas e teorias nas pesquisas de IPO	14
FIGURA 5 - Segmentos de listagem B3.....	16
FIGURA 6 - Etapas para abertura de capital no Brasil.....	18
FIGURA 7 - Resumo processo IPO.....	19
FIGURA 8 - Neurônio humano e artificial	25
FIGURA 9 - Funções de ativação de RNAs	26
FIGURA 10 - Múltiplas camadas e vários neurônios humano e artificial	27
FIGURA 11 - Blocos de aprendizagem supervisionada	28
FIGURA 12 - Exemplo <i>Backpropagation</i>	28
FIGURA 13 - Derivadas das funções de ativação	29
FIGURA 14 - Validação Cruzada com <i>k-folds</i>	30
FIGURA 15 - Ciclo de aprendizado	31
FIGURA 16 - Etapas e tarefas da pesquisa	38
FIGURA 17 - Estrutura da Revisão de Literatura	40
FIGURA 18 - <i>Outliers</i> em um <i>boxplot</i>	47
FIGURA 19 - <i>Grid Search</i> com <i>Stratified K-Fold</i>	53
FIGURA 20 - Filtro para equilíbrio da base de dados.....	57
FIGURA 21 - Matriz de confusão para combinação 14 do teste	92

LISTA DE GRÁFICOS

GRÁFICO 1 - Volume captado x quantidade de IPOS de ações.....	17
GRÁFICO 2 - Distribuição geográfica das sedes das empresas: cidades e estados	58
GRÁFICO 3 - Idade das empresas no momento do IPO	58
GRÁFICO 4 - Participação anual dos coordenadores líderes	60
GRÁFICO 5 - Distribuição dos setores por ano.....	61
GRÁFICO 6 - Variabilidade de preços iniciais dos IPOS ao longo dos anos	62
GRÁFICO 7 - Variabilidade do volume financeiro registrado	63
GRÁFICO 8 - Distribuição de IPOS em anos eleitorais e em meses	66
GRÁFICO 9 - Média de pontuação da Bolsa de Valores no Fechamento e Contagem de IPOS por ano	67
GRÁFICO 10 - Média da Meta Selic anual e da Expectativa da Inflação (próximos 12 meses) no dia do IPO e Contagem de IPOS por ano	68
GRÁFICO 11 - Índices de confiança e Contagem de IPOS por ano	69
GRÁFICO 12 - Correlação das variáveis selecionadas.....	73
GRÁFICO 13 – Intervalo de confiança da acurácia da Validação Cruzada por combinação.....	79
GRÁFICO 14 - Acurácia para cada <i>fold</i> da combinação 15	82
GRÁFICO 15 - Acurácia para cada <i>fold</i> da combinação 10	82
GRÁFICO 16 - Acurácia média nos dados de treinamento (e validação) e teste	84
GRÁFICO 17 - Dispersão entre acurácias da validação e teste.....	85
GRÁFICO 18 - Intervalo de confiança da acurácia do teste.....	86
GRÁFICO 19 - ROC AUC da combinação 14 no teste	91
GRÁFICO 20 - Erro entre valores preditos da combinação 14.....	93

LISTA DE QUADROS

QUADRO 1 - Etapas com participantes do processo de IPO	20
QUADRO 2 - Características do ser humano e do sistema inteligente	22
QUADRO 3 - Definições de Inteligência Artificial	23
QUADRO 4 - Síntese de trabalhos relacionados: RNA e IPO	36
QUADRO 5 - Definição da variável alvo e sua codificação	41
QUADRO 6 - Dados coletados	42
QUADRO 7 - Variáveis explicativas identificadas	45
QUADRO 8 - Índices financeiros	49
QUADRO 9 - Elementos da RNA	53
QUADRO 10 - Métricas de avaliação RNA classificatória	55
QUADRO 11 - Amostra em quantidade	57
QUADRO 12 - Diversidades de setores por ano	60
QUADRO 13 – IPOs com volume registrado em bilhões e em percentual	63
QUADRO 14 - Variáveis selecionadas para cada uma das 55 combinações	74
QUADRO 15 - Capital suposto para aplicação na simulação	94
QUADRO 16 - Síntese de acertos da RNA x Classificador Aleatório	100
QUADRO 17 – Possíveis variáveis identificadas para o contexto de IPOs brasileiro	117
QUADRO 18 - Rol de empresas e setores da amostra	119

LISTA DE TABELAS

TABELA 1 - Segmentos de Governança em quantidade e percentual.....	59
TABELA 2 - Coordenadores líderes em quantidade e percentual.....	59
TABELA 3 - Situação econômico-financeira por setores.....	64
TABELA 4 - Correlação das variáveis explicativas com a alvo	69
TABELA 5 - VIF	70
TABELA 6 - VIF das variáveis selecionadas	71
TABELA 7 - Características das combinações com melhores acurácias na validação	80
TABELA 8 - Características das combinações com melhores acurácias no teste ...	87
TABELA 9 - Métricas das melhores combinações do teste.....	89
TABELA 10 - Resumo de retornos: simulação com aplicações em todos os IPOs do conjunto de teste	95
TABELA 11 - Resumo de retornos: simulação com aplicações em IPOs aleatórios do conjunto de teste	96
TABELA 12 - Resumo de retornos: simulação com aplicações a partir da RNA em IPOs do conjunto de teste	98
TABELA 13 - Descrição detalhada de índices financeiros em cada setor.....	120

SUMÁRIO

1 INTRODUÇÃO	2
1.1 SITUAÇÃO PROBLEMA	3
1.2 QUESTÃO DE PESQUISA.....	3
1.3 OBJETIVOS	4
1.3.1 Objetivo geral	4
1.3.2 Objetivos específicos.....	4
1.4 JUSTIFICATIVA	5
1.5 ENQUADRAMENTO METODOLÓGICO.....	7
1.5.1 Abordagem e estratégia de pesquisa	7
1.5.2 Organização da dissertação.....	7
2 REFERENCIAL TEÓRICO.....	9
2.1 DECISÃO FINANCEIRA: <i>INITIAL PUBLIC OFFERING</i> (IPO).....	9
2.1.1 Descrição de pesquisas gerais sobre IPO.....	13
2.2 IPO NO BRASIL	14
2.2.1 Procedimentos para abertura de capital no Brasil.....	15
2.2.1.1 Segmentos de listagem	15
2.2.1.2 Captação de recursos.....	16
2.2.1.3 Processo.....	17
2.2.1.4 Participantes.....	19
2.3 INVESTIMENTO EM IPO	21
2.4 SISTEMAS INTELIGENTES	22
2.4.1 Inteligência Artificial (IA)	23
2.4.2 Rede Neural Artificial (RNA).....	24
2.4.2.1 RNAs e finanças.....	31
3 MATERIAIS E MÉTODO.....	37
3.1 ETAPAS DA PESQUISA.....	37
3.1.1 Etapa 1: Introdução	39
3.1.2 Etapa 2: Fundamentação	40
3.1.3 Etapa 3: Coleta e pré-processamento dos dados	42
3.1.4 Etapa 4: Exploração dos dados e definição da RNA.....	48
3.1.4.1 Elementos constitutivos da RNA	51
3.1.5 Etapa 5: Execução da RNA e verificações	54

4 ANÁLISE DOS DADOS E DISCUSSÃO DOS RESULTADOS	56
4.1 ANÁLISE DOS DADOS	56
4.1.1 Descrição	56
4.1.1.1 Empresas e momento de IPO	56
4.1.1.2 Conjuntura geral	66
4.1.2 Explorações das variáveis	69
4.2 DISCUSSÃO DOS RESULTADOS DA RNA	78
4.2.1 Avaliação geral do modelo: validação cruzada das combinações	78
4.2.2 Consistência e capacidade de generalização	83
4.2.2.1 Simulações de investimentos com a melhor combinação do teste	93
4.2.2.2 Viabilidade do modelo para tomada de decisões financeiras	101
5 CONSIDERAÇÕES FINAIS	103
5.1 CONTRIBUIÇÕES	103
5.1.1 Práticas	104
5.1.2 Teóricas	104
5.1.3 Sociais	105
5.2 LIMITAÇÕES E POSSÍVEIS PESQUISAS FUTURAS	105
5.3 SÍNTESE DO ESTUDO	106
REFERÊNCIAS	107
APÊNDICES	117

1 INTRODUÇÃO

Desde 1986, os métodos quantitativos e a Inteligência Artificial (IA) têm se estabelecido como ferramentas essenciais na tomada de decisão estratégica em finanças. Essas abordagens têm ganhado ainda mais destaque após 2016 (GOODELL *et al.*, 2021; WARIN; STOJKOV, 2021), desempenhando um papel transformador na pesquisa científica (RAY, 2023). A IA, por sua vez, não se restringe a um setor específico e tem sido amplamente adotada em diversos contextos, moldando a interação cotidiana com produtos e tecnologias. Um exemplo recente é o ChatGPT da OpenAI, uma ferramenta que tem recebido atenção crescente nos anos de 2022 e 2023 (GARCÍA-PEÑALVO, 2023).

No âmbito financeiro e empresarial, o avanço tecnológico¹ tem impulsionado o surgimento e crescimento das *fintechs*² (SUN; LI; WANG, 2023), empresas que utilizam tecnologia de forma inovadora para oferecer serviços financeiros. Essas empresas, visando expandir suas operações e captar recursos, recorrem à abertura de capital. Desde 2018, essa tendência tem renovado o interesse em pesquisas sobre Ofertas Públicas Iniciais de ações (*Initial Public Offering* - IPO) (JOSHIPURA; MATHUR; GWALANI, 2023). Nesse contexto, os estudos que envolvem IPOs podem também se beneficiar cada vez mais do uso de tecnologias avançadas, como a IA, para análise e avaliação dessas operações.

As Redes Neurais Artificiais (RNAs), como parte da IA, surgem como uma das ferramentas promissoras para lidar com os desafios envolvidos na análise de IPOs. Esses modelos de IA possuem a capacidade de processar dados complexos, imprecisos e volumosos (PAIVA, 2014), proporcionando *insights* para auxiliar na tomada de decisões estratégicas no mercado financeiro. Ao incorporar técnicas de econometria, probabilidades e programação dinâmica, as RNAs tornam-se uma ferramenta robusta para lidar com uma ampla gama de fontes de dados financeiros, como documentos contábeis, séries temporais, notícias e publicações em mídias sociais (GOODELL *et al.*, 2021). Além disso, esses modelos minimizam viés ideológico ou preferências associadas a determinadas escolas de pensamento, tornando-os mais imparciais na análise dos dados (WARIN; STOJKOV, 2021).

¹ Incluindo a IA e *Blockchain* (ALAASSAR; MENTION; AAS, 2023).

² Abreviatura de “finanças e tecnologia”.

Assim, esta pesquisa tem o foco de explorar o potencial das RNAs na tomada de decisão estratégica em finanças, concentrando-se na análise de IPOs na Bolsa de Valores do Brasil (B3).

1.1 SITUAÇÃO PROBLEMA

Compreender as informações disponíveis nos mercados em desenvolvimento pode ser desafiador devido a possíveis instabilidades no acesso aos dados (NGUYEN; PARSONS, 2021). Além disso, fatores comportamentais, como traços de irracionalidade, também podem influenciar o processo decisório (ROSS *et al.*, 2015).

No caso específico de IPOs, a assimetria de informações é um problema complexo. Isso ocorre principalmente devido à ausência de um histórico amplo do desempenho financeiro das empresas nos prospectos divulgados ao público em geral. Esses documentos são limitados a apenas três anos, conforme exigência da Instrução Normativa 400/03 da Comissão de Valores Mobiliários (CVM). A situação descrita pode resultar em dificuldades na emissão de relatórios precisos e na percepção de maior risco e volatilidade dos investimentos em IPO, conforme relatado pelo site Valor (LEWGOY, 2021), *The Cap Renda Variável* (VOGLINO, 2021), *Estadão* (DYNIEWICZ; PEREIRA, 2021) e pela coluna *e|investidor* (GUERATTO, 2021).

Diante dessa situação, a decisão de investir em IPO requer uma análise detalhada e abrangente de variáveis diversas, incluindo o cenário macroeconômico, político e financeiro. Essa dificuldade enfatiza a necessidade de encontrar métodos menos suscetíveis às emoções do mercado financeiro para abordar esse problema e permitir o desenvolvimento de soluções de pesquisa nessa área. No entanto, é importante ressaltar que ainda existem limitações em relação à capacidade de análise e à racionalidade, que não podem ser completamente eliminadas, mesmo com o uso de métodos essencialmente quantitativos.

1.2 QUESTÃO DE PESQUISA

A tomada de decisão em relação à compra de ações de IPO representa um desafio, especialmente para os investidores de pequeno porte. Com o propósito de fornecer apoio nesse processo, o presente estudo visa explorar o potencial das RNAs como uma ferramenta para embasar a decisão de compra de ações de IPO no

contexto do mercado brasileiro. Dessa forma, a pergunta de pesquisa que norteia este estudo é: **Como as RNAs supervisionadas podem ser aplicadas para a construção da decisão de compra no curto prazo de ações de IPO no Brasil?**

Para responder a essa questão, são consideradas a complexidade e as incertezas inerentes aos IPOs, utilizando-se variáveis disponíveis no momento da publicação do prospecto definitivo. É importante ressaltar que o objetivo da ferramenta não consiste em tomar a decisão em substituição ao investidor, mas sim em fornecer indícios que o auxiliem no processo decisório.

1.3 OBJETIVOS

1.3.1 Objetivo geral

Desenvolver um modelo supervisionado de RNA com capacidade preditiva para auxiliar investidores na construção da decisão de compra de ações de IPO no curto prazo no Brasil, integrando a tecnologia de IA ao problema de investimento³ em IPO.

1.3.2 Objetivos específicos

1. Verificar quais os dados socioeconômico-financeiros que estão disponíveis juntamente à divulgação do prospecto definitivo dos IPOs e identificar as fontes desses dados, bem como as formas de coletá-los;
2. Preparar os dados e conjuntos de entrada para o modelo, realizando os tratamentos necessários para a sua utilização;
3. Definir os elementos constitutivos da RNA;
4. Simular combinações de aportes a partir das predições do modelo de RNA e comparar com um cenário de decisão “aleatória”;

³ Nesta dissertação, seguindo trabalhos da área de gestão, o termo “investimento” é considerado como sinônimo de “aplicação financeira” e os “investidores” equivalentes a “agentes poupadores (aplicadores)”. Portanto, não será considerado o investimento como formação bruta de capital fixo somado às variações de estoque de produção, como ocorre nas Ciências Econômicas.

1.4 JUSTIFICATIVA

A compreensão das características das empresas que abrem capital é um tema relevante na literatura financeira (MCDONALD, 2022). No entanto, apesar dos esforços para a compreensão de aspectos posteriores à abertura de capital, a decisão de compra em um processo de IPO de empresas ainda está em desenvolvimento (CARBONE *et al.*, 2021).

Esse foco dos estudos *ex-post-facto* agrega à literatura de finanças sob diversas óticas, tais quais a partir das ofertas de ações já ocorridas e com dados disponíveis de cotações (por exemplo: DA SILVA; FAMÁ, 2011; DIODATO, 2020; TORRES, 2021), bem como avaliando a subvalorização (HUIBERS, 2020) ou sobrevalorização (DA SILVA; FAMÁ, 2011) dos preços do IPO, ou estimando os retornos esperados (possibilidades de ganhos de capital) (por exemplo: AHMAD-ZALUKI; BADRU, 2021; CARAMICO, 2012; CHEN, Y. S.; CHENG, 2012), ou comparando os sentimentos⁴ do investidor a situações como a Covid-19 depois do IPO (MAZUMDER; SAHA, 2021), ou analisando o antes e depois da pandemia nos desempenhos de IPOs já ocorridos (BAIG; CHEN, 2021).

No entanto, considerar as informações disponíveis *antes* do IPO pode oferecer uma perspectiva adicional sobre o potencial de crescimento da empresa. Embora a falta de informações completas possa gerar incertezas, a utilização de técnicas de processamento de dados não estruturados pode ajudar a diminuir essa limitação, permitindo a previsão do desempenho das ações. Assim, a aplicação de RNAs pode ser relevante para tomar decisões mais informadas e reduzir o risco envolvido em investimentos em IPOs.

Logo, esta investigação pode ser especialmente útil para as empresas que desejam abrir o capital, permitindo que elas ajustem sua situação e melhorem suas perspectivas antes de iniciar o processo junto à CVM. Além disso, considerando o aumento de cadastros de pessoas físicas na B3⁵, os investidores podem ter uma outra perspectiva sobre participar ou não da oferta de ações.

⁴ Entendido como expectativas irracionais de risco-retorno não baseadas nas informações disponíveis (SRIVASTAVA; ZHANG; EACHEMPATI, 2021).

⁵ Em janeiro de 2022, a B3 atingiu a marca de 5 milhões de contas de 4,2 milhões investidores pessoas físicas cadastradas (B3, 2022).

Ressalta-se que existe uma grande dificuldade em mensurar risco e retorno, da mesma maneira que fatores comportamentais e emocionais tendem a afetar a ponderação do que se deve considerar em um investimento. Assim, a aplicação de RNAs pode ajudar a modelar neurônios artificiais que analisem estritamente os dados disponíveis.

Dessa forma, é necessário investigar como as RNAs podem ser aplicadas para a decisão de compra em um processo de IPO. É importante ressaltar a abundância de estudos sobre RNAs⁶ aplicadas a diferentes objetos, assim como especificamente sobre IPOs⁷, mas a combinação⁸ desses temas ainda está em desenvolvimento no âmbito acadêmico.

Além disso, o presente estudo visa atender às chamadas de trabalhos específicos, como os propostos por Hearn; Filatotchev (2019) para controle do investimento em IPO, Yan *et al.* (2019) para estudos com novas metodologias e O'leary (2009) para temas de finanças sob RNAs. Esta pesquisa também está alinhada com a sugestão de Goodell *et al.*, (2021) no agrupamento de IA, Aprendizado de Máquina e *Fintech*, mais especificamente no que se refere à consultoria financeira sem intervenção humana significativa.

Dessa forma, a presente pesquisa pode contribuir para avanços na aplicação de RNAs na área de finanças, especialmente em economias em desenvolvimento (CHAHINE *et al.*, 2018), como o Brasil⁹, e nascentes (ZHANG *et al.*, 2022).

⁶ 1.871.475 trabalhos encontrados na plataforma *Web of Science*: "Artificial Neural network*" OR "Artificial Neural" OR "Neural" OR "RNA" OR "ANN" (tópico - até dezembro de 2022). Em termos mais gerais foram encontrados 3.826.127 estudos: "Neural Network" OR "Artificial Neural Network*" OR "AI" OR "artificial intelligence" OR "machine learning" OR "data mining" OR "data science" OR "algorithm" OR "analytics" OR "robot" OR "automation" OR "big data" OR "text mining" OR "natural language processing" OR "soft computing" (tópico - até dezembro de 2022).

⁷ 18.727 trabalhos encontrados na plataforma *Web of Science*: "IPO*" OR "initial public offering*" OR "capital opening*" OR "initial funding" OR "open capital" OR "Decision to go public" OR "Pric* of IPO*" OR "going public" (tópico - até dezembro de 2022)

⁸ 858 trabalhos encontrados na plataforma *Web of Science*: "Artificial Neural network*" OR "Artificial Neural" OR "Neural" OR "RNA" OR "ANN" AND "IPO*" OR "initial public offering*" OR "capital opening*" OR "initial funding" OR "open capital" OR "Decision to go public" OR "Pric* of IPO*" OR "going public" (tópico - até dezembro de 2022)

⁹ Mercado financeiro importante e ainda com poucas pesquisas (MARTINEZ *et al.*, 2009).

1.5 ENQUADRAMENTO METODOLÓGICO

A metodologia adotada nesta pesquisa é a de modelagem e simulação (TURRIONI; MELLO, 2012), utilizando modelos quantitativos baseados em variáveis específicas (BERTRAND; FRANSOO, 2002) e o *design* como limiar entre a teoria descritiva e a aplicação prática (VAN AKEN, 2004).

Por mais que não seja possível prever uma estratégia perfeita para compra de ações em um IPO, pois a realidade é muito dinâmica, a utilização de RNAs tem se destacado como método de modelagem disponível para capturar as não linearidades de um sistema sem intervenção humana (LAWRENCE, 1997) - e poder ser aplicado a inúmeras situações da área financeira (GOODELL *et al.*, 2021).

1.5.1 Abordagem e estratégia de pesquisa

A pesquisa utiliza uma abordagem quantitativa e dedutiva para estudar os processos de IPO e a capacidade preditiva das RNAs na decisão de compra, reconhecendo a constante mudança do mercado financeiro. É uma pesquisa aplicada e descritiva, que busca aprofundar o conhecimento do objeto de estudo e descobrir associações entre as variáveis.

1.5.2 Organização da dissertação

A pesquisa está organizada em cinco capítulos, conforme FIGURA 1, sendo esta **introdução** o primeiro, que contempla os objetivos, justificativa e enquadramento metodológico. O segundo é o **referencial teórico**, que define os conceitos e relaciona a pesquisa a outras já realizadas.

FIGURA 1 - Organização da dissertação

Nº	CAPÍTULO	CONTEÚDO ABORDADO
1	INTRODUÇÃO	Situação problema; Questão de pesquisa; Objetivos; Justificativa; Enquadramento metodológico.
2	REFERENCIAL TEÓRICO	Processo Decisório em Finanças; Decisão de Investimento em IPO; Sistemas Inteligentes e Redes Neurais Artificiais.
3	MATERIAIS E MÉTODO	Etapas e tarefas da pesquisa, incluindo desde a definição do objeto de estudo, coleta e o pré-processamento dos dados até a definição da arquitetura e topologia da Rede Neural Artificial.
4	ANÁLISE DOS DADOS E DISCUSSÃO DOS RESULTADOS	Aspectos descritivos da amostra incluída na pesquisa e discussão dos resultados obtidos da decisão de investimento em IPO a partir da Rede Neural Artificial.
5	CONSIDERAÇÕES FINAIS	Contribuições; Limitações; Pesquisas futuras.

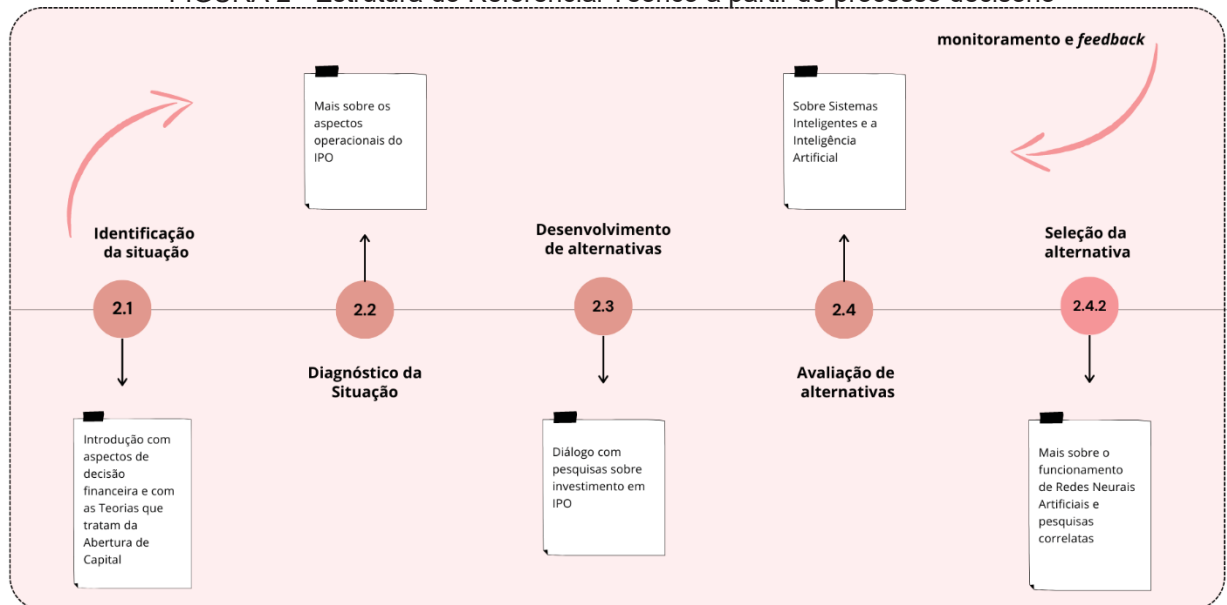
Fonte: A autora (2023)

O terceiro capítulo detalha os **materiais e métodos**, incluindo a coleta e pré-processamento dos dados, bem como os elementos da RNA. O quarto capítulo apresenta e discute os **resultados** obtidos. Já o quinto e último capítulo contém as **considerações finais**, incluindo as contribuições da pesquisa e limitações, bem como sugestões para trabalhos futuros.

2 REFERENCIAL TEÓRICO

Este capítulo segue a estrutura do processo decisório, que envolve a identificação e diagnóstico da situação, o desenvolvimento e avaliação de alternativas, a seleção e implementação da melhor opção, bem como o monitoramento contínuo. No referencial teórico, portanto, são explorados temas diversos ao contexto de processos de IPO, desde as teorias que permeiam a decisão de abertura de capital até uma explanação concisa sobre as possibilidades de aplicação de sistemas inteligentes em finanças, conforme representado na FIGURA 2:

FIGURA 2 - Estrutura do Referencial Teórico a partir do processo decisório



Fonte: A autora (2023) baseada em (SOBRAL; PECI, 2014)

2.1 DECISÃO FINANCEIRA: *INITIAL PUBLIC OFFERING* (IPO)

As decisões financeiras corporativas representam um desafio para os gestores, principalmente quando se trata de investimento, financiamento e distribuição de dividendos. Tais decisões, segundo Damodaran (2002), são cruciais para garantir a sustentabilidade do negócio (ROSS *et al.*, 2015), tendo em vista que geram riscos e retornos que precisam ser equilibrados e administrados (MUJIATUN; FERINA, 2021). No caso de investimento, é preciso priorizar a alocação de recursos em ativos que ofereçam retorno superior ao custo de capital. Já em relação ao financiamento, há diferentes teorias que buscam explicar a melhor forma de captação de recursos financeiros.

A *Pecking Order Theory*, por exemplo, sugere uma hierarquia na escolha das fontes de financiamento (MYERS; MAJLUF, 1984), enquanto a teoria do *Trade-off* procura equilibrar os benefícios e problemas existentes na escolha de se alavancar (ABDELJAWAD; MAT NOR, 2017). Por outro lado, a teoria de Modigliani; Miller (1963) busca explicar a estrutura ótima de capital para uma empresa, considerando que o valor da firma não é afetado pela maneira como ela é financiada em um mundo com mercados perfeitos e sem impostos.

Nesse contexto, as empresas precisam considerar diferentes opções de obtenção de recursos para alcançar uma combinação estratégica de capital (ASSAF NETO, 2000). É necessário avaliar os custos e benefícios envolvidos, assim como as características das diversas fontes de financiamento (NGA; LONG, 2021). Uma opção viável para concretizar estratégias de longo prazo, aumentar a captação de recursos e democratizar a participação dos pequenos acionistas é a abertura de capital, que pode ser realizada por meio de IPO no mercado de capitais (ROSS *et al.*, 2015).

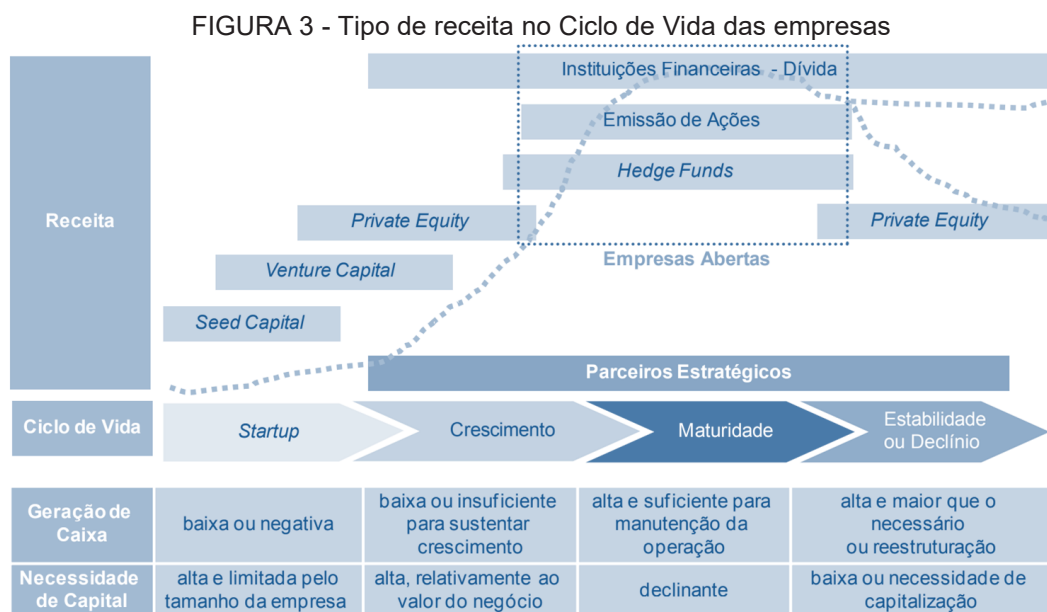
Essa decisão pode trazer diversas vantagens, tais como a redução do risco financeiro, a profissionalização das decisões administrativas, a diversificação das carteiras de investimento dos controladores, o aumento da liquidez e a diminuição do custo de capital (CALS DE OLIVEIRA; MARTELANC, 2014; MELLO; PARSONS, 2000; PAGANO; PANETTA; ZINGALES, 1998; BRAU; FAWCETT, 2006; OLIVEIRA *et al.*, 2013). Além disso, a abertura de capital pode reforçar o caixa, garantir perenidade, proporcionar visibilidade e governança, adequar a estrutura de capital e facilitar fusões e aquisições (B3, 2022).

No entanto, para que esses benefícios se concretizem, é necessário que todos os agentes envolvidos se interessem pelo processo, principalmente os investidores que devem estar dispostos a comprar ações do IPO (BOMFIM; SANTOS; JÚNIOR, 2006). A decisão de investimento é baseada nas informações fornecidas pelas empresas, mas existe uma assimetria de informações entre os agentes, especialmente os pequenos investidores, que possuem menos acesso aos dados necessários para tomada de decisão (COLAK; FU; HASAN, 2020). Além disso, a racionalidade limitada das pessoas também influencia nas decisões, resultando em contratos incompletos e impossibilidade de prever todas as contingências (JENSEN; MECKLING, 1976).

A assimetria de informações pode levar à seleção adversa e à subvalorização dos produtos financeiros de melhor qualidade (AKERLOF, 1978; NÓBREGA;

JURUBEBA, 2020; ROCK, 1986), mas a divulgação de IPOs ou a disponibilização dos dados nos prospectos pode minimizar esse problema (BYARD; DARROUGH; SUH, 2021; GARANINA; DUMAY, 2017). Por outro lado, a subvalorização pode ser vista como um ponto positivo para o acesso dos pequenos investidores ao mercado de capitais (SCHWARTZ, 2016).

Outras teorias, como a *Disclosure*, *Life Cycle*, *Market Timing* e *Changing Composition of IPO Issuers*, tentam explicar o IPO (RITTER; WELCH, 2002). A *Disclosure* incentiva a abertura de capital em momentos em que as avaliações públicas estão acima do esperado. A Teoria do *Life Cycle* afirma que a abertura de capital só ocorre quando a organização está mais madura. Já a *Changing Composition of IPO Issuers* apregoa que o tipo de estrutura de capital escolhida pelas empresas que abrem capital varia ao longo do tempo, e a decisão de abertura de capital deve considerar as condições de mercado (LOUGHRAN; RITTER, 1995) combinadas ao ciclo de vida mais propício (RITTER; WELCH, 2002), como pode ser visto no exemplo da FIGURA 3:



Fonte: Adaptado de BRASIL (2014)

À vista disso, a Teoria do *Market Timing* sugere que os gestores são capazes de identificar momentos em que a emissão de ações se torna menos onerosa do que outros tipos de financiamento (ELLIOTT; KOËTER-KANT; WARR, 2008). Essa teoria sugere que as empresas conduzem o IPO conforme as condições econômicas do

momento (STEFFEN, 2012), o que faz surgir uma oportunidade durante toda a fase de expansão e afluência no mercado (JAHANZEB *et al.*, 2013).

Para mensurar o *Market Timing* de um IPO, Rossi Jr.; Marotta (2010) afirmam que o volume em um determinado período é um indicador. Essa abordagem pressupõe que quanto mais IPOs ocorrem em um determinado momento, mais propício será para novas aberturas de capital, indicando um possível alinhamento entre as avaliações de mercado e as decisões de financiamento das empresas (ROSSI JR.; MAROTTA, 2010). Essa noção encontra respaldo em estudos que evidenciam como as empresas tendem a emitir ações quando seus valores de mercado estão altos em relação aos valores contábeis e históricos, e a recomprar ações quando os valores de mercado estão em baixa (BAKER; WURGLER, 2002).

A Teoria do *Market Timing* apresenta uma distinção em relação à *Pecking Order* (JAHANZEB *et al.*, 2013), a qual se encontra fundamentada na eficiência de mercado na forma semiforte. Enquanto a *Pecking Order* postula que as empresas seguem uma ordem de preferência no financiamento de projetos, priorizando recursos internos, dívida e, por fim, capital próprio, o *Market Timing* adota uma perspectiva mais flexível.

O cerne do *Market Timing* reside na ideia de que as empresas almejam maximizar o valor organizacional, procurando, ao mesmo tempo, minimizar os custos de capital (NGOC; TIEN; THU, 2021). Diante disso, os gestores são incentivados a ajustar suas decisões de financiamento em consonância com as flutuações temporárias nos custos de diferentes fontes de financiamento. Em outras palavras, eles adaptam suas estratégias financeiras com base nas condições cambiantes do mercado e buscam otimizar suas decisões de captação de recursos, contribuindo assim para a composição final da estrutura de capital da empresa (BAKER; WURGLER, 2002).

Diante disso, as empresas tendem a entrar no mercado de capitais quando as taxas de juros estão baixas (GRAHAM; HARVEY, 2001) ou, ainda, em situações de inflexão da política macroeconômica, ou seja, quando há juros reais maiores, câmbio sobrevalorizado, etc. (BOMFIM; SANTOS; JÚNIOR, 2006). Ademais, a relação entre as emissões primárias e as taxas de juros cria momentos oportunos para captação de recursos (MATSUO; EID JR, 2004).

2.1.1 Descrição de pesquisas gerais sobre IPO

Diversos estudos foram realizados para investigar as variáveis que explicam o desempenho operacional pós-IPO. Entre essas variáveis, destacam-se a qualidade da alta gestão (CHEMMANUR *et al.*, 2021), a atenção dos investidores em empreendedores apoiados ou não por Venture Capital (RAGOZZINO; BLEVINS, 2021), conexões governamentais para o subfaturamento (YANG; MA; DOTY, 2020), e outras.

Outro tema de interesse é a influência da abertura de capital de uma empresa em outros IPOs, principalmente dos concorrentes (AGHAMOLLA; THAKOR, 2021), e a relação entre a emissão de ações e a liquidez do mercado (HANSELAAR; STULZ; VAN DIJK, 2019). A mídia também é objeto de estudo, desde o papel da cobertura sobre o desempenho do IPO (GULDIKEN *et al.*, 2017) até o tom utilizado em momentos de crises financeiras no pré-IPO (BHARDWAJ; IMAM, 2019).

Além disso, o desempenho operacional pós-IPO tem sido investigado a partir da cultura nacional (GUPTA; VELIYATH; GEORGE, 2018; ZHOU; TANG; WU, 2022), da política local (COLAK *et al.*, 2021) e da concentração do mercado de capitais (BAE; BAILEY; KANG, 2021).

Diante disso, vê-se que o desempenho do IPO é influenciado por diversos fatores, desde a forma de captação do capital até impactos políticos e culturais (CHEMMANUR; HULL; KRISHNAN, 2021; RAGOZZINO; BLEVINS, 2021; YANG; MA; DOTY, 2020). A mídia também é considerada um fator de influência no sucesso ou fracasso de curto prazo da abertura de capital de empresas, além da importância de um mercado de capitais ativo para o desenvolvimento econômico e aumento do emprego (AGHAMOLLA; THAKOR, 2021; BAE; BAILEY; KANG, 2021; BHARDWAJ; IMAM, 2019; COLAK; FU; HASAN, 2022; GULDIKEN *et al.*, 2017; GUPTA; VELIYATH; GEORGE, 2018; HANSELAAR; STULZ; VAN DIJK, 2019; ZHOU; TANG; WU, 2022).

Ao relacionar as temáticas abordadas por esses estudos, percebe-se que a pesquisa sobre IPO e seu desempenho tem sido realizada considerando diversos constructos temáticos que se interconectam. Essas temáticas incluem sobrevivência e capital de risco, cultura e política, instituições, educação e conduta, conjuntura sustentável e mídia, como ilustrado na FIGURA 4. Essa abordagem multidisciplinar

permite uma compreensão mais holística e aprofundada do fenômeno do IPO e de seus impactos no mercado financeiro.

FIGURA 4 - Resumo de temas e teorias nas pesquisas de IPO¹⁰



Fonte: A autora (2023)

Para abordar esses assuntos, foram utilizadas diversas bases teóricas e perspectivas acadêmicas, destacando-se as teorias que tratam da sinalização e da assimetria de informações. A primeira se refere ao momento ideal para a abertura de capital (LOUGHRAN; RITTER, 1995), enquanto a segunda, em resumo, demonstra a dificuldade em se obter acesso irrestrito às informações relevantes para uma decisão (AKERLOF, 1978).

2.2 IPO NO BRASIL

Esta seção tem como objetivo abordar as particularidades dos procedimentos existentes para IPO no Brasil, visando oferecer uma análise mais ampla sobre o tema.

¹⁰ Pesquisas a partir de 2017.

2.2.1 Procedimentos para abertura de capital no Brasil

Para realizar a abertura de capital na B3, que consiste na primeira venda de ações no mercado de capitais brasileiro (CASAGRANDE NETO; SOUZA; ROSSI, 2000), a empresa precisa ser constituída como Sociedade Anônima e obter registro na CVM (ASSAF NETO, 2000). Adicionalmente, deve ter pelo menos três anos¹¹ de balanço auditado por auditor independente registrado na CVM, contar com diretor de relações institucionais estatutário e ter conselho de administração. Além disso, Identificar o segmento de listagem e definir a estratégia da oferta antes de protocolar o pedido são etapas fundamentais para garantir que a empresa esteja em conformidade com as exigências do mercado de capitais (B3, 2022).

2.2.1.1 Segmentos de listagem

A Governança Corporativa (GC) é um conjunto de mecanismos e princípios que tem como objetivo minimizar os problemas de agência e assimetria de informações no mundo empresarial, ao mesmo tempo em que busca solucionar os desafios de coordenação de ações coletivas em organizações (BECHT; BOLTON; RÖELL, 2003; MIRANDA; MELO; MARTINS, 2021). A GC é composta por diversos fatores, tais como transparência, auditoria, uniformidade das informações e equidade. Além disso, são contemplados a prestação de contas e a responsabilidade corporativa (BECHT; BOLTON; RÖELL, 2003)

Cada organização tem um tipo específico de governança que é mais eficaz, dependendo das atividades desempenhadas. Para garantir governanças diferenciadas, existem segmentos de listagem com regras específicas. Na B3 (2022b), conforme FIGURA 5, atualmente existem os segmentos de listagem Bovespa Mais, Bovespa Mais Nível 2, Novo Mercado, Nível 2 e Nível 1.

¹¹ Embora haja situações em que a idade da empresa é inferior a esses três anos e, por isso, torna-se impossível apresentar o histórico completo de demonstrações financeiras padronizadas.

FIGURA 5 - Segmentos de listagem B3



Fonte: A autora (2023) baseada em (B3, 2022)

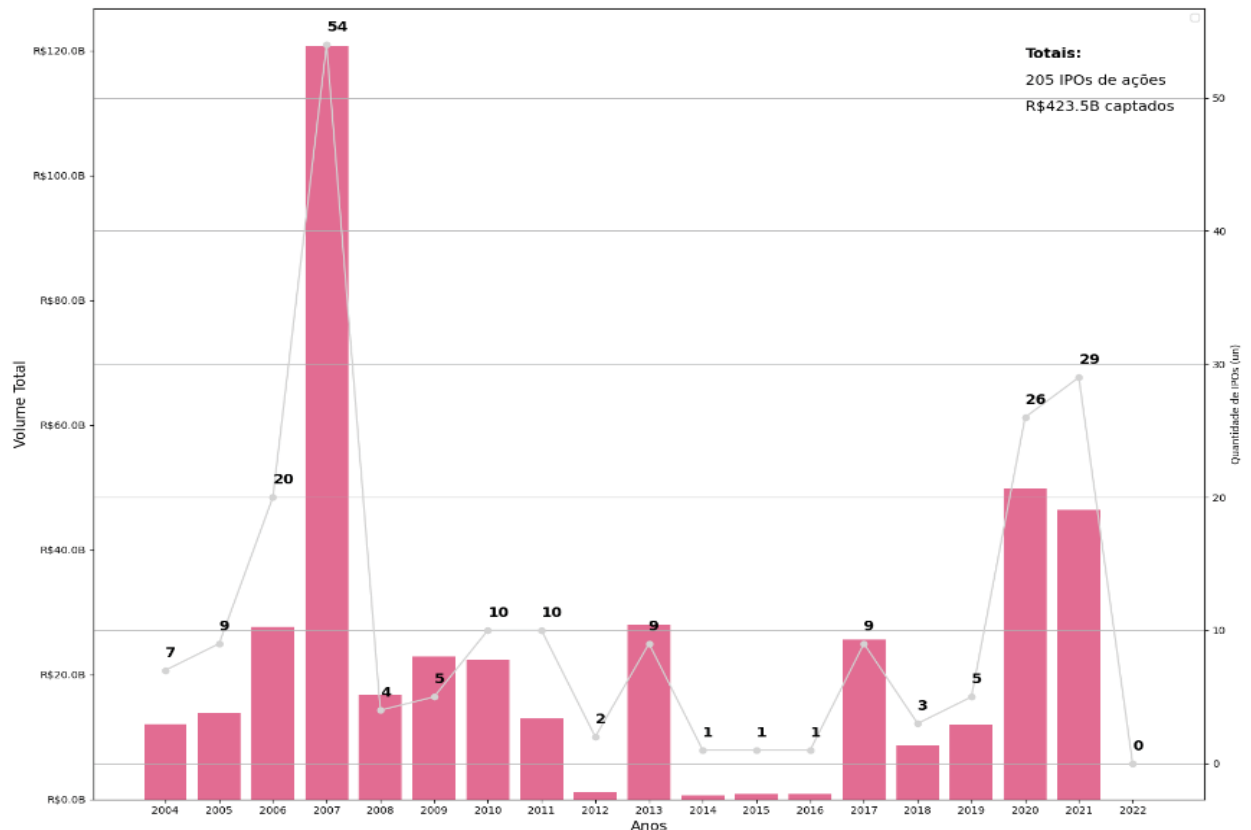
Os diferentes segmentos de listagem da B3 vão além¹² das prescrições da legislação das sociedades anônimas, englobando aspectos como composição do capital social, conselho de administração, divulgação de informações, auditoria e *compliance* (B3, 2022b). A adesão a esses segmentos pode trazer vantagens para as organizações, pois as boas práticas adotadas contribuem para reduzir os custos de capital e aumentar a liquidez da empresa (MARTINS; SILVA; NARDI, 2006), tornando o mercado de capitais um meio viável para captação de recursos.

2.2.1.2 Captação de recursos

A abertura de capital permite à empresa captar recursos junto ao público para investimentos em projetos e expansão, direcionando o dinheiro arrecadado com a venda das ações diretamente para o caixa da companhia (B3, 2022). O GRÁFICO 1 apresenta o volume de recursos captados e a quantidade de processos de IPO de ações finalizados na B3 entre 2004 e 2022:

¹² Há à disposição o Código das melhores práticas de governança corporativa, incluindo aspectos de natureza socioambiental, no site <https://conhecimento.ibgc.org.br/>

GRÁFICO 1 - Volume captado x quantidade de IPOS de ações



Fonte: A autora (2023) com dados da B3 (2022) sobre os IPOS de ações regidos pela ICVM 400 (CVM, 2003). Os valores monetários foram inflacionados pelo IPCA (2022 como ano-base) disponível no IPEADATA (2022).

Devido ao número limitado¹³ de IPOS no Brasil, cada um deles é considerado um momento importante e notável (LEAL, 2005).

