



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS AGRÁRIAS**  
**DEPARTAMENTO DE ECONOMIA AGRÍCOLA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA RURAL**

**IVAN DE OLIVEIRA HOLANDA FILHO**

**ENSAIOS SOBRE O MERCADO FINANCEIRO BRASILEIRO: APLICAÇÃO DE  
MODELOS PREDITIVOS E ANÁLISE DE ANOMALIAS**

**FORTALEZA**

**2024**

IVAN DE OLIVEIRA HOLANDA FILHO

ENSAIOS SOBRE O MERCADO FINANCEIRO BRASILEIRO: APLICAÇÃO DE  
MODELOS PREDITIVOS E ANÁLISE DE ANOMALIAS

Dissertação apresentada ao Programa de Pós-Graduação em Economia Rural do Departamento de Economia Agrícola da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Economia Rural. Linha de Pesquisa: Economia Aplicada ao Agronegócio.

Orientador: Prof. Dr. Vitor Hugo M. C. Silva

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

H669e Holanda Filho, Ivan de Oliveira.  
Ensaio sobre o mercado financeiro brasileiro: aplicação de modelos preditivos e análise de anomalias /  
Ivan de Oliveira Holanda Filho. – 2024.  
96 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Programa de  
Pós-Graduação em Economia Rural, Fortaleza, 2024.  
Orientação: Prof. Dr. Vitor Hugo Miro Couto Silva.

1. Machine Learning. 2. Anomalias. 3. Ações. 4. Mercado Financeiro. 5. Ibovespa. I. Título.  
CDD 338.1

---

IVAN DE OLIVEIRA HOLANDA FILHO

ENSAIOS SOBRE O MERCADO FINANCEIRO BRASILEIRO: APLICAÇÃO DE  
MODELOS PREDITIVOS E ANÁLISE DE ANOMALIAS

Dissertação apresentada ao Programa de Pós-Graduação em Economia Rural do Departamento de Economia Agrícola da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Economia Rural. Linha de Pesquisa: Economia Aplicada ao Agronegócio.

Aprovada em:25/01/2024.

BANCA EXAMINADORA

---

Prof. Dr.Vitor Hugo Miro Couto Silva (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr.Francisco José Silva Tabosa (Membro interno)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr.Vitor Borges Monteiro (Membro externo)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Nicolino Trompieri Neto (Membro externo)  
Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE)

## AGRADECIMENTOS

Agradeço a toda minha família, em especial aos meus pais, Ivan de Oliveira Holanda e Irismar Maria Xavier da Silva Holanda, que, ao longo dos anos, me incentivaram e acreditaram no desenvolvimento do meu potencial acadêmico e profissional.

Este trabalho, como qualquer outro que eu tenha me dedicado, não seria possível de ser feito sem o amor, a confiança e o apoio de minha amada esposa, Patrícia Pimentel Pereira.

A minha filha, Sophia Pimentel Holanda, que trouxe alegria imensurável em nossas vidas com seu lindo sorriso. Espero que, em algum dia, saiba o quanto ajudou nesse processo.

Ao Professor Vitor Hugo Miro Couto Silva, pela paciência, pelo apoio, pela confiança, pelos apontamentos e pelas correções sempre tão pontuais e precisas, além da busca incansável de melhorar o trabalho.

Aos Professores, Francisco José Silva Tabosa, Vitor Borges Monteiro e Nicolino Trompieri Neto, por aceitarem o convite de participação na banca e pelas valiosas sugestões de aperfeiçoamento deste trabalho.

A todos os amigos, pelo apoio e incentivo para a realização de mais esta etapa acadêmica na minha vida e, em especial, aos amigos Marcos Paulo e Moíses.

Aos colegas da turma de Mestrado, Rubens, Trícia, Jayane, Medna, Aline, Nicole, Ingrid, Guilherme e Gescilene pela troca de experiências e pelas percepções sempre caminhando juntos, mesmo com temáticas distintas de trabalhos.

À Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (Funcap), pelo apoio financeiro com a manutenção da bolsa de auxílio.

À Universidade Federal do Ceará, excelente instituição de aprendizado.

Por fim, ao Programa de Pós-Graduação em Economia Rural, pela grande contribuição em minha formação profissional.