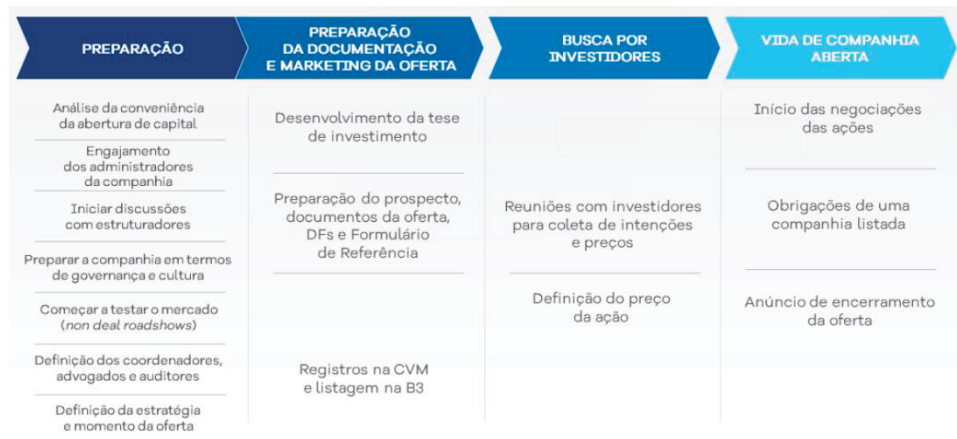
2.2.1.3 Processo

A partir da ICVM nº 480/09 (CVM, 2009), o processo eletrônico para abertura de capital perpassa quatro etapas¹⁴, segundo a B3 (2022), que podem variar o prazo total, porém giram em torno de 10 semanas: (i) preparação; (ii) preparação da documentação e *marketing* da oferta; (iii) busca por investidores e; (iv) vida da companhia aberta, conforme FIGURA 6:

¹³ Principalmente se comparar às pesquisas com dados dos Estados Unidos, por exemplo: 173.643 observações de trimestres (MCDONALD, 2022) ou 5.288 empresas analisadas (HARRIS, 2018)

¹⁴ Há o sistema eletrônico Empresas.NET “que deve ser utilizado pelas companhias registradas ou que pretendam se registrar na CVM para gerar e enviar os Formulários Cadastral, de Referência, ITR e DFP à CVM e à B3, no caso de companhias listadas nos mercados por ela administrados”(CVM, 2022).

FIGURA 6 - Etapas para abertura de capital no Brasil



Fonte: (B3, 2022)

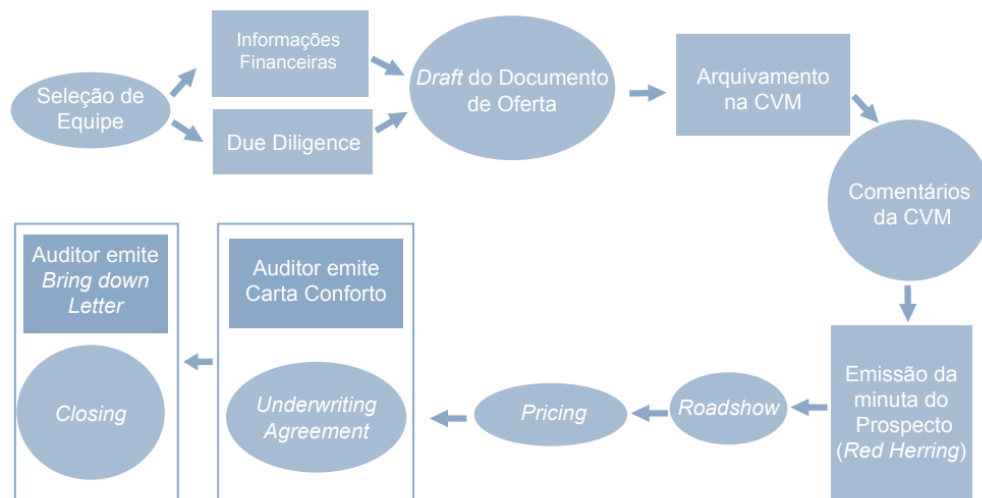
Após a análise dos documentos pela CVM e pela B3, inicia-se o prazo (20 dias úteis (du) + 40 du) para cumprir as exigências dessas instituições. Em seguida, é realizado o 2º protocolo, que abre um novo prazo (10 du + 10 du) para correção de vícios sanáveis. A partir daí, ocorre a precificação e assinatura dos contratos (d+42), possibilitando a publicação do anúncio de início e do prospecto definitivo - o principal documento da divulgação da oferta - com todas as informações relevantes (B3, 2022).

Ressalta-se, ainda, que “todas as informações precisam ser verdadeiras, completas, consistentes e que não induzam os investidores ao erro”, conforme Resolução nº80/2022 e suas alterações (CVM, 2022):

Art. 15. O emissor deve divulgar informações verdadeiras, completas, consistentes e que não induzam o investidor a erro. Art. 16. Todas as informações divulgadas pelo emissor devem ser escritas em linguagem simples, clara, objetiva e concisa. Art. 17. O emissor deve divulgar informações de forma abrangente, equitativa e simultânea para todo o mercado. Art. 18. As informações fornecidas pelo emissor devem ser úteis à avaliação dos valores mobiliários por ele emitidos (CVM, 2022)

O processo de IPO inclui também o período de *roadshow*, como mostrado na FIGURA 7. Nessa etapa, após a divulgação do prospecto preliminar, a empresa apresenta-se a possíveis investidores institucionais com o objetivo de coletar intenções de compra e preços. O *roadshow* é essencial para que a empresa possa ter uma ideia das demandas por suas ações e, assim, definir a estratégia de preço para o IPO.

FIGURA 7 - Resumo processo IPO



Fonte: Adaptado de BRASIL (2014)

Após o *roadshow*, é realizado o *bookbuilding*, que tem como principal objetivo a definição do preço de lançamento no IPO. Durante essa etapa, as reservas de ações são coletadas para a construção do livro de ofertas, que é um registro de todas as ordens de compra e venda de ações. Com base nessas informações, o preço das ações é definido e, posteriormente, divulgado ao mercado (B3, 2022). A liquidação das ações e o anúncio de encerramento finalizam o processo de abertura de capital.

Em resumo, a abertura de capital no Brasil é um processo complexo e regulamentado que envolve diversas etapas. É fundamental que a empresa forneça informações transparentes e úteis para os investidores, e que utilize o *roadshow* e o *bookbuilding* para definir uma estratégia de preço que reflita adequadamente o valor de suas ações no mercado.

2.2.1.4 Participantes

Durante a fase de pré-IPO, as empresas e seus consultores desempenham um papel crucial, sendo responsáveis por conduzir a *due diligence*, preparar as informações financeiras e prospectos exigidos para a oferta. As instituições financeiras, por sua vez, atuam como assessores financeiros, colaborando na definição de preço, estrutura da oferta e prospecção de investidores. A CVM, como entidade reguladora, é responsável por aprovar a oferta e fiscalizar a transparência das informações divulgadas, garantindo a proteção dos investidores e a integridade do mercado de capitais.

Na fase de *bookbuilding*, as instituições financeiras lideram o processo de prospecção de investidores e alocação de ações. Os coordenadores da oferta assumem o papel de intermediários entre as empresas e os investidores, enquanto os assessores legais da companhia e dos coordenadores prestam suporte na elaboração dos contratos e na revisão da documentação.

Os auditores independentes são responsáveis por revisar as demonstrações financeiras e garantir a exatidão e transparência das informações divulgadas. Os investidores, por fim, são os compradores das ações e, portanto, possuem papel crucial para o sucesso da oferta.

Em síntese, a realização de um IPO envolve diversos agentes, cada um com suas responsabilidades específicas que são fundamentais para o sucesso da operação, conforme QUADRO 1. A participação conjunta e coordenada desses agentes é essencial para garantir a efetividade do processo e atender às expectativas dos investidores e do mercado financeiro.

QUADRO 1 - Etapas com participantes do processo de IPO

Etapas	Partes Envolvidas				
	Empresas e Consultores	Instituições Financeiras	CVM	Bolsa de Valores	Investidores
1ª Preparações Preliminares					
2ª Contratação da Instituição Coordenadora					
3ª Definição do preço, volume e características					
4ª Autorização da Assembleia Geral					
5ª Processos de registro junto à CVM					
6ª Processo de registro em Bolsa					
7ª Formação do <i>pool</i> de distribuição					
8ª Marketing da operação					
9ª Anúncio de início da distribuição pública					
10ª Subscrição e Liquidação Financeira					
11ª Anúncio de encerramento da distribuição					

Fonte: Adaptado de (CASAGRANDE NETO; SOUZA; ROSSI, 2000)

Contudo, cabe ressaltar que a decisão de investimento em uma oferta pública de ações é um processo que envolve diversas variáveis, tais como o perfil de risco do investidor, as condições do mercado e a avaliação das perspectivas de valorização da companhia. Nesse sentido, na próxima seção será abordada brevemente a tomada de decisão de investimento em IPOs, com ênfase nas principais estratégias e fatores que podem ser considerados pelos investidores.

2.3 INVESTIMENTO EM IPO

Investir é comprometer recursos no presente com a expectativa de obter retornos no futuro (BODIE; KANE; MARCUS, 2014). Embora o risco da aplicação seja diretamente proporcional, aplicar em renda variável oferece a possibilidade de ganhos de capital maiores do que na renda fixa (ROSS et al., 2015).

No mercado de IPO a escolha é dificultada por causa da indisponibilidade do histórico operacional amplo das empresas (BADRU; AHMAD-ZALUKI; WAN-HUSSIN, 2017), o que resulta em menos conhecimento sobre as empresas por parte dos investidores (BHARDWAJ; IMAM, 2019). Portanto, a decisão de investimento variará conforme a quantidade de dados à disposição e o número de possibilidades identificadas (SCHIFFMAN; KANUK, 2007).

Existem alguns dados mais determinantes à decisão de aplicação em IPO, tais como a filosofia e o desempenho da empresa, a perspectiva do futuro e as notícias relacionadas (KHATRI, 2017), assim como a expectativa de valorização do capital e a segurança da aplicação (SRINIVAS; VENKATA RAO, 2017).

À vista disso, no estudo realizado por Leal (2005), utilizando dados de 1979 a 1992 e índices financeiros, foram identificadas informações indicativas de risco para IPOs brasileiros. Constatou-se que os investidores exigem um Retorno Sobre Ativos (*Return On Assets* - ROA) acima de 6% e o Índice de Endividamento inferior a 20%. Adicionalmente, foi verificado que o prazo de permanência com os papéis de IPO não deve ser superior a alguns dias, com média de 2 a 5 dias, para evitar que os retornos iniciais se transformem em perdas de longo prazo em um período de 3 anos.

As informações disponíveis nos prospectos podem fornecer oportunidades aos investidores e permitir uma análise fundamentalista com análise vertical e horizontal, diagrama de índices e indicadores econômico-financeiros (BODIE; KANE; MARCUS, 2014; ASSAF NETO, 2000). Além disso, a importância dos conselhos das corretoras no investimento em IPO foi constatada (KHATRI, 2017).

O investidor deve ser capaz de filtrar e analisar uma grande quantidade de informações e prever aspectos que possam afetar sua escolha (LISTYARTI; SURYANI, 2016). Essa complexidade é refletida na opinião de que o investimento em IPO é difícil, como indicado pela maioria dos 300 entrevistados por Khatri (2017) em Surat City (Índia).

Isso ocorre porque as decisões são tomadas com base em uma racionalidade limitada, e não perfeita (SIMON, 1979)¹⁵, uma vez que os seres humanos são agentes passionais, com diferentes graus de experiência e conhecimento (BYRNE, 2005), podendo cometer erros. Por isso, sistemas inteligentes, como ferramentas de apoio à decisão financeira, podem ser eficientes para auxiliar¹⁶.

2.4 SISTEMAS INTELIGENTES

O processamento de dados relacionados à decisão de investimento é complexo e os avanços computacionais, incluindo sistemas inteligentes, podem ajudar a solucionar esse problema. No QUADRO 2 pode-se visualizar uma comparação entre as características do ser humano e as do sistema:

QUADRO 2 - Características do ser humano e do sistema inteligente

SER HUMANO	SISTEMA INTELIGENTE
Utiliza o conhecimento sob a forma de regras ou heurísticas para resolver problemas em um domínio restrito	Processa o conhecimento sob a forma de regras e usa o raciocínio simbólico para resolver problemas
Compila o conhecimento	Separa conhecimento do processamento
Explica a linha de raciocínio e fornece os detalhes	Define a resolução de problemas e explica como uma determinada conclusão foi alcançada
Melhora a qualidade de resolução de problemas por meio de anos de aprendizado e treinamento prático. Esse processo é lento, ineficiente e caro	Melhora a qualidade de resolução de problemas por meio da adição de novas regras ou adapta as antigas na base de conhecimento. Quando os novos conhecimentos são adquiridos, as mudanças são fáceis

Fonte: Adaptado de (NEGNEVITSKY, 2011)

Logo, esses atributos levaram ao aprofundamento de estudos sobre a IA. Ela tem sido utilizada de maneira universal e com resultados significativos em várias áreas de conhecimento, incluindo finanças (GOODELL *et al.*, 2021; INJADAT *et al.*, 2021-; NEGNEVITSKY, 2011). A IA tem se tornado relevante para qualquer tarefa intelectual (RUSSELL; NORVIG, 2013). Em função disso, os próximos tópicos tratam da IA e dos seus desdobramentos na condução desta dissertação.

¹⁵ Embora exista uma vertente que defenda a existência de um agente racional nas suas escolhas, seja para montar uma carteira ótima (MARKOWITZ, 1952) ou para precificar o quanto que é justo pagar por um ativo (SHARPE, 1964).

¹⁶ Não substituir.

2.4.1 Inteligência Artificial (IA)

A inteligência pode ser entendida como o conjunto de processos internos que buscam resultados externos em ambientes complexos e dinâmicos (DE SPIEGELEIRE; MAAS; SWEIJS, 2017). Entretanto, as pessoas não são capazes de processar grandes volumes de dados, o que impulsionou o desenvolvimento da IA (RUSSELL; NORVIG, 2021-).

A IA é capaz de replicar habilidades mentais humanas e possibilita a tomada de decisões mais informadas por meio da análise resultante do processamento artificial (INJADAT *et al.*, 2021). Portanto, a IA é um ramo da ciência da computação que busca automatizar o comportamento inteligente (LUGER, 2004) e pode ser categorizada em quatro tipos, conforme QUADRO 3: (i) sistemas que pensam como seres humanos; (ii) sistemas que pensam racionalmente; (iii) sistemas que agem como seres humanos e (iv) sistemas que agem racionalmente (RUSSELL; NORVIG, 2013):

QUADRO 3 - Definições de Inteligência Artificial

Pensando como um humano	Pensando racionalmente
“O novo e interessante esforço para fazer os computadores pensarem (...) máquinas com mentes, no sentido total e literal.” (Haugeland, 1985) “[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado...” (Bellman, 1978)	“O estudo das faculdades mentais pelo uso de modelos computacionais.” (Charniak e McDermott, 1985) “O estudo das computações que tornam possível perceber, raciocinar e agir.” (Winston, 1992)
Agindo como seres humano	Agindo racionalmente
“A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas.” (Kurzweil, 1990) “O estudo de como os computadores podem fazer tarefas que hoje são melhor desempenhadas pelas pessoas.” (Rich and Knight, 1991)	“Inteligência Computacional é o estudo do projeto de agentes inteligentes.” (Poole et al., 1998) “Al... está relacionada a um desempenho inteligente de artefatos.” (Nilsson, 1998)

Fonte: (RUSSELL; NORVIG, 2013)

Neste trabalho, o objetivo é utilizar a IA para simular a capacidade humana de automatizar atividades, especialmente a tomada de decisões financeiras. Para isso, serão utilizadas técnicas de aprendizado de máquina, em particular as RNAs. Essas técnicas são relevantes para tomada de decisões em tempo real, com nível de confiança e precisão (SRIVASTAVA; ZHANG; EACHEMPATI, 2021). Nesse sentido, no próximo tópico, serão explicadas as RNAs.

2.4.2 Rede Neural Artificial (RNA)

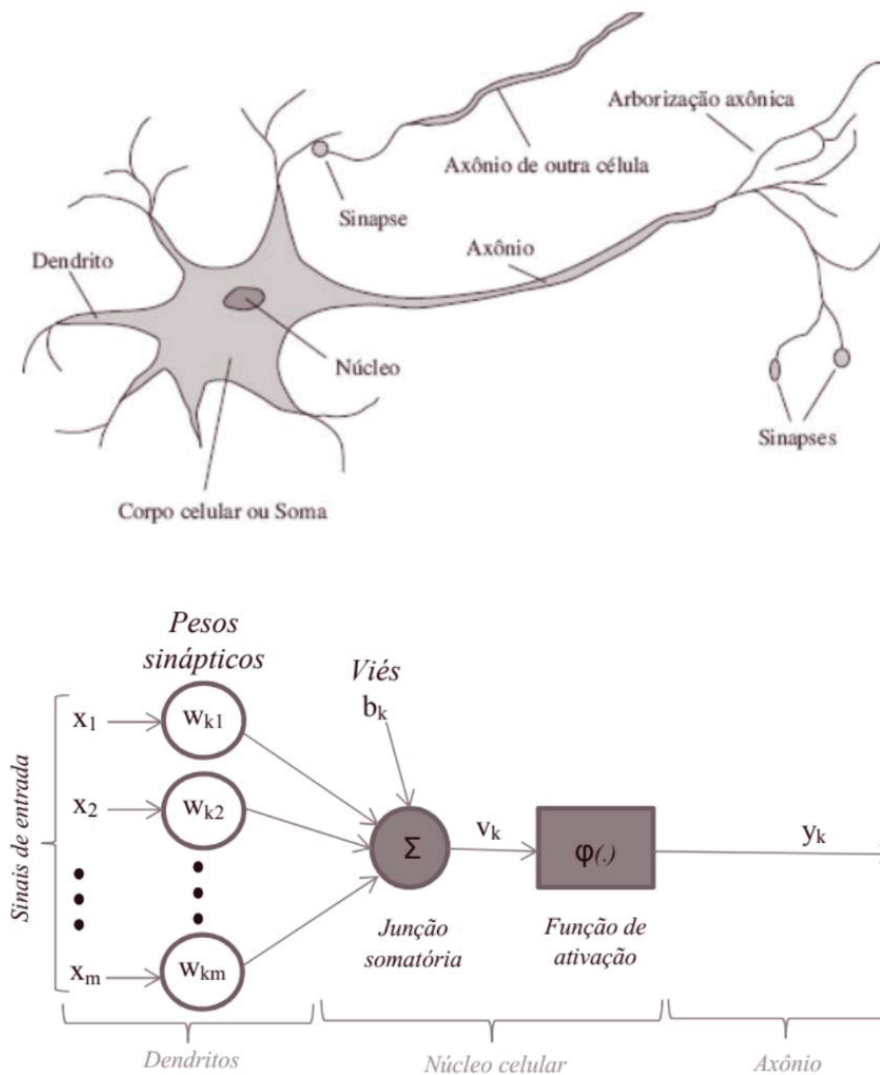
As RNAs (ou *Neural Networks*, ou redes neuronais, ou neuro computadores, ou redes connexionistas, ou processadores paralelamente distribuídos) compõem o rol de técnicas da IA na categoria de aprendizado de máquina.

Elas se destacam pelo fato de terem a habilidade de aprender, minimizando os erros (SWALES; YOON, 1992) e são, também, métodos de reconhecimento de padrões, classificação, previsão (KAASTRAA; BOYDB, 1996) e resumo de dados (WONG; SELVI, 1998). À vista disso, a definição de Rede Neural é: “um processador maciçamente paralelamente distribuído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso” (HAYKIN, 2001, p. 28).

A RNA se assemelha ao cérebro humano em dois aspectos: “o conhecimento é adquirido a partir do seu ambiente por meio de um processo de aprendizagem, além de existirem forças de conexão entre neurônios, conhecidas como pesos sinápticos, que servem para armazenar o conhecimento adquirido” (HAYKIN, 2001, p. 28).

O cérebro humano processa as informações por meio de neurônios e com a RNA não é diferente. Existem funções de ativação em cada neurônio humano, como se pode ver na primeira parte da FIGURA 8, sendo que os dendritos recebem sinais de outros neurônios; o corpo celular processa a informação; o axônio transmite a saída do neurônio e; as sinapses fazem conexões com outros neurônios. Assim como ocorre nos neurônios artificiais, apesar de a conexão entre estes neurônios passar a ocorrer por meio de um peso (sináptico) que se ajusta durante o processo de aprendizagem:

FIGURA 8 - Neurônio humano e artificial

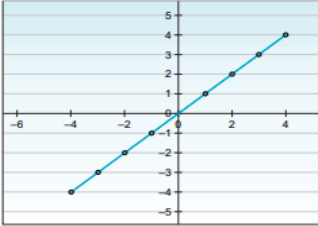
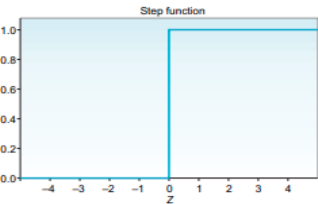
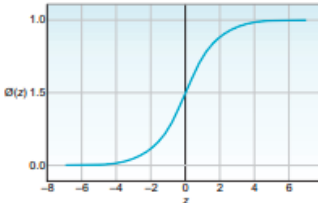
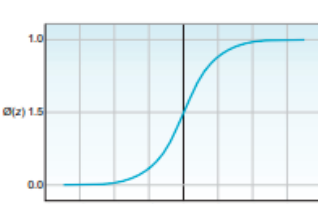
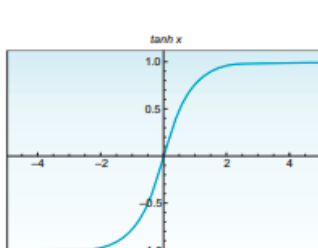
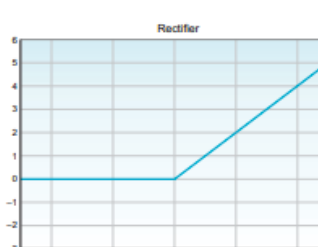
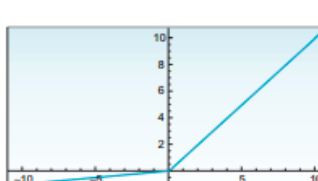


Fonte: Adaptado de (HAYKIN, 2001; MENDONÇA NETO, 2014)

Diante disso, os sinais de entrada são os dados do problema real, os pesos vão descrever as forças de conexão, a junção somatória resulta no nível de ativação e, por fim, a função de ativação calcula o resultado final (LUGER, 2004). Essa é uma simplificação do funcionamento de uma RNA, mas cabe destacar, todavia, que a arquitetura dela variará conforme o tipo de problema e definição do sub-ramo e algoritmo escolhido, ou seja, na imagem anterior (FIGURA 8) foi apresentado um exemplo mais similar à Rede *Feedforward*.

De qualquer maneira, a melhor função de ativação dependerá do tipo de problema que se busca resolver. Na FIGURA 9 estão algumas das funções possíveis:

FIGURA 9 - Funções de ativação de RNAs

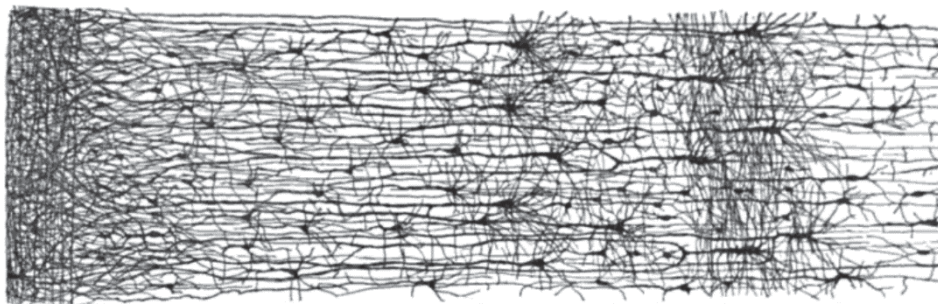
Activation function	Description	Plot	Equation
Linear transfer function (identity function)	The signal passes through it unchanged. It remains a linear function. Almost never used.		$f(x) = x$
Heaviside step function (binary classifier)	Produces a binary output of 0 or 1. Mainly used in binary classification to give a discrete value.		$\text{output} = \begin{cases} 0 & \text{if } w \cdot x + b \leq 0 \\ 1 & \text{if } w \cdot x + b > 0 \end{cases}$
Sigmoid/logistic function	Squishes all the values to a probability between 0 and 1, which reduces extreme values or outliers in the data. Usually used to classify two classes.		$\sigma(z) = \frac{1}{1 + e^{-z}}$
Softmax function	A generalization of the sigmoid function. Used to obtain classification probabilities when we have more than two classes.		$\sigma(x_j) = \frac{e^{x_j}}{\sum_i e^{x_i}}$
Hyperbolic tangent function (tanh)	Squishes all values to the range of -1 to 1. Tanh almost always works better than the sigmoid function in hidden layers.		$\begin{aligned} \tanh(x) &= \frac{\sinh(x)}{\cosh(x)} \\ &= \frac{e^x - e^{-x}}{e^x + e^{-x}} \end{aligned}$
Rectified linear unit (ReLU)	Activates a node only if the input is above zero. Always recommended for hidden layers. Better than tanh.		$f(x) = \max(0, x)$
Leaky ReLU	Instead of having the function be zero when $x < 0$, leaky ReLU introduces a small negative slope (around 0.01) when (x) is negative.		$f(x) = \max(0.01x, x)$

Fonte: (ELGENDY, 2020)

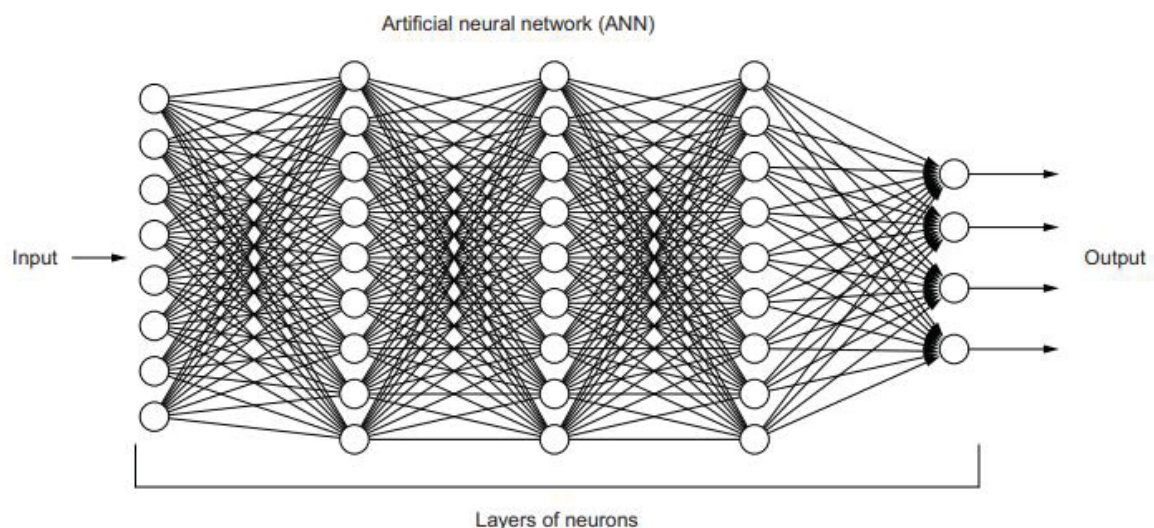
A função “Passo ou degrau” é utilizada em problemas de classificação binária, similar à “Sigmoidal”, mas esta serve para calcular a probabilidade de uma classe. Já a função “Leaky ReLU” tem bom desempenho em camadas ocultas, já a função “Softmax” é usada em problemas de classificação quando existem mais do que duas classes (ELGENDY, 2020).

A partir disso, deve-se reforçar a importância da arquitetura da Rede Neural, ou seja, a definição do número de camadas ocultas: única, em que existe apenas uma camada (*single-layer*) que recebe os dados de entrada e transforma em resultados; múltiplas (*multilayer*) que propiciam tanto o processamento entre os neurônios quanto entre camadas ocultas até a saída (MENDONÇA NETO, 2014), como mostrado na segunda parte da FIGURA 10:

FIGURA 10 - Múltiplas camadas e vários neurônios humano e artificial



Ramon y Cajal (public domain). Reproduced from https://en.wikipedia.org/wiki/Cerebral_cortex



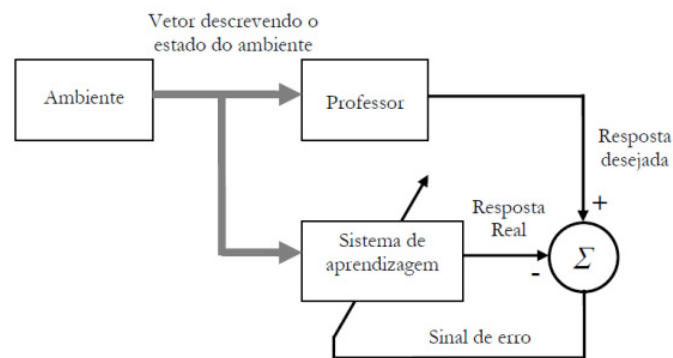
Fonte: (ELGENDY, 2020; GÉRON, 2022)

A camada de entrada distribui os padrões por toda a rede, a intermediária (oculta) processa as informações e atua como um conjunto de detectores associativos

e a de saída gera o resultado (FLETCHER; GOSS, 1993). Portanto, o resultado (saída) da rede dependerá do aprendizado dela, isto é, a RNA precisa ter a habilidade de aprender e de melhorar o desempenho por meio da aprendizagem. Em outros termos, pode-se criar um processo de estimulação para que as RNAs se modifiquem e aperfeiçoem os resultados (HAYKIN, 2001).

O processo de aprendizagem pode ser *supervisionado* ou *não supervisionado*. Este caso pressupõe apenas a entrada de dados, pois a saída não é conhecida (ZURADA, 1992); já aquele é muito utilizado em finanças (GOODELL *et al.*, 2021), em que um agente externo fornecerá tanto os dados de entrada, quanto os de saída para a RNA aprender, conforme processo da FIGURA 11:

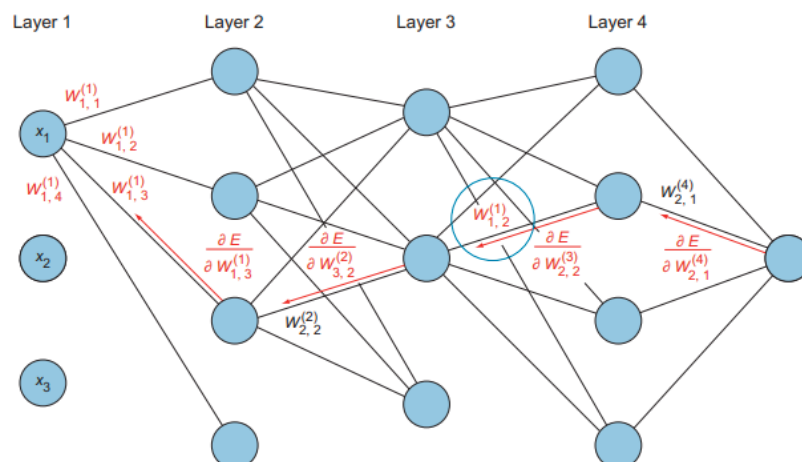
FIGURA 11 - Blocos de aprendizagem supervisionada



Fonte: (HAYKIN, 2001, p. 88)

Um algoritmo muito utilizado é o *backpropagation* que ajusta os erros da RNA para melhorar o aprendizado, conforme FIGURA 12:

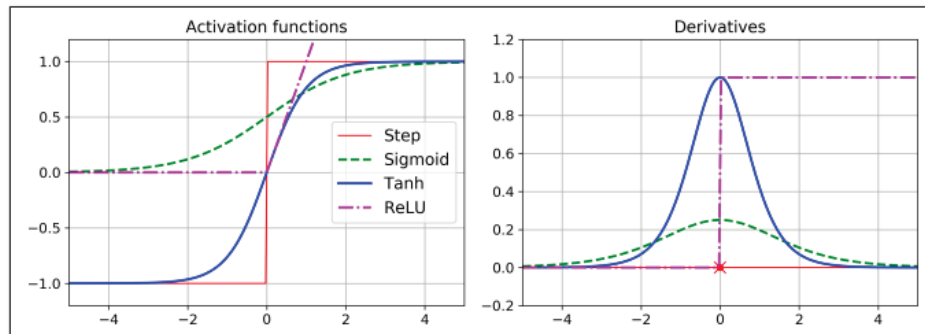
FIGURA 12 - Exemplo *Backpropagation*



Fonte: (ELGENDY, 2020)

Logo, o aprendizado por retro propagação significa propagar derivadas do erro, com respeito a cada erro específico, para que o aprendizado seja melhorado a cada iteração subsequente (ELGENDY, 2020; HAYKIN, 2001). As derivadas em formato gráfico de algumas das funções de ativação podem ser visualizadas na FIGURA 13:

FIGURA 13 - Derivadas das funções de ativação

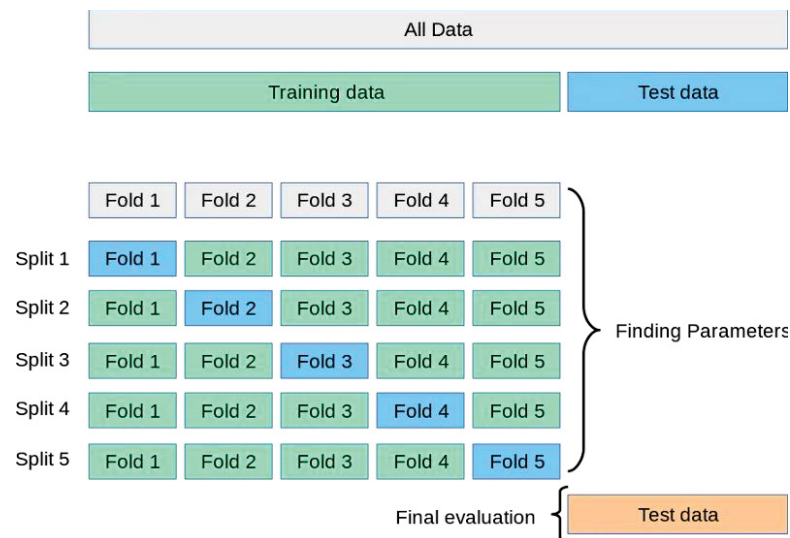


Fonte: (GÉRON, 2022)

Além disso, existem vários *solvers*, ou otimizadores, disponíveis para melhorar o modelo, tais como o método do Gradiente Descendente Estocástico (SGD) e suas variantes, como o *Momentum*, *Nesterov Accelerated Gradient* (NAG) e o *Adaptive Moment Estimation* (Adam), além dos métodos do gradiente conjugado (CG), *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) e o algoritmo L-BFGS (GOODFELLOW; BENGIO; COURVILLE, 2016). Esses *solvers* são responsáveis por ajustar os pesos das conexões entre os neurônios da rede, a fim de minimizar a função de custo ou perda associada ao problema de aprendizado.

Além disso, para obter um desempenho mais consistente, é comum a divisão dos dados em conjuntos de treinamento, validação e teste, permitindo que o modelo seja ajustado, validado e testado em conjuntos distintos (GOODFELLOW; BENGIO; COURVILLE, 2016).

Uma estratégia de validação utilizada é a *cruzada*, que divide o conjunto de dados em k partes iguais e treina o modelo k vezes, usando $k-1$ partes como treinamento e a parte restante como validação. Esse “ k ” geralmente é igual a 5 ou 10, porém não existe uma regra formal (KUHN; JOHNSON, 2013). De qualquer modo, os resultados são combinados para obter uma estimativa mais robusta do desempenho do modelo (GÉRON, 2022), como exemplificado na FIGURA 14:

FIGURA 14 - Validação Cruzada com *k-folds*

Fonte: (SCIKIT LEARN, [s. d.])

Tal abordagem de validação é uma das implementações disponíveis na linguagem de programação *Python* para aprendizado de máquina no pacote *scikit-learn*, que também oferece funções para divisão de dados em conjuntos de treinamento, validação e teste, bem como implementações de várias técnicas de validação cruzada, como a *stratified k fold cross validation* (PEDREGOSA *et al.*, 2011).

Ainda assim, a aplicação de RNAs em finanças é desafiadora devido à sensibilidade dos dados financeiros, sendo um dos principais desafios encontrar um equilíbrio entre o *overfitting* e a generalização dos resultados. O *overfitting* ocorre quando o modelo é ajustado demais aos dados de treinamento, reduzindo a capacidade de generalização e levando a um desempenho ruim em dados não vistos (GOODFELLOW; BENGIO; COURVILLE, 2016).

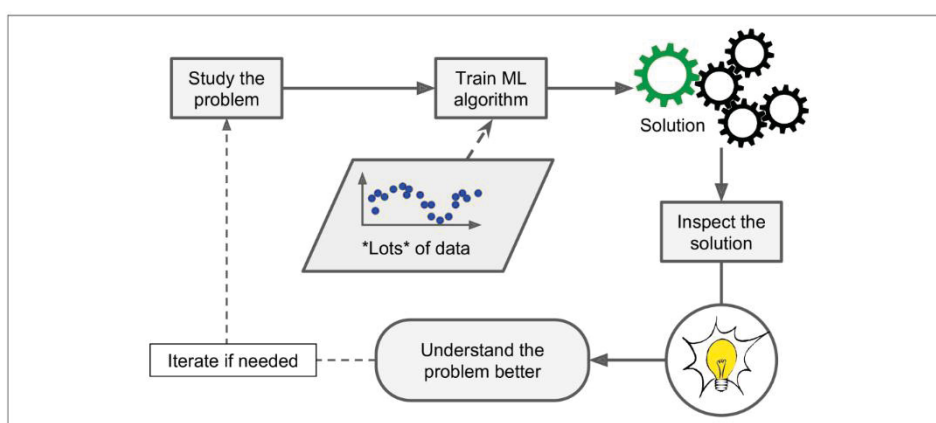
Existem duas técnicas principais de regularização utilizadas em modelos de aprendizado de máquina para mitigar o *overfitting*: a regularização L1 e L2 (HASTIE *et al.*, 2009). A regularização L1, ou Lasso, adiciona uma penalização proporcional à soma dos valores absolutos dos pesos do modelo, reduzindo assim os coeficientes dos pesos para zero e eliminando recursos irrelevantes do modelo. A técnica L2, ou Ridge, adiciona uma penalização proporcional à soma dos quadrados dos pesos do modelo, reduzindo todos os pesos do modelo, mas de forma menos acentuada do que a técnica L1. Ambas as formas de regularização ajudam a reduzir a complexidade do modelo e limitam os valores dos pesos dos recursos.

Além das técnicas de regularização e divisão dos dados, é crucial avaliar adequadamente as previsões dos modelos classificatórios para garantir um

desempenho confiável em dados não vistos. Existem várias métricas para avaliar a qualidade das previsões, como acurácia, precisão, *recall* e *F1 score* (ALPAYDIN, 2020). A *acurácia* mede a proporção de previsões corretas em relação ao total de previsões. A *precisão* mede a proporção de verdadeiros positivos (previsões positivas corretas) em relação ao total de previsões positivas. O *recall* mede a proporção de verdadeiros positivos em relação ao total de valores verdadeiros positivos. O *F1 score* é uma média harmônica da precisão e recall. Essas métricas podem ajudar a identificar pontos fortes e fracos dos modelos e auxiliar na tomada de decisões.

No contexto de aplicação de modelos em finanças, a literatura tem se dedicado a explorar as possibilidades do uso de RNAs no estudo do mercado financeiro (WARIN; STOJKOV, 2021), visando a melhorar a capacidade de previsão e tomada de decisão. Lembrando que, conforme FIGURA 15, o algoritmo compõe um ciclo de aprendizado do próprio ser humano.

FIGURA 15 - Ciclo de aprendizado



Fonte: (GÉRON, 2022)

Com isso, na próxima subseção, serão apresentadas algumas pesquisas que utilizaram RNAs em finanças, mais especificamente em IPOs.

2.4.2.1 RNAs e finanças

As RNAs são frequentemente utilizadas na área de negócios (TKÁČ; VERNER, 2016), sendo aplicáveis em diversos casos como previsão de desempenho de ações, negociação de títulos, previsão de preços ou índices de mercado, estimativa de saúde financeira, análise de demonstrações financeiras, previsão de taxas, gestão de carteiras, entre outros (WONG; SELVI, 1998). Diante dessa ampla gama de possibilidades, há diversas pesquisas que se concentram na aplicação de RNAs no

mercado financeiro, principalmente aquelas que envolvem previsão e classificação (JAIN; NAG, 1995).

Um estudo realizado por Chen, W. Sen; DU (2009) na Bolsa de Valores de Taiwan concluiu que a IA foi mais adequada do que as estatísticas tradicionais para prever potenciais dificuldades financeiras de organizações, utilizando 37 índices e 68 empresas. Já Chung; Tan; Holdsworth (2008) verificaram na Nova Zelândia que as RNAs foram satisfatórias na previsão de insolvência no setor financeiro, considerando diferenças no nível de lucratividade e liquidez das empresas falidas.

Martins; Mette; Macedo (2008) utilizaram uma RNA para prever os preços das ações da Petrobras (PETR4) com base em 1821 observações diárias no período de 2001 a 2008, obtendo alto grau de aderência em períodos curtos. Por outro lado, Oliveira; Nobre; Zárate (2013) concluíram que a janela temporal de t-3 meses apresentou melhor acerto de direção (*Prediction of Change in Direction* - POCID) de 93,62% e *Mean Absolute Percentage Error* (MAPE) de 5,45% para a PETR4 no período de 2000 a 2011.

De Pauli; Kleina; Bonat (2020) testaram as ações de códigos BOVA11, ITUB4, PETR4, VALE3, B3SA3 e BBDC4 negociadas no período de 2019 e 2020 na B3 por meio de cinco arquiteturas de RNAs (*Multilinear Regression*-MLR, Elman, Jordan, *Radial Basis Function*-RBF, MLP). Eles concluíram que todas as arquiteturas apresentaram bom desempenho na previsão de preços de um dia anterior, exceto a *Radial Basis Function*-RBF.

No contexto da previsão de retorno de IPOs, em 1995, foi desenvolvida uma nova abordagem para identificar a subprecificação de 552 IPOs ocorridos entre 1980 e 1990, conforme relatado por Jain; Nag (1995). Essa metodologia consistiu na construção de uma RNA com uma camada oculta, que utilizava dados econômicos primários e índices financeiros como entradas, tais como a reputação do banqueiro de investimento, o registro de receitas brutas levantadas no IPO, a extensão da propriedade retida pelos empreendedores originais na empresa pós-IPO, entre outros. A escolha da RNA foi devido às tentativas anteriores de modelagem paramétrica (como regressões de mínimos quadrados ordinários ou ponderados) que não conseguiram explicar de forma satisfatória a precificação dos IPOs (JAIN; NAG, 1995).

Em 2009, para prever os preços de fechamento do primeiro dia pós-IPO de pequenas e médias empresas na China durante os anos de 2006 e 2007, Xuemei; Yangbo (2009) modelaram uma RNA composta por nove fatores de influência. Esses

fatores foram: participação líquida, taxa de participação líquida, taxa líquida de ativos, taxa de lucro de ativos líquidos, lucro por ação, ativo líquido por ação, variação mensal e variação trimestral. A RNA apresentou convergência rápida e erro de previsão baixo tanto no treinamento quanto na simulação (XUEMEI; YANGBO, 2009).

Para analisar o desempenho operacional de longo prazo de novos empreendimentos, Jain; Nag (1998) também utilizaram RNAs. A análise foi baseada em 461 IPOs ocorridos entre 1976 e 1988. Para essa classificação, eles consideraram uma janela temporal de quatro anos, um ano antes até três anos depois da abertura de capital, e usaram o retorno operacional sobre ativos como medida de desempenho. As variáveis escolhidas foram: tamanho da oferta de IPO, reputação do banqueiro de investimento, preço de emissão, subcotação inicial, risco da empresa, fração da empresa retida pós-IPO pelos gerentes e empresários e despesas de capital deflacionadas pelos ativos no ano anterior ao IPO. A conclusão foi que as RNAs são uma alternativa viável para problemas completos de classificação, com um modelo de sete nós na camada oculta atingindo 68,65% de precisão (JAIN; NAG, 1998).

Já Robertson et al. (1998) propuseram o uso de RNAs para prever os retornos de curto prazo de empresas tecnológicas e não tecnológicas, baseando-se no processo de revisão de preços pré-IPO nos EUA de 1075 empresas entre 1989 e 1994. Os modelos de RNA foram mais eficazes do que os de regressão, apresentando uma melhoria de 5,72% na amostra de empresas não tecnológicas. O R^2 ajustado nas RNAs desenvolvidas foi de 29,1% para ofertas não tecnológicas e 16,5% para ofertas tecnológicas (ROBERTSON *et al.*, 1998).

Reber; Berry; Toms (2005) fizeram um modelo de RNA para prever preços de IPO com variáveis baseadas em agentes econômicos assimetricamente informados. O modelo atingiu um R^2 ajustado de 27,19% nas associações lineares. As variáveis usadas foram baseadas em sinais (desempenho passado da empresa, previsões, quantidade de capital levantado, patrimônio líquido retido, quantidade de ações na emissão geral, risco, características da equipe de gestão, reputação do investidor) e no mercado (retorno esperado, volatilidade, novas emissões) (REBER; BERRY; TOMS, 2005).