“A melhor forma de prever o futuro  
é criá-lo.” (Peter Drucker)

## RESUMO

Esta dissertação é constituída por dois ensaios que têm como tema o mercado brasileiro de ações. O objetivo do primeiro ensaio é aplicar as técnicas de *Machine Learning* para prever movimentos do Índice Bovespa (Ibovespa), uma vez que é o mais importante indicador do desempenho médio das ações negociadas na bolsa brasileira, a B3 (Brasil, Bolsa, Balcão). Para cumprir com objetivos específicos, aplicou-se diferentes modelos e algoritmos de aprendizado supervisionado com propósito de avaliar e estabelecer um comparativo do desempenho preditivo desses modelos. Tendo como referência os diversos estudos na literatura internacional, o trabalho analisa movimentos diários do Ibovespa no período entre os anos de 2012 e 2022. Para tais feitos, foi empregada a linguagem de programação *python* e foi formatada uma base de dados provenientes de diferentes fontes de dados que incluem o *Yahoo Finance*, o Banco Central do Brasil e o IBGE. Os movimentos do Ibovespa foram modelados com a aplicação de modelos Logit, LASSO, *Support Vector Machine*, *Random Forest* e Redes Neurais Artificiais. Adotou-se uma combinação de variáveis predictoras inspiradas em métodos de análise técnica e fundamentalista. Os resultados destacam o desempenho dos modelos Lasso, Logit e do algoritmo de Redes Neurais Artificiais, que apresentam bom desempenho com acurácias próximas de 75%. Por sua vez, o segundo ensaio dedica-se a analisar e fazer inferências a respeito do efeito de anomalias de calendário no mercado de ações no Brasil. Apresenta-se uma revisão de diferentes tipos de anomalias relacionadas ao calendário das quais se destacam os efeitos de ciclo político e os denominados efeito Halloween, efeito janeiro e efeito dia da semana. Com a proposta de realizar as inferências sobre a significância estatística desses efeitos foram aplicados os modelos de regressão logística, para estudar os efeitos sobre movimentos diários, e o modelo de séries temporais ARIMA e ARIMAX para investigar os possíveis efeitos sobre os retornos calculados com base no Ibovespa. Os resultados mostram as evidências de que as variáveis de calendário não apresentam os efeitos significantes sobre os movimentos diários do índice. Por sua vez, diferentes versões dos modelos estimados para avaliar a correlação dessas variáveis sobre os retornos mostraram os resultados diferenciados, os quais são significantes para alguns meses do ano e dias da semana. Nesse aspecto, o conhecimento produzido sobre esse tema torna-se vital. Em um mundo cada vez mais globalizado, em que os mercados e as

ferramentas de investimento estão mais acessíveis aos cidadãos, torna-se importante verificar a ocorrência de tais eventos e produzir as evidências para que os agentes possam definir as melhores estratégias de atuação nos mercados.

**Palavras-chave:** mercado financeiro; machine learning; anomalia financeira.

## ABSTRACT

This dissertation is composed of two essays that have the Brazilian stock market as their central theme. The objective of the first test is to apply Machine Learning techniques to predict movements in the Bovespa Index (Ibovespa), which is the most important indicator of the average performance of shares traded on the Brazilian stock exchange, B3 (Brasil, Bolsa, Balcão). To meet specific objectives, different supervised learning models and algorithms were applied with the purpose of evaluating and establishing a comparison of the predictive performance of these models. Using several studies in international literature as a reference, the work analyzes daily movements of the Ibovespa in the period between 2012 and 2022. For these purposes, the *Python* programming language was used and a database was formatted from different data sources, which include *Yahoo Finance*, the Central Bank of Brazil and IBGE. Ibovespa movements were modeled using Logit, LASSO, Support Vector Machine, Random Forest and Artificial Neural Networks models, adopting a combination of predictor variables inspired by technical and fundamental analysis methods. The results highlighted the performance of the Lasso, Logit models and the Artificial Neural Networks algorithm, which showed good performance with accuracies close to 75%. In turn, the second essay is dedicated to analyzing and making inferences regarding the effect of calendar anomalies on the stock market in Brazil. The study is dedicated to reviewing different types of anomalies related to the calendar, of which the political cycle effects and the so-called Halloween effect, January effect and day of the week effect stand out. With the aim of making inferences about the statistical significance of these effects, logistic regression models were applied to study the effects on daily movements, and the ARIMA and ARIMAX time series models were applied to investigate possible effects on returns calculated based on the Ibovespa. The results show evidence that calendar variables do not have significant effects on the daily movements of the index. In turn, different versions of the models estimated to evaluate the correlation of these variables on returns showed different results, which were significant for some months of the year and days of the week. In this aspect, the knowledge produced on this topic becomes vital. In an increasingly globalized world in which markets and investment tools are more accessible to citizens, it is important to verify the occurrence of these events and produce evidence so that agents can define better operating strategies in these markets.