Foi utilizado um modelo de RNA para prever o índice de IPOs austríacos, com base em 500 observações de 15 de dezembro de 1992 a 19 de dezembro de 1994 para estimar os parâmetros do modelo, bem como 110 observações para testes (de 20 de dezembro de 1994 até 31 de maio de 1995). Haefke; Helmenstein (1996)

constatarem, portanto, que o uso de RNAs aumentou o desempenho em termos de variância-média e resultou em maior retorno em comparação com modelos lineares, devido às características específicas do mercado austríaco (HAEFKE; HELMENSTEIN, 1996).

Recentemente, Han; Kim (2021) desenvolveram um modelo de RNA *Multilayer Perceptron* (MLP) em conjunto com o Método de Avaliação Múltipla (MVM), a fim de fornecer critérios mais precisos de avaliação de IPOs com base em preços. Eles utilizaram dados de 216 IPOs listados na KOSDAQ entre 2014 e 2018 e concluíram que o erro percentual absoluto melhorou significativamente com a utilização do MLP (HAN; KIM, 2021). Além disso, Wang et al. (2018) propuseram o uso de modelos linguísticos, como a Rede Neural *Fuzzy*, para prever os preços dos IPOs. Esse método oferece suporte de decisão interpretável e confiável aos investidores (WANG et al., 2018).

Alguns trabalhos acadêmicos brasileiros podem ser citados em relação ao uso de RNAs para prever o desempenho de IPOs. Almeida (2013) utilizou uma Rede *Multilayer Perceptron* com duas camadas ocultas no *software* MATLAB® e obteve uma taxa de acerto de até 86,96% na predição da valorização de ações de IPO. Correia (2015) também o MATLAB®, mas com uma arquitetura de rede multicamadas alimentada diretamente, com funções sigmoidais e algoritmo *backpropagation*, obtendo uma taxa de acerto de até 78% na predição da valorização da ação no primeiro dia de pregão.

Alguns dos estudos relacionados a IPOs utilizando RNAs estão apresentados no QUADRO 4. Destaca-se, contudo, que até o momento, não foram encontradas pesquisas sobre IPOs brasileiros utilizando RNAs publicadas em revistas científicas¹⁷, a não ser de forma implícita como Rodrigues; Rodrigues; Rocha Rodrigues (2021). Eles utilizaram alguns modelos de classificação binária, incluindo a Rede Neural, para prever o sucesso de *startups* por meio da abertura de capital como *proxy* para recuperação do investimento. Esses modelos incluem *Regressão Logística*, *Decision Tree*, *Random Forest*, *Extreme Gradiente Boosting* e *Support Vector Machine*.

Em complemento aos estudos anteriores, esta pesquisa buscou explorar a combinação de dados socioeconômico-financeiros apontados pelo Referencial

¹⁷ Nos sites de busca: *Web of Science*, *Scielo*, *Scopus* e *Google Acadêmico*.

Teórico, coletados no momento da publicação do prospecto definitivo dos IPOs. Além disso, foram definidos e combinados elementos-chave da RNA, com o objetivo de aperfeiçoar as previsões para auxiliar na tomada de decisões de compra, ressaltando que o trabalho não se trata de uma prescrição, mas sim da apresentação de indícios de variáveis explicativas. Ademais, foram realizadas simulações de combinações de investimentos com base nas previsões do modelo de RNA em comparação a um cenário de decisão ingênua.

Por consequência, no próximo capítulo será apresentado o passo a passo da pesquisa e, posteriormente, o projeto de RNA desta dissertação será discutido com mais detalhes.

QUADRO 4 - Síntese de trabalhos relacionados: RNA e IPO

Geral				Rede Neural Artificial						Referência
Variável de interesse	Objetivo	Amostra	Período da amostra	Arquitetura	Algoritmo	Função	Treino	Teste	Iterações	
Preço	construir modelos de rede neural para precificar IPOs	552 IPOs	1980-1990	Feedforward 1 camada oculta	Backpropagation N pesos aleatórios	Generalized Delta-Rule	50% (276), de forma aleatória	50% (276), de forma aleatória	40.000	(JAIN; NAG, 1995)
Desempenho operacional de longo prazo	determinar a eficácia das redes neurais na previsão de qualquer medida de desempenho operacional	461 IPOs	1976-1988	Feedforward Regra Delta 1 camada oculta com 6 ou 7 nós	Backpropagation	função tangente hiperbólica	+50% (230), de forma aleatória	+50% (231), de forma aleatória	40.000	(JAIN; NAG, 1998)
Preço	Fornecer uma ferramenta de previsão melhorada (incluindo a Teoria de Assimetria de Informação)	172 IPOs	1992-1996	multicamada	Backpropagation & aprendizagem de gradiente conjugado	soma dos erros ao quadrado	86	43	x	(REBER; BERRY; TOMS, 2005)
Preço	Comparar as previsões em conjunto de RNAs e Método de Avaliação Múltipla	216 IPOs	2014-2018	1 camada oculta com 2 nós	x	seleção da a partir das funções sigmoide, tanh e ReLu	x	x	x	(HAN; KIM, 2021)

Fonte: A autora (2023)

3 MATERIAIS E MÉTODO

O presente capítulo descreve o processo de pesquisa adotado e apresenta em detalhes o método utilizado para a análise. A discussão inicia com uma revisão do projeto de pesquisa, abordando a introdução (capítulo 1) e a fundamentação teórica (capítulo 2), seguida pela descrição da forma de exploração e análise dos dados, que culminam na análise do próximo capítulo (capítulo 4).

3.1 ETAPAS DA PESQUISA

Para desenvolver um modelo de RNA para prever a decisão de compra em processos de IPO na B3, várias etapas foram realizadas. Essas etapas abrangeram tarefas desde a definição e aperfeiçoamento das bases conceituais da pesquisa até o alcance dos objetivos específicos. A FIGURA 16 ilustra a estratégia adotada.

Em resumo, após a imersão na literatura e aprofundamento em pesquisas correlatas (etapas 1 e 2), foi necessário verificar a disponibilidade dos dados socioeconômico-financeiros no momento de publicação do prospecto definitivo dos IPOs, identificar suas fontes e determinar os métodos de coleta. Além disso, os dados foram preparados e tratados para uso no modelo (etapa 3), e os elementos constituintes da RNA foram definidos (etapa 4). Por fim, foram realizadas simulações de combinações de investimentos com base nas previsões do modelo de RNA (etapa 5), comparando-as com um cenário de decisão ingênua.

FIGURA 16 - Etapas e tarefas da pesquisa

DECISÃO DE INVESTIMENTO COM REDE NEURAL ARTIFICIAL (RNA)		Tarefa 1	Tarefa 2	Tarefa 3	Tarefa 4	Tarefa 5
Etapa 1: Introdução	Etapa 2: Fundamentação	Imersão em trabalhos acadêmicos da área de finanças e inteligência artificial	Definição do objeto de pesquisa	Definição do problema de pesquisa relacionando as áreas de interesse	Definição da unidade de análise específica ao contexto brasileiro	Definição dos objetivos específicos a serem alcançados ao longo da pesquisa
		Planejamento e condução de uma busca sistematizada de trabalhos acadêmicos para compreender o objeto de pesquisa	Revisão de Literatura	Seleção de trabalhos a partir da base Scopus e Web Of Science	Utilização do VOSviewer e do Bibliometrix (R Studio) para compreender relações das pesquisas	Ajuste dos objetivos específicos e re-definição da variável alvo da RNA
		Sistematização dos dados necessários para condução da pesquisa	Busca pelos dados em fontes institucionais como as do Banco Central do Brasil, IBGE, Ipea, FGV, entre outras	Coleta e organização dos dados	Tratamento dos dados ausentes e inflacionamento das informações financeiras, assim como identificação dos preços outliers (e exclusão)	Codificação dos dados categóricos e padronização da base de dados (StandardScaler)
		Cálculo dos retornos obtidos nos primeiros dias após os IPOs e dos índices financeiros	Análises exploratórias e Descrição dos dados	Embaralhamento dos dados para balanceamento da base (50%/50% para as classes - retorno e não retorno)	Cálculo de correlação entre as variáveis independentes e a alvo, assim como do Fator de inflação da Variância entre as variáveis definidas como explicativas	Divisão do conjunto de dados em 50 combinações aleatórias e outras 5 com variáveis selecionadas e Definição da arquitetura e topologia da RNA
		Execução da RNA e cálculo das métricas do modelo para cada combinação como, por exemplo, acurácia, precisão, F1, etc.	Verificação das cinco melhores acurácias da validação cruzada e comparação com as acurácias obtidas no conjunto de teste	Análise da melhor configuração da RNA obtida e das variáveis mais importantes para a predição da classificação	A partir da melhor acurácia do conjunto de teste, simulação de aportes e cálculo dos ganhos efetivos. Comparação dos resultados com uma situação de aleatoriedade	Verificação da capacidade preditiva da RNA construída e discussão a partir da literatura encontrada

Fonte: A autora (2023)

3.1.1 Etapa 1: Introdução

No contexto da primeira etapa da pesquisa, foram realizadas cinco tarefas. A primeira consistiu na **imersão em trabalhos científicos** da área de finanças e de inteligência artificial, com o objetivo de compreender aspectos que poderiam ser relacionados ao processo decisório.

Em seguida, foi realizada a definição clara do **objeto de pesquisa**, que se concentra em IPOs do Brasil. A análise revelou que a assimetria de informações e a ausência de um histórico amplo de desempenho financeiro das empresas nos prospectos divulgados ao público são desafios a serem enfrentados. Essa definição do objeto de pesquisa permitiu delimitar o escopo da investigação e identificar uma lacuna a ser preenchida.

A terceira etapa envolveu a formulação precisa do **problema de pesquisa**, que consiste em compreender como as RNAs supervisionadas podem ser aplicadas para a construção da decisão de compra no curto prazo de ações de IPO no Brasil. O uso de RNAs pode auxiliar na análise de dados disponíveis antes do IPO, oferecendo uma perspectiva do potencial de crescimento das empresas.

Posteriormente, foi realizada a definição adequada da **unidade de análise**, que se concentra nas características de IPOs no Brasil e nos dados disponíveis no momento da publicação do prospecto definitivo. A pesquisa inclui todas as empresas que realizaram IPO na B3 durante o período de 2004 a 2022¹⁸, conforme estabelecido pela ICVM 400 (CVM, 2003). No entanto, a amostra é limitada a empresas que ofertaram ações e com pelo menos três anos de dados financeiros¹⁹ disponíveis nos prospectos definitivos das fontes oficiais, assim como De Jong; Kabir; Nguyen (2008) e Arosa; Richie; Schuhmann (2014) conduziram seus trabalhos. Essa definição permitiu estabelecer os limites da pesquisa e selecionar as fontes de dados apropriadas para a coleta de informações relevantes, incluindo a revisão de literatura sobre o tema.

Por fim, foram estabelecidos os **objetivos específicos** da pesquisa. Esses objetivos incluem verificar os dados socioeconômico-financeiros disponíveis nos

¹⁸ Em 2022 não ocorreram IPOs no Brasil.

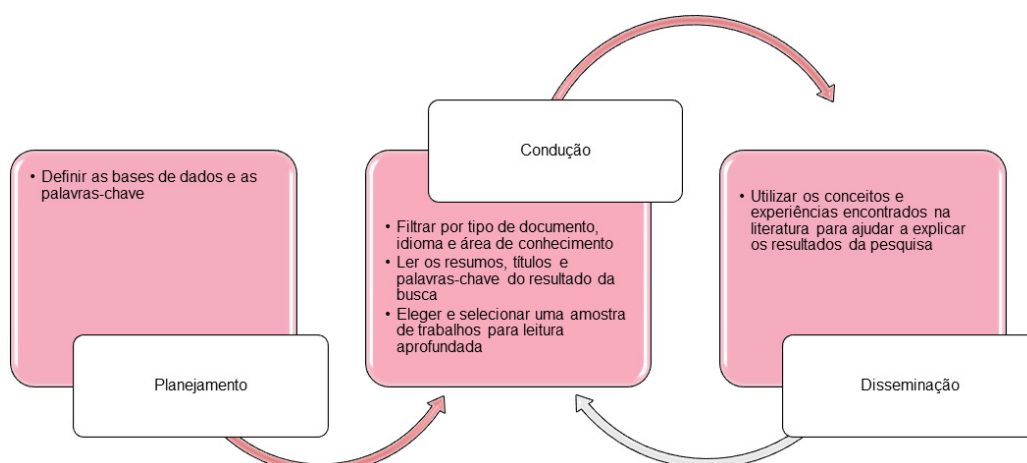
¹⁹ Resumo / Sumário das Demonstrações Financeiras Padronizadas.

prospectos definitivos dos IPOs, preparar os dados para a utilização no modelo de RNA, definir os elementos necessários do modelo e simular aportes em IPOs.

3.1.2 Etapa 2: Fundamentação

Na segunda etapa deste estudo, foram realizadas tarefas com o objetivo de compreender melhor o objeto de pesquisa e definir as possíveis variáveis explicativas relevantes. Inicialmente, foi conduzido um **planejamento** detalhado para realizar uma **busca sistematizada** de trabalhos acadêmicos, a fim de obter uma visão abrangente sobre o tema em questão. A **revisão da literatura** seguiu as etapas de planejamento, condução e disseminação propostas por Tranfield; Denyer; Smart (2003) e apresentadas na FIGURA 17:

FIGURA 17 - Estrutura da Revisão de Literatura



Fonte: A autora (2023) baseada em (TRANFIELD; DENYER; SMART, 2003)

A **seleção dos trabalhos** foi realizada a partir da base de dados Scopus, que é amplamente reconhecida por compilar trabalhos na área de finanças (GOODELL *et al.*, 2021). Além disso, foram consideradas também a *Web of Science*, o Google Acadêmico e o Scielo para abranger periódicos multidisciplinares.

A busca considerou um conjunto de palavras-chave relacionadas ao tema, com foco em artigos revisados por pares nas áreas de *Business*, *Finance*, *Management* e *Economics*, utilizando-se termos como "IPO*", "Artificial Neural network*" e suas variações:

"IPO*" OR "initial public offering*" OR "capital opening*" OR "initial funding" OR "open capital" OR "Decision to go public" OR "Pric* of IPO*" OR "going public" AND "Artificial Neural network*" OR "Artificial Neural" OR "Neural" OR "RNA" OR "ANN"²⁰

Essa abordagem permitiu obter um conjunto de trabalhos que abordavam a temática em questão. Para analisar a estrutura e as relações entre os trabalhos selecionados, foram utilizadas ferramentas como o *VOSviewer* e o *Bibliometrix* (*R Studio*). Essas **ferramentas** auxiliaram na identificação de constructos e nas **conexões entre os trabalhos**, fornecendo *insights* para a etapa de elegibilidade. Durante esse processo, trabalhos duplicados ou com pouca relevância para o tema foram excluídos.

Por meio dessas etapas, foi possível realizar um mapeamento dos principais pontos encontrados na revisão da literatura, como os objetivos da pesquisa, as teorias abordadas, as conclusões, as contribuições, as limitações e as sugestões de trabalhos futuros. Essa análise e composição da revisão da literatura forneceram a base para o desenvolvimento desta pesquisa, incluindo a **redefinição da variável alvo de predição**, conforme QUADRO 5:

QUADRO 5 - Definição da variável alvo e sua codificação

Variável “y”, “alvo”, “explicada”, “dependente”	
1 (um)	0 (zero)
compra, pois o retorno positivo será acima de 1%	não compra, porque não haverá retorno positivo acima de 1%
no <u>primeiro dia</u> de negociações da ação de IPO	

Fonte: A autora (2023)

Portanto, a variável que se busca prever foi definida como binária em relação ao “1 = compra, pois o retorno positivo será acima de 1%” ou “0 = não compra, porque não haverá retorno positivo acima de 1%” no primeiro dia de negociações da ação de IPO, seguindo a tendência acadêmica de focar em um período específico (por exemplo: Robertson et al. (1998), Xuemei; Yangbo (2009); Almeida (2013); Correia (2015))

²⁰ Nos sites de busca Google Acadêmico e *Scielo* as palavras-chave foram utilizadas isoladas ou combinadas em português ou espanhol ou francês.

3.1.3 Etapa 3: Coleta e pré-processamento dos dados

No terceiro estágio da pesquisa, uma série de tarefas foi conduzida para adquirir, organizar e preparar os dados necessários para posterior utilização na RNA. Inicialmente, houve a **sistematização** minuciosa dos dados requeridos a partir da literatura pertinente, abrangendo informações socioeconômicas e financeiras. Em seguida, foi realizada a **busca nas fontes institucionais**, incluindo o Banco Central do Brasil (BCB), Instituto Brasileiro de Geografia e Estatística (IBGE), Instituto de Pesquisa Econômica Aplicada (Ipea), Fundação Getúlio Vargas (FGV) e outras fontes relevantes. Adicionalmente, documentos públicos disponíveis nos *sites* da CVM, da B3 e das empresas que passaram pelo processo de abertura de capital foram consultados, fornecendo dados adicionais. Vale ressaltar que a base de dados utilizada por Almeida (2013) e Correia (2015)²¹, referente ao período de 2004 a 2014, também foi incorporada.

Com efeito, as fontes documentais estão elencadas no QUADRO 6:

QUADRO 6 - Dados coletados

Relacionado	Dado	Link referência
Datas	Ano/mês abertura	https://www.gov.br/cvm/pt-br/centrais-de-conteudo/publicacoes/series-historicas/distribuicoes-publicas/acoes-c/distribuicoes-de-acoes-cdas-e-bdrs-planilha-individualizada.ods/view
Datas	Data de Registro / Dispensa	https://www.gov.br/cvm/pt-br/centrais-de-conteudo/publicacoes/series-historicas/distribuicoes-publicas/acoes-c/distribuicoes-de-acoes-cdas-e-bdrs-planilha-individualizada.ods/view
Datas	Data início da negociação	https://www.b3.com.br/pt_br/produtos-e-servicos/solucoes-para-emissores/ofertas-publicas/estatisticas/
Identificação da empresa	Nome do Emissor	https://www.gov.br/cvm/pt-br/centrais-de-conteudo/publicacoes/series-historicas/distribuicoes-publicas/acoes-c/distribuicoes-de-acoes-cdas-e-bdrs-planilha-individualizada.ods/view
Identificação da empresa	CNPJ do Emissor	https://dados.cvm.gov.br/dataset/oferta-distrib/resource/c70a97f3-8e3d-4ada-9d3b-b0005e9d1fcb
Identificação da empresa	Setor de atividade	https://dados.cvm.gov.br/dataset/cia_aberta-cad/resource/f51d2be3-f521-4b79-ac76-d8319c86c86c
Identificação da empresa	Unidade Federativa da sede	https://dados.cvm.gov.br/dataset/cia_aberta-cad/resource/f51d2be3-f521-4b79-ac76-d8319c86c86c
Identificação da empresa	Município da sede	https://dados.cvm.gov.br/dataset/cia_aberta-cad/resource/f51d2be3-f521-4b79-ac76-d8319c86c86c
Identificação da empresa	Ano constitutivo da empresa	https://dados.cvm.gov.br/dataset/cia_aberta-cad/resource/f51d2be3-f521-4b79-ac76-d8319c86c86c
Identificação da empresa	Denominação social	https://dados.cvm.gov.br/dataset/cia_aberta-cad/resource/f51d2be3-f521-4b79-ac76-d8319c86c86c
Identificação da empresa	Nome pregão	https://www.b3.com.br/pt_br/produtos-e-servicos/solucoes-para-emissores/ofertas-publicas/estatisticas/

²¹ Setor, Segmento, Código, Coordenador Líder, Preço de Lançamento (R\$), Data da primeira negociação, Data de Abertura, Governança, Volume R\$ milhões, Natureza da Oferta, Meta Selic (% a.a.), Taxa Selic (% a.a.), Pontuação Ibovespa, IPCA Mens, IPCA Acumulado 12 meses, Índice Confiança da Indústria (ICI) c/ ajuste sazonal.

Identificação da empresa	Segmento de governança	https://www.b3.com.br/pt_br/produtos-e-servicos/solucoes-para-emissores/ofertas-publicas/estatisticas/
Identificação da empresa	Classificação setorial	https://www.b3.com.br/pt_br/produtos-e-servicos/solucoes-para-emissores/ofertas-publicas/estatisticas/
Identificação do processo de IPO	Processo	https://www.gov.br/cvm/pt-br/centrais-de-conteudo/publicacoes/series-historicas/distribuicoes-publicas/acoes-c/distribuicoes-de-acoes-cdas-e-bdrs-planilha-individualizada.ods/view
Identificação do processo de IPO	Tipo de Oferta	https://www.gov.br/cvm/pt-br/centrais-de-conteudo/publicacoes/series-historicas/distribuicoes-publicas/acoes-c/distribuicoes-de-acoes-cdas-e-bdrs-planilha-individualizada.ods/view
Identificação do processo de IPO	Código da CVM	https://dados.cvm.gov.br/dataset/cia_aberta-cad/resource/f51d2be3-f521-4b79-ac76-d8319c86c86c
Pós-IPO	Cotações históricas	Almeida (2013) e Correia (2015) e https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/historico/mercado-a-vista/series-historicas/
Situação conjuntura	Expectativa IPCA 12 meses	https://dadosabertos.bcb.gov.br/dataset/expectativas-mercado/resource/7130de18-9052-452c-a8eb-a7dae25d1e2a
Situação conjuntura	Expectativa taxa Selic	https://dadosabertos.bcb.gov.br/dataset/expectativas-mercado/resource/7130de18-9052-452c-a8eb-a7dae25d1e2a
Situação conjuntura	Meta Selic anual	http://api.bcb.gov.br/dados/serie/bcdata
Situação conjuntura	Acumulado do IPCA 12 meses	https://sidra.ibge.gov.br/tabela/1737
Situação conjuntura	Variação % IPCA	https://sidra.ibge.gov.br/tabela/1737
Situação conjuntura	Fechamento IBOV	https://br.financas.yahoo.com/quote/%5EBVSP/history?p=%5EBVSP
Situação conjuntura	Índice de Confiança do Consumidor (ICC)	http://api.bcb.gov.br/dados/serie/bcdata
Situação conjuntura	Anos eleitorais	https://www.tse.jus.br/eleicoes/eleicoes-anteriores
Situação conjuntura	Indicador de incerteza da economia – componente mídia	https://extra-ibre.fgv.br/IBRE/sitefgvdados/visualizaconsulta.aspx
Situação conjuntura	Índice de Confiança Empresarial (ICE) com ajuste sazonal	https://extra-ibre.fgv.br/IBRE/sitefgvdados/visualizaconsulta.aspx
Situação conjuntura	Índice de Confiança da Indústria (ICI) com ajuste Sazonal	https://extra-ibre.fgv.br/IBRE/sitefgvdados/visualizaconsulta.aspx
Situação conjuntura	Volume diário negociado na Bolsa de Valores.	https://extra-ibre.fgv.br/IBRE/sitefgvdados/visualizaconsulta.aspx
Situação empresa	Ativo total (AT)	<i>Prospectos das empresas</i>
Situação empresa	Ativo Circulante (AC)	<i>Prospectos das empresas</i>
Situação empresa	Ativo Não Circulante (ANC)	<i>Prospectos das empresas</i>
Situação empresa	Passivo Circulante (PC)	<i>Prospectos das empresas</i>
Situação empresa	Passivo Não Circulante (PNC)	<i>Prospectos das empresas</i>
Situação empresa	Patrimônio Líquido (PNC)	<i>Prospectos das empresas</i>
Situação empresa	Passivo Total (PT)	<i>Prospectos das empresas</i>
Situação empresa	Receita Líquida (RL)	<i>Prospectos das empresas</i>
Situação empresa	Custo da Mercadoria Vendida (CMV)	<i>Prospectos das empresas</i>

Situação empresa	Resultado Bruto (RB)	Prospectos das empresas
Situação empresa	Lucro Operacional (LO)	Prospectos das empresas
Situação empresa	Resultado antes dos impostos e contribuição (RAT)	Prospectos das empresas
Situação empresa	Resultado Líquido (RL)	Prospectos das empresas
Situação empresa	Ebitda (ajustado)	Prospectos das empresas
Situação empresa	Dias de Lock-up (DL)	Prospectos das empresas
Sobre o IPO	Coordenador da oferta (líder)	https://www.gov.br/cvm/pt-br/centrais-de-conteudo/publicacoes/series-historicas/distribuicoes-publicas/acoes-c/distribuicoes-de-acoes-cdas-e-bdrs-planilha-individualizada.ods/view
Sobre o IPO	Quantidade de títulos (exclusive o lote suplementar)	https://www.gov.br/cvm/pt-br/centrais-de-conteudo/publicacoes/series-historicas/distribuicoes-publicas/acoes-c/distribuicoes-de-acoes-cdas-e-bdrs-planilha-individualizada.ods/view
Sobre o IPO	Quantidade de títulos no lote suplementar	https://www.gov.br/cvm/pt-br/centrais-de-conteudo/publicacoes/series-historicas/distribuicoes-publicas/acoes-c/distribuicoes-de-acoes-cdas-e-bdrs-planilha-individualizada.ods/view
Sobre o IPO	Volume financeiro total captado	https://www.b3.com.br/pt_br/produtos-e-servicos/solucoes-para-emissores/ofertas-publicas/estatisticas/
Sobre o IPO	Preço d0	https://www.b3.com.br/pt_br/produtos-e-servicos/solucoes-para-emissores/ofertas-publicas/estatisticas/
Sobre o IPO	Tickers	https://investpy.readthedocs.io/_api/stocks.html

Fonte: A autora (2023)

Após a fase de busca, **os dados coletados** das 205 empresas que abriram capital de 2004 a 2022 foram armazenados em uma planilha eletrônica utilizando o *Microsoft Excel* 2019. Com isso, foi possível detectar 42 possíveis variáveis explicativas que comporiam a pesquisa, excetuando aquelas referentes às contas das demonstrações contábeis, e criar a síntese do QUADRO 7.

As variáveis escolhidas como potenciais para explicarem do IPO no Brasil representam constructos teóricos relevantes em duas abordagens principais: "*bottom-up*" e "*top-down*", para obter uma visão abrangente dos fatores que podem afetar o desempenho das IPOs no mercado financeiro.

Na abordagem "*bottom-up*", que se concentra nos aspectos internos da empresa emissora do IPO, foram selecionadas variáveis como a data de lançamento do IPO (ano_ipo; mes) e o preço registrado no momento do lançamento (P0). Além disso, foram consideradas variáveis relacionadas à Educação & Conduta, como o segmento de governança, e Instituições, como o setor e o tipo de oferta da empresa. Essas escolhas foram baseadas especificamente nas pesquisas de Almeida (2013), Correia (2015), Martins *et al.* (2006), e outros autores que discutiram a importância desses fatores para o desempenho das empresas no contexto do IPO.

QUADRO 7 - Variáveis explicativas identificadas²²

Abreviações	Variáveis identificadas	Abreviações	Variáveis identificadas
Ano_ipo; mes	Ano e mês de lançamento do IPO	InvestK	Investimento em Capital
P0	Preço registrado (lançamento)	AqSoc	Aquisições Societárias
exp_selic_ano_ipo_ant	Expectativa mercado da Taxa Selic p/ ANO IPO no ano anterior	EstrutK	Estrutura de Capital
exp_selic_DIA_ipo	Expectativa mercado da Taxa Selic no DIA do IPO para o ANO vigente	CapitaK	Capital de Giro
exp_selic_mes_pos_ipo	Expectativa mercado da Taxa Selic para o MÊS posterior ao IPO	Efic	Eficiência Operacional
exp_inflacao_12_meses	Expectativas de mercado para inflação p/ os próximos 12 meses	EmSecund	Emissão Secundária
ipca_acum_12_meses	IPCA acumulado últimos 12 meses	paPE	participação <i>Private Equity</i>
var_mensal_ipca_ipo	Variação % mensal IPCA data do IPO	cap_terc_cap_prio	Capital de terceiros / Capital próprio
meta_selic_dia_reg_ipo	Meta Taxa Selic vigente (anual) no DIA IPO	comp_endividamento	Composição do endividamento
pont_ibov_fech_reg_ipo	Pontuação IBOV (fechamento do último DIA disponível anterior ao IPO)	imob_cap_prio	Imobilização do capital próprio
vol_diario_ibov_reg_ipo	Volume diário total negociado na Bolsa de São Paulo - Ibovespa - no DIA DO REGISTRO DO IPO	imob_rec_ nao_cor	Imobilização dos recursos não-correntes
ice	Índice de Confiança na Economia	gaf	Grau de alavancagem financeira (GAF)
ici	Índice de Confiança na Indústria	LC	Liquidez Corrente
icc	Índice de Confiança do Consumidor	GA	Giro do Ativo
uf	Unidade Federativa da sede da empresa	MB	Margem Bruta
segmento	Segmento de Governança	MgOp	Margem Operacional
setor	Setor	ML	Margem Líquida
tipo_oferta	Tipo de Oferta	MkGb	Mark-up global
indic_incerteza_eco	Indicador de Incerteza da Economia Brasil - Componente Média (Mídia - IIE-Br)	ROE	Retorno sobre o Patrimônio Líquido (ROE)
ano_eleitoral	Ano eleitoral	ROA	Retorno sobre o Ativo (ROA)
idade_emp_ipo	Idade empresa IPO	PL	P/L
vol_financeiro	Volume financeiro total registrado	EbitdaLuc	Ebitda/Lucro Líquido

Fonte: A autora (2023)

Por outro lado, na abordagem "*top-down*", que examina os fatores externos que podem influenciar o desempenho dos IPOs, foram selecionadas variáveis como as expectativas de mercado para a taxa Selic (exp_selic_ano_ipo_ant, exp_selic_dia_ipo, exp_selic_mes_pos_ipo), o IPCA acumulado nos últimos 12

²² A referência acadêmica que baseou a escolha de cada uma delas está no QUADRO 17.

meses (*ipca_acum_12_meses*) e o indicador de incerteza econômica da média (*indic_incerteza_eco*). Essas variáveis foram embasadas nas pesquisas de Almeida (2013), Correia (2015), Khatri (2017), Agrawal (2021), Bhardwaj e Imam (2019), e outros, que discutiram a relevância desses indicadores econômicos e políticos para compreender o cenário macroeconômico e suas influências nos IPOs.

A seleção das possíveis variáveis também levou em conta estudos anteriores sobre os fatores que impactam a sobrevivência das empresas após a realização do IPO, e, portanto, foram incluídas variáveis relacionadas a Sobrevivência & Capital, como idade da empresa no momento do IPO, volume financeiro total registrado, investimento em capital, aquisições societárias, estrutura de capital, entre outras. Essas variáveis foram fundamentadas nas pesquisas de Beatty e Ritter (1986), Ritter e Welch (2002), Adebisi *et al.* (2012), Assaf Neto (2000), Jain e Nag (1995, 1998), Leal (2005), e outros.

Porém, cabe destacar que as *variáveis finais* do modelo só foram definidas após os procedimentos de tratamento dos dados coletados²³. Inicialmente, os valores foram inflacionados, tendo como referência o ano de 2022, disponível no IPEADATA (2022), assim como foram deletadas as possíveis variáveis explicativas que estavam com mais dados ausentes²⁴,

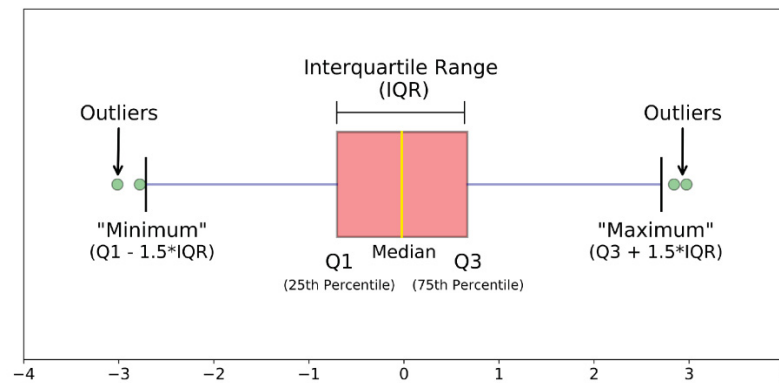
Essas medidas visaram estabelecer uma estrutura inicial para o desenvolvimento da RNA. Para dar continuidade ao processo de organização dos dados, foi utilizado o ambiente do *Google Colab*, que proporcionou infraestrutura adequada para a utilização da linguagem de programação *Python* na versão 3.10.12²⁵.

Diante disso, com o objetivo de garantir a qualidade dos dados e fortalecer a análise, os **preços atípicos** das ações de IPO foram excluídos, seguindo o cálculo da busca por *outliers* mostrado na FIGURA 18:

²³ Restaram 27 variáveis e elas serão apresentadas na TABELA 6, no próximo capítulo, em que os resultados dos filtros realizados são discutidos.

²⁴ A saber: 'MB', 'ML', 'PL'

²⁵ Além disso, as bibliotecas e suas respectivas versões utilizadas foram: pandas (1.5.3); numpy (1.22.4); scipy (1.10.1); python-bcb (0.1.8); b3fileparser (0.1.4); yfinance (0.2.18); sidrapy (0.1.4); wget (3.2); investpy (1.0.8); seaborn (0.12.2); matplotlib (3.7.1); scikit-learn (1.2.2).

FIGURA 18 - *Outliers em um boxplot*

Fonte: (KDNUGETS, 2023)

Adicionalmente, foi realizada a **codificação** dos dados categóricos presentes na base de dados. Essa etapa foi necessária, uma vez que muitos algoritmos de aprendizado de máquina requerem dados numéricos como entrada. Para realizar essa codificação²⁶, foi utilizado o método conhecido como *Label Encoder*, em que é atribuído um valor numérico único para cada categoria presente na variável categórica.

Equação 1

$$\text{LabelEncoder}(x) = \begin{cases} i, & \text{se } x_i = x \\ \text{NaN}, & \text{caso contrário} \end{cases}$$

No caso desta pesquisa, o *Label Encoder* atribui um valor numérico único a cada categoria presente nas variáveis categóricas "líder", "setor", "tipo_oferta", "segmento" e "uf".

Além dessa codificação, foram utilizados outros métodos específicos para codificação de variáveis particulares. A variável categórica "ano_eleitoral" foi codificada de forma binária, atribuindo o valor "1" quando a empresa estava em um ano eleitoral e "0" caso contrário. A variável categórica "segmento de governança" foi codificada de forma ordinal (HANCOCK; KHOSHGOFTAAR, 2020), atribuindo valores numéricos crescentes com base em uma ordem predefinida que reflete a importância relativa dos segmentos de governança.

Em seguida, foi realizada a padronização da base de dados utilizando o *StandardScaler*:

²⁶ Além das utilizadas nesta dissertação, existem outras, tais quais *Catboost encoding*, *Helmert encoding*, *target encoding*.

Equação 2

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean
 σ = Standard Deviation

Esse método é utilizado para transformar as variáveis numéricas contínuas em uma escala padrão, com média igual a zero e desvio padrão igual a um. A padronização é importante para evitar que variáveis com unidades de medida diferentes tenham pesos desproporcionais durante o treinamento do modelo.

3.1.4 Etapa 4: Exploração dos dados e definição da RNA

Após a organização dos dados, procedeu-se ao **cálculo** dos **retornos** das ações nos períodos imediatamente subsequentes à abertura de capital, conforme:

Equação 3

$$R_t = \left(\frac{A_t - A_{t-1}}{A_{t-1}} \right)$$

R_t = retorno do ativo no período t ;
 A_t, A_{t-1} = valor do ativo no período t e $t-1$ (valor do registro = d0),
 respectivamente;

Fonte: (ASSAF NETO, 2000)

Além disso, foi realizada a **análise vertical** (av) referente às contas das demonstrações contábeis, assim como foram obtidos alguns **índices financeiros** utilizando as fórmulas e aspectos de estrutura de capital, de solvência, lucratividade e retorno, como as apresentadas no QUADRO 8, considerando as informações disponíveis do último ano nos prospectos definitivos das empresas:

Na sequência, foram realizadas **análises exploratórias** dos dados. Essas análises tiveram o objetivo de identificar possíveis relações entre as variáveis independentes e a variável alvo, bem como avaliar a presença de *outliers* ou padrões incomuns nos dados. Com isso, foi possível realizar uma **descrição** das características principais das empresas da amostra selecionada.

QUADRO 8 - Índices financeiros

Índice	Abreviação	Fórmula
Capital de terceiros em relação ao Capital próprio	cap_terc_cap_prio	(Passivo Circulante + Passivo não circulante) / Patrimônio Líquido
Capital de terceiros em relação ao Capital total (ou endividamento geral)	cap_terc_cap_total	(Passivo Circulante + Passivo não circulante) / Ativo Total
Composição do endividamento	comp_endividamento	Passivo Circulante / (Passivo Circulante + Passivo não circulante)
Imobilização do capital próprio	imob_cap_prio	Ativo não circulante / Patrimônio Líquido
Imobilização dos recursos não-correntes	imob_rec_ao_cor	Ativo não circulante / (Passivo não circulante + Patrimônio Líquido)
Multiplicador de Alavancagem Financeira	gaf	Ativo total / Patrimônio Líquido
Liquidez Corrente	LC	Ativo circulante / Passivo Circulante
Giro do Ativo	GA	Receita Líquida / Saldo médio* do Ativo Total
Retorno sobre o Patrimônio Líquido	ROE	Lucro líquido / Saldo médio* do Patrimônio Líquido
Retorno sobre o Ativo	ROA	Lucro líquido / Ativo Total

*O saldo médio foi calculado considerando os últimos anos disponíveis no prospecto;

Fonte: A autora (2023) baseada em (ASSAF NETO, 2000)

Ademais, foi realizado o **embaralhamento** das observações para balanceamento da base de dados, que ficou desequilibrada após as etapas anteriores (restaram 88 valores “1” na variável alvo e 80 “0”). Para sanar essa diferença, embora baixa, foi aplicada a técnica de amostragem aleatória estratificada a partir da variável alvo para efetuar o embaralhamento da base:

Equação 4

$$n_i = \frac{N_i}{N} \cdot n$$

n_i representa o tamanho da amostra a ser selecionada do estrato S_i , N_i é o tamanho total do estrato S_i no conjunto de dados, N é o tamanho total do conjunto de dados e n é o tamanho desejado da amostra final

Essa técnica garantiu que as classes de “compra” e “não compra” fossem igualmente representadas na base de dados, com uma proporção de 50% para cada classe (80 a 80), logo a **amostra final para a RNA** contemplou 160 empresas. Isso foi importante para evitar um viés no treinamento do modelo.

Outra tarefa realizada foi o cálculo de **correlação** das variáveis independentes com a variável alvo (Y) por meio do coeficiente de Pearson:

Equação 5

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

X_i e Y_i representam os valores individuais das variáveis X e Y , enquanto $\bar{X}$ e $\bar{Y}$ são as medidas dos valores das variáveis X e Y

Assim como foi calculado o Fator de Inflação da Variância (*Variance Inflation Factor*) entre as possíveis variáveis independentes:

Equação 6

$$VIF_i = \frac{1}{1 - R_i^2}$$

VIF_i representa o Fator de Inflação da Variância para a variável independente i , e R_i^2 é o coeficiente de determinação (R-quadrado) resultante da regressão linear entre a variável independente i e as demais variáveis independentes.

As medidas de correlação das possíveis variáveis independentes com a variável alvo e o cálculo do VIF desempenharam um papel crucial na seleção das variáveis relevantes para o modelo de RNA. Ao identificar relações fortes entre as variáveis e detectar a presença de multicolinearidade, foi possível evitar que variáveis altamente correlacionadas tivessem um peso indevido. Isso ajudou a direcionar o estudo para as **variáveis mais relevantes**, reduzindo a redundância de informações e melhorando a eficiência do modelo. Com isso, a base de dados foi dividida de forma “aleatória”²⁷ em 50 **combinações** dessas variáveis, seguindo a abordagem de Paiva (2014), e outros 5 com variáveis selecionadas.

Com efeito, essa divisão permitiu testar diferentes relações de variáveis e avaliar a robustez do modelo. Cada combinação foi gerada de forma “aleatória” permitindo repetição, embora cada uma possa ter sido gerada com um número variável de neurônios na camada oculta com base em frações aleatórias das variáveis de entrada. Além disso, foram geradas conciliações específicas que compartilhavam colunas comuns entre o conjunto de dados principal e outros conjuntos de dados relacionados. Essa abordagem possibilitou a exploração de diversas combinações.

²⁷ Todas as gerações “aleatórias” desta pesquisa foram por realizadas por meio da função *random* do *Python*, ou seja, utilizando combinações pseudoaleatórias.

3.1.4.1 Elementos constitutivos da RNA

Tendo em vista as combinações criadas aleatoriamente das possíveis variáveis explicativas e com base nas diretrizes propostas por Kuhn; Johnson (2013), foi utilizada a estratégia de **dividir** os dados em **conjuntos de treinamento (com validação) (80%)**, que foi utilizado para criar e ajustar o modelo, **e teste (20%)**, que foi mantido separado durante todo o processo.

Além disso, para garantir uma divisão adequada dos dados e evitar distorções devido à concentração de dados em determinados anos ou a sequências de “1” ou “0” na variável alvo, foi realizada novamente uma amostragem aleatória e estratificada. Essa abordagem garantiu que as proporções das classes (amostras com a variável alvo igual a “1” e amostras com a variável alvo igual a “0”) fossem mantidas em proporção de 50% e 50%, tanto no conjunto de treinamento quanto no conjunto de teste, minimizando possíveis enviesamentos e mantendo os conjuntos os mais homogêneos possíveis (KUHN; JOHNSON, 2013).

A **topologia** de uma RNA é composta por arquitetura, algoritmo e funções. Apesar da ampla utilização de RNAs na área financeira, não há consenso de qual arquitetura que deve ser utilizada em cada tipo de problema (COAKLEY; BROWN, 2000). Optou-se, portanto, por seguir a tendência dos trabalhos acadêmicos e utilizar *Multilayer Perceptron* (MLP), ou seja, uma Rede estruturada com **camada** oculta. Assim como ajustar os pesos de modo a minimizar a diferença das saídas, por meio do **algoritmo Backpropagation**, uma vez que possui a capacidade de aprender e memorizar padrões com precisão (CHI, 2009), além de ser bastante empregado na área financeira para calcular os gradientes da função de perda com relação aos pesos da RNA.

Entre a camada de entrada e a de saída há a **oculta** que, por sua vez, tem o objetivo de identificar atributos especiais dos dados (MALIK; NASEREDDIN, 2006) e capturar a não linearidade (PAO, 2008). Contudo, a adição de várias delas tende a não agregar ao modelo (SALCHENBERGER; CINAR; LASH, 1992), por isso apenas uma camada oculta será utilizada nesta pesquisa, assim como Jain; Nag (1995), Jain; Nag, (1998), Han; Kim (2021).

A quantidade de **nós** na camada de entrada é igual ao número de variáveis e na de saída corresponde ao resultado esperado, ou seja, nesta pesquisa é igual a um, representando a compra (“1”) ou não (“0”) da ação do IPO.

Além disso, para otimizar o desempenho da RNA, foram utilizadas técnicas de busca de hiper parâmetros, como o **Grid Search**. O *Grid Search* consiste em uma abordagem sistemática para explorar diferentes combinações de parâmetros de um modelo de aprendizado de máquina, a fim de encontrar a combinação que resulte em um desempenho ótimo (HASTIE *et al.*, 2009).

No caso da RNA, foram definidos - após diversas simulações e testes prévios se baseando nas pesquisas apresentadas no Referencial Teórico - os seguintes **parâmetros** para o *Grid Search*: quantidade de neurônios na camada oculta, função de ativação, solver, número de iterações, coeficiente de regularização e taxa de aprendizado.

A quantidade de **neurônios** na **camada oculta** foi definida como variando de 50% a 75% em relação ao número de variáveis de entrada do modelo, seguindo a sugestão de Salchenberger; Cinar; Lash (1992). Isso, pois quando existem muitas conexões intermediárias, a RNA acaba memorizando dados de entrada ao invés de ser capaz de ter capacidade de generalização. Em outros termos, quanto mais nós na camada intermediária mais a rede será capaz de classificar apenas no conjunto de treinamento - e menos nos demais (SALCHENBERGER; CINAR; LASH, 1992).

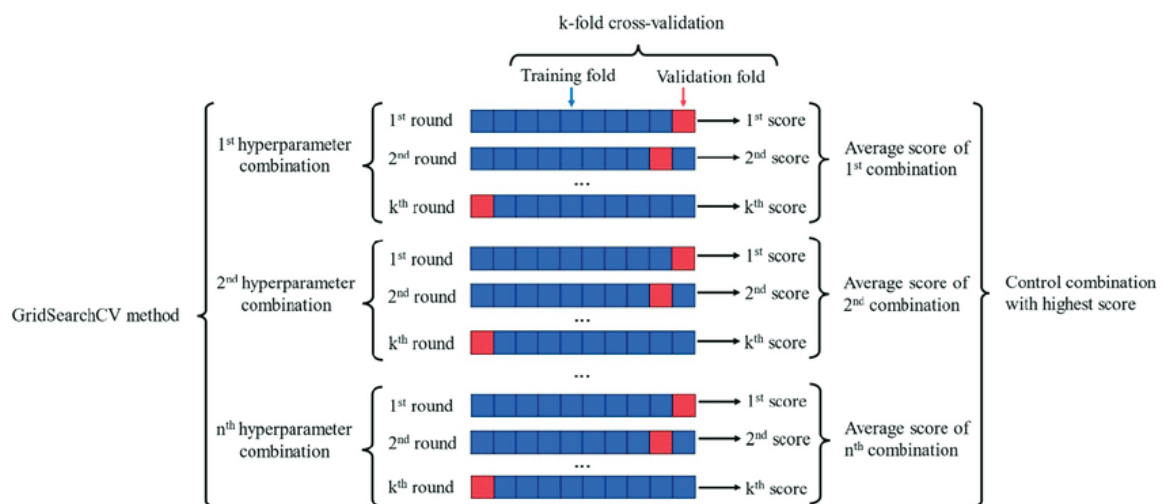
Os outros parâmetros definidos foram: a **função de ativação** 'logistic', 'relu' ou 'tanh' (HAN; KIM (2021)); o **solver**, que pode ser 'lbfgs', 'adam' ou 'sgd'; o número máximo de **iterações**, definido por [1000, 5000, 10000, 40000]; o coeficiente de regularização alpha, para tentar diminuir o efeito de *overfitting*, com valores [0.0001, 0.001, 0.01]; e a **taxa de aprendizado**, que pode ser 'constant' ou 'adaptive'. Essas escolhas foram baseadas em considerações teóricas e experimentais relacionadas ao desempenho e capacidade de generalização da RNA (GOODFELLOW; BENGIO; COURVILLE, 2016).

Após a definição dos parâmetros utilizando o *Grid Search*, para evitar a avaliação enviesada do modelo, foi empregada a técnica de **validação cruzada Stratified K-Fold** com 80% dos dados. Nesse processo, o conjunto de treinamento foi dividido em "k = 5" partes, conhecidas como "dobras" (*folds*). A RNA foi treinada "k" vezes, cada vez utilizando k-1 dobras para treinamento e a dobra restante para validação. Nesse processo, todas as observações contidas nos 80% dos dados selecionados para treinamento, foram utilizadas tanto para treinamento propriamente dito quanto para validação em algum momento (HASTIE *et al.*, 2009).

A representação da combinação do *Grid Search* com a validação cruzada de 5 folds, empregando o *Stratified K-Fold*, utilizada em cada combinação desta pesquisa, pode ser visualizada na FIGURA 19. Essa abordagem assegura a preservação da proporção de cada classe em cada dobra, garantindo que as distribuições das classes sejam mantidas tanto no conjunto de treinamento quanto no conjunto de validação.

Além disso, para avaliar a performance da RNA em dados não vistos durante o treinamento (e validação), foi realizado um teste final utilizando o conjunto previamente separado com 20% dos dados. Esse conjunto de teste contém dados que não foram utilizados em nenhuma fase de treinamento ou validação, permitindo uma avaliação imparcial e realista do modelo em condições reais.

FIGURA 19 - *Grid Search* com *Stratified K-Fold*



Fonte: SHATNAWI *et al.*, 2022.

Em síntese, os elementos da RNA utilizada na pesquisa podem ser visualizados no QUADRO 9:

QUADRO 9 - Elementos da RNA

Elemento	Definição na RNA
Divisão dos dados	Conjunto de treinamento (80%) e conjunto de teste (20%)
Amostragem dos dados	Aleatória e estratificada para garantir proporções balanceadas
Arquitetura	<i>Multilayer Perceptron</i> (MLP) com uma camada oculta
Tipo de Aprendizado	Supervisionado
Quantidade de neurônios na camada de entrada	Conforme quantidade de variáveis definidas em cada combinação
Quantidade de neurônios na camada oculta	De 50% a 75% da quantidade de variáveis de entrada
Quantidade de neurônios na camada de saída	Apenas um representando a decisão de pertencer a uma das duas classes possíveis
Algoritmo de treinamento	<i>Backpropagation</i>

Validação cruzada	<i>Stratified K-Fold</i> (k = 5) utilizando o conjunto de treinamento
Função de ativação	'logistic', 'relu' ou 'tanh'
Solver	'lbfgs', 'adam' ou 'sgd'
Número de iterações	1000, 5000, 10000 ou 40000
Coefficiente de regularização	0.0001, 0.001 ou 0.01
Taxa de aprendizado	'constant' ou 'adaptive'
Técnica de busca	<i>Grid Search</i>
Teste final	Apenas com conjunto de teste

Fonte: A autora (2023)

Com efeito, o *script* desenvolvido em *Python* para a subsequente execução da RNA pode ser resumido como um conjunto estruturado de sub-etapas. Inicialmente, foi estabelecida a “semente” para garantir a reprodutibilidade dos resultados. Em seguida, foram estabelecidas as configurações do *Grid Search*, englobando parâmetros relacionados à ativação, solver, número máximo de iterações, alpha e *learning rate*. Essas configurações permitem explorar amplamente as combinações possíveis do classificador *Perceptron* Multicamadas (MLP).

Posteriormente, ocorreu a divisão do conjunto de dados em treinamento (e validação) e teste. A técnica de validação cruzada *Stratified K-Fold* foi empregada, viabilizando o treinamento e validação em diferentes partições dos dados. Dentro de cada combinação de variáveis, realizou-se o *Grid Search* para identificar a melhor configuração do classificador MLP. As métricas de desempenho foram avaliadas em cada *fold*, proporcionando uma análise detalhada do desempenho do modelo.

Por fim, o modelo treinado foi avaliado no conjunto de teste, e todas as métricas de desempenho foram novamente calculadas. Ao longo desse processo, os resultados foram registrados em um *DataFrame*.

3.1.5 Etapa 5: Execução da RNA e verificações

Nesta quinta etapa, foram executadas as RNAs em cada uma das combinações definidas previamente, com o intuito de realizar a predição da classificação das ações de IPO. A execução das RNAs permitiu o cálculo das métricas do modelo para cada combinação, a fim de avaliar seu desempenho.

As **métricas** utilizadas para avaliar a performance do modelo foram selecionadas com base em estudos anteriores (HASTIE *et al.*, 2009) e foram listadas no QUADRO 10. Dentre as métricas comumente adotadas, destacam-se a acurácia (*accuracy*), que mensura a proporção de predições corretas em relação ao total de amostras; a precisão (*precision*), que indica a proporção de verdadeiros positivos em

relação ao total de predições positivas; a revocação (*recall*), que representa a proporção de verdadeiros positivos em relação ao total de amostras positivas; a medida *F1 Score*, que é a média harmônica entre precisão e revocação e; ROC AUC (*Area Under the Receiver Operating Characteristic Curve*).

QUADRO 10 - Métricas de avaliação RNA classificatória

Métrica	Fórmula
Acurácia (Accuracy)	$(\text{Verdadeiros Positivos} + \text{Verdadeiros Negativos}) / \text{Total de Amostras}$
Precisão (Precision)	$\text{Verdadeiros Positivos} / (\text{Verdadeiros Positivos} + \text{Falsos Positivos})$
Revocação (Recall)	$\text{Verdadeiros Positivos} / (\text{Verdadeiros Positivos} + \text{Falsos Negativos})$
Medida F1 (F1 Score)	$2 * (\text{Precisão} * \text{Revocação}) / (\text{Precisão} + \text{Revocação})$

Fonte: A autora (2023) baseada em (HASTIE *et al.*, 2009)

Além disso, foram **verificadas as cinco melhores acurácias** obtidas por meio da técnica de validação cruzada. Em seguida, essas acurácias foram comparadas com as acurácias obtidas no conjunto de teste, a fim de verificar a consistência do desempenho do modelo e sua capacidade de generalização para dados não vistos anteriormente.

Também foi realizada a identificação da **melhor configuração** da RNA obtida, assim como das **variáveis** do conjunto de dados que **apresentaram maior importância** na predição da classificação das ações de IPO. Posteriormente, foram simulados **aportes financeiros** com base nas previsões do modelo e calculados os ganhos efetivos resultantes desses investimentos. Essa simulação permitiu a avaliação da capacidade do modelo em gerar retornos financeiros positivos quando utilizado como base para a tomada de decisões de investimento em IPO.

Com o intuito de comparação, **os resultados obtidos pelo modelo foram confrontados com uma situação aleatória na decisão de compra das ações de IPO**. Essa comparação teve como objetivo verificar se o modelo apresentava uma capacidade preditiva superior à mera aleatoriedade na seleção de ações.

Por fim, com base em todas as etapas anteriores, foi realizada uma breve **análise** para determinar se o modelo desenvolvido apresentou um **desempenho** satisfatório na tarefa de classificação das ações de IPO, com o objetivo de avaliar sua viabilidade como uma ferramenta útil na tomada de decisões de investimento nesse contexto específico. Os resultados e conclusões derivados da avaliação da capacidade preditiva serão discutidos no próximo capítulo, logo após a apresentação das características das empresas amostradas e da exploração das variáveis de entrada do modelo.

4 ANÁLISE DOS DADOS E DISCUSSÃO DOS RESULTADOS

Neste capítulo, serão abordados e discutidos os resultados da pesquisa, com foco nos achados obtidos nas etapas 4 e 5 mencionadas no capítulo anterior. O objetivo principal é apresentar as conclusões em resposta à pergunta de pesquisa: “Como as RNAs supervisionadas podem ser aplicadas para a construção da decisão de compra no curto prazo de ações de IPO no Brasil?”.

Para uma melhor contextualização dos resultados obtidos, este capítulo inicia com uma **descrição** pormenorizada da base de dados, fornecendo informações relevantes sobre a amostra utilizada na pesquisa. Em seguida, são **exploradas** as escolhas das **variáveis** de entrada para que, posteriormente, possam ser **discutidas as métricas** do desempenho do modelo, seguida por uma **análise** da **consistência** e capacidade de generalização dele.

Adicionalmente, são conduzidas **simulações** de **investimentos** com base nas **previsões** geradas, permitindo a mensuração dos ganhos efetivos decorrentes dessas estratégias. Por fim, é realizada uma breve **avaliação** da **viabilidade** do modelo desenvolvido como uma ferramenta útil na tomada de decisões financeiras, levando em consideração suas especificidades e particularidades.

4.1 ANÁLISE DOS DADOS

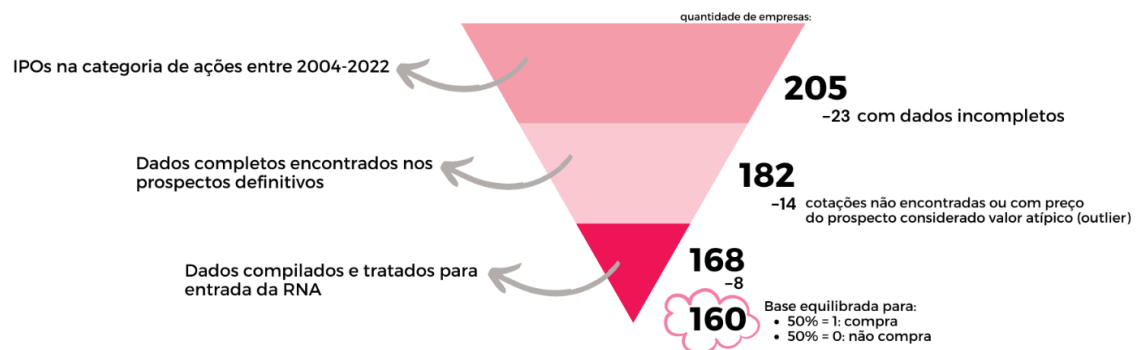
4.1.1 Descrição

4.1.1.1 Empresas e momento de IPO

A amostra utilizada nesta pesquisa, após o tratamento dos dados, consiste em um total de 168 empresas²⁸. Porém, para garantir a preservação do equilíbrio da variável alvo (50% / 50%), foram excluídas de forma aleatória 8 observações antes de alimentar a RNA com os dados. A FIGURA 20 mostra os resultados dos filtros explicados na Etapa 3 do método descrito no capítulo anterior:

²⁸ Foram listados seus respectivos nomes e setores no APENDICE no QUADRO 18.

FIGURA 20 - Filtro para equilíbrio da base de dados



Fonte: A autora (2023)

Diante disso, os detalhes da filtragem estão apresentados no QUADRO 11:

QUADRO 11 - Amostra em quantidade

Ano	Quantidade inicial de IPOs de ações (ordinárias e preferenciais)	Empresas Excluídas no 1º filtro*	Quantidade de IPOs de ações após 1º filtro	Empresas Excluídas no 2º filtro**	Quantidade da amostra de IPOs de ações após os filtros
2004	7	1: ALL AMER LAT	6	2: NATURA; GRENDENE	4
2005	9	1: COSAN	8	1: NOSSA CAIXA	7
2006	20	3: BRASILAGRO; MMX MINER; ABYARA	17	1: TOTVS	16
2007	54	4: CR2; INVEST TUR; SPRINGS; BMF; CRUZEIRO SUL	49	2: BR BROKERS; PX ENERGIA	47
2008	4	1: OGX PETROLEO	3	0	3
2009	5	2: TIVIT; DIRECIONAL	3	0	3
2010	10	3: MULTIPLUS; BR INSURANCE; RAIA	7	2: OSX BRASIL; HRT PETROLEO	5
2011	10	3: AUTOMETAL; QGEP PART; BR PHARMA	7	0	7
2012	2	0	2	0	2
2013	9	1: BB SEGURIDADE	8	0	8
2014	1	0	1	0	1
2015	1	0	1	1: PARCORRETORA	0
2016	1	0	1	0	1
2017	9	1: OMEGA GER	8	1: BK BRASIL	7
2018	3	1: INTER BANCO	2	1: INTERMEDICA	1
2019	5	1: BANCO BMG	4	1: CENTAURO	3
2020	26	0	26	0	26
2021	29	0	29	2: MOSAICO; FOCUS ON	27
2022	0	0	0	0	0
Total	205	23	182	14	168

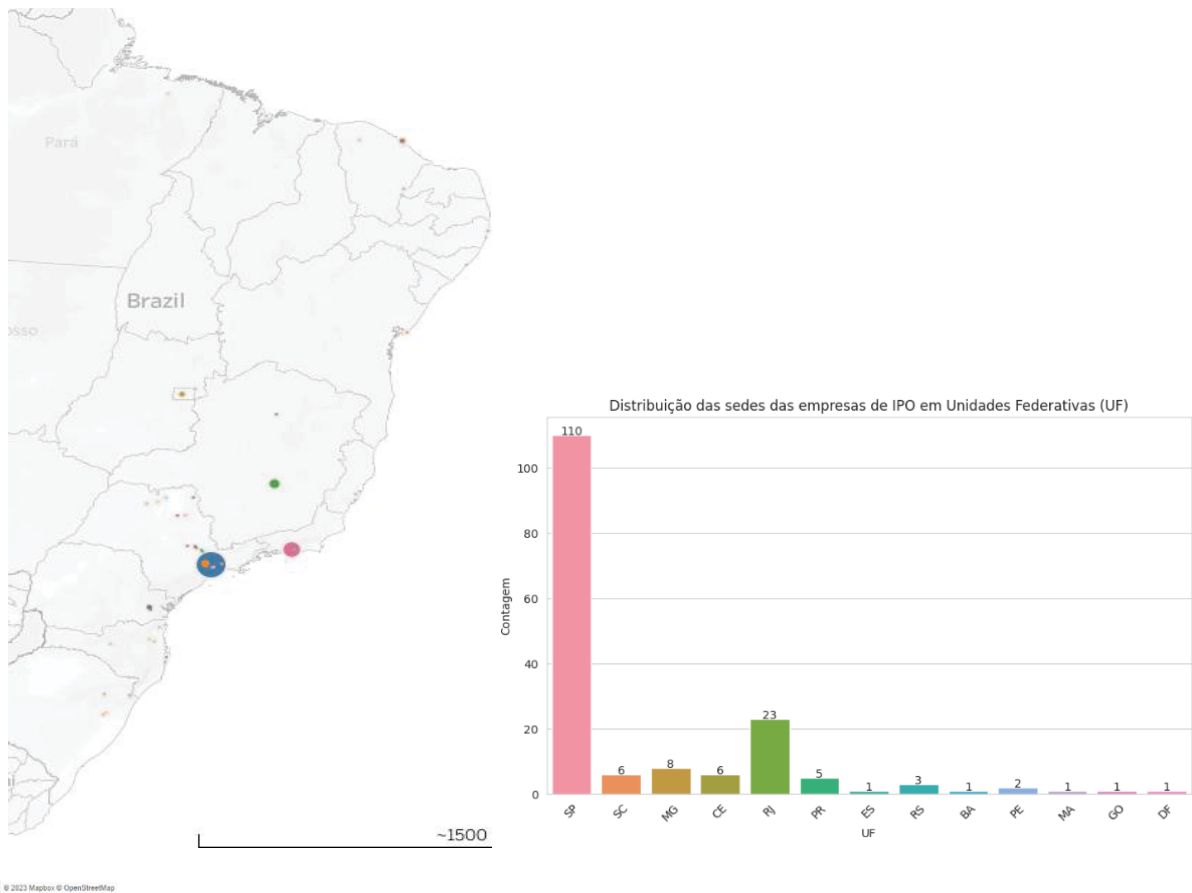
*Não foi encontrado o histórico financeiro completo no prospecto

** Não foi encontrado o preço do primeiro dia (P1) de negociação pós-IPO ou o preço (P0) divulgado no prospecto foi considerado como *outlier* no conjunto de dados.

Fonte: A autora (2023) com dados da B3 (2022) sobre os IPOs de ações regidos pela ICVM 400 (CVM, 2003)

A maioria das empresas da amostra está sediada no estado de São Paulo (SP), representando mais de 50% do total. Em seguida, tem-se o estado do Rio de Janeiro (RJ) com pouco mais de 20 ocorrências, como se vê no GRÁFICO 2:

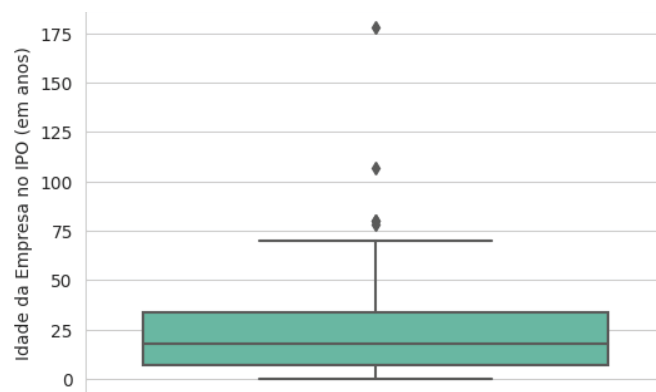
GRÁFICO 2 - Distribuição geográfica das sedes das empresas: cidades e estados



Fonte: A autora (2023)

Quanto à idade das empresas no ato de abertura de capital, tendo em vista a Teoria do *Life Cycle* (RITTER; WELCH, 2002), observa-se que a mediana está próxima a 20 anos, sendo que 75% delas têm até 75 anos. No entanto, algumas companhias apresentam idade entre 75 e 175 anos, como ilustrado no GRÁFICO 3:

GRÁFICO 3 - Idade das empresas no momento do IPO



Fonte: A autora (2023)

No contexto da governança, é importante destacar o segmento NM como uma referência de boas práticas de governança. Nesse segmento, estão concentradas 150 empresas da amostra da pesquisa – 89,3% como mostra a TABELA 1:

TABELA 1 - Segmentos de Governança em quantidade e percentual

Segmento	Quantidade	% do Total
NM	150	89.3
N2	9	5.4
N1	7	4.2
MA	2	1.2
Total	168	100

Fonte: A autora (2023)

O segmento NM se destaca e tem como característica exigir maior comprometimento das empresas com a adoção de práticas transparentes, éticas e de responsabilidade corporativa.

Independentemente do nível de governança, para que ocorram os IPOs, são necessários bancos coordenadores líderes, entre os quais se destacam o Banco BTG Pactual, Itaú e *Credit Suisse*, com percentuais de participação de 28%, 31,4% e 15,5%, respectivamente, conforme indicado na TABELA 2:

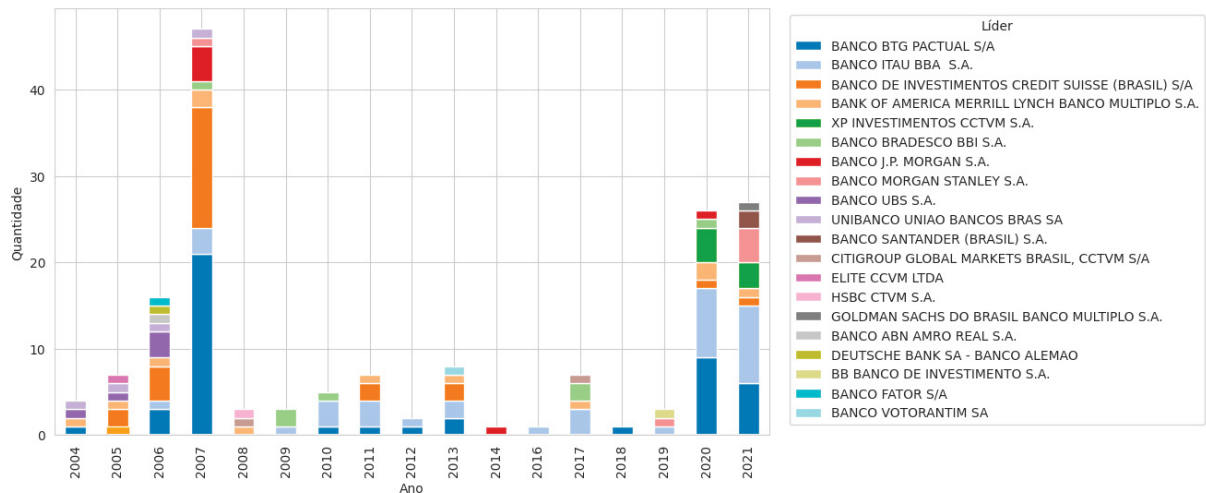
TABELA 2 - Coordenadores líderes em quantidade e percentual

Líder	Quantidade	% do Total
BANCO BTG PACTUAL S/A	47	28.0
BANCO ITAU BBA S.A.	36	21.4
BANCO DE INVESTIMENTOS CREDIT SUISSE (BRASIL) S/A	26	15.5
BANK OF AMERICA MERRILL LYNCH BANCO MULTIPLO S.A.	12	7.1
XP INVESTIMENTOS CCTVM S.A.	7	4.2
BANCO BRADESCO BBI S.A.	7	4.2
BANCO J.P. MORGAN S.A.	6	3.6
BANCO MORGAN STANLEY S.A.	6	3.6
BANCO UBS S.A.	5	3.0
UNIBANCO UNIAO BANCOS BRAS SA	4	2.4
CITIGROUP GLOBAL MARKETS BRASIL, CCTVM S/A	2	1.2
BANCO SANTANDER (BRASIL) S.A.	2	1.2
BANCO FATOR S/A	1	0.6
DEUTSCHE BANK SA - BANCO ALEMAO	1	0.6
BANCO ABN AMRO REAL S.A.	1	0.6
HSBC CTVM S.A.	1	0.6
BANCO VOTORANTIM SA	1	0.6
BB BANCO DE INVESTIMENTO S.A.	1	0.6
ELITE CCVM LTDA	1	0.6
GOLDMAN SACHS DO BRASIL BANCO MULTIPLO S.A.	1	0.6
TOTAL	168	100

Fonte: A autora (2023)

Durante o período de análise de 17 anos, foi observada a participação do Banco BTG Pactual em 10 anos, do Itaú em 12 anos e do *Credit Suisse* em 7 anos distintos, conforme evidenciado pelo GRÁFICO 4:

GRÁFICO 4 - Participação anual dos coordenadores líderes



Fonte: A autora (2023)

Ao analisar os setores de atuação das empresas listadas, identifica-se a presença de 11 setores diferentes ao longo dos anos, separados no QUADRO 12:

QUADRO 12 - Diversidades de setores por ano

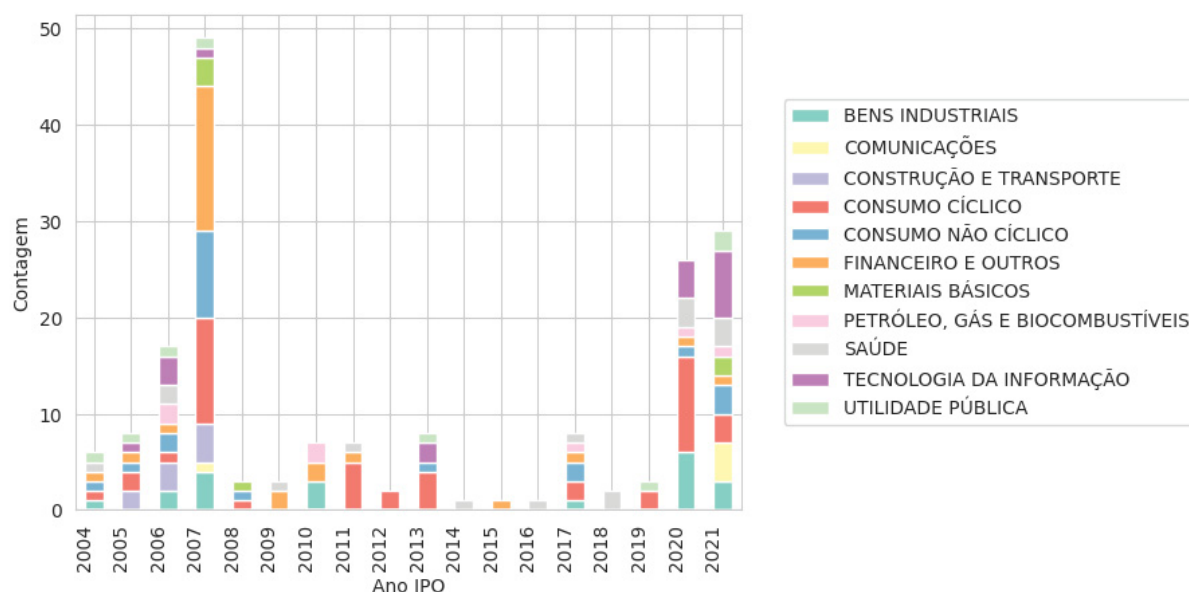
ano_ipo	total_ipos	diversidade_setores	lista_setores
2004	4	4	BENS INDUSTRIAIS, UTILIDADE PÚBLICA, SAÚDE, FINANCEIRO E OUTROS
2005	7	5	CONSUMO NÃO CÍCLICO, CONSUMO CÍCLICO, CONSTRUÇÃO E TRANSPORTE, UTILIDADE PÚBLICA, TECNOLOGIA DA INFORMAÇÃO
2006	16	9	UTILIDADE PÚBLICA, CONSTRUÇÃO E TRANSPORTE, CONSUMO CÍCLICO, BENS INDUSTRIAIS, PETRÓLEO, GÁS E BIOCOMBUSTÍVEIS, TECNOLOGIA DA INFORMAÇÃO, CONSUMO NÃO CÍCLICO, SAÚDE, FINANCEIRO E OUTROS
2007	47	8	CONSUMO CÍCLICO, CONSTRUÇÃO E TRANSPORTE, FINANCEIRO E OUTROS, CONSUMO NÃO CÍCLICO, COMUNICAÇÕES, BENS INDUSTRIAIS, MATERIAIS BÁSICOS, TECNOLOGIA DA INFORMAÇÃO
2008	3	3	MATERIAIS BÁSICOS, CONSUMO CÍCLICO, CONSUMO NÃO CÍCLICO
2009	3	2	FINANCEIRO E OUTROS, SAÚDE
2010	5	2	FINANCEIRO E OUTROS, BENS INDUSTRIAIS
2011	7	3	FINANCEIRO E OUTROS, CONSUMO CÍCLICO, SAÚDE
2012	2	1	CONSUMO CÍCLICO
2013	8	4	TECNOLOGIA DA INFORMAÇÃO, CONSUMO NÃO CÍCLICO, CONSUMO CÍCLICO, UTILIDADE PÚBLICA
2014	1	1	SAÚDE
2016	1	1	SAÚDE
2017	7	6	CONSUMO CÍCLICO, SAÚDE, BENS INDUSTRIAIS, CONSUMO NÃO CÍCLICO, FINANCEIRO E OUTROS, PETRÓLEO, GÁS E BIOCOMBUSTÍVEIS
2018	1	1	SAÚDE
2019	3	2	UTILIDADE PÚBLICA, CONSUMO CÍCLICO

2020	26	7	CONSUMO CÍCLICO, TECNOLOGIA DA INFORMAÇÃO, BENS INDUSTRIAIS, SAÚDE, FINANCEIRO E OUTROS, CONSUMO NÃO CÍCLICO, PETRÓLEO, GÁS E BIOCOMBUSTÍVEIS
2021	27	10	CONSUMO NÃO CÍCLICO, TECNOLOGIA DA INFORMAÇÃO, CONSUMO CÍCLICO, BENS INDUSTRIAIS, COMUNICAÇÕES, UTILIDADE PÚBLICA, MATERIAIS BÁSICOS, SAÚDE, FINANCEIRO E OUTROS, PETRÓLEO, GÁS E BIOCOMBUSTÍVEIS

Fonte: A autora (2023)

As análises realizadas revelaram que os anos de 2021 e 2006 apresentaram as maiores diversidades em termos de quantidade de setores participando de aberturas de capital, totalizando 9 e 10 setores, respectivamente. Entre esses setores, a distribuição anual pode ser observada no GRÁFICO 5. Cabe ressaltar que o setor de "consumo cíclico" se destacou ao longo de todo o período de 2004 a 2022 com 40 empresas ao longo do período, assim como o segmento "financeiro e outros" teve uma contagem expressiva (27).

GRÁFICO 5 - Distribuição dos setores por ano



Fonte: A autora (2023)

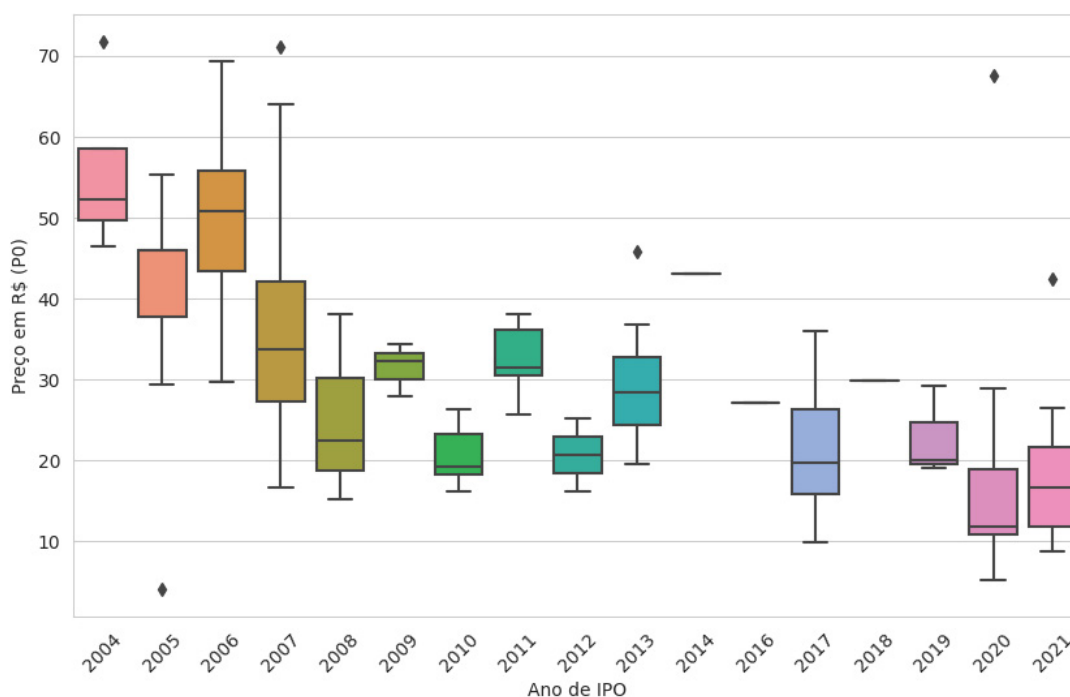
Ainda em relação ao GRÁFICO 5, pode-se observar uma tendência identificada em que empresas do mesmo setor de atuação influenciam as aberturas de capital umas das outras, conforme já apontado por Aghamolla; Thakor (2021). No entanto, é importante ressaltar que no Brasil setores como o financeiro apresentaram um aumento significativo em 2007, mesmo que nos anos anteriores não tenha havido uma quantidade expressiva de aberturas de capital nesse setor. Por outro lado, os setores de bens industriais e consumo cíclico demonstraram uma constância em suas aberturas de capital ao longo do período analisado. Já nos anos de 2020 e 2021, o

setor de tecnologia da informação começou a se destacar, assim como no último ano houve um destaque para o setor de comunicações em quantidade de IPOs.

Ademais, pode-se inferir a existência de um *Market Timing* de um IPO à medida que aumentam as quantidades de empresas em um determinado período, como já apontado por Rossi Jr.; Marotta (2010). No Brasil, os anos com mais aberturas de capital foram 2007, 2021, 2020 com 47, 27 e 26 empresas, respectivamente. Entretanto, além disso, a estrutura de capital das empresas pode auxiliar na compreensão do IPO, uma vez que elas buscam minimizar seus custos de capital e maximizar seus ganhos (NGOC; TIEN; THU, 2021).

Com isso, ao analisar o preço inicial (P0) registrado no prospecto definitivo das ações, constatou-se uma variabilidade ao longo dos anos. Entre 2008 e 2021, os preços das ações não ultrapassaram R\$ 40,00 (valor já corrigido para o ano base de 2022), com exceção de três *outliers*. Os maiores preços iniciais ocorreram nos anos de 2004 e 2007, sendo este último o ano com a maior diferença entre o valor mínimo e máximo, como é mostrado no GRÁFICO 6.

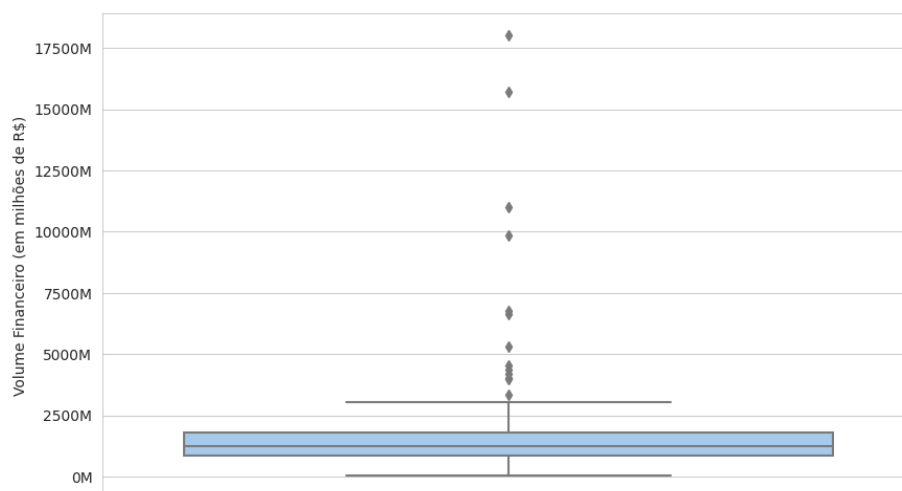
GRÁFICO 6 - Variabilidade de preços iniciais dos IPOs ao longo dos anos



Fonte: A autora (2023)

Diante desses resultados, o GRÁFICO 7 apresenta a distribuição do volume financeiro registrado no prospecto definitivo das empresas, ou seja, a dispersão do montante médio que as empresas buscaram captar por meio da abertura de capital.

GRÁFICO 7 - Variabilidade do volume financeiro registrado



Fonte: A autora (2023)

Diante disso, quatro anos concentraram mais de 50% do volume financeiro registrado, como se observa no QUADRO 13: 2007 (32,81%), 2020 (12,08%) e 2021 (11,38%).

QUADRO 13 – IPOs com volume registrado em bilhões e em percentual

ano_ipo	contagem_ipo	vol_financeiro_bilhoes	%_vol_financeiro
2004	4	7,64	2,60
2005	7	9,4	3,20
2006	16	22,77	7,74
2007	47	96,48	32,81
2008	3	1,99	0,68
2009	3	21,26	7,23
2010	5	8,7	2,96
2011	7	8,66	2,94
2012	2	1,33	0,45
2013	8	8,95	3,04
2014	1	0,67	0,23
2016	1	1,04	0,35
2017	7	22,41	7,62
2018	1	4,2	1,43
2019	3	9,58	3,26
2020	26	35,52	12,08
2021	27	33,47	11,38
Total	168	294,08	100,00

Fonte: A autora (2023) com dados da B3 (2022) sobre os IPOs de ações regidos pela ICVM 400 (CVM, 2003). Os valores monetários foram inflacionados pelo IPCA (2022 como ano-base) disponível no IPEADATA (2022).

Essas informações financeiras são essenciais para entender o desempenho individual das empresas, mas cabe compreender a situação econômico-financeira em cada um dos setores. Para isso, calculou-se a média, desvio padrão e quartis de alguns indicadores, que incluem: proporção de capital próprio, endividamento, risco

financeiro, investimento em ativos imobilizados, liquidez e rentabilidade, apresentados na íntegra no APÊNDICE²⁹. Na TABELA 3 seguem apenas as medianas (med):

TABELA 3 - Situação econômico-financeira por setores

	cap_terc_ cap_prio	cap_ter c_cap_t otal	comp_en dividame nto	imob_c ap_prio	imob_rec _nao_cor	GA F	LC	GA	ROE	RO A
BENS INDUSTRIAIS	2,44	0,76	0,48	2,10	0,86	3,35	1,21	1,12	0,26	0,06
COMUNICAÇÕES	4,79	0,88	0,39	2,98	1,02	3,93	0,80	0,82	0,33	0,07
CONSTRUÇÃO E TRANSPORTE	6,14	0,80	0,55	1,52	0,67	7,53	1,80	0,72	0,26	0,06
CONSUMO CÍCLICO	2,47	0,71	0,67	1,53	0,75	3,49	1,18	0,90	0,25	0,05
CONSUMO NÃO CÍCLICO	1,96	0,66	0,55	1,46	0,76	2,96	1,53	1,03	0,21	0,05
FINANCEIRO E OUTROS	3,08	0,83	0,70	1,58	0,90	4,08	1,12	0,22	0,21	0,03
MATERIAIS BÁSICOS	1,77	0,64	0,32	1,56	0,80	2,77	1,76	0,63	0,09	0,03
PETRÓLEO, GÁS E BIOCOMBUSTÍVEIS	1,24	0,55	0,23	1,79	0,83	2,24	1,92	0,53	- 0,10	- 0,03
SAÚDE	3,14	0,75	0,45	1,82	0,84	4,54	1,36	1,15	0,25	0,06
TECNOLOGIA DA INFORMAÇÃO	1,28	0,57	0,72	0,96	0,65	2,21	1,39	1,54	0,24	0,06
UTILIDADE PÚBLICA	1,66	0,66	0,29	2,20	1,02	2,66	0,95	0,51	0,10	0,01

Fonte: A autora (2023)

O setor com maior dependência de capital de terceiros em relação ao próprio é o de Construção e Transporte (med 6,14), seguido pelo de Comunicações (med 4,79). Isso quer dizer que no setor de Construção e Transporte, para cada 1 real do próprio capital que a empresa possui, ela depende de 6,14 reais de capital de terceiros para financiar suas atividades. No setor de Comunicações, para cada 1 real do próprio capital que a empresa possui, ela depende de 4,79 reais de capital de terceiros para financiar suas atividades.

No setor de consumo não cíclico, o valor de med 0,67 na composição do endividamento indica que o passivo circulante representa aproximadamente 67% do passivo total da empresa desse setor. No setor de tecnologia da informação, o valor de med 0,72 indica que o passivo circulante corresponde a cerca de 72% do passivo total da empresa desse setor.

No setor de comunicações, o valor de med 2,98 na imobilização do capital próprio indica que o ativo não circulante é aproximadamente 2,98 vezes maior que o patrimônio líquido da empresa desse setor. Isso significa que o valor dos ativos de longo prazo, como propriedades, equipamentos, investimentos de longo prazo e

²⁹ Na TABELA 13.

outros bens duráveis, é substancialmente maior do que o patrimônio líquido. No setor de tecnologia da informação, o valor de med 0,96 indica que o ativo não circulante é aproximadamente 0,96 vezes o tamanho do patrimônio líquido da empresa desse setor. Isso sugere que o valor dos ativos de longo prazo é quase igual ao patrimônio líquido.

A imobilização de recursos não correntes, por sua vez, no setor de comunicações e utilidade pública tem med 1,02, indicando que o ativo não circulante é aproximadamente igual à soma do passivo não circulante e do patrimônio líquido da empresa desse setor. O valor dos ativos de longo prazo, como propriedades, equipamentos, investimentos de longo prazo e outros bens duráveis, é aproximadamente equivalente à combinação das obrigações de longo prazo e do patrimônio líquido. Logo, as empresas desses setores possuem em média uma estrutura de financiamento equilibrada, em que os ativos de longo prazo estão alinhados com as fontes de financiamento de longo prazo.

A métrica "gaf" é usada para avaliar o multiplicador da alavancagem financeira de uma empresa, medindo a proporção dos ativos totais em relação ao patrimônio líquido. Valores mais elevados indicam uma maior dependência de ativos em relação ao patrimônio líquido, o que pode indicar um maior risco financeiro. No setor de construção e transporte, em que o "gaf" é igual a 7,53, a empresa depende significativamente de ativos em relação ao patrimônio líquido, o que pode indicar uma maior necessidade de financiamento externo para sustentar suas operações. Já no setor de tecnologia da informação, onde o "gaf" é igual a 2,21, a dependência de ativos em relação ao patrimônio líquido é menor, sugerindo uma estrutura financeira relativamente mais equilibrada.

Agora, a "LC" é utilizada para avaliar a capacidade de uma empresa em cumprir suas obrigações de curto prazo com seus ativos líquidos disponíveis. Valores maiores que 1 indicam que a empresa possui mais ativos circulantes do que obrigações de curto prazo, o que é considerado uma posição financeira favorável. No setor de comunicações, a med "LC" é igual a 0,80, o valor indica que a empresa possui ativos circulantes inferiores às suas obrigações de curto prazo, o que pode sugerir uma posição financeira relativamente menos favorável. Já no setor de utilidade pública, a med "LC" é igual a 0,95, a empresa possui ativos circulantes quase equivalentes às suas obrigações de curto prazo, o que indica uma posição financeira mais equilibrada.

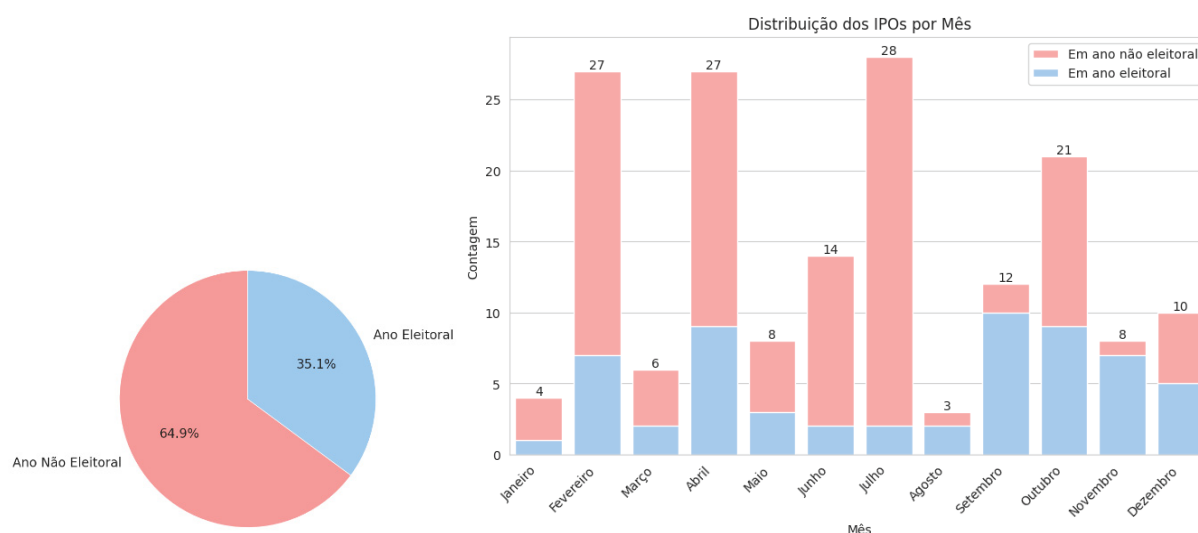
O ROE é uma métrica importante para os investidores, pois indica a eficiência da empresa em utilizar seus recursos financeiros e a rentabilidade que ela está obtendo em relação ao capital investido pelos acionistas. Diante disso, vê-se que a maioria dos setores apresentam a med do ROE próxima a 0,20. Por sua vez, o ROA, é utilizado para avaliar a eficiência da empresa em utilizar seus ativos para gerar lucros. Quanto maior o ROA, melhor é a capacidade da empresa de gerar lucro em relação aos seus ativos. No setor de comunicações, por exemplo, as empresas estão gerando um retorno de cerca de 7% sobre o seu ativo total.

Com isso, observa-se que cada setor apresenta características distintas em relação aos indicadores financeiros.

4.1.1.2 Conjuntura geral

No que concerne ao momento de realização de IPOs, é possível observar que 64,9% dessas operações ocorreram em anos que não são marcados por eleições (com mais incertezas). Além disso, os meses de fevereiro, abril, julho e outubro destacam-se como períodos de maior incidência de IPOs, sendo que neste último mês ocorreu uma divisão quase igualitária entre anos eleitorais e não eleitorais, conforme indicado pela combinação do GRÁFICO 8.

GRÁFICO 8 - Distribuição de IPOs em anos eleitorais e em meses

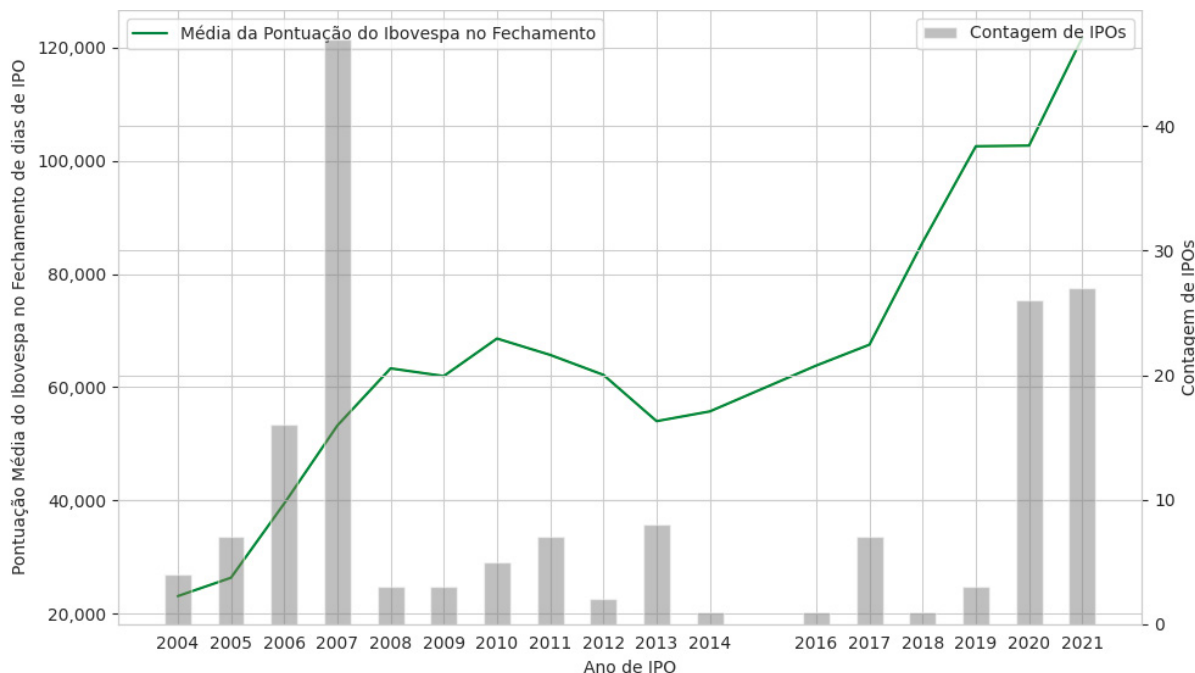


Fonte: A autora (2023)

Em compensação, ao analisar o mercado representado pela pontuação da Bolsa de Valores, pode-se observar uma tendência de alta na pontuação média

durante os dias de abertura de capital em comparação ao ano anterior no GRÁFICO 9. No entanto, é importante ressaltar que entre os anos de 2014 e 2016, apesar do aumento dessa pontuação, a quantidade de IPOs permaneceu estável. Da mesma forma, entre os anos de 2017 e 2019, houve uma redução nas aberturas de capital:

GRÁFICO 9 - Média de pontuação da Bolsa de Valores no Fechamento e Contagem de IPOs por ano

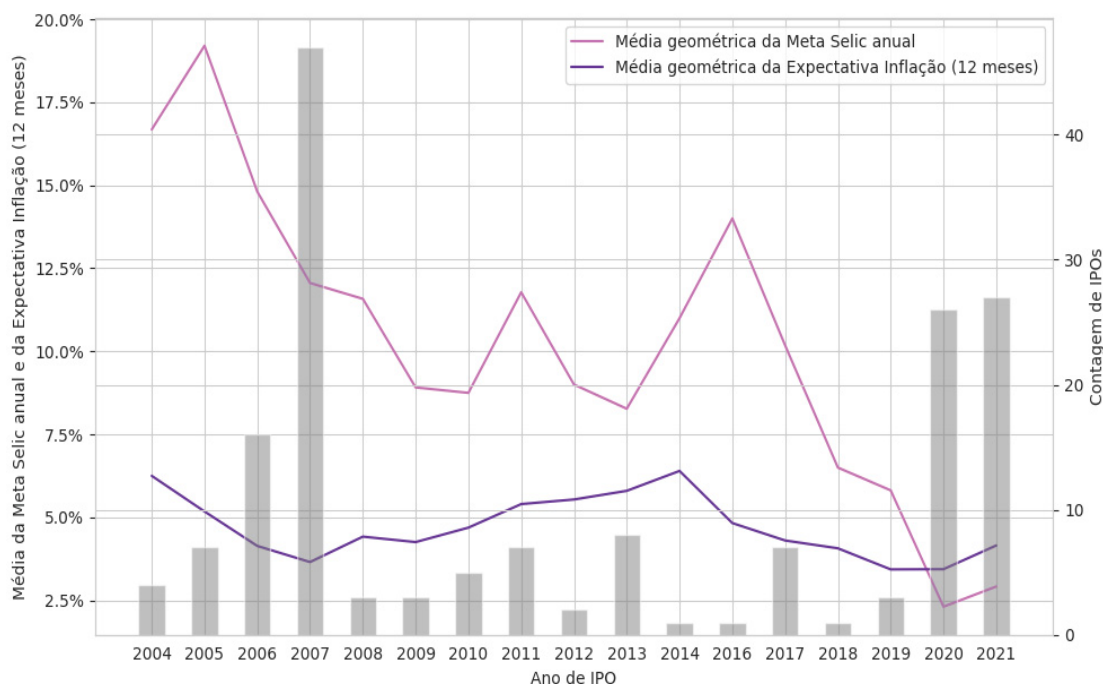


Fonte: A autora (2023)

De acordo com estudos realizados por Graham; Harvey (2001) e Matsuo; Eid Jr (2004), é comum que as empresas busquem acessar o mercado de capitais quando as taxas de juros estão em níveis baixos. Nesse contexto, a análise do GRÁFICO 10 evidencia que a média da meta anual da taxa Selic nos dias de abertura de capital ao longo dos anos corrobora essa afirmação. Verifica-se que nos períodos de 2004-2005 e 2014-2016, em que houve menos aberturas de capital, as metas da taxa básica da economia apresentaram momentos de crescimento.

No entanto, é relevante destacar que, durante os períodos de queda da taxa, como observado em 2008-2010 e 2011-2012, não houve um aumento significativo de IPOs. Por outro lado, nos anos de 2020 e 2021, mesmo em um contexto marcado pela pandemia da Covid-19 e com expectativas de aumento da inflação nos próximos 12 meses, ocorreu um aumento significativo no número de aberturas de capital.

GRÁFICO 10 - Média da Meta Selic anual e da Expectativa da Inflação (próximos 12 meses) no dia do IPO e Contagem de IPOs por ano

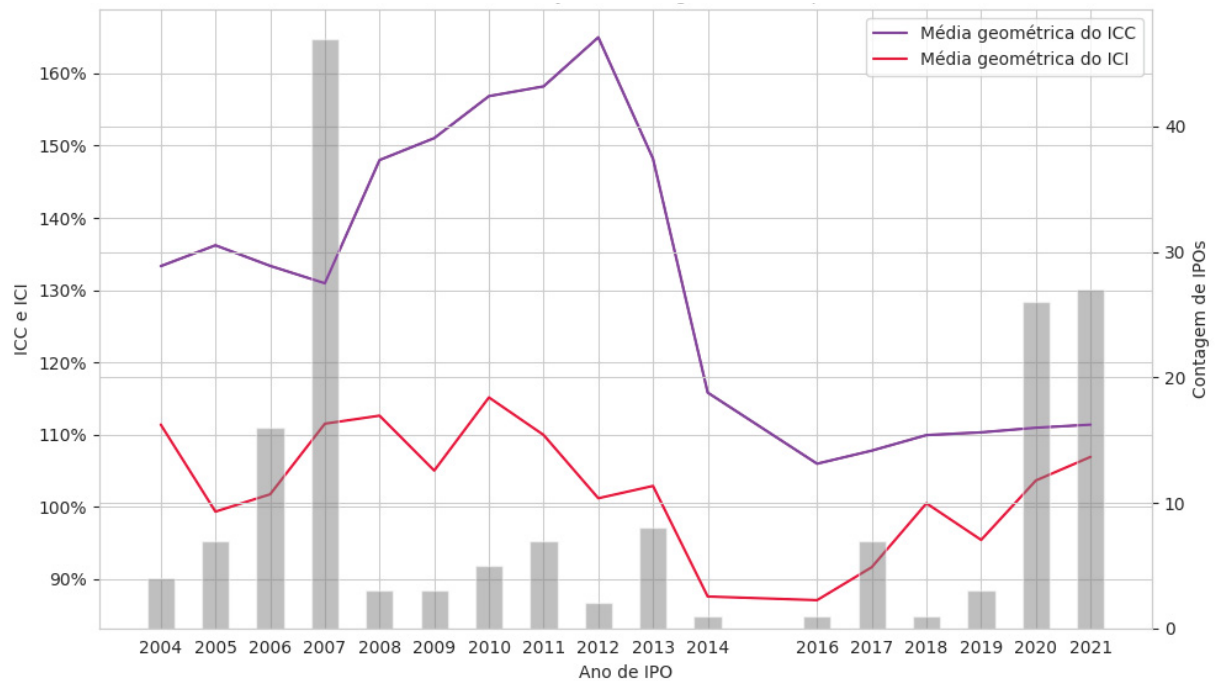


Fonte: A autora (2023)

Isso evidencia que, além da análise macroeconômica estritamente focada, os índices de confiança específicos de setores em relação ao cenário econômico do país podem oferecer informações esclarecedoras sobre as motivações por trás das aberturas de capital. A confiança na situação econômica do país pode ser exemplificada pelas expectativas atuais e futuras de consumidores e empresários, por meio do Índice de Confiança do Consumidor (ICC), e pelo Índice de Confiança da Indústria (ICI), que é um indicador-síntese proveniente da Sondagem Conjuntural da Indústria de Transformação. No GRÁFICO 11 podem ser visualizadas as médias das confianças nos momentos de abertura de capital divididas por ano.

Ao analisar o GRÁFICO 11, é possível notar que a confiança do consumidor e da indústria apresentaram variações significativas ao longo dos anos, embora nem sempre estejam diretamente relacionadas ao número de aberturas de capital. Apesar disso, desde 2019, observou-se um sinal de crescimento da confiança e um aumento expressivo de IPOs no Brasil, coincidindo com um período de taxas de juros baixas e expectativas de inflação crescente.

GRÁFICO 11 - Índices de confiança e Contagem de IPOs por ano



Fonte: A autora (2023)

Com base nisso, os dados coletados fornecem uma visão geral para justificar a definição das variáveis de entrada do modelo proposto. Na próxima subseção, portanto, será explorada a seleção e definição dessas variáveis de entrada da RNA, considerando suas propriedades e interações.

4.1.2 Explorações das variáveis

Após a aplicação dos procedimentos de tratamento e codificação mencionados no capítulo 3, foram mantidas 37 variáveis com variabilidade maior que zero ao longo das observações. Com o objetivo de verificar a relação dessas variáveis com a variável alvo (compra ou não compra das ações), foi calculada a correlação entre elas. A TABELA 4 revela que apenas 9 das possíveis variáveis explicativas apresentaram uma relação superior a 10% com a variável dependente.

TABELA 4 - Correlação das variáveis explicativas com a alvo

id	Variável	Correlação com Y	id	Variável	Correlação com Y
1	ROE	0,190808	19	LC	0,05507
2	ROA	0,176406	20	avAC	0,048649
3	P0	0,157059	21	avPNC	0,041677
4	idade_emp_ipo	0,130889	22	avANC	0,040265

5	vol_financeiro	0,115299	23	avPC	0,030704
6	ipca_acum_12_meses	0,111015	24	tipo_oferta	0,030183
7	gaf	0,110725	25	meta_selic_dia_reg_ipo	0,028042
8	cap_terc_cap_prio	0,109003	26	comp_endividamento	0,026342
9	icc	0,10048	27	avPL	0,02591
10	imob_cap_prio	0,083884	28	qtd_titulos	0,023935
11	vol_diario_ibov_reg_ipo	0,082703	29	ano_ipo	0,021648
12	ano_eleitoral	0,078621	30	exp_selic_mes_pos_ipo	0,02089
13	ICE	0,077544	31	imob_rec_nao_cor	0,020086
14	lider	0,066947	32	segmento	0,01099
15	indic_incerteza_eco	0,063596	33	GA	0,008102
16	pont_ibov_fech_reg_ipo	0,056158	34	setor	0,006563
17	mes	0,055982	35	exp_inflacao_12_meses	0,004769
18	ici	0,055177	36	uf	0,004482
			37	cap_terc_cap_total	0,001111

Fonte: A autora (2023)

Além de examinar a relação das variáveis explicativas com a variável dependente, é importante considerar também a existência de relações entre as próprias variáveis explicativas. A presença de multicolinearidade entre elas pode representar um desafio. Portanto, foi realizado o cálculo do VIF, mostrado na TABELA 5:

TABELA 5 - VIF

id	Variável	VIF	id	Variável	VIF
22	avANC	233,329511	30	exp_selic_mes_pos_ipo	6,051977
20	avAC	227,183666	9	icc	5,66559
7	gaf	83,151458	5	vol_financeiro	5,599063
29	ano_ipo	58,984257	28	qtd_titulos	4,962407
16	pont_ibov_fech_reg_ipo	40,317207	1	ROE	4,349638
8	cap_terc_cap_prio	34,616323	11	vol_diario_ibov_reg_ipo	3,924312
25	meta_selic_dia_reg_ipo	27,458967	3	P0	3,305402
13	ICE	20,503803	2	ROA	2,803472
23	avPC	19,06168	31	imob_rec_nao_cor	2,530907
10	imob_cap_prio	17,378249	33	GA	2,420179
37	cap_terc_cap_total	13,295615	17	mes	2,418977
21	avPNC	11,614763	12	ano_eleitoral	1,901275
15	indic_incerteza_eco	9,833877	24	tipo_oferta	1,540782
18	ici	9,183281	14	lider	1,523683
35	exp_inflacao_12_meses	8,097395	34	setor	1,490969
6	ipca_acum_12_meses	8,063461	19	LC	1,395118
26	comp_endividamento	7,523974	32	segmento	1,286851
27	avPL	6,94408	36	uf	1,274615
			4	idade_emp_ipo	1,222246

Fonte: A autora (2023)

Os resultados indicam que algumas variáveis apresentam uma alta correlação, especialmente devido ao método de cálculo, que envolve valores de origem iguais das demonstrações financeiras. Para efeitos de teste, foram eliminadas as variáveis com VIF acima de 20 e, posteriormente, acima de 10, seguindo a abordagem apresentada por Gujarati; Porter (2011). Com isso, **restaram as 27 variáveis** elencadas da TABELA 6:

TABELA 6 - VIF das variáveis selecionadas

id	Variável	VIF	id	Variável	VIF
15	indic_incerteza_eco	6,420076	2	ROA	2,097697
21	avPNC	5,473024	33	GA	1,954248
26	comp_endividamento	4,883195	27	avPL	1,94366
5	vol_financeiro	4,654247	12	ano_eleitoral	1,853619
28	qtd_titulos	4,373809	18	ici	1,838284
35	exp_inflacao_12_meses	4,144844	17	mes	1,664085
9	icc	3,902267	31	imob_rec_nao_cor	1,52756
6	ipca_acum_12_meses	3,775535	24	tipo_oferta	1,321485
11	vol_diario_ibov_reg_ipo	3,245423	19	LC	1,313422
30	exp_selic_mes_pos_ipo	3,22857	34	setor	1,297319
37	cap_terc_cap_total	3,01386	14	lider	1,274383
3	P0	2,685192	36	uf	1,232846
1	ROE	2,251332	4	idade_emp_ipo	1,165423
			32	segmento	1,162322

Fonte: A autora (2023)

Dentre essas variáveis finais, destacam-se algumas de natureza macroeconômica, como "indic_incerteza_eco" e "exp_inflacao_12_meses", que refletem o ambiente econômico e as expectativas inflacionárias, influenciando o desempenho das empresas (ALMEIDA, 2013; CORREIA, 2015; KHATRI, 2017; AGRAWAL, 2021).

Adicionalmente, variáveis relacionadas à estrutura financeira das empresas, como "comp_endividamento" e "cap_terc_cap_total", são fundamentais para compreender a saúde financeira e a alavancagem das companhias (ADEBIYI *et al.*, 2012; ASSAF NETO, 2000; JAIN; NAG, 1995). Outras variáveis, como "vol_financeiro" e "qtd_titulos", estão associadas ao mercado de capitais e podem impactar o interesse dos investidores nas ofertas públicas (ALMEIDA, 2013; CORREIA, 2015).

Aspectos internos das empresas também são relevantes, como indicam as variáveis "ROE" e "ROA", que medem a eficiência operacional e a rentabilidade dos ativos, enquanto "setor" e "segmento" fornecem informações sobre o contexto

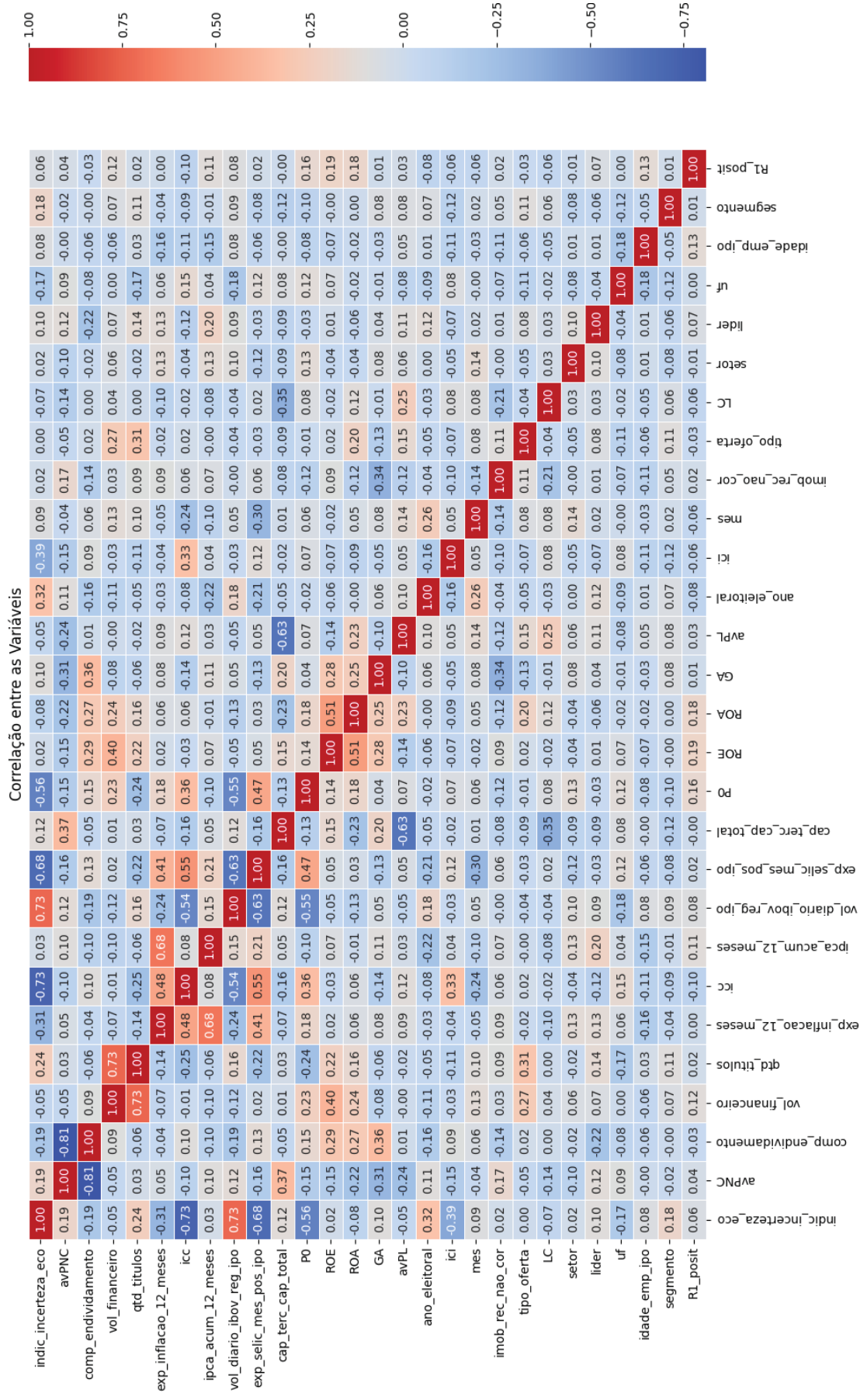
empresarial (ALMEIDA, 2013; CORREIA, 2015; JAIN; NAG, 1998; LEAL, 2005; MARTINS, V. A.; SILVA; NARDI, 2006).

Além disso, variáveis específicas como "vol_diario_ibov_reg_ipo" e "exp_selic_mes_pos_ipo" capturam a volatilidade do mercado acionário e as expectativas em relação à taxa Selic, respectivamente, influenciando o cenário das IPOs (ALMEIDA, 2013; CORREIA, 2015; KHATRI, 2017; AGRAWAL, 2021).

Com efeito, "ano_eleitoral", "uf", "idade_emp_ipo", "P0", "ici", "mes", "imob_rec_nao_cor", "tipo_oferta", "LC" e "lider" também são variáveis que complementam a pesquisa e fornecem uma análise abrangente sobre os fatores que podem influenciar o sucesso e a sobrevivência das empresas após a realização de um IPO (COLAK *et al.*, 2021; MARTINS *et al.*, 2006; BEATTY; RITTER, 1986).

A partir disso, as correlações entre diferentes variáveis pode ser visualizada no GRÁFICO 12, cujo destaque pode ser dado em resumo para: "indic_incerteza_eco" tem uma forte correlação negativa com "icc" (-0,73) e uma correlação negativa moderada com "exp_inflacao_12_meses" (-0,30); "icc" tem uma forte correlação positiva com "exp_selic_mes_pos_ipo" (0,55) e uma correlação positiva moderada com "exp_inflacao_12_meses" (0,48); "ROE" tem uma correlação positiva moderada com "ROA" (0,51) e com "vol_financeiro" (0,40) e uma correlação positiva fraca com "GA" (0,046); "ano_eleitoral" tem uma correlação positiva com "mes" (0,26); "LC" tem uma correlação negativa com "cap_terc_cap_total" e nula com "qtde_titulos"; "ROA", "ROE" e "com_endividamento" não apresentam correlação (0,000) com "segmento".

GRÁFICO 12 - Correlação das variáveis selecionadas



Fonte: A autora (2023)

Com efeito, a partir dessa seleção, as 27 variáveis foram combinadas (Comb) de 55 maneiras diferentes. A partir delas, foram elencadas as quantidades de neurônios que a camada oculta da RNA teria, conforme apresentado no QUADRO 14.

QUADRO 14 - Variáveis selecionadas para cada uma das 55 combinações

Comb nº	Variaveis_Combinacao	Entradas	Neuronios_Camada
0	['segmento', 'ROE', 'idade_emp_ipo', 'P0', 'setor', 'imob_rec_ao_cor', 'comp_endividamento', 'avPL', 'exp_inflacao_12_meses', 'LC', 'lider', 'vol_financeiro', 'exp_selic_mes_pos_ipo', 'cap_terc_cap_total']	14	(10,)
1	['cap_terc_cap_total', 'setor', 'vol_financeiro', 'exp_selic_mes_pos_ipo', 'ROA', 'P0', 'idade_emp_ipo', 'qtd_titulos', 'mes', 'exp_inflacao_12_meses', 'ici', 'GA', 'lider', 'ano_eleitoral', 'uf', 'ROE']	16	(12,)
2	['qtd_titulos', 'ipca_acum_12_meses', 'lider', 'exp_selic_mes_pos_ipo', 'vol_financeiro', 'segmento', 'comp_endividamento', 'setor', 'icc', 'imob_rec_ao_cor', 'avPNC', 'avPL']	12	(8,)
3	['icc', 'vol_diario_ibov_reg_ipo', 'mes', 'lider', 'ipca_acum_12_meses', 'indic_incerteza_eco', 'avPL', 'P0', 'uf', 'vol_financeiro', 'exp_inflacao_12_meses', 'comp_endividamento']	12	(8,)
4	['qtd_titulos', 'P0', 'exp_selic_mes_pos_ipo', 'comp_endividamento', 'ROA', 'segmento', 'ici', 'mes', 'uf', 'indic_incerteza_eco', 'tipo_oferta', 'ano_eleitoral', 'exp_inflacao_12_meses', 'avPL', 'ROE', 'setor', 'ipca_acum_12_meses', 'avPNC', 'vol_financeiro', 'idade_emp_ipo', 'vol_diario_ibov_reg_ipo', 'icc', 'imob_rec_ao_cor', 'lider', 'cap_terc_cap_total', 'GA', 'LC']	27	(15,)
5	['GA', 'exp_inflacao_12_meses', 'cap_terc_cap_total', 'lider', 'comp_endividamento', 'ici', 'ano_eleitoral', 'imob_rec_ao_cor', 'ipca_acum_12_meses', 'setor', 'qtd_titulos']	11	(8,)
6	['vol_diario_ibov_reg_ipo', 'GA', 'exp_selic_mes_pos_ipo', 'ano_eleitoral', 'qtd_titulos', 'ipca_acum_12_meses', 'uf', 'comp_endividamento', 'setor', 'tipo_oferta', 'ici', 'indic_incerteza_eco', 'imob_rec_ao_cor', 'ROA', 'mes', 'icc', 'ROE', 'LC', 'idade_emp_ipo']	19	(13,)
7	['segmento', 'comp_endividamento', 'qtd_titulos', 'P0', 'avPL', 'lider', 'icc', 'ipca_acum_12_meses', 'exp_inflacao_12_meses', 'GA', 'imob_rec_ao_cor']	11	(6,)
8	['LC', 'ROE', 'uf', 'icc', 'mes', 'exp_inflacao_12_meses', 'lider']	7	(5,)
9	['exp_selic_mes_pos_ipo', 'ici', 'exp_inflacao_12_meses', 'icc', 'comp_endividamento', 'LC', 'setor', 'cap_terc_cap_total']	8	(5,)
10	['GA', 'ROA', 'lider', 'exp_selic_mes_pos_ipo', 'P0', 'segmento', 'mes', 'vol_financeiro', 'LC', 'idade_emp_ipo', 'ipca_acum_12_meses', 'ROE', 'icc', 'imob_rec_ao_cor', 'ici', 'setor', 'ano_eleitoral', 'tipo_oferta', 'avPNC', 'qtd_titulos', 'comp_endividamento']	21	(12,)
11	['qtd_titulos', 'GA', 'icc', 'avPNC', 'comp_endividamento', 'ROA', 'uf', 'cap_terc_cap_total', 'vol_financeiro', 'indic_incerteza_eco', 'tipo_oferta', 'ipca_acum_12_meses', 'avPL', 'vol_diario_ibov_reg_ipo', 'ici', 'ano_eleitoral', 'ROE', 'exp_selic_mes_pos_ipo', 'setor', 'mes', 'lider', 'exp_inflacao_12_meses', 'imob_rec_ao_cor', 'idade_emp_ipo', 'LC', 'segmento', 'P0']	27	(14,)

12	['mes', 'imob_rec_nao_cor', 'ipca_acum_12_meses', 'vol_financeiro', 'avPNC', 'uf']	6	(5,)
13	['uf', 'qtd_titulos', 'GA', 'cap_terc_cap_total', 'idade_emp_ipo', 'vol_diario_ibov_reg_ipo', 'ROE', 'P0', 'exp_selic_mes_pos_ipo', 'ici', 'ROA', 'indic_incerteza_eco', 'mes', 'segmento', 'exp_inflacao_12_meses', 'LC']	16	(11,)
14	['indic_incerteza_eco', 'avPL', 'comp_endividamento', 'uf', 'ipca_acum_12_meses', 'ano_eleitoral', 'exp_inflacao_12_meses', 'qtd_titulos', 'idade_emp_ipo', 'vol_diario_ibov_reg_ipo', 'exp_selic_mes_pos_ipo', 'segmento', 'ici', 'LC', 'mes', 'lider', 'avPNC', 'ROE']	18	(10,)
15	['ipca_acum_12_meses', 'setor', 'segmento', 'lider', 'comp_endividamento', 'imob_rec_nao_cor', 'avPNC', 'P0', 'avPL', 'vol_diario_ibov_reg_ipo', 'ROA', 'vol_financeiro', 'indic_incerteza_eco', 'uf', 'exp_selic_mes_pos_ipo', 'ici', 'qtd_titulos', 'ano_eleitoral', 'mes']	19	(15,)
16	['cap_terc_cap_total', 'ici', 'exp_selic_mes_pos_ipo', 'setor', 'LC', 'P0', 'exp_inflacao_12_meses', 'GA', 'idade_emp_ipo', 'ano_eleitoral', 'segmento', 'indic_incerteza_eco', 'lider', 'ROE', 'ROA', 'avPL', 'imob_rec_nao_cor']	17	(9,)
17	['segmento', 'setor', 'vol_diario_ibov_reg_ipo', 'imob_rec_nao_cor', 'exp_inflacao_12_meses']	5	(4,)
18	['mes', 'GA', 'P0', 'icc', 'ipca_acum_12_meses', 'tipo_oferta', 'idade_emp_ipo', 'comp_endividamento', 'ano_eleitoral', 'imob_rec_nao_cor', 'setor', 'exp_inflacao_12_meses', 'exp_selic_mes_pos_ipo', 'avPNC', 'LC', 'qtd_titulos', 'cap_terc_cap_total', 'vol_financeiro', 'ici', 'ROE', 'vol_diario_ibov_reg_ipo', 'lider', 'avPL', 'ROA', 'indic_incerteza_eco']	25	(17,)
19	['GA', 'vol_financeiro', 'cap_terc_cap_total', 'lider', 'ano_eleitoral', 'qtd_titulos', 'ici', 'uf', 'exp_selic_mes_pos_ipo', 'segmento', 'idade_emp_ipo', 'avPNC', 'imob_rec_nao_cor', 'indic_incerteza_eco', 'tipo_oferta', 'icc', 'ipca_acum_12_meses', 'ROE', 'avPL', 'setor', 'comp_endividamento', 'P0', 'ROA', 'LC', 'exp_inflacao_12_meses']	25	(18,)
20	['indic_incerteza_eco', 'LC', 'comp_endividamento', 'cap_terc_cap_total', 'exp_selic_mes_pos_ipo', 'ici', 'tipo_oferta', 'icc', 'ROE', 'lider', 'uf', 'setor', 'mes', 'imob_rec_nao_cor', 'avPNC', 'avPL', 'ROA', 'ano_eleitoral', 'P0', 'vol_diario_ibov_reg_ipo']	20	(11,)
21	['LC', 'qtd_titulos', 'exp_inflacao_12_meses', 'setor', 'comp_endividamento', 'idade_emp_ipo']	6	(5,)
22	['ipca_acum_12_meses', 'imob_rec_nao_cor', 'GA', 'LC', 'setor', 'ROE', 'avPNC', 'avPL', 'lider', 'P0', 'mes', 'tipo_oferta', 'vol_diario_ibov_reg_ipo', 'comp_endividamento', 'ROA', 'indic_incerteza_eco', 'qtd_titulos']	17	(12,)
23	['indic_incerteza_eco', 'exp_inflacao_12_meses', 'ano_eleitoral', 'idade_emp_ipo', 'exp_selic_mes_pos_ipo', 'comp_endividamento', 'ROA', 'vol_financeiro', 'ici', 'qtd_titulos', 'uf', 'mes', 'LC', 'imob_rec_nao_cor', 'avPNC', 'vol_diario_ibov_reg_ipo', 'GA', 'segmento', 'ROE', 'tipo_oferta', 'ipca_acum_12_meses']	21	(14,)
24	['ipca_acum_12_meses', 'exp_selic_mes_pos_ipo', 'lider', 'imob_rec_nao_cor', 'comp_endividamento', 'cap_terc_cap_total', 'idade_emp_ipo', 'LC', 'setor', 'ROA', 'exp_inflacao_12_meses', 'indic_incerteza_eco', 'avPL', 'GA', 'segmento', 'tipo_oferta', 'qtd_titulos', 'vol_diario_ibov_reg_ipo', 'P0']	19	(11,)

25	['imob_rec_nao_cor', 'avPL', 'icc', 'avPNC', 'exp_selic_mes_pos_ipo', 'tipo_oferta', 'vol_diario_ibov_reg_ipo', 'ici', 'qtd_titulos', 'ROA', 'ROE']	11	(8,)
26	['ipca_acum_12_meses', 'tipo_oferta', 'exp_inflacao_12_meses', 'qtd_titulos', 'LC', 'vol_financeiro', 'P0', 'ROA', 'ici', 'imob_rec_nao_cor', 'segmento', 'icc', 'mes', 'indic_incerteza_eco', 'avPL', 'ROE', 'uf', 'GA', 'avPNC', 'vol_diario_ibov_reg_ipo', 'ano_eleitoral', 'idade_emp_ipo']	22	(15,)
27	['mes', 'imob_rec_nao_cor', 'exp_selic_mes_pos_ipo', 'comp_endividamento', 'indic_incerteza_eco', 'ROE', 'LC', 'ici', 'ipca_acum_12_meses', 'qtd_titulos', 'lider', 'tipo_oferta', 'GA', 'idade_emp_ipo', 'icc', 'uf', 'avPL', 'setor', 'vol_diario_ibov_reg_ipo', 'ano_eleitoral']	20	(14,)
28	['imob_rec_nao_cor', 'GA', 'qtd_titulos', 'comp_endividamento', 'ici', 'exp_selic_mes_pos_ipo', 'exp_inflacao_12_meses', 'P0', 'ROA', 'lider', 'vol_diario_ibov_reg_ipo', 'ipca_acum_12_meses', 'indic_incerteza_eco']	13	(9,)
29	['lider', 'segmento', 'GA', 'P0', 'uf', 'indic_incerteza_eco', 'comp_endividamento', 'idade_emp_ipo', 'mes', 'ici', 'qtd_titulos']	11	(6,)
30	['avPNC', 'GA', 'P0', 'tipo_oferta', 'ROA', 'mes', 'icc', 'uf', 'LC', 'imob_rec_nao_cor', 'lider', 'indic_incerteza_eco', 'ipca_acum_12_meses', 'cap_terc_cap_total']	14	(10,)
31	['P0', 'uf', 'ROA', 'setor', 'exp_inflacao_12_meses', 'imob_rec_nao_cor', 'avPL', 'idade_emp_ipo', 'cap_terc_cap_total', 'lider', 'tipo_oferta', 'comp_endividamento']	12	(7,)
32	['exp_inflacao_12_meses', 'lider', 'cap_terc_cap_total', 'mes', 'P0', 'avPNC', 'LC', 'ROA', 'idade_emp_ipo', 'setor', 'comp_endividamento']	11	(8,)
33	['cap_terc_cap_total', 'indic_incerteza_eco', 'ROE', 'segmento', 'P0', 'mes', 'vol_diario_ibov_reg_ipo']	7	(5,)
34	['GA', 'comp_endividamento', 'vol_financeiro', 'avPNC', 'vol_diario_ibov_reg_ipo', 'ano_eleitoral', 'avPL', 'ipca_acum_12_meses', 'ROE', 'P0', 'exp_inflacao_12_meses', 'icc', 'imob_rec_nao_cor', 'mes', 'lider', 'qtd_titulos']	16	(9,)
35	['avPNC', 'idade_emp_ipo', 'exp_selic_mes_pos_ipo', 'icc', 'ipca_acum_12_meses', 'qtd_titulos', 'imob_rec_nao_cor', 'lider', 'ano_eleitoral', 'setor', 'comp_endividamento', 'tipo_oferta', 'segmento']	13	(9,)
36	['avPNC', 'indic_incerteza_eco', 'ipca_acum_12_meses', 'ROE', 'cap_terc_cap_total', 'exp_inflacao_12_meses', 'imob_rec_nao_cor', 'tipo_oferta', 'icc']	9	(5,)
37	['uf', 'mes', 'avPL', 'lider', 'vol_diario_ibov_reg_ipo', 'ici', 'indic_incerteza_eco', 'tipo_oferta', 'GA', 'segmento', 'ROE', 'ano_eleitoral', 'ROA']	13	(9,)
38	['ici', 'imob_rec_nao_cor', 'vol_diario_ibov_reg_ipo', 'uf', 'LC']	5	(3,)
39	['lider', 'LC', 'uf', 'setor', 'exp_selic_mes_pos_ipo', 'exp_inflacao_12_meses', 'idade_emp_ipo']	7	(6,)
40	['icc', 'LC']	2	(2,)
41	['LC', 'cap_terc_cap_total', 'GA', 'ano_eleitoral', 'icc', 'imob_rec_nao_cor', 'qtd_titulos', 'vol_diario_ibov_reg_ipo', 'indic_incerteza_eco', 'uf', 'setor', 'ici', 'tipo_oferta', 'mes', 'avPL', 'ipca_acum_12_meses', 'exp_inflacao_12_meses', 'vol_financeiro', 'ROA', 'ROE', 'segmento', 'exp_selic_mes_pos_ipo', 'idade_emp_ipo', 'lider']	24	(14,)
42	['cap_terc_cap_total', 'setor', 'exp_inflacao_12_meses', 'idade_emp_ipo', 'vol_financeiro', 'GA', 'comp_endividamento', 'icc', 'qtd_titulos', 'vol_diario_ibov_reg_ipo', 'avPL', 'segmento', 'ici', 'tipo_oferta', 'ano_eleitoral', 'ipca_acum_12_meses']	22	(16,)

	'exp_selic_mes_pos_ipo', 'uf', 'ROE', 'avPNC', 'indic_incerteza_eco', 'P0']		
43	['qtd_titulos', 'ROA', 'icc', 'ROE', 'tipo_oferta', 'setor', 'ano_eleitoral', 'vol_financeiro', 'imob_rec_nao_cor', 'ici', 'indic_incerteza_eco', 'mes', 'LC', 'P0', 'lider']	15	(12,)
44	['uf', 'P0', 'avPNC', 'exp_selic_mes_pos_ipo', 'cap_terc_cap_total', 'imob_rec_nao_cor', 'segmento', 'ROE', 'indic_incerteza_eco', 'exp_inflacao_12_meses', 'LC', 'ici', 'vol_financeiro', 'vol_diario_ibov_reg_ipo', 'avPL', 'qtd_titulos', 'ROA', 'ano_eleitoral', 'lider', 'mes', 'ipca_acum_12_meses', 'GA', 'icc', 'setor']	24	(13,)
45	['GA', 'imob_rec_nao_cor', 'exp_inflacao_12_meses', 'LC', 'ROE', 'idade_emp_ipo', 'qtd_titulos', 'tipo_oferta', 'comp_endividamento', 'mes', 'vol_financeiro', 'vol_diario_ibov_reg_ipo', 'ipca_acum_12_meses', 'segmento', 'P0', 'ici', 'indic_incerteza_eco', 'avPNC', 'uf', 'ano_eleitoral', 'setor', 'cap_terc_cap_total', 'ROA', 'lider', 'icc', 'exp_selic_mes_pos_ipo']	26	(18,)
46	['uf', 'P0', 'mes', 'lider', 'vol_diario_ibov_reg_ipo', 'exp_selic_mes_pos_ipo', 'avPNC', 'qtd_titulos', 'cap_terc_cap_total', 'indic_incerteza_eco', 'setor', 'icc', 'ipca_acum_12_meses', 'ici', 'comp_endividamento', 'GA', 'tipo_oferta', 'exp_inflacao_12_meses', 'ROA', 'ano_eleitoral', 'ROE', 'imob_rec_nao_cor', 'idade_emp_ipo', 'segmento']	24	(15,)
47	['lider', 'vol_financeiro', 'ici', 'P0', 'comp_endividamento', 'avPL', 'qtd_titulos', 'mes', 'imob_rec_nao_cor', 'tipo_oferta', 'icc', 'idade_emp_ipo', 'LC', 'setor', 'cap_terc_cap_total', 'avPNC', 'uf', 'exp_inflacao_12_meses', 'ROE', 'GA', 'ipca_acum_12_meses', 'vol_diario_ibov_reg_ipo', 'indic_incerteza_eco', 'ano_eleitoral', 'exp_selic_mes_pos_ipo', 'ROA', 'segmento']	27	(19,)
48	['ano_eleitoral', 'lider', 'P0', 'exp_selic_mes_pos_ipo', 'exp_inflacao_12_meses', 'ici', 'LC', 'qtd_titulos', 'ROE', 'GA', 'avPNC', 'uf', 'ipca_acum_12_meses', 'setor', 'indic_incerteza_eco', 'avPL', 'mes', 'ROA', 'icc', 'imob_rec_nao_cor', 'idade_emp_ipo', 'vol_financeiro']	22	(16,)
49	['ROE', 'P0', 'setor', 'imob_rec_nao_cor', 'ano_eleitoral', 'segmento', 'comp_endividamento', 'GA', 'exp_inflacao_12_meses', 'idade_emp_ipo', 'qtd_titulos', 'lider']	12	(9,)
50	['avPL', 'tipo_oferta', 'icc', 'ano_eleitoral', 'uf', 'ROE']	6	(4,)
51	['exp_selic_mes_pos_ipo', 'comp_endividamento', 'tipo_oferta', 'ROA', 'avPL', 'cap_terc_cap_total', 'uf', 'LC']	8	(5,)
52	['icc', 'comp_endividamento', 'tipo_oferta', 'segmento', 'vol_financeiro', 'avPNC']	6	(5,)
53	['imob_rec_nao_cor', 'qtd_titulos', 'vol_diario_ibov_reg_ipo']	3	(2,)
54	['mes', 'lider', 'vol_financeiro', 'ici', 'P0', 'ipca_acum_12_meses', 'ROA', 'GA', 'comp_endividamento', 'ROE', 'avPL', 'ano_eleitoral', 'segmento', 'qtd_titulos']	14	(11,)

Fonte: A autora (2023)

Essa abordagem, explicada na Etapa 4 da pesquisa no capítulo 3, permite explorar diversas combinações de variáveis para identificar quais delas apresentam um desempenho mais relevante na predição da variável alvo, bem como para avaliar a eficácia e a precisão do modelo proposto.

Na próxima seção, a discussão dos resultados da RNA propriamente ditos, será possível analisar a aplicação do modelo, explorando métricas de desempenho, validade e acurácia.

4.2 DISCUSSÃO DOS RESULTADOS DA RNA

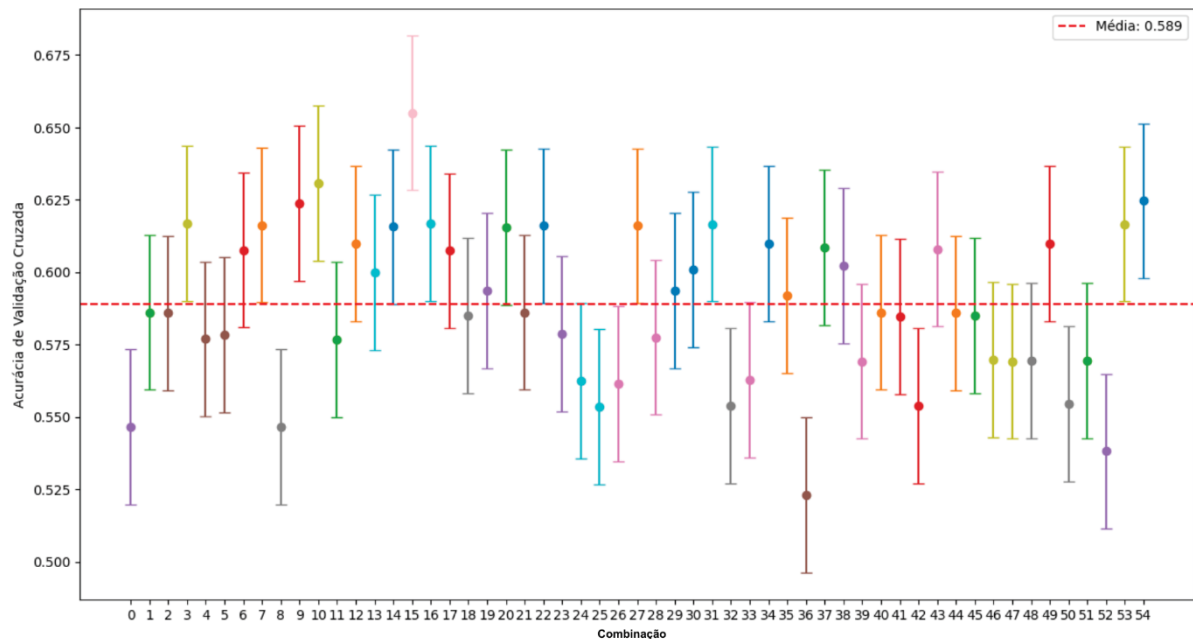
Após a seleção e combinação de 27 variáveis nas 55 combinações distintas, obteve-se um conjunto de dados estruturado para a RNA. Foram utilizadas 128 empresas para o treinamento/validação, correspondendo a aproximadamente 80% do total, enquanto outras 32 empresas foram destinadas para o teste, representando cerca de 20% do total de observações.

O tempo total de execução computacional necessário para processar a RNA foi registrado em aproximadamente 4,14 horas. Esse período engloba o processamento desde as combinações e as iterações para cada um até o cálculo das métricas de desempenho. Em vista disso, esta seção está subdividida em duas partes principais: (i) Avaliação geral do modelo e (ii) Consistência e capacidade de generalização da melhor configuração.

4.2.1 Avaliação geral do modelo: validação cruzada das combinações

Ao realizar a avaliação global do modelo a partir dos dados de treinamento (e validação) foi observada uma variação entre 0.5230 e 0.6550 da melhor acurácia em cada combinação, conforme mostrado no GRÁFICO 13. Ainda neste gráfico, pode-se visualizar em cada combinação um intervalo de confiança de 95% sobre a incerteza associada à precisão do modelo, em termos de sua capacidade de generalização para dados não utilizados durante o treinamento, sendo que a média de todas as acurácias foi igual a 0,589.

GRÁFICO 13 – Intervalo de confiança da acurácia da Validação Cruzada por combinação



Fonte: A autora (2023)

É importante destacar que todas as configurações da RNA alcançaram mais de 50% de acertos. A melhor combinação durante o procedimento de validação cruzada foi a de número 15, com uma acurácia de 0,6550, utilizando as seguintes características:

a) 19 variáveis de entrada: 'GA', 'ROA', 'lider', 'exp_selic_mes_pos_ipo', 'P0', 'segmento', 'mes', 'vol_financeiro', 'LC', 'idade_emp_ipo', 'ipca_acum_12_meses', 'ROE', 'icc', 'imob_rec_nao_cor', 'ici', 'setor', 'ano_eleitoral', 'tipo_oferta', 'avPNC', 'qtd_titulos', 'comp_endividamento';

b) 15 neurônios na camada oculta;

c) Parâmetros específicos: 'activation': 'relu', 'alpha': 0.0001, 'learning_rate': 'constant', 'max_iter': 5000, 'solver': 'sgd'

As cinco melhores combinações, considerando apenas a acurácia da validação cruzada, foram as listadas na TABELA 7: 15, 10, 54, 9, e 3.

TABELA 7 - Características das combinações com melhores acurácias na validação

Comb	Variaveis	Num_Variaveis	Neuronios	Configuracoes	Acuracia
15	['ipca_acum_12_meses', 'setor', 'segmento', 'lider', 'comp_endividamento', 'imob_rec_nao_cor', 'avPNC', 'P0', 'avPL', 'vol_diario_ibov_reg_ipo', 'ROA', 'vol_financeiro', 'indic_incerteza_eco', 'uf', 'exp_selic_mes_pos_ipo', 'ici', 'qtd_titulos', 'ano_eleitoral', 'mes']	19	(15,)	{'activation': 'relu', 'alpha': 0.0001, 'learning_rate': 'constant', 'max_iter': 5000, 'solver': 'sgd'}	0,655077
10	['GA', 'ROA', 'lider', 'exp_selic_mes_pos_ipo', 'P0', 'segmento', 'mes', 'vol_financeiro', 'LC', 'idade_emp_ipo', 'ipca_acum_12_meses', 'ROE', 'icc', 'imob_rec_nao_cor', 'ici', 'setor', 'ano_eleitoral', 'tipo_oferta', 'avPNC', 'qtd_titulos', 'comp_endividamento']	21	(12,)	{'activation': 'tanh', 'alpha': 0.0001, 'learning_rate': 'constant', 'max_iter': 1000, 'solver': 'lbfgs'}	0,630769
54	['mes', 'lider', 'vol_financeiro', 'ici', 'P0', 'ipca_acum_12_meses', 'ROA', 'GA', 'comp_endividamento', 'ROE', 'avPL', 'ano_eleitoral', 'segmento', 'qtd_titulos']	14	(11,)	{'activation': 'relu', 'alpha': 0.01, 'learning_rate': 'constant', 'max_iter': 1000, 'solver': 'lbfgs'}	0,624615
9	['exp_selic_mes_pos_ipo', 'ici', 'exp_inflacao_12_meses', 'icc', 'comp_endividamento', 'LC', 'setor', 'cap_terc_cap_total']	8	(5,)	{'activation': 'logistic', 'alpha': 0.0001, 'learning_rate': 'constant', 'max_iter': 1000, 'solver': 'lbfgs'}	0,623692
3	['icc', 'vol_diario_ibov_reg_ipo', 'mes', 'lider', 'ipca_acum_12_meses', 'indic_incerteza_eco', 'avPL', 'P0', 'uf', 'vol_financeiro', 'exp_inflacao_12_meses', 'comp_endividamento']	12	(8,)	{'activation': 'relu', 'alpha': 0.0001, 'learning_rate': 'constant', 'max_iter': 1000, 'solver': 'adam'}	0,616923

Fonte: A autora (2023)

A TABELA 7 exemplifica a variabilidade de resultados que pode ocorrer em uma RNA, considerando as melhores acurácias dentre todas as combinações realizadas na pesquisa. No que tange aos parâmetros, observa-se que o número máximo de iterações para convergir ao resultado foi igual a 5.000, diferentemente de

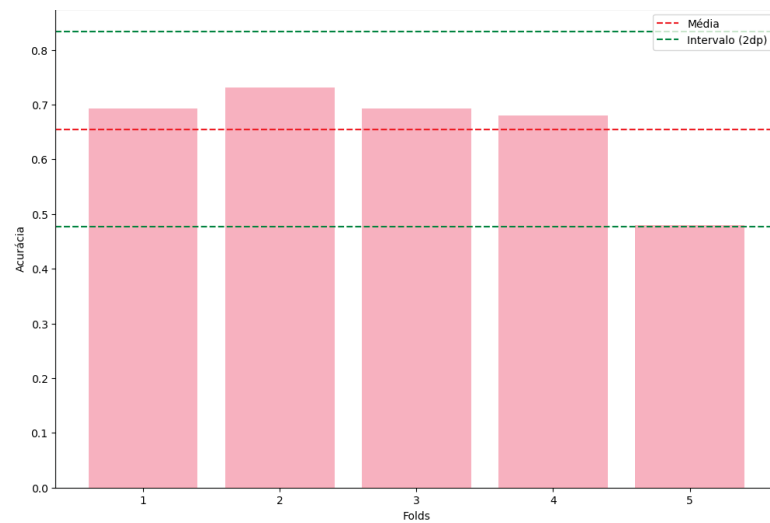
Jain; Nag, (1995) e Jain; Nag (1998) que precisaram de 40.000 iterações. Essa diferença se justifica principalmente pela quantidade de observações. Aqui foram 168, no caso desses pesquisadores foram 552 e 461 IPOs, respectivamente. Logo, a diferença da quantidade de empresas que compõem a amostra pode alterar significativamente os resultados.

Além disso, a função de ativação 'relu' foi utilizada em três dos cinco casos, embora também tenham sido relevantes a 'logistic' e 'tanh', tal qual fora testado por Han; Kim (2021). Já a taxa de aprendizado 'learning_rate' foi definida como 'constant' em todas as configurações.

O destaque para as variáveis de entrada recai sobre 'comp_endividamento' (quatro ocorrências), 'ipca_acum_12_meses' (três ocorrências), 'líder' (três ocorrências), 'vol_financeiro' (três ocorrências) e 'mês' (três ocorrências). Essas variáveis foram mencionadas com maior frequência nas cinco melhores combinações.

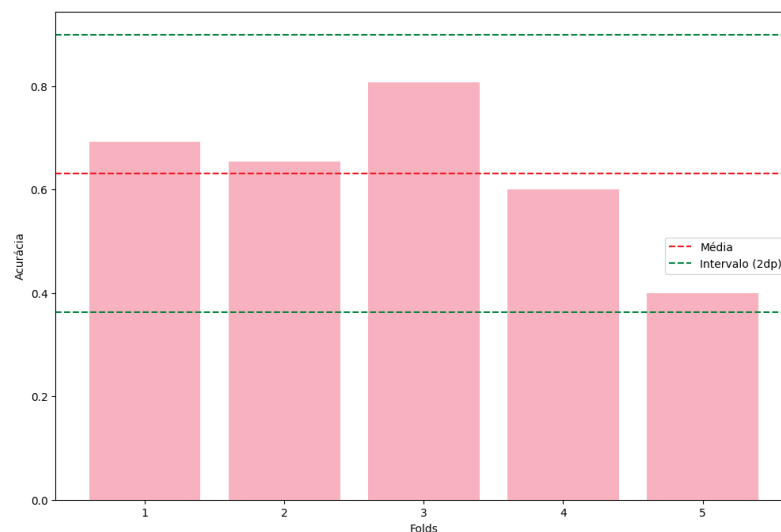
Quando se analisa individualmente cada combinação listado na TABELA 7, a possibilidade de enfrentar dificuldades na coleta de dados específicos suscita a consideração de que alcançar uma acurácia média de 62% utilizando apenas oito variáveis (combinação 9) pode ser mais vantajoso em comparação à utilização de 19 ou 21 variáveis distintas (combinação 15 ou 10). Esses achados indicam que a estratégia de empregar um número reduzido de variáveis pode ser preferível.

Além disso, é importante ressaltar que esses resultados são médias obtidas a partir de cada "fold" da validação. Portanto, na combinação 15, por exemplo, a acurácia variou de 47,67% a 83,35% (GRÁFICO 14), considerando dois desvios padrões acima e abaixo da média de 65,51%: no fold 1, a acurácia foi de 69,23%, seguida por 73,08% no fold 2, 69,23% no fold 3, 68% no fold 4 e 48% no fold 5. Essa ampla variação nos resultados indica a sensibilidade do modelo e destaca a necessidade de considerar a faixa de acurácia possível.

GRÁFICO 14 - Acurácia para cada *fold* da combinação 15

Fonte: A autora (2023)

A combinação 10 (GRÁFICO 15) também apresentou uma diferença substancial entre os “folds” da validação cruzada: no fold 1, a acurácia foi de 69,23%, seguida por 65,38% no fold 2, 80,77% no fold 3, 60% no fold 4 e 40% no fold 5. Essa variação nos resultados destaca a natureza estocástica da análise utilizando RNAs.

GRÁFICO 15 - Acurácia para cada *fold* da combinação 10

Fonte: A autora (2023)

A acurácia média da combinação 10 foi de 63,08%, com um intervalo de 36,27% a 89,88%, demonstrando a incerteza associada à precisão do modelo em termos de sua capacidade de generalização para dados não utilizados no treinamento. Esses resultados evidenciam a importância de considerar a probabilidade de acerto ao tomar decisões baseadas em RNA.

Investir em um ativo com uma chance de acerto inferior a 50% ou com quase 90% de possibilidade de decisão correta são situações bem distintas. Portanto, é importante ressaltar que a utilização de RNAs não permite afirmações absolutas, especialmente quando se trata de embasar decisões financeiras. É essencial considerar não apenas o resultado, como uma alta acurácia, mas também levar em conta os intervalos de incerteza associados.

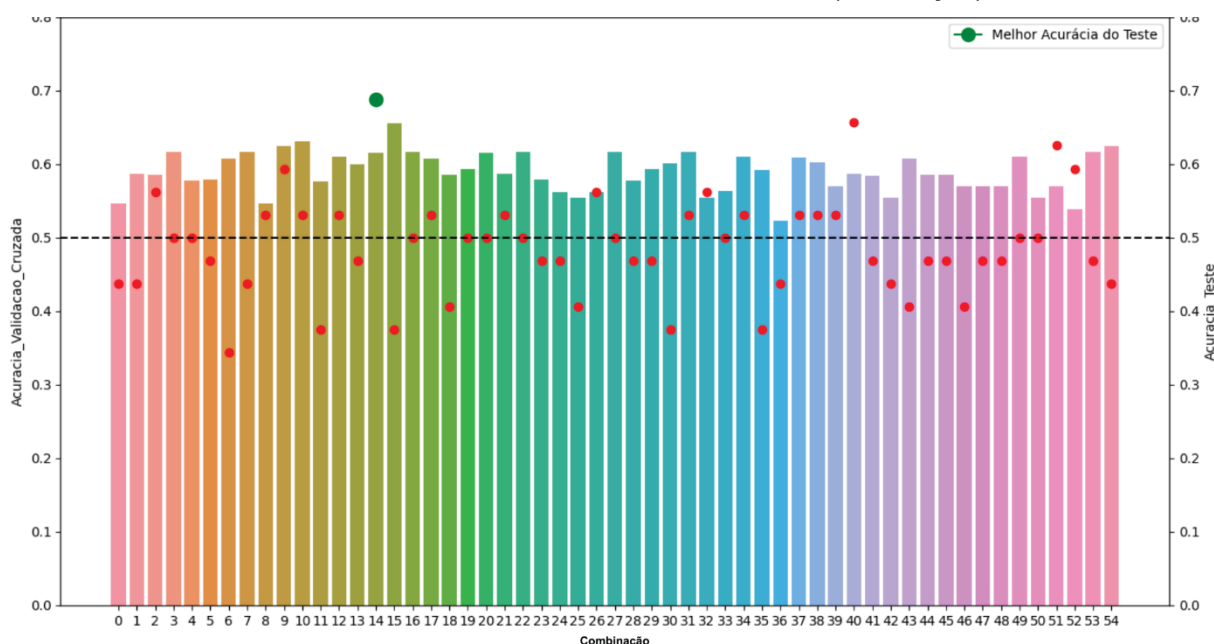
Todavia, a compreensão dos resultados da validação cruzada desempenha um papel relevante na definição de configurações mais adequadas para testes futuros. Essa análise possibilita a avaliação da consistência do modelo e contribui para aprimorar sua performance ao ser submetido a novos conjuntos de dados. Portanto, é por meio desse processo iterativo que se busca alcançar uma maior confiabilidade e robustez no desempenho da RNA em geral. Essa etapa será abordada e discutida na próxima subseção.

4.2.2 Consistência e capacidade de generalização

Após realizar a análise da validação cruzada e examinar as acurácias obtidas para as diferentes combinações, é pertinente estender essa avaliação aos dados de teste, a fim de investigar a capacidade de generalização do modelo e identificar eventuais problemas, como o fenômeno do *overfitting*. Nesta seção, serão discutidos os resultados dos testes utilizando as melhores configurações de cada combinação da etapa de validação, com ênfase na discrepância entre as acurácias.

Para uma apreensão mais ampla e abrangente dos resultados obtidos, o GRÁFICO 16 fornece uma visualização concisa das acurácias associadas às diferentes combinações, sendo as barras referentes aos resultados da validação cruzada e os pontos aos do teste.

GRÁFICO 16 - Acurácia média nos dados de treinamento (e validação) e teste



Fonte: A autora (2023)

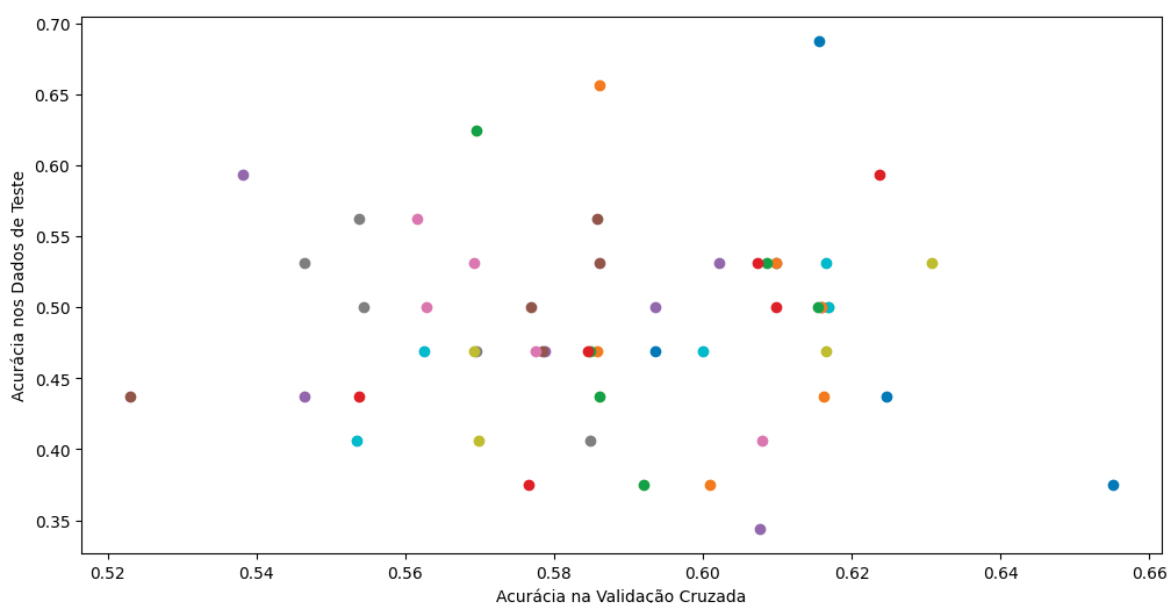
A maioria das combinações (38) não demonstrou um desempenho satisfatório (acima de 50%) durante a avaliação com os dados de teste. Agora, considerando acurácias superiores a 60% o destaque ocorre nas combinações de número 14 e 40. A combinação de número 14 obteve a maior acurácia no teste, alcançando 0,6875, enquanto a combinação de número 40 obteve uma acurácia de 0,6562. Entretanto, é importante destacar que, além de observar as taxas de acerto mais altas, é igualmente relevante examinar se ocorreram discrepâncias significativas entre os resultados obtidos na validação e no teste (GOODFELLOW; BENGIO; COURVILLE, 2016).

Nesse contexto, cabe ressaltar que a combinação de número 26 apresentou a menor diferença entre os períodos de validação e teste, com apenas 0,000962 de variação. Em outras palavras, durante a fase de validação, essa combinação alcançou uma acurácia de 0,5615, enquanto no teste obteve uma acurácia de 0,5625. Por outro lado, as combinações de número 14 e 40 também registraram uma diferença relativamente baixa (aproximadamente 0,07), embora ainda maior do que a observada na combinação de número 26.

Ao analisar todas as combinações, verificou-se que a média da diferença entre a acurácia do teste e a acurácia da validação cruzada foi de -0,10. Em termos percentuais, constatou-se que a acurácia do teste foi, em média, cerca de 10% inferior à acurácia obtida na validação cruzada.

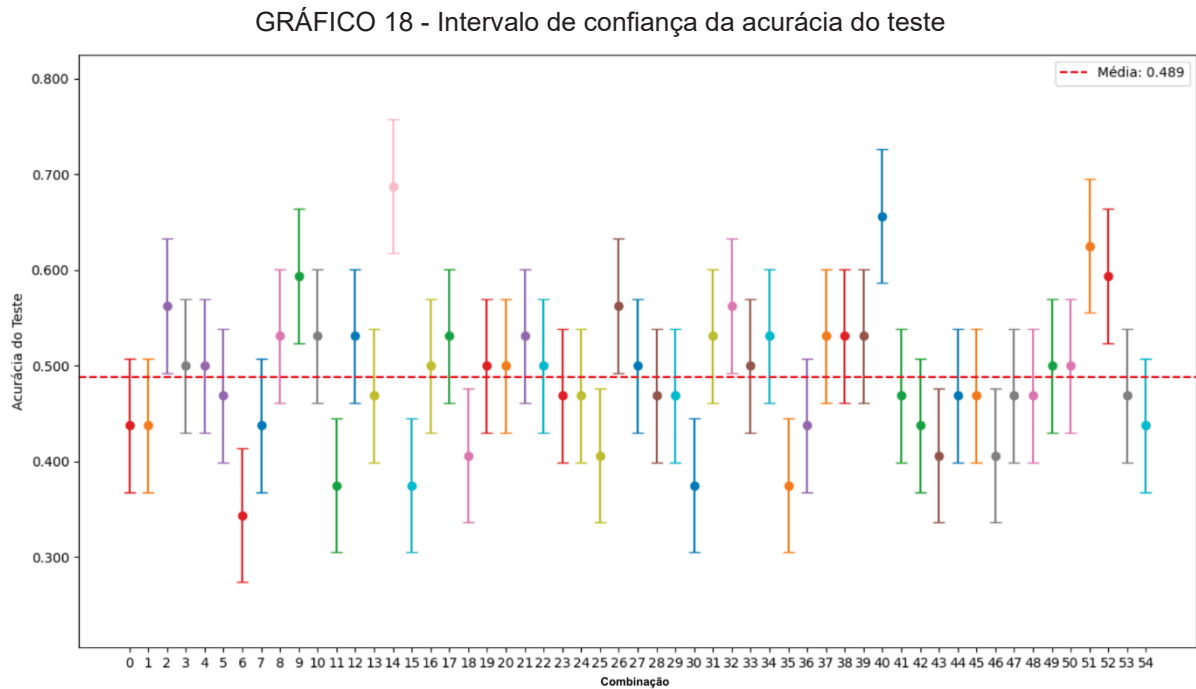
Para verificar a dispersão das acurácias, o GRÁFICO 17 apresenta uma distribuição relativa das combinações, com uma tendência de concentração ao redor da média. O objetivo ideal seria alcançar acurácias elevadas tanto na validação quanto no teste, resultando em uma concentração de pontos no quadrante superior direito. No entanto, também são observados casos em que os pontos estão posicionados mais à direita e abaixo desse quadrante, indicando uma falha na capacidade de o modelo de generalizar corretamente em todos os casos.

GRÁFICO 17 - Dispersão entre acurácias da validação e teste



Fonte: A autora (2023)

No que concerne aos intervalos de confiança de 95% para os dados de teste, constatou-se que a combinação numerada como 14 apresentou a maior taxa de acerto, alcançando um valor de 0,6875 com a possibilidade de chegar até 80% de acertos da predição. Essa constatação está alinhada com os resultados obtidos por Correia (2015) em seu estudo de previsão da valorização da ação de IPO no primeiro dia no Brasil, no qual foram utilizadas também 32 empresas para o teste, considerando variáveis relacionadas à situação econômica e às finanças das empresas. De forma semelhante, a combinação de número 40 apresentou um bom desempenho em termos de acurácia, alcançando 0,6562, o que pode gerar uma taxa de acertos de até 75%. Esses resultados podem ser visualizados no GRÁFICO 18.



À vista disso, a TABELA 8 elenca as cinco combinações com as melhores acurácias no conjunto de teste. A combinação de número 14, por exemplo, utiliza 18 variáveis de entrada e possui uma configuração de dez neurônios na camada oculta. Os hiper parâmetros adotados para essa combinação incluem uma função de ativação "tanh", um coeficiente de regularização (alpha) de 0.01, uma taxa de aprendizado (learning_rate) constante, um número máximo de iterações de 1000 e um algoritmo de otimização "lbfgs". Essa combinação obteve uma acurácia de 0,615692 na validação cruzada e 0,6875 no teste.

Agora, ao analisar as configurações dos hiper parâmetros mais frequentes nas combinações apresentadas na TABELA 8, é possível observar uma variedade de valores adotados. Dentre as configurações mais comuns, destaca-se o uso da função de ativação 'tanh' e 'relu', sendo a primeira predominante (21 ocorrências) ao considerar a busca de melhores parâmetros no total de 55 combinações. Quanto aos valores de alpha, eles variam entre 0.0001, 0.001 e 0.01, indicando diferentes níveis de regularização utilizados nas combinações. A taxa de aprendizado constante é adotada em todos os casos, enquanto os solvers 'lbfgs', 'adam' e 'sgd' são empregados.

TABELA 8 - Características das combinações com melhores acurácias no teste

Com b	Variaveis	Num_Variave is	Configuraco es	Acuracia_Validacao_Cruz ada	Acuracia_Tes te
14	['indic_incerteza_eco', 'avPL', 'comp_endividamento', 'uf', 'ipca_acum_12_meses', 'ano_eleitoral', 'exp_inflacao_12_mese s', 'qtd_titulos', 'idade_emp_ipo', 'vol_diario_ibov_reg_ip o', 'exp_selic_mes_pos_ip o', 'segmento', 'ici', 'LC', 'mes', 'lider', 'avPNC', 'ROE']	18	{'activation': 'tanh', 'alpha': 0.01, 'learning_rate' : 'constant', 'max_iter': 1000, 'solver': 'lbfgs'}	0,615692	0,6875
40	['icc', 'LC']	2	{'activation': 'relu', 'alpha': 0.0001, 'learning_rate' : 'constant', 'max_iter': 1000, 'solver': 'adam'}	0,586154	0,65625
51	['exp_selic_mes_pos_i po', 'comp_endividamento', 'tipo_oferta', 'ROA', 'avPL', 'cap_terc_cap_total', 'uf', 'LC']	8	{'activation': 'tanh', 'alpha': 0.001, 'learning_rate' : 'constant', 'max_iter': 1000, 'solver': 'lbfgs'}	0,569538	0,625
9	['exp_selic_mes_pos_i po', 'ici', 'exp_inflacao_12_mese s', 'icc', 'comp_endividamento', 'LC', 'setor', 'cap_terc_cap_total']	8	{'activation': 'logistic', 'alpha': 0.0001, 'learning_rate' : 'constant', 'max_iter': 1000, 'solver': 'lbfgs'}	0,623692	0,59375
52	['icc', 'comp_endividamento', 'tipo_oferta', 'segmento', 'vol_financeiro', 'avPNC']	6	{'activation': 'relu', 'alpha': 0.0001, 'learning_rate' : 'adaptive', 'max_iter': 1000, 'solver': 'sgd'}	0,538154	0,59375

Fonte: A autora (2023)

Na combinação 14, o conjunto de variáveis abrange desde indicadores econômicos, como "indic_incerteza_eco" e "ipca_acum_12_meses", até informações específicas sobre empresas, como "avPL", "comp_endividamento" e "segmento".

Além disso, são incorporados dados geográficos, como a variável "uf", e elementos temporais, como "ano_eleitoral" e "mes". Essa abordagem permite considerar múltiplos aspectos e resultou na maior acurácia dos dados de teste (mais de 68% de acertos).

Por outro lado, a combinação 40 apresentou uma estratégia mais simplificada, utilizando apenas duas variáveis: "icc" e "LC". Essa abordagem pode ser um exemplo para simplificar o modelo e focar em variáveis específicas. Embora essa redução no número de variáveis possa agilizar o processo de treinamento e validação do modelo, também pode levar a uma perda de informações importantes presentes em outras variáveis. De qualquer modo, apresentou mais de 65% de acerto, mostrando que a percepção do consumidor a respeito da econômica juntamente ao aspecto de liquidez da empresa podem explicar se um IPO trará ou não resultados positivos no primeiro dia de negociação na B3.

Na combinação 51, observa-se a inclusão de variáveis relacionadas a indicadores econômicos ("exp_selic_mes_pos_ipo"), características específicas das empresas ("tipo_oferta", "ROA", "avPL", "cap_terc_cap_total", "LC") e dados regionais ("uf"). Essa combinação de variáveis busca capturar diferentes aspectos para a classificação, abrangendo tanto fatores econômicos quanto características específicas das empresas.

A combinação 9 incorpora indicadores econômicos ("exp_selic_mes_pos_ipo", "ici", "exp_inflacao_12_meses") e informações sobre empresas ("icc", "comp_endividamento", "LC", "setor", "cap_terc_cap_total"). Além disso, é importante destacar que essa combinação apresentou uma diferença baixa entre a acurácia do teste e da validação (-0,029), sugerindo um bom desempenho de generalização do modelo. Isso indica que o modelo foi capaz de manter um desempenho consistente ao ser aplicado em novos conjuntos de dados, corroborando a eficácia das variáveis selecionadas.

Já a combinação 59 utiliza variáveis como "icc", "comp_endividamento", "tipo_oferta", "segmento", "vol_financeiro" e "avPNC". Essa combinação sugere a consideração de informações relacionadas ao desempenho financeiro das empresas, aspectos setoriais e características específicas das ofertas resultou na mesma acurácia da combinação 9 (próximo a 60%), porém uma diferença maior em relação ao que foi obtido na validação (mais ou menos 54%).

Ademais, é necessário examinar as métricas de precisão, *recall* e F1 *Score* para compreender ainda mais o desempenho dessas 5 melhores combinações, conforme TABELA 9:

TABELA 9 - Métricas das melhores combinações do teste

Comb	Precision	Recall	F1 Score
14	0,625	0,9375	0,75
40	0,631579	0,75	0,685714
51	0,6	0,75	0,666667
9	0,578947	0,6875	0,628571
52	0,571429	0,75	0,648649

Fonte: A autora (2023)

A precisão avalia a proporção de casos classificados como positivos que são realmente positivos, enquanto o *recall* avalia a proporção de casos positivos que foram corretamente identificados pelo modelo. No caso da combinação 14, a precisão de 0,625 significa que aproximadamente 62,5% dos casos identificados como positivos por essa combinação são de fato positivos. Por outro lado, o *recall* de 0,9375 indica que cerca de 93,75% dos casos positivos foram corretamente identificados por essa combinação.

Em outras palavras, a precisão enfatiza a capacidade do modelo em evitar falsos positivos, ou seja, classificar erroneamente um caso negativo como positivo. Já o *recall* enfatiza a capacidade do modelo em identificar corretamente os casos positivos, evitando falsos negativos, ou seja, não classificar erroneamente um caso positivo como negativo.

A métrica F1 *Score* combina precisão e *recall* em uma única medida e oferece uma visão mais abrangente do desempenho global das combinações. A combinação 14 registra um valor de F1 de 0,75, evidenciando um equilíbrio entre a precisão e a capacidade de identificação correta dos casos positivos.

A combinação 40 registrou uma precisão de 0,631579, indicando que aproximadamente 63,16% dos casos identificados como positivos por essa combinação são realmente positivos. O *recall*, por sua vez, alcançou o valor de 0,75, o que significa que cerca de 75% dos casos positivos foram corretamente identificados por esse arranjo. O valor de F1, que combina a precisão e o *recall*, foi de 0,685714, indicando um equilíbrio razoável entre a precisão e a capacidade de identificação correta dos casos positivos.

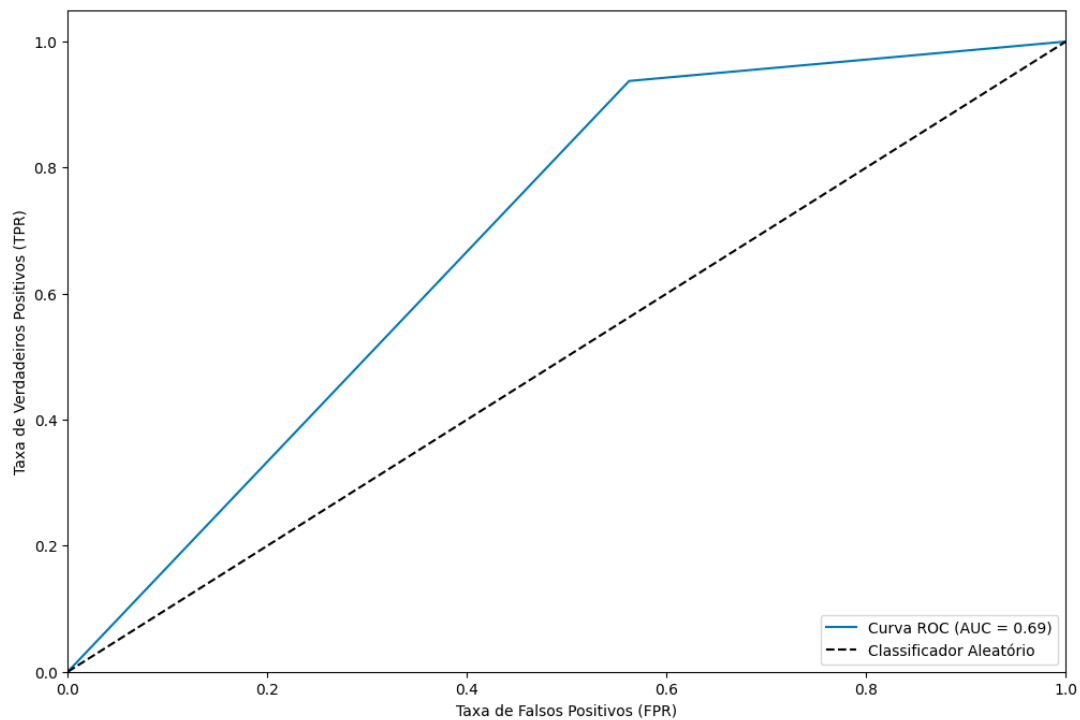
A combinação 51 apresentou uma precisão ligeiramente inferior, com um valor de 0,6, indicando que aproximadamente 60% dos casos identificados como positivos por essa combinação são realmente positivos. No entanto, o *recall* foi semelhante à combinação 40, com um valor de 0,75, mostrando que cerca de 75% dos casos positivos foram corretamente identificados. O valor de F1 foi de 0,666667, evidenciando um equilíbrio semelhante entre a precisão e o *recall* em relação à combinação 40.

A combinação 9 mostrou uma precisão um pouco menor, com um valor de 0,578947, indicando que aproximadamente 57,89% dos casos identificados como positivos por essa combinação são realmente positivos. O *recall*, no entanto, foi de 0,6875, o que significa que cerca de 68,75% dos casos positivos foram corretamente identificados. O valor de F1 foi de 0,628571, refletindo um equilíbrio entre a precisão e a capacidade de identificação correta dos casos positivos, embora ligeiramente inferior às combinações 40 e 51.

A combinação 52, por sua vez, apresentou uma precisão de 0,571429, indicando que aproximadamente 57,14% dos casos identificados como positivos por essa combinação são realmente positivos. O *recall* foi de 0,75, indicando que cerca de 75% dos casos positivos foram corretamente identificados. O valor de F1 foi de 0,648649, evidenciando um equilíbrio similar à combinação 9 em relação à precisão e ao *recall*.

Focando especificamente na combinação 14, a curva ROC gerada revelou uma área considerável sob a curva, indicando um bom desempenho na classificação dos casos positivos, conforme GRÁFICO 19. A taxa de verdadeiros positivos alcançou um valor elevado de aproximadamente 0,9, enquanto a taxa de falsos positivos ficou próxima de 0,6. Esses resultados sugerem que a combinação 14 apresentou uma performance superior ao de um classificador aleatório, que dividiria o quadrante em partes iguais.

GRÁFICO 19 - ROC AUC da combinação 14 no teste

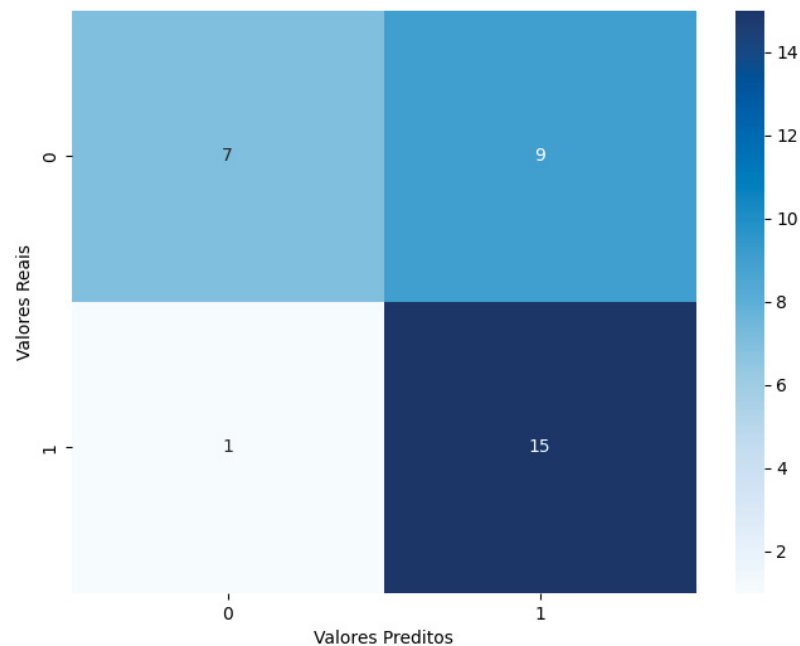


Fonte: A autora (2023)

A alta taxa de verdadeiros positivos demonstra a capacidade da combinação 14 em identificar corretamente a maioria dos casos positivos, enquanto a taxa de falsos positivos moderada indica que houve uma taxa razoável de classificações incorretas de casos negativos como positivos. Essa análise ressalta que o desempenho da combinação 14 na curva ROC foi positivo, superando um classificador aleatório e evidenciando sua habilidade em distinguir entre casos positivos e negativos.

Além da análise da curva ROC, a matriz de confusão oferece uma visão mais detalhada do desempenho da combinação 14 na classificação dos casos positivos e negativos. A matriz de confusão para a combinação 14 apresentou os seguintes resultados: dos 16 casos de “0: não compra”, sete foram classificados corretamente como positivos, enquanto 9 foram erroneamente classificados como “compra”. Em relação aos 16 casos de “1: compra”, apenas um foi incorretamente classificado como “não compra”, enquanto os outros 15 foram corretamente identificados como “compra”, conforme FIGURA 21:

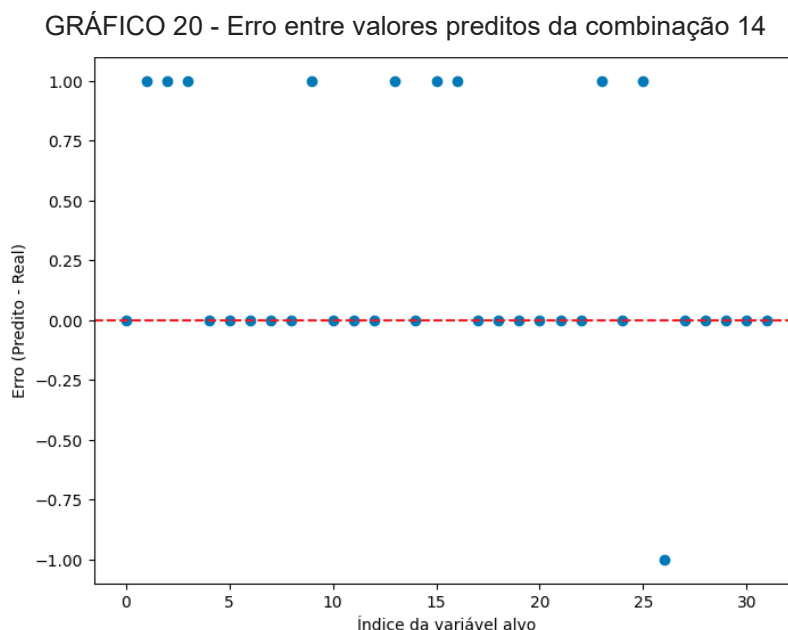
FIGURA 21 - Matriz de confusão para combinação 14 do teste



Fonte: A autora (2023)

Esses resultados evidenciam que o desempenho da combinação 14 na classificação dos casos positivos apresentou uma taxa de erro significativa, conforme número de falsos negativos, ou seja, casos de "não compra" erroneamente classificados como "compra". Por outro lado, essa combinação demonstrou uma capacidade satisfatória em identificar corretamente os casos negativos, uma vez que o número de falsos positivos, ou seja, casos de "compra" erroneamente classificados como "não compra", foi baixo.

Uma outra abordagem para visualizar os resultados obtidos é por meio do GRÁFICO 20, que ilustra a discrepância entre as classificações preditas e os valores reais. Nesse gráfico de erro, a linha horizontal pontilhada no ponto zero divide o quadrante em duas partes, representando as classificações corretas. Os pontos acima dessa linha pontilhada indicam os erros nos quais casos que deveriam ser classificados como "não compra" (0) foram erroneamente classificados como "compra" (1), enquanto os pontos abaixo representam os erros nos quais casos que deveriam ser classificados como "compra" (1) foram erroneamente classificados como "não compra" (0).



Fonte: A autora (2023)

Portanto, ao aplicar a combinação 14 em dados não utilizados durante as etapas de treinamento e validação, constatou-se a ocorrência de 9 casos nos quais a classificação atribuiu incorretamente a condição de "compra" a situações que deveriam ter sido classificadas como "não compra". A extensão desses erros de classificação pode acarretar perdas financeiras consideráveis, as quais variam de acordo com o perfil do investidor e o volume de transações realizadas.

Nesse contexto, cabe reconhecer que as simulações baseadas em dados históricos e em parâmetros de uma RNA não asseguram um acerto idêntico em cenários futuros. Elas devem ser interpretadas como um recurso informativo para a análise da performance da RNA. Diante disso e com o intuito de determinar a magnitude do ganho potencial que teria ocorrido com os resultados da combinação 14, simulações de investimentos serão realizadas na próxima subseção.

4.2.2.1 Simulações de investimentos com a melhor combinação do teste

O conjunto de teste foi composto por 32 empresas que realizaram ofertas públicas iniciais (IPOs), e a seleção das empresas foi feita de forma estratificada, resultando em 16 empresas com retorno acima de 1% e 16 empresas com retorno abaixo de 1% no primeiro dia de negociação.

No que diz respeito às previsões da RNA utilizando a combinação 14, foi observado um total de 10 previsões incorretas, representando uma taxa de erro de

31,3%, enquanto 22 previsões foram corretas, totalizando uma taxa de acerto de 68,7%. Para mensurar o potencial de ganho que teria sido obtido ao seguir estritamente esses resultados da RNA, supõe-se um investidor com R\$ 1000 disponíveis para investir em cada IPO.

Dessa forma, considerando que o investidor aportasse em todos os IPOs selecionados, conforme a quantidade mencionada em cada ano no conjunto de teste, seria necessário dispor de diferentes montantes para cada período, considerando os valores mínimos apenas, e totalizaria R\$ 32000. Para os anos de 2005 e 2006, nos quais ocorreram dois IPOs, o investimento total exigido seria de R\$ 2000 para cada ano. No ano de 2007, com dez IPOs, o investidor precisaria dispor de R\$ 10000. Em 2008, com dois IPOs, o montante necessário seria de R\$ 2000. Para os anos de 2010, 2017 e 2018, nos quais houve apenas um IPO em cada período, o investidor teria que dispor de R\$ 1000 para cada ano. Nos anos de 2013 e 2020, com dois IPOs e seis IPOs, respectivamente, seriam necessários investimentos de R\$ 2000 e R\$ 6000. Por fim, levando em consideração os cinco IPOs ocorridos em 2021, o investimento necessário seria de R\$ 5000, conforme apresentado no QUADRO 15.

QUADRO 15 - Capital suposto para aplicação na simulação

ano_ipo	empresa	mínimo_disp	ano_ipo	empresa	mínimo_disp
2005	RENAR	R\$ 2.000,00	2010	ALIANSCCE	R\$ 1.000,00
	ENERGIAS BR		2013	LINX	R\$ 2.000,00
2006	POSITIVO INF	R\$ 2.000,00		CPFL RENOVAV	
	ODONTOPREV		2017	MOVIDA	R\$ 1.000,00
2007	FER HERINGER	R\$ 10.000,00	2018	HAPVIDA	R\$ 1.000,00
	TENDA		2020	QUERO-QUERO	R\$ 6.000,00
	BOVESPA HLD			PLANOEPLANO	
	PINE			3R PETROLEUM	
	JBS			ESTAPAR	
	IGUATEMI			HIDROVIAS	
	SAO MARTINHO			NEOGRID	
	CREMER		2021	OCEANPACT	R\$ 5.000,00
	BEMATECH			BRISANET	
	JHSF PART			GPS	
	NUTRIPLANT			CSNMINERACAO	
2008	LE LIS BLANC	R\$ 2.000,00		CBA	

Fonte: A autora (2023)

Portanto, ao considerar a aplicação em todos os IPOs selecionados, o investidor teria adquirido um total de 1986 ações ao longo do período analisado, ao investir um montante de R\$ 31.975,44, ligeiramente abaixo dos R\$ 32.000 destinados às compras. Esse investimento teria gerado um retorno total de R\$ 32.656,70, correspondente a uma taxa de 2,13%. No entanto, ao examinar a TABELA 10, verifica-se que em 2007 um único investimento resultou em um retorno de 52% (BOVESPA HLD), e em 2017, um retorno de 20% (MOVIDA). No cômputo geral, a taxa média de retorno por ano teria sido de 2,20%:

TABELA 10 - Resumo de retornos: simulação com aplicações em todos os IPOs do conjunto de teste

ano_ip o	nome	p0	p1	retorn o	acoe s	compra	venda	retorn o
2005	RENAR	4,09	4,14	1%	244	R\$ 997,96	R\$ 1.010,16	R\$ 12,20
2005	ENERGIAS BR	45,9 7	51,1	11%	22	R\$ 1.011,34	R\$ 1.124,20	R\$ 112,86
2006	POSITIVO INF	58,1 9	57,3 2	-1%	17	R\$ 989,23	R\$ 974,44	-R\$ 14,79
2006	ODONTOPREV	69,3 3	80,2	16%	14	R\$ 970,62	R\$ 1.122,80	R\$ 152,18
2007	FER HERINGER	40,3	40,4 1	0%	25	R\$ 1.007,50	R\$ 1.010,25	R\$ 2,75
2007	TENDA	21,3 3	21,3 3	0%	47	R\$ 1.002,51	R\$ 1.002,51	R\$ -
2007	BOVESPA HLD	54,5 2	82,9 4	52%	18	R\$ 981,36	R\$ 1.492,92	R\$ 511,56
2007	PINE	45,0 4	44,0 9	-2%	22	R\$ 990,88	R\$ 969,98	-R\$ 20,90
2007	JBS	18,9 6	16,5 9	-13%	53	R\$ 1.004,88	R\$ 879,27	-R\$ 125,61
2007	IGUATEMI	71,1 1	79,6 4	12%	14	R\$ 995,54	R\$ 1.114,96	R\$ 119,42
2007	SAO MARTINHO	47,4 1	56,0 8	18%	21	R\$ 995,61	R\$ 1.177,68	R\$ 182,07
2007	CREMER	41,4 8	41,4 8	0%	24	R\$ 995,52	R\$ 995,52	R\$ -
2007	BEMATECH	35,5 6	39,8 5	12%	28	R\$ 995,68	R\$ 1.115,80	R\$ 120,12
2007	JHSF PART	18,9 6	19,6 7	4%	53	R\$ 1.004,88	R\$ 1.042,51	R\$ 37,63
2008	NUTRIPLANT	22,3 8	23,5	5%	45	R\$ 1.007,10	R\$ 1.057,50	R\$ 50,40
2008	LE LIS BLANC	15,1 1	12,0 9	-20%	66	R\$ 997,26	R\$ 797,94	-R\$ 199,32
2010	ALIANSC	18,2 3	17,7 7	-3%	55	R\$ 1.002,65	R\$ 977,35	-R\$ 25,30
2013	LINX	45,8 2	54,3 1	19%	22	R\$ 1.008,04	R\$ 1.194,82	R\$ 186,78
2013	CPFL RENOVAV	21,2 3	20,4	-4%	47	R\$ 997,81	R\$ 958,80	-R\$ 39,01
2017	MOVIDA	9,88	11,8 2	20%	101	R\$ 997,88	R\$ 1.193,82	R\$ 195,94
2018	HAPVIDA	29,8 3	7,07	-76%	34	R\$ 1.014,22	R\$ 240,38	-R\$ 773,84
2020	QUERO-QUERO	14,7 3	14,8 9	1%	68	R\$ 1.001,64	R\$ 1.012,52	R\$ 10,88
2020	PLANOEPLANO	10,9 4	10,4 8	-4%	91	R\$ 995,54	R\$ 953,68	-R\$ 41,86
2020	3R PETROLEUM	24,4 5	24,2 2	-1%	41	R\$ 1.002,45	R\$ 993,02	-R\$ 9,43
2020	ESTAPAR	12,2 3	11,4 1	-7%	82	R\$ 1.002,86	R\$ 935,62	-R\$ 67,24
2020	HIDROVIAS	8,8	8,64	-2%	114	R\$ 1.003,20	R\$ 984,96	-R\$ 18,24
2020	NEOGRID	5,24	5,75	10%	191	R\$ 1.000,84	R\$ 1.098,25	R\$ 97,41
2021	OCEANPACT	11,8	11,8	0%	85	R\$ 1.003,00	R\$ 1.003,00	R\$ -

2021	BRISANET	14,7 3	14,4	-2%	68	R\$ 1.001,64	R\$ 979,20	-R\$ 22,44
2021	GPS	12,6 9	13,8 5	9%	79	R\$ 1.002,51	R\$ 1.094,15	R\$ 91,64
2021	CSNMINERACA O	8,99	9,83	9%	111	R\$ 997,89	R\$ 1.091,13	R\$ 93,24
2021	CBA	11,8 5	12,5 9	6%	84	R\$ 995,40	R\$ 1.057,56	R\$ 62,16
Total	32			2,13%	1.986	R\$ 31.975,4 4	R\$ 32.656,7 0	R\$ 681,26

Fonte: A autora (2023)

No entanto, embora o investidor tivesse obtido um retorno positivo ao longo das aplicações mencionadas, é essencial ressaltar que o investimento não estaria isento de riscos. O risco, neste contexto, foi mensurado pelo desvio padrão, que se mostrou significativamente elevado, alcançando 18,9% durante o período analisado. Essa alta volatilidade pode ser uma preocupação para os investidores, uma vez que indica a possibilidade de flutuações consideráveis nos retornos das ações de IPOs selecionadas ao longo do tempo sem critério algum.

Dando continuidade à análise, foi realizada uma simulação considerando a aplicação aleatória dos recursos, ou seja, o investidor não aportaria em todos os IPOs, mas faria investimentos de forma aleatória. Nessa simulação, o investidor aplicaria um montante de R\$ 18.013, comprando um total de 1.169 ações, em contraste com o investimento total de R\$ 31.975,44 no cenário anterior. Como resultado, haveria um retorno superior, alcançando 4,23% (R\$ 762,57), conforme TABELA 11, porém, mesmo nesse cenário, o risco continuaria presente, com um desvio padrão de 7%:

TABELA 11 - Resumo de retornos: simulação com aplicações em IPOs aleatórios do conjunto de teste

ano_ipo	nome	real	aleatorio	resultado	acoes	compra	venda	%retorno
2005	RENAR	1	0	errou	0	R\$ -	R\$ -	0%
2005	ENERGIAS BR	1	1	acertou	22	R\$ 1.011,34	R\$ 1.124,20	11%
2006	POSITIVO INF	0	0	acertou	0	R\$ -	R\$ -	0%
2006	ODONTOPREV	1	0	errou	0	R\$ -	R\$ -	0%
2007	FER HERINGER	0	0	acertou	0	R\$ -	R\$ -	0%
2007	TENDA	0	0	acertou	0	R\$ -	R\$ -	0%
2007	BOVESPA HLD	1	0	errou	0	R\$ -	R\$ -	0%
2007	PINE	0	1	errou	23	R\$ 990,88	R\$ 969,98	-2%

2007	JBS	0	1	errou	53	R\$ 1.004,88	R\$ 879,27	-13%
2007	IGUATEMI	1	1	acertou	15	R\$ 995,54	R\$ 1.114,96	12%
2007	SAO MARTINHO	1	0	errou	0	R\$ -	R\$ -	0%
2007	CREMER	0	1	errou	25	R\$ 995,52	R\$ 995,52	0%
2007	BEMATECH	1	1	acertou	29	R\$ 995,68	R\$ 1.115,80	12%
2007	JHSF PART	1	1	acertou	53	R\$ 1.004,88	R\$ 1.042,51	4%
2008	NUTRIPLANT	1	0	errou	0	R\$ -	R\$ -	0%
2008	LE LIS BLANC	0	0	acertou	0	R\$ -	R\$ -	0%
2010	ALIANSC	0	1	errou	55	R\$ 1.002,65	R\$ 977,35	-3%
2013	LINX	1	1	acertou	22	R\$ 1.008,04	R\$ 1.194,82	19%
2013	CPFL RENOVAV	0	0	acertou	0	R\$ -	R\$ -	0%
2017	MOVIDA	1	1	acertou	102	R\$ 997,88	R\$ 1.193,82	20%
2018	HAPVIDA	0	0	acertou	0	R\$ -	R\$ -	0%
2020	QUERO-QUERO	1	1	acertou	68	R\$ 1.001,64	R\$ 1.012,52	1%
2020	PLANOEPLANO	0	1	errou	92	R\$ 995,54	R\$ 953,68	-4%
2020	3R PETROLEUM	0	0	acertou	0	R\$ -	R\$ -	0%
2020	ESTAPAR	0	1	errou	82	R\$ 1.002,86	R\$ 935,62	-7%
2020	HIDROVIAS	0	0	acertou	0	R\$ -	R\$ -	0%
2020	NEOGRID	1	1	acertou	191	R\$ 1.000,84	R\$ 1.098,25	10%
2021	OCEANPACT	0	1	errou	85	R\$ 1.003,00	R\$ 1.003,00	0%
2021	BRISANET	0	1	errou	68	R\$ 1.001,64	R\$ 979,20	-2%
2021	GPS	1	1	acertou	79	R\$ 1.002,51	R\$ 1.094,15	9%
2021	CSNMINERACAO	1	1	acertou	112	R\$ 997,89	R\$ 1.091,13	9%
2021	CBA	1	0	errou	0	R\$ -	R\$ -	0%
Total	32	16	18		1169	R\$ 18.013,21	R\$ 18.775,78	4,23

Fonte: A autora (2023)

Esses resultados indicam que a utilização de um classificador aleatório ainda resultaria em um retorno financeiro positivo e um menor risco³⁰. Essa constatação é relevante, uma vez que converge com os princípios da Teoria da Eficiência de Mercado. Em outras palavras, sugere-se que o mercado tende a refletir toda a

³⁰ Embora a aleatoriedade realizada nesta pesquisa represente apenas uma opção, talvez outras rodadas pudessem gerar resultados distintos.

informação disponível, de modo que mesmo uma estratégia aleatória de seleção de IPOs pode gerar ganhos, considerando-se o período analisado. Esta observação corrobora a hipótese de que, em um mercado eficiente, todas as informações relevantes estão prontamente refletidas nos preços dos ativos, dificultando a obtenção de vantagens sistemáticas por meio de abordagens seletivas.

Entretanto, é importante salientar que a gestão do risco ainda é um fator crítico a ser considerado no processo de investimento. Embora a aplicação aleatória tenha demonstrado resultados positivos, a volatilidade é inerente ao mercado de capitais, sujeito a incertezas e oscilações imprevistas. Nesse contexto, uma abordagem de diversificação, combinando investimentos em IPOs com alocações em carteiras de renda fixa, por exemplo, poderia oferecer uma estratégia mais robusta para gerenciar o risco e preservar o capital.

Diante disso, é importante também considerar a previsão realizada pela RNA e seu impacto nos resultados. A rede neural apresentou um total de 22 acertos para a compra de ações, resultando em 1.319 ações. Nesse caso, o investidor desembolsaria um montante maior, em comparação com a estratégia aleatória, totalizando R\$ 23.966,72 (TABELA 12). No entanto, vale ressaltar que essa abordagem resultou em uma proporção ligeiramente maior de rendimento, atingindo R\$1.085,97 (4,53%), todavia um risco alto de 18,2% (desvio padrão do % de retorno) ao longo de todo período.

TABELA 12 - Resumo de retornos: simulação com aplicações a partir da RNA em IPOs do conjunto de teste

ano_ipo ³¹	nome	real	RNA	resultado	acoes	compra	retorno	%retorno
2005	RENAR	1	0	errou	0	R\$ -	R\$ -	0%
2005	ENERGIAS BR	1	1	acertou	22	R\$ 1.011,34	R\$ 112,86	11%
2006	POSITIVO INF	0	1	errou	17	R\$ 989,23	-R\$ 14,79	-1%
2006	ODONTOPREV	1	1	acertou	14	R\$ 970,62	R\$ 152,18	16%
2007	FER HERINGER	0	1	errou	25	R\$ 1.007,50	R\$ 2,75	0%
2007	TENDA	0	0	acertou	0	R\$ -	R\$ -	0%
2007	BOVESPA HLD	1	1	acertou	18	R\$ 981,36	R\$ 511,56	52%
2007	PINE	0	1	errou	22	R\$ 990,88	-R\$ 20,90	-2%

³¹ Os anos foram aqui classificados em ordem crescente apenas para fins de apresentação dos resultados, porém em todos os conjuntos de teste e treinamento eles estavam embaralhados.

2007	JBS	0	0	acertou	0	R\$ -	R\$ -	0%
2007	IGUATEMI	1	1	acertou	14	R\$ 995,54	R\$ 119,42	12%
2007	SAO MARTINHO	1	1	acertou	21	R\$ 995,61	R\$ 182,07	18%
2007	CREMER	0	1	errou	24	R\$ 995,52	R\$ -	0%
2007	BEMATECH	1	1	acertou	28	R\$ 995,68	R\$ 120,12	12%
2007	JHSF PART	1	1	acertou	53	R\$ 1.004,88	R\$ 37,63	4%
2008	NUTRIPLANT	1	1	acertou	45	R\$ 1.007,10	R\$ 50,40	5%
2008	LE LIS BLANC	0	0	acertou	0	R\$ -	R\$ -	0%
2010	ALIANSC	0	0	acertou	0	R\$ -	R\$ -	0%
2013	LINX	1	1	acertou	22	R\$ 1.008,04	R\$ 186,78	19%
2013	CPFL RENOVAV	0	0	acertou	0	R\$ -	R\$ -	0%
2017	MOVIDA	1	1	acertou	101	R\$ 997,88	R\$ 195,94	20%
2018	HAPVIDA	0	1	errou	34	R\$ 1.014,22	-R\$ 773,84	-76%
2020	QUERO-QUERO	1	1	acertou	68	R\$ 1.001,64	R\$ 10,88	1%
2020	PLANOEPLANO	0	1	errou	91	R\$ 995,54	-R\$ 41,86	-4%
2020	3R PETROLEUM	0	0	acertou	0	R\$ -	R\$ -	0%
2020	ESTAPAR	0	1	errou	82	R\$ 1.002,86	-R\$ 67,24	-7%
2020	HIDROVIAS	0	0	acertou	0	R\$ -	R\$ -	0%
2020	NEOGRID	1	1	acertou	191	R\$ 1.000,84	R\$ 97,41	10%
2021	OCEANPACT	0	1	errou	85	R\$ 1.003,00	R\$ -	0%
2021	BRISANET	0	1	errou	68	R\$ 1.001,64	-R\$ 22,44	-2%
2021	GPS	1	1	acertou	79	R\$ 1.002,51	R\$ 91,64	9%
2021	CSNMINERACAO	1	1	acertou	111	R\$ 997,89	R\$ 93,24	9%
2021	CBA	1	1	acertou	84	R\$ 995,40	R\$ 62,16	6%
Total		16	24		1319	R\$ 23.966,72	R\$ 1.085,97	4,53%

Fonte: A autora (2023)

Além disso, é importante mencionar que, com os recursos não utilizados nas aquisições de IPOs, o investidor teria a opção de direcioná-los para outros ativos financeiros até que novas oportunidades surgissem no mercado de capitais. Essa flexibilidade permitiria ao investidor otimizar seu portfólio ou, caso desejasse, aumentar o aporte em cada IPO selecionado.

Essas considerações evidenciam certa efetividade da previsão realizada pela RNA, a depender do perfil de risco do investidor, resultando em um rendimento

superior em relação à estratégia aleatória – sem considerar os proventos que ainda poderiam ser auferidos. Além disso, a disponibilidade de recursos não utilizados abre possibilidades adicionais para o investidor explorar outras opções de investimento.

Cabe ressaltar que o impacto financeiro dessas previsões pode depender do contexto³² e de imprevistos econômicos que não podem ser captados pela rede neural. Desse modo, no QUADRO 16 é possível observar o número de acertos da RNA em comparação ao classificador aleatório:

QUADRO 16 - Síntese de acertos da RNA x Classificador Aleatório

ano_ipo	empresa	RNA	aleatório	real	resultado
2005	RENAR	0	0	1	Empate: erro
2005	ENERGIAS BR	1	1	1	Empate: acerto
2006	POSITIVO INF	1	0	0	VantagemAleatório
2006	ODONTOPREV	1	0	1	VantagemRNA
2007	FER HERINGER	1	0	0	VantagemAleatório
2007	TENDA	0	0	0	Empate: acerto
2007	BOVESPA HLD	1	0	1	VantagemRNA
2007	PINE	1	1	0	Empate: erro
2007	JBS	0	1	0	VantagemRNA
2007	IGUATEMI	1	1	1	Empate: acerto
2007	SAO MARTINHO	1	0	1	VantagemRNA
2007	CREMER	1	1	0	Empate: erro
2007	BEMATECH	1	1	1	Empate: acerto
2007	JHSF PART	1	1	1	Empate: acerto
2008	NUTRIPLANT	1	0	1	VantagemRNA
2008	LE LIS BLANC	0	0	0	Empate: acerto
2010	ALIANSCCE	0	1	0	VantagemRNA
2013	LINX	1	1	1	Empate: acerto
2013	CPFL RENOVAV	0	0	0	Empate: acerto
2017	MOVIDA	1	1	1	Empate: acerto
2018	HAPVIDA	1	0	0	VantagemAleatório
2020	QUERO-QUERO	1	1	1	Empate: acerto
2020	PLANOEPLANO	1	1	0	Empate: erro
2020	3R PETROLEUM	0	0	0	Empate: acerto
2020	ESTAPAR	1	1	0	Empate: erro
2020	HIDROVIAS	0	0	0	Empate: acerto
2020	NEOGRID	1	1	1	Empate: acerto
2021	OCEANPACT	1	1	0	Empate: erro
2021	BRISANET	1	1	0	Empate: erro
2021	GPS	1	1	1	Empate: acerto
2021	CSNMINERACAO	1	1	1	Empate: acerto
2021	CBA	1	0	1	VantagemRNA
Total	32 empresas	22 acertos	18 acertos	16 retornos positivos	22 empates (15 acertos e 7 erros); 7 Vantagens RNA e 3 Vantagens do aleatório

Fonte: A autora (2023)

³² Assim como da aleatoriedade envolvida no momento de execução do modelo. Portanto, cabe destacar que os resultados dessas simulações poderiam ter sido diferentes, considerando os intervalos de confiança das acurácias discutidas anteriormente.

Constata-se que houve empate em várias ocasiões entre as previsões da RNA e a estratégia aleatória. Entretanto, em sete situações específicas, a RNA obteve vantagem preditiva em relação à estratégia aleatória, enquanto em três a vantagem ficou para a abordagem aleatória.

4.2.2.2 Viabilidade do modelo para tomada de decisões financeiras

Ao comparar o desempenho da RNA com uma estratégia aleatória, constatou-se que a RNA proporcionou um rendimento superior. No entanto, é importante destacar que, em termos de risco, medido pelo desvio padrão, a RNA também demonstrou um resultado mais elevado em relação à estratégia aleatória.

Essa diferença no desvio padrão entre a RNA e a estratégia aleatória indica que a utilização da RNA para previsão de compra de ações de IPOs implica em uma maior variabilidade nos retornos. Em outras palavras, a RNA pode gerar resultados mais expressivos, tanto positivos quanto negativos, em comparação com uma abordagem aleatória.

A maior variabilidade nos retornos pode ser um fator de relevância para os investidores, pois implica em um nível de risco potencialmente maior ao optar pela utilização da RNA como suporte à tomada de decisão. Enquanto a estratégia aleatória pode resultar em retornos mais consistentes, a utilização da RNA pode oferecer a possibilidade de rendimentos mais atrativos, porém acompanhados de uma maior exposição ao risco.

De qualquer modo, é relevante ressaltar que, ao avaliar a efetividade da RNA, é crucial considerar não apenas o percentual de ganho obtido nas simulações aqui apresentadas, mas também a taxa de acertos em relação à estratégia aleatória. Essa taxa de acertos é um indicador importante, pois reflete a capacidade da RNA em tomar decisões acertadas de compra de ações de IPO, levando em conta as variáveis e informações disponíveis no momento da decisão.

Além disso, uma vantagem adicional da aplicação das RNAs no contexto de investimento é a possibilidade de explorar outras opções de alocação dos recursos não utilizados. Essa flexibilidade permite aos investidores diversificar suas carteiras de investimento, aumentando o potencial de retorno financeiro. Ao aguardar o surgimento de novas oportunidades no mercado de capitais, os investidores podem tomar decisões mais embasadas e com maior probabilidade de sucesso.

Portanto, considerando os resultados obtidos e a capacidade das RNAs de auxiliarem na tomada de decisões de investimento, conclui-se que essa ferramenta tem viabilidade para ser aplicada na construção da decisão de compra no curto prazo de ações de IPO no Brasil, a depender do perfil de risco do investidor.

5 CONSIDERAÇÕES FINAIS

No contexto brasileiro, caracterizado por particularidades e dinâmicas próprias no mercado de IPOs, compreender como as RNAs podem ser utilizadas como ferramentas de suporte à tomada de decisões financeiras pode proporcionar benefícios significativos aos investidores.

Ao investigar a viabilidade do uso de RNAs supervisionadas, é imprescindível considerar aspectos como a seleção apropriada das variáveis relevantes, a configuração dos hiper parâmetros, a escolha adequada da função de ativação e a avaliação do desempenho do modelo utilizando métricas como precisão, *recall* e F1.

A integração de indicadores econômicos, características das empresas emissoras e dados financeiros revelou-se promissora na classificação das ações de IPO. Esse achado reforça a importância de considerar múltiplas variáveis em conjunto, em vez de analisar características isoladas, para compreender o potencial retorno positivo no primeiro dia de negociação. Adicionalmente, esses resultados estão em consonância com os construtos teóricos explicativos dos IPOs encontrados na literatura (FIGURA 4).

No entanto, é importante ressaltar que a utilização de RNAs supervisionadas para tomada de decisões financeiras deve ser encarada como uma fonte de informação, uma vez que não é possível garantir a precisão absoluta das previsões em cenários futuros. Portanto, cabe aos investidores considerar as informações fornecidas pelo modelo como mais um elemento a ser ponderado em conjunto com outras análises e estratégias de investimento.

Em suma, a aplicação de RNAs supervisionadas para a construção da decisão de compra no curto prazo de ações de IPO no Brasil apresenta-se como uma possibilidade promissora para auxiliar os investidores, embora tenha que ser considerado o risco envolvido nessa opção.

5.1 CONTRIBUIÇÕES

Esta pesquisa fornece *insights* e contribuições, tanto teóricas quanto aplicadas, que podem ser úteis para pesquisadores, profissionais do mercado financeiro e pequenos investidores.

Uma das principais contribuições desta pesquisa é o desenvolvimento de um modelo adaptável de RNAs supervisionadas para a tomada de decisões financeiras em IPOs e a combinação de variáveis relacionadas às empresas e à conjuntura. Ao considerar a complexidade e a volatilidade desse mercado, a utilização de RNAs oferece uma abordagem flexível, capaz de aprender padrões e tendências dos dados históricos, auxiliando os investidores na identificação de ações com potencial de retorno positivo no curto prazo (primeiro dia de negociação).

5.1.1 Práticas

A pesquisa também oferece contribuições práticas ao propor diretrizes e melhores práticas para a implementação e utilização efetiva das RNAs supervisionadas no contexto de investimentos em IPOs. Essas diretrizes abrangem desde a seleção e preparação adequada dos dados até a avaliação do desempenho do modelo, garantindo que os investidores possam utilizar essa ferramenta de forma consistente.

Portanto, o modelo desenvolvido na dissertação possibilita a utilização pelos aplicadores que buscam alternativas de investimentos combinando diversas variáveis que, comumente, são consideradas incomparáveis por causa de suas características únicas.

5.1.2 Teóricas

Do ponto de vista acadêmico, esta pesquisa contribui para o avanço do conhecimento no campo das finanças com utilização de IA, bem como atende à sugestão de Goodell *et al.*, (2021) no que se refere a mais pesquisas sobre consultoria financeira sem intervenção humana significativa. Ao explorar a aplicação das RNAs supervisionadas em um contexto específico e desafiador como o mercado de IPOs no Brasil, são abertas novas possibilidades de investigação e aprofundamento teórico na área.

5.1.3 Sociais

Por meio da abertura de capital e decisões de investimento baseadas em dados robustos, a sociedade tende a se desenvolver mais, embora o estudo de empresas listadas em bolsa de valores não reflita totalmente o desempenho da economia real (BORISOV; ELLUL; SEVILIR, 2021).

5.2 LIMITAÇÕES E POSSÍVEIS PESQUISAS FUTURAS

Apesar dos esforços para contribuir de forma prática, teórica e social, é importante reconhecer as limitações desta dissertação, que podem servir como estímulos para pesquisas futuras. Uma dessas limitações é o foco exclusivo no cenário brasileiro, abrangendo apenas uma bolsa de valores e um período relativamente curto (2004-2022). Embora essa abordagem contribua para um entendimento mais aprofundado da realidade brasileira, é recomendável que estudos futuros realizem comparações entre países em desenvolvimento, a fim de investigar similaridades e diferenças na escolha de investimentos em IPO. Sugere-se também que pesquisadores identifiquem aspectos brasileiros que possam ser contrastados com nações desenvolvidas, além de considerar a utilização de variáveis do meio *online*, como *Trends Google*, para identificação de *Market Timing*.

Em segundo lugar, destaca-se a ausência de uma teoria consolidada que oriente o *design* das RNAs, como apontado anteriormente por Jain e Nag (1998). A escolha de utilizar apenas um tipo específico de RNA e um conjunto particular de variáveis constitui uma limitação que abre caminho para estudos futuros, com o intuito de testar diferentes abordagens e buscar maior acurácia. Além disso, é importante ressaltar que a replicabilidade dos resultados pode não ser exatamente igual, devido à aleatoriedade presente no processamento. Portanto, é necessário considerar a possibilidade de ocorrência de generalizações inadequadas (*overfitting*).

Em terceiro lugar, constatou-se que a RNA não conseguiu superar, em termos de risco (representado pelo desvio padrão), a estratégia simulada de escolha aleatória de investimentos. Essa limitação pode ser atribuída, em grande medida, à forma de seleção da variável alvo utilizada neste estudo. Especificamente, a RNA foi treinada considerando apenas o critério de retorno positivo (acima de 1%) no primeiro dia de negociação das ações adquiridas por meio de IPOs. No entanto, seria pertinente que

a análise considerasse também retornos iguais ou superiores à média das empresas já estabelecidas no mercado acionário, proporcionando uma comparação mais abrangente do desempenho das ações selecionadas.

Ademais, um possível aprimoramento na metodologia consistiria em incorporar uma variável adicional que refletisse o desvio padrão dos retornos das demais empresas pertencentes ao mesmo setor das ações adquiridas por IPO. Isso permitiria uma melhor avaliação do aspecto de risco associado à aplicação, levando em consideração a volatilidade relativa do setor em que as empresas atuam.

Em quarto lugar, a utilização da RNA demandará um conhecimento da ferramenta e um processamento computacional que, possivelmente, não está disponível a todo e qualquer investidor. Então, pesquisas futuras poderão desenvolver *softwares* e/ou *sites* com interface de fácil compreensão, a fim de que o grande público possa utilizar.

5.3 SÍNTESE DO ESTUDO

Em resumo, esta dissertação buscou investigar a viabilidade da aplicação de RNAs supervisionadas na tomada de decisões financeiras em IPOs no Brasil. Ao desenvolver um modelo adaptável, identificar variáveis-chave e propor diretrizes para sua implementação efetiva, esta pesquisa contribui tanto no âmbito prático, oferecendo uma ferramenta que auxilia investidores na identificação de ações com potencial de retorno no curto prazo, quanto no âmbito teórico, ao avançar o conhecimento sobre o uso de inteligência artificial nas finanças.

Apesar das limitações identificadas, como a necessidade de pesquisas futuras abrangendo diferentes contextos e a falta de uma teoria consolidada para o *design* das RNAs, os resultados obtidos fornecem uma base para novas investigações e possíveis aplicações em mercados financeiros e áreas similares. Espera-se que este estudo contribua para o aprimoramento das estratégias de investimento em IPOs, impulsionando novas abordagens e expandindo ainda mais o conhecimento sobre a interseção entre inteligência artificial e finanças.

REFERÊNCIAS

- ABDELJAWAD, Islam; MAT NOR, Fauzias. The capital structure dynamics of Malaysian firms: timing behavior vs adjustment toward the target. **International Journal of Managerial Finance**, [s. l.], v. 13, n. 3, p. 226–245, 2017.
- AGHAMOLLA, Cyrus; THAKOR, Richard T. IPO peer effects. **Journal of Financial Economics**, [s. l.], n. xxxx, p. 1–21, 2021. Disponível em: <https://doi.org/10.1016/j.jfineco.2021.05.055>.
- AHMAD-ZALUKI, N A; BADRU, B O. Intended use of IPO proceeds and initial returns. **JOURNAL OF FINANCIAL REPORTING AND ACCOUNTING**, Univ Utara Malaysia, Othman Yeop Abdullah Grad Sch Business, Sintok, Kedah, Malaysia, v. 19, n. 2, p. 184–210, 2021.
- AKERLOF, George A. The market for “lemons”: Quality uncertainty and the market mechanism. In: **UNCERTAINTY IN ECONOMICS**. [S. l.]: Elsevier, 1978. p. 235–251.
- ALAASSAR, Ahmad; MENTION, Anne-Laure; AAS, Tor Helge. Facilitating innovation in FinTech: a review and research agenda. **Review of Managerial Science**, [s. l.], v. 17, n. 1, p. 33–66, 2023.
- ALMEIDA, Camila de Paula. **APLICAÇÃO DE REDES NEURAIS ARTIFICIAIS NA PREVISÃO DO DESEMPENHO DAS OFERTAS PÚBLICAS INICIAIS**. 2013. Trabalho de conclusão de curso - Universidade Federal do Paraná, Curitiba, 2013.
- ALPAYDIN, Ethem. **Introduction to machine learning**. [S. l.]: MIT press, 2020.
- AROSA, Clara Maria Verduch; RICHIE, Nivine; SCHUHMANN, Peter W. The impact of culture on market timing in capital structure choices. **Research in International Business and Finance**, [s. l.], v. 31, p. 178–192, 2014. Disponível em: <http://dx.doi.org/10.1016/j.ribaf.2013.06.007>.
- ASSAF NETO, Alexandre. Mercado financeiro. [s. l.], 2000.
- B3. **Abertura de Capital**. [S. l.], 2022. Disponível em: https://www.b3.com.br/pt_br/produtos-e-servicos/solucoes-para-emissores/abertura-de-capital/como-abrir-o-capital/. Acesso em: 15 ago. 2022.
- B3. **B3 atinge 5 milhões de contas de investidores em renda variável em janeiro**. [S. l.], 2022. Disponível em: https://www.b3.com.br/pt_br/noticias/5-milhoes-de-contas-de-investidores.htm. Acesso em: 1 maio 2022.
- B3. **Segmentos de listagem**. [S. l.], 2022. Disponível em: https://www.b3.com.br/pt_br/produtos-e-servicos/solucoes-para-emissores/segmentos-de-listagem/sobre-segmentos-de-listagem/. Acesso em: 27 nov. 2022.
- BADRU, B O; AHMAD-ZALUKI, N A; WAN-HUSSIN, W N. Board characteristics and the amount of capital raised in the Malaysian IPO market. **JOURNAL OF MULTINATIONAL FINANCIAL MANAGEMENT**, Univ Utara Malaysia, Othman Yeop Abdullah Grad Sch Business, Sintok, Kedah, Malaysia, v. 42–43, p. 37–55, 2017.
- BAE, K H; BAILEY, W; KANG, J. Why is stock market concentration bad for the economy?. **JOURNAL OF FINANCIAL ECONOMICS**, York Univ, Schulich Sch Business, Toronto, ON, Canada, v. 140, n. 2, 2021.
- BAIG, Ahmed S.; CHEN, Mengxi. Did the COVID-19 pandemic (really) positively impact the IPO Market? An Analysis of information uncertainty. **Finance Research Letters**, [s. l.], n. August, p. 102372, 2021. Disponível em: <https://doi.org/10.1016/j.frl.2021.102372>.
- BAKER, Malcolm; WURGLER, Jeffrey. Market timing and capital structure.

Journal of Finance, [s. l.], v. 57, n. 1, p. 1–32, 2002.

BECHT, Marco; BOLTON, Patrick; RÖELL, Ailsa. Corporate governance and control. *In: HANDBOOK OF THE ECONOMICS OF FINANCE*. [S. l.]: Elsevier, 2003. v. 1, p. 1–109.

BERTRAND, J. Will M.; FRANSOO, Jan C. Operations management research methodologies using quantitative modeling. **International Journal of Operations and Production Management**, [s. l.], v. 22, n. 2, p. 241–264, 2002. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/01443570210414338/full/html>.

BHARDWAJ, A; IMAM, S. The tone and readability of the media during the financial crisis: Evidence from pre-IPO media coverage. **INTERNATIONAL REVIEW OF FINANCIAL ANALYSIS**, Univ Warwick, Warwick Business Sch, Coventry CV4 7AL, W Midlands, England, v. 63, p. 40–48, 2019.

BODIE, Zvi; KANE, Alex; MARCUS, Alan J. Investments, 10e. **Maidenhead: McGraw-Hill**, [s. l.], 2014.

BOMFIM, Livia Pontes de Miranda; SANTOS, Carolina Macagnani dos; JÚNIOR, Tabajara Pimenta. Processos de abertura de capital de empresas brasileiras em 2004 e 2005: razões e percepções. **Revista de Administração - RAUSP**, [s. l.], v. 42, p. 524–534, 2006.

BRASIL, PwC. **Como abrir o capital da sua empresa no Brasil (IPO): início de uma nova década de crescimento**. [S. l.]: PwCIL, 2015. *E-book*. Disponível em: <https://www.pwc.com.br/pt/publicacoes/assets/brasil-ipo-guide-2014.pdf>.

BRAU, James C.; FAWCETT, Stanley E. Initial public offerings: An analysis of theory and practice. **Journal of Finance**, [s. l.], v. 61, n. 1, p. 399–436, 2006.

BYARD, D; DARROUGH, M; SUH, J. Re-examining the impact of mandatory IFRS adoption on IPO underpricing. **REVIEW OF ACCOUNTING STUDIES**, CUNY, Baruch Coll, New York, NY 10021 USA, v. 26, n. 4, p. 1344–1389, 2021.

BYRNE, Kathleen. How do consumers evaluate risk in financial products?. **Journal of Financial Services Marketing**, [s. l.], v. 10, n. 1, p. 21–36, 2005.

CALS DE OLIVEIRA, Bruno; MARTELANC, Roy. IPO Determinants of Brazilian Companies. **RBFIn - Brazilian Review of Finance**, [s. l.], v. 12, n. 2, p. 135–161, 2014.

CARAMICO, Augusto. PROPOSIÇÃO DE UM MODELO PREDITIVO DO IBOVESPA UTILIZANDO REDES NEURAIS ARTIFICIAIS. **CAD**, São Paulo, v. 7, n. 2, p. 21–42, 2012. Disponível em: <https://revistas.pucsp.br/caadm/article/view/27917>.

CARBONE, Emmadonata *et al.* IPO in family business: A systematic review and directions for future research. **Journal of Family Business Strategy**, [s. l.], n. xxxx, p. 100433, 2021. Disponível em: <https://doi.org/10.1016/j.jfbs.2021.100433>.

CASAGRANDE NETO, Humberto; SOUZA, Lucy A; ROSSI, Maria Cecília. Abertura do capital de empresas no Brasil: um enfoque prático. **São Paulo: Atlas**, [s. l.], 2000.

CHAHINE, S *et al.* The dark side of venture capital syndication and IPO firm performance: The impact of different institutional environments. *In: THE OXFORD HANDBOOK OF IPOS*. [S. l.: s. n.], 2018. p. 430–459. *E-book*. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85061756621&doi=10.1093%2Foxfordhb%2F9780190614577.013.4&partnerID=40&md5=43ab61c542826cc62c9d85eb5c6618e5>.

CHEMMANUR, T J *et al.* The relationship between venture capital backing and the top management team quality of firms going public and implications for initial public offerings. **JOURNAL OF BUSINESS VENTURING**, Boston Coll, Carroll Sch Management, 336 Fulton Hall, Chestnut Hill, MA 02467 USA, v. 36, n. 6, 2021.

CHEMMANUR, T J; HULL, T J; KRISHNAN, K. Cross-Border LBOs, Human Capital, and Proximity: Value Addition through Monitoring in Private Equity Investments. **JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS**, Boston Coll, Carroll Sch Management, Chestnut Hill, MA 02167 USA, v. 56, n. 3, p. 1023–1063, 2021.

CHEN, You Shyang; CHENG, Ching Hsue. A soft-computing based rough sets classifier for classifying IPO returns in the financial markets. **Applied Soft Computing Journal**, [s. l.], v. 12, n. 1, p. 462–475, 2012. Disponível em: <http://dx.doi.org/10.1016/j.asoc.2011.08.023>.

CHEN, Wei Sen; DU, Yin Kuan. Using neural networks and data mining techniques for the financial distress prediction model. **Expert Systems with Applications**, [s. l.], v. 36, n. 2 PART 2, p. 4075–4086, 2009. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2008.03.020>.

CHI, Li Chiu. Do transparency and disclosure predict firm performance? Evidence from the Taiwan market. **Expert Systems with Applications**, [s. l.], v. 36, n. 8, p. 11198–11203, 2009. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2009.02.099>.

CHUNG, Kim Choy; TAN, Shin Shin; HOLDSWORTH, David K. Insolvency prediction model using multivariate discriminant analysis and artificial neural network for the finance industry in New Zealand. **International journal of business and management**, [s. l.], v. 39, n. 1, p. 19–28, 2008.

COAKLEY, James R.; BROWN, Carol E. Artificial neural networks in accounting and finance: modeling issues. **International Journal of Intelligent Systems in Accounting, Finance & Management**, [s. l.], v. 9, n. 2, p. 119–144, 2000.

COLAK, G *et al.* Political power, local policy uncertainty and IPO pricing. **JOURNAL OF CORPORATE FINANCE**, NYU, Stern Sch Business, New York, NY USA, v. 67, 2021.

COLAK, G; FU, M; HASAN, I. On modeling IPO failure risk. **Economic Modelling**, [s. l.], v. 109, 2022. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85124421011&doi=10.1016%2Fj.econmod.2022.105790&partnerID=40&md5=e6303331335f8d46f6558abcb7d7e438>.

COLAK, G; FU, M C; HASAN, I. Why are some Chinese firms failing in the US capital markets? A machine learning approach. **PACIFIC-BASIN FINANCE JOURNAL**, Hanken Sch Econ, Helsinki, Finland, v. 61, 2020.

CORREIA, Lucas Rubino. **UTILIZAÇÃO DE REDES NEURAIS NA IDENTIFICAÇÃO DE TENDÊNCIAS DE DESEMPENHO DAS OFERTAS PÚBLICAS INICIAIS**. 2015. Trabalho de conclusão de curso Curitiba, 2015.

CVM. **Empresas.NET**. [S. l.], 2022. Disponível em: <https://conteudo.cvm.gov.br/menu/regulados/companhias/prog-empnet.html>. Acesso em: 27 nov. 2022.

CVM. Instrução CVM 480. Dispõe sobre o registro de emissores de valores mobiliários admitidos à negociação em mercados regulamentados de valores mobiliários. **ICVM 480**, Brasília, Comissão de Valores Mobiliários, 2009. Disponível em: <https://conteudo.cvm.gov.br/legislacao/instrucoes/inst480.html>.

CVM. INSTRUÇÃO CVM Nº 400. Dispõe sobre as ofertas públicas de distribuição de valores mobiliários, nos mercados primário ou secundário, e revoga a Instrução CVM nº 13, de 30 de setembro de 1980, e a Instrução CVM nº 88, de 3 de novembro de 1988. **ICVM 400**, Brasília, Comissão de Valores Mobiliários, 2003. Disponível em: <https://conteudo.cvm.gov.br/legislacao/instrucoes/inst400.html>.

CVM. **Resolução CVM 80**. Dispõe sobre o registro e a prestação de informações periódicas e eventuais dos emissores de valores mobiliários admitidos à negociação em mercados regulamentados de valores mobiliários. Brasília, 2022. Disponível em: <https://conteudo.cvm.gov.br/legislacao/resolucoes/resol080.html>.

DA SILVA, José Milton Almeida; FAMÁ, Rubens. Evidências de retornos anormais nos processos de IPO na Bovespa no período de 2004 a 2007: um estudo de evento. **Revista de Administração**, São Paulo, v. 46, n. 2, p. 178–190, 2011. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0080210716301996>.

DE JONG, Abe; KABIR, Rezaul; NGUYEN, Thuy Thu. Capital structure around the world: The roles of firm- and country-specific determinants. **Journal of Banking and Finance**, [s. l.], v. 32, n. 9, p. 1954–1969, 2008.

DE PAULI, Suellen; KLEINA, Mariana; BONAT, Wagner. Comparing Artificial Neural Network Architectures for Brazilian Stock Market Prediction. **Annals of Data Science**, [s. l.], v. 7, n. 4, p. 613–628, 2020. Disponível em: <https://doi.org/10.1007/s40745-020-00305-w>.

DE SPIEGELEIRE, Sthepan; MAAS, Mattijs;; SWEIJS, Tim. What is Artificial Intelligence?. **Hague Centre for Strategic Studies**, [s. l.], 2017. Disponível em: <https://www.jstor.org/stable/pdf/resrep12564.7.pdf>.

DIODATO, RAINNA LEITE. **PREVISÃO DE COTAÇÕES HISTÓRICAS DE AÇÕES UTILIZANDO REDES NEURAIS ARTIFICIAIS**. 2020. 2–20 f. Trabalho de conclusão de curso - Centro Universitário Doutor Leão Sampaio, Juazeiro do Norte, 2020. Disponível em: https://sis.unileao.edu.br/uploads/3/ADMINISTRACAO-2/RAINNA_LEITE_DIODATO.pdf.

DYNIWICZ, Luciana; PEREIRA, Renée. **O material jornalístico produzido pelo Estadão é protegido por lei. As regras têm como objetivo proteger o investimento feito pelo Estadão na qualidade constante de seu jornalismo**. [S. l.], 2021. Disponível em: <https://www.estadao.com.br/infograficos/economia,ipo-o-que-mudou-no-dia-a-dia-das-empresas-que-chegaram-a-bolsa-de-valores,1177649>. Acesso em: 31 out. 2021.

ELGENDY, Mohamed. **Deep learning for vision systems**. [S. l.]: Simon and Schuster, 2020.

ELLIOTT, William B.; KOËTER-KANT, Johanna; WARR, Richard S. Market timing and the debt-equity choice. **Journal of Financial Intermediation**, [s. l.], v. 17, n. 2, p. 175–197, 2008.

FLETCHER, Desmond; GOSS, Ernie. Forecasting with neural networks. An application using bankruptcy data. **Information and Management**, [s. l.], v. 24, n. 3, p. 159–167, 1993.

GARANINA, T; DUMAY, J. Forward-looking intellectual capital disclosure in IPOs Implications for intellectual capital and integrated reporting. **JOURNAL OF INTELLECTUAL CAPITAL**, St Petersburg Univ, Grad Sch Management, St Petersburg, Russia, v. 18, n. 1, p. 128–148, 2017.

GARCÍA-PEÑALVO, Francisco José. La percepción de la Inteligencia Artificial en contextos educativos tras el lanzamiento de ChatGPT: disrupción o pánico. **Education in the Knowledge Society (EKS)**, [s. l.], v. 24, n. SE-Editorial, p. e31279, 2023. Disponível em: <https://revistas.usal.es/tres/index.php/eks/article/view/31279>.

GÉRON, Aurélien. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**. [S. l.]: “O’Reilly Media, Inc.”, 2022.

GOODELL, John W. *et al.* Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis.

Journal of Behavioral and Experimental Finance, [s. l.], v. 32, p. 100577, 2021. Disponível em: <https://doi.org/10.1016/j.jbef.2021.100577>.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. [S. l.]: MIT press, 2016.

GRAHAM, John R.; HARVEY, Campbell R. The theory and practice of corporate finance: Evidence from the field. **Journal of Financial Economics**, [s. l.], v. 60, n. 2–3, p. 187–243, 2001.

GUERATTO, FABRIZIO. **O que o investidor pode aprender com o IPO da Intelbras (INTB3)**. [S. l.], 2021. Disponível em: <https://einvestidor.estadao.com.br/colunas/fabrizio-gueratto/o-que-o-investidor-pode-aprender-com-o-ipo-da-intelbras-intb3/>. Acesso em: 31 out. 2021.

GUJARATI, Damodar N; PORTER, Dawn C. **Econometria básica-5**. [S. l.]: Amgh Editora, 2011.

GULDIKEN, O *et al.* The impact of media coverage on IPO stock performance. **JOURNAL OF BUSINESS RESEARCH**, Manhattan Coll, Sch Business, Bronx, NY 10471 USA, v. 72, p. 24–32, 2017.

GUPTA, D R; VELIYATH, R; GEORGE, R. Influence of national culture on IPO activity. **JOURNAL OF BUSINESS RESEARCH**, Indian Inst Management Bangalore, Bannerghatta Rd, Bangalore 560076, Karnataka, India, v. 90, p. 226–246, 2018.

HAEFKE, Christian; HELMENSTEIN, Christian. Forecasting Austrian IPOs: An application of linear and neural network error-correction models. **Journal of Forecasting**, [s. l.], v. 15, n. 3, p. 237–251, 1996.

HAN, J J; KIM, H J. Stock price prediction using multiple valuation methods based on artificial neural networks for KOSDAQ IPO companies. **INVESTMENT ANALYSTS JOURNAL**, Seoul Sch Integrated Sci & Technol aSSIST, Seoul Business Sch, Seoul 03767, South Korea, v. 50, n. 1, p. 17–31, 2021.

HANCOCK, John T.; KHOSHGOFTAAR, Taghi M. Survey on categorical data for neural networks. **Journal of Big Data**, [s. l.], v. 7, n. 1, 2020. Disponível em: <https://doi.org/10.1186/s40537-020-00305-w>.

HANSELAAR, R M; STULZ, R M; VAN DIJK, M A. Do firms issue more equity when markets become more liquid?. **JOURNAL OF FINANCIAL ECONOMICS**, Aegon, The Hague, Netherlands, v. 133, n. 1, p. 64–82, 2019.

HARRIS, O. The impact of industrial districts on the pricing of IPOs. **QUARTERLY REVIEW OF ECONOMICS AND FINANCE**, East Carolina Univ, Greenville, NC 27858 USA, v. 69, p. 274–285, 2018.

HASTIE, Trevor *et al.* **The elements of statistical learning: data mining, inference, and prediction**. [S. l.]: Springer, 2009. v. 2

HAYKIN, Simon. **Redes Neurais: princípios e prática**. 2ªed. São Paulo: Artmed Editora S.A., 2001.

HEARN, B; FILATOTCHEV, I. Founder retention as CEO at IPO in emerging economies: The role of private equity owners and national institutions. **JOURNAL OF BUSINESS VENTURING**, Univ Southampton, Dept Accounting, CRAAG, Southampton Business Sch, Bldg 2, Univ Rd, Highfield Campus, Southampton SO17 1BJ, Hants, England, v. 34, n. 3, p. 418–438, 2019.

HUIBERS, Fred E. Towards an Optimal IPO Mechanism. **Journal of Risk and Financial Management**, [s. l.], v. 13, n. 6, p. 115, 2020.

INJADAT, Mohammad Noor *et al.* **Machine learning towards intelligent systems: applications, challenges, and opportunities**. [S. l.]: Springer Netherlands, 2021-. ISSN 15737462. v. 54 Disponível em: <https://doi.org/10.1007/s10462-020-09948-w>.

IPEADATA. **Banco de Dados do Instituto de Pesquisa em Economia Aplicada (IPEA)**. [S. l.]: Estatísticas Macroeconômicas, 2022. Disponível em: <http://www.ipeadata.gov.br/Default.aspx>.

JAHANZEB, Agha *et al.* Trade-Off Theory, Pecking Order Theory and Market Timing Theory: A Comprehensive Review of Capital Structure Theories. **International Journal of Management and Commerce Innovations (IJMCI)**, [s. l.], v. 1, n. 1, p. 11–18, 2013.

JAIN, Bharat A.; NAG, Barin N. A neural network model to predict long-run operating performance of new ventures. **Annals of Operations Research**, [s. l.], v. 78, p. 83–110, 1998.

JAIN, Bharat A.; NAG, Barin N. Artificial Neural Network Models for Pricing Initial Public Offerings. **Decision Sciences**, [s. l.], v. 26, n. 3, p. 283–302, 1995.

JENSEN, Michael; MECKLING, William. Theory of the firm: Managerial behavior, agency costs, and ownership structure. **Journal of Financial Economics**, [s. l.], v. 3, n. 4, p. 305–360, 1976.

JOSHIPURA, Mayank; MATHUR, Sachin; GWALANI, Hema. Decrypting IPO pricing: an integrated bibliometric and content analysis approach. **Managerial Finance**, [s. l.], v. 49, n. 1, p. 135–162, 2023.

KAASTRAA, Ieabeling; BOYDB, Milton. Designing a neural network for forecasting financial time series 29. **Neurocomputing**, [s. l.], v. 10, n. 3, p. 215–236, 1996. Disponível em: [https://doi.org/10.1016/0925-2312\(95\)00039-9](https://doi.org/10.1016/0925-2312(95)00039-9).

KDNUGGETS. **Understanding Boxplots**. [S. l.], 2023. Disponível em: <https://www.kdnuggets.com/2019/11/understanding-boxplots.html>.

KHATRI, Namrata N. Factors Influencing Investors Investment in Initial Public Offering. **International Journal of Management and Applied Science**, [s. l.], v. 3, n. 7, p. 64–72, 2017.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. [S. l.]: Springer, 2013. v. 26

LAWRENCE, Ramon. Using Neural Networks to Forecast Stock Market Prices. **Methods**, [s. l.], p. 1–21, 1997. Disponível em: <http://people.ok.ubc.ca/rlawrenc/research/Papers/nn.pdf>.

LEAL, Ricardo P.C. Using accounting information in prospectuses to invest in Brazilian IPOs during high inflation years. **Latin American Business Review**, [s. l.], v. 5, n. 3, p. 65–90, 2005.

LEWGOY, Julia. **Reserva de ações da Brisanet acaba hoje. Vale comprar na oferta?**. [S. l.], 2021. Disponível em: <https://valorinveste.globo.com/mercados/renda-variavel/noticia/2021/07/26/reserva-de-acoes-da-brisanet-acaba-hoje-vale-comprar-na-oferta.ghtml>. Acesso em: 31 out. 2021.

LISTYARTI, Indra; SURYANI, Tatik. Determinant Factors of Investors ' Behavior in Investment Decision in Indonesian Capital Markets Determinant Factors of Investors ' Behavior in Investment Decision in Indonesian Capital Markets. **Journal of Economics, Business, and Accountancy Ventura**, [s. l.], v. 17, n. April, p. 45–54, 2016.

LOUGHRAN, TIM; RITTER, JAY R. The New Issues Puzzle. **The Journal of Finance**, [s. l.], v. 50, n. 1, p. 23–51, 1995.

LUGER, George F. **Inteligência Artificial-: Estruturas e estratégias para a solução de problemas complexos**. [S. l.]: Bookman, 2004.

MALIK, Farooq; NASEREDDIN, Mahdi. Forecasting output using oil prices: A cascaded artificial neural network approach. **Journal of Economics and Business**,

[s. l.], v. 58, n. 2, p. 168–180, 2006.

MARKOWITZ, Harry M. Portfolio Selection. **Journal of Finance**, [s. l.], v. 7, n. 1, 1952.

MARTINEZ, Leonardo C. *et al.* From an artificial neural network to a stock market day-trading system: A case study on the BM&F BOVESPA. **Proceedings of the International Joint Conference on Neural Networks**, [s. l.], p. 2006–2013, 2009.

MARTINS, Marco dos Santos; METTE, Frederike; MACEDO, Guilherme Ribeiro de. A Utilização de Redes Neurais Artificiais para a Estimação dos Preços da Petrobrás PN na Bovespa. **ConTexto**, [s. l.], v. 8, n. 14, p. 1–16, 2008.

MARTINS, Vinícius Aversari; SILVA, Ricardo Luiz Menezes da; NARDI, Paula Carolina Ciampaglia. Governança corporativa e liquidez das ações. **Anais**, [s. l.], 2006.

MATSUO, Alexandre Kazuma; EID JR, W. Influência de fatores macroeconômicos nas emissões primárias do mercado brasileiro. **ENCONTRO BRASILEIRO DE FINANÇAS, Rio de Janeiro**, [s. l.], v. 4, 2004.

MAZUMDER, Sharif; SAHA, Pritam. COVID-19: Fear of pandemic and short-term IPO performance. **Finance Research Letters**, [s. l.], n. September 2020, p. 101977, 2021. Disponível em: <https://doi.org/10.1016/j.frl.2021.101977>.

MCDONALD, M B. The shrinking stock market. **JOURNAL OF FINANCIAL MARKETS**, Fairfield Univ, 1073 N Benson Rd, Fairfield, CT 06824 USA, v. 58, 2022.

MELLO, Antonio S.; PARSONS, John E. Hedging and Liquidity. **Review of Financial Studies**, [s. l.], v. 13, n. 1, p. 127–153, 2000.

MENDONÇA NETO, João Nunes de. **Fractais e redes neurais artificiais aplicados à previsão de retorno de ativos financeiros brasileiros**. 2014. Tese - Universidade de São Paulo, São Paulo, 2014. Disponível em: <https://teses.usp.br/teses/disponiveis/12/12136/tde-02092014-194147/pt-br.php>.

MIRANDA, Kléber Formiga; MELO, Jefferson Ricardo do Amaral; MARTINS, Orleans Silva. Firms' legitimation through corporate governance and its association with risk and return in Brazil. **RAUSP Management Journal**, [s. l.], v. 56, n. 1, p. 55–70, 2021. Disponível em: <https://doi.org/10.1108/RAUSP-05-2020-0087>.

MODIGLIANI, Franco; MILLER, Merton H. Corporate Income Taxes and the Cost of Capital : A Correction Author (s): Franco Modigliani and Merton H . Miller Source : The American Economic Review , Vol . 53 , No . 3 (Jun . , 1963), pp . 433-443 Published by : American Economic Association Stable. **The American Economic Review**, [s. l.], v. 53, n. 3, p. 433–443, 1963.

MUJIATUN, Siti; FERINA, Dwita. Effect of Profitability and Asset Structure on Capital Structure (In Sharia Based Manufacturing Companies In Indonesia Stock Exchange In 2016-2019 Period). **International Journal of Economics**, [s. l.], v. 2, n. 2, p. 383–391, 2021. Disponível em: <http://jurnal.umsu.ac.id/index.php/insis/article/view/6325>.

MYERS, Stewart C.; MAJLUF, Nicholas S. Corporate financing and investment decisions when firms have information that investors do not have. **Journal of Financial Economics**, [s. l.], v. 13, n. 2, p. 187–221, 1984.

NEGNEVITSKY, Michael. **Artificial intelligence: a guide to intelligent systems**. 3rd. ed. [S. l.]: Pearson Prentice Hall, 2011.

NGA, Nguyen Thi Viet; LONG, Giang Ngoc. The choice of capital structure: A study on energy industry in a developing country. **Accounting**, [s. l.], v. 7, n. 2, p. 289–294, 2021.

NGOC, Nguyen Minh; TIEN, Nguyen Hoang; THU, To Huynh. The Impact of Capital Structure on Financial Performance of Logistic Service Providers Listed on Ho

Chi Minh City Stock Exchange. **PalArch's Journal of Archaeology of Egypt/Egyptology**, [s. l.], v. 18, n. 2, p. 688–719, 2021.

NGUYEN, James; PARSONS, Richard. A Study of Market Efficiency in Emerging Markets Using Improved Statistical Techniques. **Emerging Markets Finance and Trade**, [s. l.], v. 00, n. 00, p. 1–13, 2021. Disponível em: <https://doi.org/10.1080/1540496X.2021.1949981>.

NÓBREGA, Marcos; JURUBEBA, Diego Franco de Araújo. Assimetrias de informação na nova Lei de Licitação e o problema da seleção adversa. **Revista Brasileira de Direito Público–RBDP, Belo Horizonte**, ano, [s. l.], v. 18, p. 9–32, 2020.

O'LEARY, Daniel E. Downloads and citations in Intelligent Systems in Accounting, Finance and Management. **Intelligent Systems in Accounting, Finance & Management**, [s. l.], v. 16, n. 1–2, p. 21–31, 2009.

OLIVEIRA, Guilherme Resende *et al.* Determinants of the level of indebtedness for Brazilian firms: A quantile regression approach. **EconomiA**, [s. l.], v. 14, n. 3–4, p. 123–138, 2013. Disponível em: <http://dx.doi.org/10.1016/j.econ.2013.11.002>.

OLIVEIRA, Fagner de; NOBRE, Cristiane; ZÁRATE, Luis. Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index - Case study of PETR4, Petrobras, Brazil. **Expert Systems with Applications**, [s. l.], v. 40, n. 18, p. 7596–7606, 2013. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2013.06.071>.

PAGANO, Marco; PANETTA, Fabio; ZINGALES, Luigi. Why do companies go public? An empirical analysis. **Journal of Finance**, [s. l.], v. 53, n. 1, p. 27–64, 1998.

PAIVA, Felipe Dias. **Redes neurais para decisões no mercado de ações brasileiro**. 2014. 115 f. Tese - Universidade Federal de Lavras, Lavras, 2014. Disponível em: [http://repositorio.ufla.br/jspui/bitstream/1/2827/2/TESE_Redess neurais para decisões no mercado de ações brasileiro.pdf](http://repositorio.ufla.br/jspui/bitstream/1/2827/2/TESE_Redess%20neurais%20para%20decis%C3%B5es%20no%20mercado%20de%20a%C3%A7%C3%B5es%20brasileiro.pdf).

PAO, Hsiao Tien. A comparison of neural network and multiple regression analysis in modeling capital structure. **Expert Systems with Applications**, [s. l.], v. 35, n. 3, p. 720–727, 2008.

PEDREGOSA, Fabian *et al.* Scikit-learn: Machine learning in Python. **the Journal of machine Learning research**, [s. l.], v. 12, p. 2825–2830, 2011.

RAGOZZINO, R; BLEVINS, D P. An investigation of the attention effects of venture capitalist backing on entrepreneurial firms. **LONG RANGE PLANNING**, Univ Liverpool, Management Sch, Chatham St, Liverpool L69 7ZH, Merseyside, England, v. 54, n. 3, 2021.

RAY, Partha Pratim. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. **Internet of Things and Cyber-Physical Systems**, [s. l.], 2023.

REBER, Beat; BERRY, Bob; TOMS, Steve. Predicting mispricing of initial public offerings. **Intelligent Systems in Accounting, Finance and Management**, [s. l.], v. 13, n. 1, p. 41–59, 2005.

RITTER, Jay R.; WELCH, Ivo. A review of IPO activity, pricing, and allocations. **Journal of Finance**, [s. l.], v. 57, n. 4, p. 1795–1828, 2002.

ROBERTSON, Steven J. *et al.* Neural network models for initial public offerings. **Neurocomputing**, [s. l.], v. 18, n. 1–3, p. 165–182, 1998.

ROCK, Kevin. Why new issues are underpriced. **Journal of Financial Economics**, [s. l.], v. 15, n. 1–2, p. 187–212, 1986.

RODRIGUES, Fabiano; RODRIGUES, Francisco Aparecido; ROCHA RODRIGUES, Thelma Valéria. Modelos de machine learning para predição do

sucesso de startups. **Revista de Gestão e Projetos**, [s. l.], v. 12, n. 2, p. 28–55, 2021.

ROSS, Stephen A *et al.* **Administração financeira: Corporate Finance**. 10. ed. Porto Alegre: AMGH Editora, 2015.

ROSSI JR., José Luiz; MAROTTA, Marcelo. Equity Market Timing: Testando através dos IPOs no Mercado Brasileiro. **Brazilian Review of Finance**, [s. l.], v. 8, n. 1, p. 85, 2010.

RUSSELL, Stuart;; NORVIG, Peter. **Artificial intelligence: a modern approach**. 4th. ed. [S. l.]: Pearson Prentice Hall, 2021-. ISSN 1098-6596.

RUSSELL, Stuart;; NORVIG, Peter. **Inteligência Artificial**. 3ªed. Rio de Janeiro: Pearson Education, 2013.

SALCHENBERGER, Linda M.; CINAR, E. Mine; LASH, Nicholas A. Neural Networks: A New Tool for Predicting Thrift Failures. **Decision Sciences**, [s. l.], v. 23, n. 4, p. 899–916, 1992.

SCHIFFMAN, L G; KANUK, L L. **Purchasing behavior**. [S. l.]: Upper Saddle River, NJ: Pearson Prentice Hall, 2007.

SCHWARTZ, Kyle J. Investor participation in initial public offerings. **University of Pennsylvania Law Review**, [s. l.], v. 164, n. 2, p. 563–589, 2016.

SCIKIT LEARN. **Cross-validation: evaluating estimator performance**. [S. l.], [s. d.]. Disponível em: https://scikit-learn.org/stable/modules/cross_validation.html. Acesso em: 20 maio 2023.

SHARPE, William F. Capital Asset Prices: a Theory of Market Equilibrium Under Conditions of Risk. **The Journal of Finance**, [s. l.], v. 19, n. 3, p. 425–442, 1964.

SIMON, Herbert A. Rational decision making in business organizations. **The American economic review**, [s. l.], v. 69, n. 4, p. 493–513, 1979.

SOBRAL, Filipe; PECL, Alketa. **Administração: teoria e prática no contexto brasileiro**. 2. ed. São Paulo: Pearson Prentice Hall, 2014.

SRINIVAS, L;; VENKATA RAO, P; A Study on factors influencing IPO decisions among Retail Individual Investors in Visakhapatnam. **International Journal of Research in Economics and Social Sciences (IJRESS)**, [s. l.], v. 7, n. 9, p. 613–620, 2017.

SRIVASTAVA, Praveen Ranjan; ZHANG, Zuopeng; EACHEMPATI, Prajwal. Deep neural network and time series approach for finance systems: Predicting the movement of the Indian stock market. **Journal of Organizational and End User Computing**, [s. l.], v. 33, n. 5, p. 1–24, 2021.

STEFFEN, Helen Cristina. Abertura de Capital no Brasil : Percepções de Executivos Financeiros * Initial Public Offerings in Brazil : the Perceptions of Financial Executives. [s. l.], p. 102–115, 2012.

SUN, Yi; LI, Shihui; WANG, Rui. Fintech: From budding to explosion-an overview of the current state of research. **Review of Managerial Science**, [s. l.], v. 17, n. 3, p. 715–755, 2023.

SWALES, George S.; YOON, Young. Applying Artificial Neural Networks to Investment Analysis. **Financial Analysts Journal**, [s. l.], v. 48, n. 5, p. 78–80, 1992.

TKÁČ, Michal; VERNER, Robert. Artificial neural networks in business: Two decades of research. **Applied Soft Computing Journal**, [s. l.], v. 38, p. 788–804, 2016.

TORRES, Sérgio. **Previsão do preço de ações brasileiras utilizando redes neurais artificiais**. 2021. Dissertação (Mestrado em Administração de Empresas) - Universidade Presbiteriana Mackenzie, São Paulo, 2021. Disponível em: https://dspace.mackenzie.br/bitstream/handle/10899/28379/SERGIO_TORRES.pdf?sequence=1&isAllowed=y.

TRANFIELD, David; DENYER, David; SMART, Palminder. Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. **British Academy of Management**, [s. l.], v. 14, n. 3, p. 207–222, 2003.

TURRIONI, João Batista; MELLO, Carlos Henrique Pereira. Metodologia de pesquisa em engenharia de produção. **Itajubá: Unifei**, [s. l.], 2012.

VAN AKEN, Joan E. Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules. **Journal of Management Studies**, [s. l.], v. 41, n. 2, p. 219–246, 2004.

VOGLINO, Eduardo. **IPO da Modalmais (MODL11) Vale a Pena? Prazo, Riscos e Prospecto**. [S. l.], 2021. Disponível em: <https://comoinvestir.thecap.com.br/ipo-modal-mais-modl11-vale-a-pena/>.

WANG, Di *et al.* An interpretable neural fuzzy inference system for predictions of underpricing in initial public offerings. **Neurocomputing**, [s. l.], v. 319, p. 102–117, 2018. Disponível em: <https://doi.org/10.1016/j.neucom.2018.07.036>.

WARIN, Thierry; STOJKOV, Aleksandar. Machine Learning in Finance: A Metadata-Based Systematic Review of the Literature. **Journal of Risk and Financial Management**, Basel, v. 14, n. 7, p. 302, 2021. Disponível em: <https://www.mdpi.com/1911-8074/14/7/302>.

WONG, Bo K.; SELVI, Yakup. Neural network applications in finance: A review and analysis of literature (1990-1996). **Information and Management**, [s. l.], v. 34, n. 3, p. 129–139, 1998.

XUEMEI, Chen; YANGBO, Wu. IPO pricing of SME based on artificial neural network. **2009 International Conference on Business Intelligence and Financial Engineering, BIFE 2009**, [s. l.], p. 21–24, 2009.

YAN, Y M *et al.* Uncertainty and IPO initial returns: Evidence from the Tone Analysis of China's IPO Prospectuses. **PACIFIC-BASIN FINANCE JOURNAL**, Tianjin Univ, Coll Management & Econ, China 92 Weijin Rd, Tianjin 300072, Peoples R China, v. 57, 2019.

YANG, J; MA, J Q; DOTY, D H. Family Involvement, Governmental Connections, and IPO Underpricing of SMEs in China. **FAMILY BUSINESS REVIEW**, Univ Texas Tyler, 3900 Univ Blvd, Tyler, TX 75799 USA, v. 33, n. 2, p. 175–193, 2020.

ZHANG, Y A *et al.* Who do you take to tango? Examining pairing mechanisms between underwriters and initial public offering firms in a nascent stock market. **STRATEGIC ENTREPRENEURSHIP JOURNAL**, Rice Univ, Jesse H Jones Grad Sch Business, Houston, TX USA, v. 16, n. 1, p. 97–128, 2022.

ZHOU, X G; TANG, X M; WU, S H. The impact of national culture on IPO underpricing and its influence mechanism: A cross-border empirical research. **FINANCE RESEARCH LETTERS**, Univ Sci & Technol Beijing, Sch Econ & Management, Beijing 100083, Peoples R China, v. 47, 2022.

ZURADA, Jacek. **Introduction to artificial neural systems**. [S. l.]: West Publishing Co., 1992.

APÊNDICES

QUADRO 17 – Possíveis variáveis identificadas para o contexto de IPOs brasileiro

Nº	Análise	Constructos Teórico	Abreviações	Variáveis identificadas	Baseada em:
1	bottom-up	Conjuntura	Ano_ipo; mes	Data de lançamento do IPO	(ALMEIDA, 2013; CORREIA, 2015)
2	bottom-up	Conjuntura	P0	Preço registrado (lançamento)	(ALMEIDA, 2013; CORREIA, 2015)
3	top-down	Conjuntura	exp_selic_ano_ipo_ant	Expectativa mercado da Taxa Selic p/ ANO IPO no ano anterior	(ALMEIDA, 2013; CORREIA, 2015; KHATRI, 2017; AGRAWAL, 2021)
4	top-down	Conjuntura	exp_selic_DIA_ipo	Expectativa mercado da Taxa Selic no DIA do IPO para o ANO vigente	(ALMEIDA, 2013; CORREIA, 2015; KHATRI, 2017; AGRAWAL (2021)
5	top-down	Conjuntura	exp_selic_mes_pos_ipo	Expectativa mercado da Taxa Selic para o MÊS posterior ao IPO	(ALMEIDA, 2013; CORREIA, 2015; KHATRI, 2017; AGRAWAL (2021)
6	top-down	Conjuntura	exp_inflacao_12_meses	Expectativas de mercado para inflação p/ os próximos 12 meses	(ALMEIDA, 2013; CORREIA, 2015; KHATRI, 2017; AGRAWAL (2021)
7	top-down	Conjuntura	ipca_acum_12_meses	IPCA acumulado últimos 12 meses	(ALMEIDA, 2013; CORREIA, 2015)
8	top-down	Conjuntura	var_mensal_ipca_ipo	Var. % mensal IPCA data do IPO	(ALMEIDA, 2013; CORREIA, 2015)
9	top-down	Conjuntura	meta_selic_dia_reg_ipo	Meta Taxa Selic vigente (anual) no DIA IPO	(ALMEIDA, 2013; CORREIA, 2015)
10	top-down	Conjuntura	pont_ibov_fech_reg_ipo	Pontuação IBOV (fechamento do último DIA disponível anterior ao IPO)	(ALMEIDA, 2013; CORREIA, 2015)
11	top-down	Conjuntura	vol_diario_ibov_reg_ipo	Volume diário total negociado na Bolsa de São Paulo - Ibovespa - no DIA DO REGISTRO DO IPO	(JAHANZEB <i>et al.</i> , 2013)
12	top-down	Conjuntura	ice	ICE	(ALMEIDA, 2013; CORREIA, 2015)
13	top-down	Conjuntura	ici	ICI	(ALMEIDA, 2013; CORREIA, 2015)
14	top-down	Conjuntura	icc	ICC	(ALMEIDA, 2013; CORREIA, 2015)
15	top-down	Conjuntura	uf	UF da sede	(ALMEIDA, 2013; CORREIA, 2015)
16	bottom-up	Educação & Conduta	segmento	Segmento de Governança	(ALMEIDA, 2013; CORREIA, 2015)
17	bottom-up	Instituições	setor	Setor	(MARTINS, V. A.; SILVA; NARDI, 2006)
18	bottom-up	Instituições	tipo_oferta	Tipo de Oferta	(ALMEIDA, 2013; CORREIA, 2015)
19	top-down	Mídia	indic_incerteza_eco	Indicador de Incerteza da Economia Brasil - Componente Mídia (Mídia - IIE-Br)	(BHARDWAJ; IMAM, 2019)
20	top-down	Política e Cultura	ano_eleitoral	Ano eleitoral	(COLAK <i>et al.</i> , 2021)
21	bottom-up	Sobrevivência & Capital	idade_emp_ipo	Idade empresa IPO	(BEATTY; RITTER, 1986)
22	bottom-up	Sobrevivência & Capital	vol_financeiro	Volume financeiro total registrado	(ALMEIDA, 2013; CORREIA, 2015)
23	bottom-up	Sobrevivência & Capital	InvestK	Investimento em Capital	(RITTER; WELCH, 2002)
24	bottom-up	Sobrevivência & Capital	AqSoc	Aquisições Societárias	(RITTER; WELCH, 2002)

25	bottom-up	Sobrevivência & Capital	EstrutK	Estrutura de Capital	(RITTER; WELCH, 2002)
26	bottom-up	Sobrevivência & Capital	CapitaK	Capital de Giro	(RITTER; WELCH, 2002)
27	bottom-up	Sobrevivência & Capital	Efic	Eficiência Operacional	(RITTER; WELCH, 2002)
28	bottom-up	Sobrevivência & Capital	EmSecund	Emissão Secundária	(RITTER; WELCH, 2002)
29	bottom-up	Sobrevivência & Capital	paPE	participação <i>Private Equity</i>	(ALMEIDA, 2013; CORREIA, 2015; MINARDI; FERRARI; TAVARES, 2013)
30	bottom-up	Sobrevivência & Capital	cap_terc_cap_prio	Capital de terceiros / Capital próprio	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)
31	bottom-up	Sobrevivência & Capital	comp_endividamento	Composição do endividamento	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)
32	bottom-up	Sobrevivência & Capital	imob_cap_prio	Imobilização do capital próprio	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)
33	bottom-up	Sobrevivência & Capital	imob_rec_ao_cor	Imobilização dos recursos não-correntes	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)
34	bottom-up	Sobrevivência & Capital	gaf	Grau de alavancagem financeira (GAF)	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)
35	bottom-up	Sobrevivência & Capital	LC	Liquidez Corrente	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)
36	bottom-up	Sobrevivência & Capital	GA	Giro do Ativo	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)
37	bottom-up	Sobrevivência & Capital	MB	Margem Bruta	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)
38	bottom-up	Sobrevivência & Capital	MgOp	Margem Operacional	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)
39	bottom-up	Sobrevivência & Capital	ML	Margem Líquida	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)
40	bottom-up	Sobrevivência & Capital	MkGb	Mark-up global	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)
41	bottom-up	Sobrevivência & Capital	ROE	Retorno sobre o Patrimônio Líquido (ROE)	(ADEBIYI et al. 2012; ASSAF NETO, 2000; AGRAWAL, 2021)
42	bottom-up	Sobrevivência & Capital	ROA	Retorno sobre o Ativo (ROA)	(JAIN; NAG, 1998; LEAL, 2005)
43	bottom-up	Sobrevivência & Capital	PL	P/L	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)
44	bottom-up	Sobrevivência & Capital	EbitdaLuc	Ebitda/Lucro Líquido	(ADEBIYI et al. 2012; ASSAF NETO, 2000; JAIN; NAG, 1995)

Fonte: A autora (2023)

QUADRO 18 - Rol de empresas e setores da amostra

ano_ipo	total_ipo	diversidade_s etores	lista_setores	lista_empresas_amostra
2004	4	4	BENS INDUSTRIAIS, UTILIDADE PÚBLICA, SAÚDE, FINANCEIRO E OUTROS	GOL, CPFL ENERGIA, DASA, PORTO SEGURO
2005	7	5	CONSUMO NÃO CÍCLICO, CONSUMO CÍCLICO, CONSTRUÇÃO E TRANSPORTE, UTILIDADE PÚBLICA, TECNOLOGIA DA INFORMAÇÃO	RENAR, SUBMARINO, LOCALIZA, TAM S/A, ENERGIAS BR, OHL BRASIL, UOL
2006	16	9	UTILIDADE PÚBLICA, CONSTRUÇÃO E TRANSPORTE, CONSUMO CÍCLICO, BENS INDUSTRIAIS, PETRÓLEO, GÁS E BIOCOMBUSTÍVEIS, TECNOLOGIA DA INFORMAÇÃO, CONSUMO NÃO CÍCLICO, SAÚDE, FINANCEIRO E OUTROS	COPASA, COMPANY, GAFISA, ABNOTE, CSU CARDSYST, LUPATECH, DATASUL, MEDIAL SAÚDE, KLABINSEGALL, M.DIASBRANCO, BRASCAN RES, PROFARMA, ODONTOPREV, ECODESEL, POSITIVO INF, LOPES BRASIL
2007	47	8	CONSUMO CÍCLICO, CONSTRUÇÃO E TRANSPORTE, FINANCEIRO E OUTROS, CONSUMO NÃO CÍCLICO, COMUNICAÇÕES, BENS INDUSTRIAIS, MATERIAIS BÁSICOS, TECNOLOGIA DA INFORMAÇÃO	TECNISA, CC DES IMOB, PDG REALT, RODOBENSIMOB, IGUATEMI, SAO MARTINHO, GVT HOLDING, EVEN, PINE, JBS, JHSF PART, BR MALLS PAR, METALFRIIO, FER HERINGER, SOFISA, BEMATECH, AGRA INCORP, CREMER, CRUZEIRO SUL, INPAR S/A, LOG-IN, PARANA, SLC AGRICOLA, MARFRIG, EZTEC, TEGMA, DAYCOVAL, REDECARD, PROVIDENCIA, INDUSVAL, MULTIPLAN, ABC BRASIL, TRIUNFO PART, MINERVA, MRV, GUARANI, GENERALSHOPP, SATIPEL, TENDA, TRISUL, BOVESPA HLD, AMIL, HELBOR, BICBANCO, MARISA, PANAMERICANO, TEMPO PART
2008	3	3	MATERIAIS BÁSICOS, CONSUMO CÍCLICO, CONSUMO NÃO CÍCLICO	NUTRIPLANT, LE LIS BLANC, HYPERMARCAS
2009	3	2	FINANCEIRO E OUTROS, SAÚDE	VISANET, CETIP, FLEURY
2010	5	2	FINANCEIRO E OUTROS, BENS INDUSTRIAIS	ALIANSCÉ, BR PROPERT, ECORODOVIAS, MILLS, JULIO SIMOES
2011	7	3	FINANCEIRO E OUTROS, CONSUMO CÍCLICO, SAÚDE	SIERRABRASIL, AREZZO CO, IMC HOLDINGS, TIME FOR FUN, MAGAZ LUIZA, QUALICORP, TECHNOS
2012	2	1	CONSUMO CÍCLICO	LOCAMERICA, UNICASA
2013	8	4	TECNOLOGIA DA INFORMAÇÃO, CONSUMO NÃO CÍCLICO, CONSUMO CÍCLICO, UTILIDADE PÚBLICA	LINX, SENIOR SOL, BIOSEV, SMILES, CPFL RENOVAV, SER EDUCAC, ANIMA, CVC BRASIL
2014	1	1	SAÚDE	OUROFINO S/A
2016	1	1	SAÚDE	ALLIAR
2017	7	6	CONSUMO CÍCLICO, SAÚDE, BENS INDUSTRIAIS, CONSUMO NÃO CÍCLICO, FINANCEIRO E OUTROS, PETRÓLEO, GÁS E BIOCOMBUSTÍVEIS	MOVIDA, IHPARDINI, AZUL*, CARREFOUR BR, IRBBRASIL RE, CAMIL, PETROBRAS BR
2018	1	1	SAÚDE	HAPVIDA
2019	3	2	UTILIDADE PÚBLICA, CONSUMO CÍCLICO	NEOENERGIA, VIVARA S.A., CEA MODAS
2020	26	7	CONSUMO CÍCLICO, TECNOLOGIA DA INFORMAÇÃO, BENS INDUSTRIAIS, SAÚDE, FINANCEIRO E OUTROS, CONSUMO NÃO CÍCLICO, PETRÓLEO, GÁS E BIOCOMBUSTÍVEIS	MITRE REALTY, MOURA DUBEUX, LOCABEW, PRINER, ESTAPAR, AMBIPAR, GRUPO SOMA, D1000VFARMA, QUERO-QUERO, PAGUE MENOS, LAVVI, PETZ, PLANOEPLANO, CURY S/A, HIDROVIAS, MELNICK, BOA VISTA, SEQUOIA LOG, GRUPO MATEUS, TRACK FIELD, MELIUIZ, ENJOEI, AERIS, 3R PETROLEUM, REDE D OR, NEOGRID
2021	27	10	CONSUMO NÃO CÍCLICO, TECNOLOGIA DA INFORMAÇÃO, CONSUMO CÍCLICO, BENS INDUSTRIAIS, COMUNICAÇÕES, UTILIDADE PÚBLICA, MATERIAIS BÁSICOS, SAÚDE, FINANCEIRO E OUTROS, PETRÓLEO, GÁS E BIOCOMBUSTÍVEIS	ESPACOLASER, INTELBRAS, MOBLY, JALLESMACHADO, BEMOBI TECH, WESTWING, OCEANPACT, ELETROMIDIA, ORIZON, CSNMINERACAO, MATER DEI, BLAU, GPS, BOA SAFRA, CAIXA SEGURIT?, PETRORECSA, GETNINJAS, SMART FIT, CBA, DESKTOP, MULTILASER, UNIFIQUE, ARMAC, TC, BRISANET, CLEARSALE, ONCOCLINICAS

Fonte: A autora (2023)

TABELA 13 - Descrição detalhada de índices financeiros em cada setor

	BENS INDUSTRIAIS	COMUNICAÇÃO S	CONSTRUÇÃO E TRANSPORT E	CONSUMO CÍCLICO	CONSUMO NÃO CÍCLICO	FINANCEIRO E OUTROS	MATERIAIS BÁSICOS	PETRÓLEO, GÁS E BIOCOMBUSTÍVEIS	SAÚDE	TECNOLOGIA DA INFORMAÇÃO	UTILIDADE PÚBLICA
	20	5	8	42	20	25	6	5	15	16	6
	contagem										
	média	2,894	7,292	6,522	3,261	3,152	1,824	2,682	2,904	2,086	0,489
	desvio padrão	1,755	6,929	5,402	23,019	3,801	1,421	2,677	1,316	3,676	4,136
	valor										
	mínimo	0,393	2,511	0,009	-104,971	0,621	0,049	0,674	1,096	-2,574	-7,636
cap_terc_cap_prio	1º quartil	1,935	2,880	2,209	0,977	1,078	0,755	1,122	1,538	0,637	0,939
	Mediana	2,440	4,794	6,138	2,471	1,960	3,082	1,242	3,137	1,278	1,658
	3º quartil	3,629	7,005	10,850	5,813	3,444	2,472	3,237	3,926	1,857	2,289
	valor										
	máximo	7,384	19,268	13,779	81,802	17,933	162,479	7,134	4,729	14,656	4,053
	média	0,715	0,899	0,692	0,746	0,652	0,543	0,625	0,705	0,652	0,705
	desvio padrão	0,166	0,203	0,317	0,361	0,154	0,277	0,191	0,107	0,351	0,244
	valor										
cap_terc_cap_total	mínimo	0,282	0,715	0,009	0,152	0,383	0,192	0,403	0,523	0,132	0,413
	1º quartil	0,671	0,742	0,657	0,582	0,509	0,440	0,529	0,605	0,480	0,623
	Mediana	0,761	0,875	0,798	0,714	0,661	0,826	0,555	0,754	0,574	0,664
	3º quartil	0,804	0,940	0,911	0,861	0,766	0,899	0,764	0,797	0,765	0,712
	valor										
	máximo	0,973	1,221	0,924	2,501	0,947	1,232	0,877	0,825	1,635	1,151
	média	0,477	0,369	0,564	0,628	0,511	0,675	0,342	0,491	0,671	0,277
	desvio padrão	0,189	0,125	0,199	0,222	0,185	0,219	0,261	0,229	0,204	0,056
	valor										
comp_endividament o	mínimo	0,102	0,168	0,319	0,143	0,186	0,173	0,135	0,127	0,244	0,170
	1º quartil	0,357	0,350	0,479	0,465	0,339	0,538	0,224	0,360	0,556	0,279
	Mediana	0,478	0,389	0,547	0,673	0,547	0,695	0,235	0,448	0,719	0,290
	3º quartil	0,578	0,441	0,586	0,775	0,635	0,821	0,322	0,572	0,813	0,301
	valor										
	máximo	0,846	0,496	1,000	1,000	0,819	1,000	0,793	0,917	0,997	0,334
	média	2,398	6,185	3,233	0,328	1,896	1,617	2,554	1,942	1,358	1,278
	desvio padrão	1,587	6,357	3,236	13,746	1,298	1,833	2,180	0,836	1,281	3,398
	valor										
imob_cap_prio	mínimo	0,391	2,434	0,555	-80,903	0,688	-3,922	0,854	0,155	-0,162	-5,386
	1º quartil	1,106	2,475	0,734	0,593	1,141	0,894	1,265	1,561	0,631	1,643
	Mediana	2,097	2,978	1,516	1,532	1,463	1,581	1,793	1,817	0,963	2,197
	3º quartil	3,417	5,749	5,817	3,039	2,200	2,140	2,579	2,478	1,583	2,739
	valor										
	máximo	6,477	17,291	8,711	21,760	6,636	6,046	6,279	3,100	4,518	4,280
	média	0,841	0,997	0,656	0,943	0,795	0,871	0,829	0,818	0,657	1,061
	desvio padrão	0,306	0,245	0,358	1,156	0,249	0,549	0,130	0,298	0,367	0,120
	valor										
imob_rec_nao_cor	mínimo	0,231	0,620	0,255	-2,511	0,425	0,000	0,679	0,094	-0,163	0,910
	1º quartil	0,697	0,940	0,327	0,525	0,607	0,495	0,742	0,679	0,470	1,008
	Mediana	0,860	1,016	0,667	0,748	0,763	0,899	0,830	0,838	0,647	1,025
	3º quartil	1,034	1,141	0,957	1,056	0,964	1,123	0,876	0,980	0,937	1,127

gaf	valor máximo	1,357	1,269	1,092	6,167	1,451	2,699	0,966	1,018	1,378	1,224	1,245
	média	3,809	7,965	7,659	1,464	4,203	11,463	2,824	3,682	3,939	3,027	1,601
	desvio padrão	1,839	7,245	5,462	39,762	3,804	32,033	1,421	2,678	1,329	3,691	4,253
	valor mínimo	1,000	3,511	1,009	-230,125	1,621	-4,311	1,049	1,674	2,096	-1,574	-6,636
	1º quartil	2,740	3,880	3,239	2,021	2,078	1,751	1,906	2,122	2,541	1,637	1,939
	Mediana	3,349	3,927	7,528	3,493	2,960	4,082	2,766	2,239	4,535	2,213	2,658
	3º quartil	4,630	8,005	11,902	6,813	4,754	8,607	3,472	4,237	4,926	2,623	3,332
	valor máximo	8,384	20,500	14,906	82,802	18,933	163,479	5,039	8,137	5,729	15,656	5,670
	média	1,322	0,866	2,143	1,493	1,661	1,153	5,962	1,828	1,486	1,530	0,964
	desvio padrão	0,838	0,217	1,403	1,001	0,648	0,743	10,698	0,696	0,734	0,744	0,285
LC	valor mínimo	0,273	0,650	0,858	0,078	0,945	0,136	1,178	0,911	0,744	0,550	0,542
	1º quartil	0,831	0,710	1,329	0,941	1,063	0,771	1,340	1,406	1,093	1,111	0,904
	Mediana	1,207	0,800	1,797	1,181	1,528	1,116	1,760	1,357	1,386	1,386	0,947
	3º quartil	1,487	0,989	2,333	1,908	2,144	1,231	2,004	2,183	1,421	1,885	1,005
	valor máximo	3,857	1,180	5,254	5,049	2,802	4,272	27,787	2,714	3,613	3,704	1,432
	média	1,407	0,905	0,788	1,130	1,579	0,391	0,743	1,027	1,492	2,151	0,495
	desvio padrão	1,127	0,435	0,432	0,909	1,475	0,377	0,621	1,226	1,050	1,803	0,284
	valor mínimo	0,214	0,463	0,211	0,127	0,000	0,000	0,001	0,003	0,582	0,617	0,106
	1º quartil	0,667	0,781	0,576	0,558	0,613	0,178	0,471	0,076	0,843	1,032	0,388
	Mediana	1,120	0,818	0,720	0,898	1,028	0,221	0,627	0,531	1,152	1,539	0,508
GA	3º quartil	1,725	0,828	0,962	1,309	1,998	0,409	0,859	1,645	1,630	2,555	0,524
	valor máximo	4,518	1,635	1,592	4,987	5,295	1,386	1,848	2,880	4,424	6,985	0,973
	média	0,241	0,143	0,703	0,369	0,310	0,547	0,096	-0,019	0,357	0,372	0,063
	desvio padrão	0,358	0,541	0,941	0,861	0,351	1,304	0,199	0,390	0,498	0,899	0,096
	valor mínimo	-0,369	-0,810	0,053	-2,161	-0,106	-0,257	-0,187	-0,445	-0,208	-1,027	-0,101
	1º quartil	0,007	0,295	0,200	0,064	0,103	0,099	-0,005	-0,132	-0,013	0,092	0,022
	Mediana	0,262	0,327	0,259	0,249	0,212	0,207	0,091	-0,104	0,247	0,239	0,096
	3º quartil	0,398	0,372	0,731	0,439	0,389	0,270	0,206	-0,032	0,415	0,320	0,130
	valor máximo	1,207	0,533	2,341	3,509	1,283	4,872	0,372	0,619	1,485	3,444	0,151
	média	0,076	0,046	0,064	0,091	0,061	0,091	0,060	0,011	0,085	0,058	0,011
ROE	desvio padrão	0,104	0,058	0,051	0,175	0,059	0,171	0,114	0,114	0,114	0,234	0,036
	valor mínimo	-0,045	-0,043	0,005	-0,179	-0,044	-0,025	-0,078	-0,081	-0,037	-0,562	-0,025
	1º quartil	0,006	0,029	0,031	0,015	0,037	0,016	-0,003	-0,039	-0,004	0,020	-0,017
	Mediana	0,060	0,066	0,059	0,053	0,051	0,028	0,028	-0,025	0,061	0,061	0,007
	3º quartil	0,115	0,070	0,083	0,145	0,082	0,063	0,131	-0,010	0,118	0,094	0,029
	valor máximo	0,371	0,111	0,156	1,048	0,225	0,637	0,231	0,209	0,368	0,657	0,069
ROA	valor mínimo	-0,045	-0,043	0,005	-0,179	-0,044	-0,025	-0,078	-0,081	-0,037	-0,562	-0,025
	1º quartil	0,006	0,029	0,031	0,015	0,037	0,016	-0,003	-0,039	-0,004	0,020	-0,017
	Mediana	0,060	0,066	0,059	0,053	0,051	0,028	0,028	-0,025	0,061	0,061	0,007
	3º quartil	0,115	0,070	0,083	0,145	0,082	0,063	0,131	-0,010	0,118	0,094	0,029
	valor máximo	0,371	0,111	0,156	1,048	0,225	0,637	0,231	0,209	0,368	0,657	0,069
	média	0,076	0,046	0,064	0,091	0,061	0,091	0,060	0,011	0,085	0,058	0,011
	desvio padrão	0,104	0,058	0,051	0,175	0,059	0,171	0,114	0,114	0,114	0,234	0,036
	valor mínimo	-0,045	-0,043	0,005	-0,179	-0,044	-0,025	-0,078	-0,081	-0,037	-0,562	-0,025
	1º quartil	0,006	0,029	0,031	0,015	0,037	0,016	-0,003	-0,039	-0,004	0,020	-0,017
	Mediana	0,060	0,066	0,059	0,053	0,051	0,028	0,028	-0,025	0,061	0,061	0,007
ROA	3º quartil	0,115	0,070	0,083	0,145	0,082	0,063	0,131	-0,010	0,118	0,094	0,029
	valor máximo	0,371	0,111	0,156	1,048	0,225	0,637	0,231	0,209	0,368	0,657	0,069
	média	0,076	0,046	0,064	0,091	0,061	0,091	0,060	0,011	0,085	0,058	0,011
	desvio padrão	0,104	0,058	0,051	0,175	0,059	0,171	0,114	0,114	0,114	0,234	0,036
	valor mínimo	-0,045	-0,043	0,005	-0,179	-0,044	-0,025	-0,078	-0,081	-0,037	-0,562	-0,025
	1º quartil	0,006	0,029	0,031	0,015	0,037	0,016	-0,003	-0,039	-0,004	0,020	-0,017
	Mediana	0,060	0,066	0,059	0,053	0,051	0,028	0,028	-0,025	0,061	0,061	0,007
	3º quartil	0,115	0,070	0,083	0,145	0,082	0,063	0,131	-0,010	0,118	0,094	0,029
	valor máximo	0,371	0,111	0,156	1,048	0,225	0,637	0,231	0,209	0,368	0,657	0,069
	média	0,076	0,046	0,064	0,091	0,061	0,091	0,060	0,011	0,085	0,058	0,011

Fonte: A autora (2023)