**Keywords:** financial market; machine learning; financial anomaly.

## LISTA DE FIGURAS

Figura 1 – Rede Neural Simples com equivalência de regressão linear.....	28
Figura 2 – Rede neural multicamadas com equivalência a regressão múltipla.....	28
Figura 3 – Matriz de Confusão.....	38
Figura 4 – Mapeamento de um espaço bidimensional para um espaço tridimensional de dados.....	42
Figura 5 – Anomalia em forma bidimensional simples.....	58
Figura 6 – Série temporal de retornos do Ibovespa (2012-2022).....	79
Figura 7 – Histograma da distribuição dos retornos (2012-2022).....	79
Figura 8 – Funções de autocorrelação e autocorrelação parcial.....	80

## LISTA DE GRÁFICOS

Gráfico 1 –	Número de investidores milhões em CPFs.....	24
Gráfico 2 –	Série do Índice Ibovespa de 2012 e 2022.....	45
Gráfico 3 –	Movimentos do Mercado de altas e baixas do Ibovespa entre 2012 e 2022.....	45
Gráfico 4 –	Retorno do Ibovespa entre 2012 e 2022.....	46
Gráfico 5 –	Curva ROC dos modelos preditivos para o índice Ibovespa na análise de 2012 e 2022.....	49

## LISTA DE TABELAS

Tabela 1 – Estatística descritiva do preço de fechamento do Ibovespa entre 2012 e 2022.....	44
Tabela 2 – Teste de Estacionariedade Dickey-Fuller e Teste de Estacionariedade Kwiatkowski-Phillips Schmidt-Shin (KPSS) na série do retorno do Ibovespa entre 2012 e 2022.....	47
Tabela 3 – Performance de modelos para análise de desempenho.....	49
Tabela 4 – Altas e baixas totalizadas por ano de 2014 e 2022.....	70
Tabela 5 – Análise do Índice Ibovespa para período de eleição para presidente do Brasil em 2022.....	71
Tabela 6 – Efeitos políticos na variação percentual ao Ibovespa.....	72
Tabela 7 – Efeito Halloween para o período de 2012 e 2022 do índice Ibovespa.....	74
Tabela 8 – Resultado do Modelo Logit para variações do Ibovespa.....	78
Tabela 9 – Teste de Estacionariedade Dickey-Fuller e Teste de Estacionariedade Kwiatkowski-Phillips Schmidt-Shin (KPSS) na série do retorno do Ibovespa entre 2012 e 2022.....	80
Tabela 10 – Resultados da estimação dos modelos ARIMA e ARIMAX para os retornos do Ibovespa.....	81

## LISTA DE QUADROS

Quadro 1 – Variáveis técnicas utilizadas para modelagem de previsão do Ibovespa.....	33
Quadro 2 – Variáveis macroeconômicas e Índice S&P 500 para previsão do Ibovespa.....	34
Quadro 3 – Métricas de desempenho para validação e comparação de modelos.....	36
Quadro 4 – Tipos de anomalias no mercado de ações.....	61

## LISTA DE SIGLAS

ADR	American Depositary Receipts
AIC	Critérios de Informação Akaike
ARIMA	Auto - Regressive Integrated Moving Average
ARIMAX	Auto- Regressive Moving Average Model including Exogenous covariates
BIC	Critérios de Informação Bayesiano
B3	Brasil, Bolsa, Balcão
CAPM	Capital Asset Pricing Model
CART	Classification and Regression Trees
CPFs	Cadastro Pessoa Físicas
HME	Hipotese do Mercado Eficiente
EMA	Médias Móveis Exponenciais
FGV	Faculdade Getúlio Vargas
IA	Inteligência Artificial
IBM	International Business Machines
IBGE	Instituto Brasileiro de Geografia e Estatística
IGP-M	Índice Geral de Preços-Mercado
INDFUT	Ibovespa Futuro
IPCA	Índice de Preços ao Consumidor Amplo
LSTM	Long ShortTerm Memory
MACD	Moving Average Converge Divergence
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
NL	Neuron Level

OMS	Organização Mundial da Saúde
RF	Random Forest
RMSE	Root Mean Square Error
RNA	Redes Neurais Artificiais
ROC	Receiving Operator Characteristics
RSI	Índice de Força Relativa
SL	Synapse Level
SMA	Médias Móveis Simples
S&P 500	Standard andPoor's 500
S&P CNX NIFTY	Indice de referência do mercado de ações da Índia
SVM	Support Vector Machine
VWAP	Preços Médio Ponderado pelo Volume

## SUMÁRIO

<b>INTRODUÇÃO GERAL.....</b>	<b>18</b>
<b>APLICAÇÃO DE MODELOS DE MACHINE LEARNING PARA PREVER MOVIMENTOS NO MERCADO FINANCEIRO.....</b>	<b>20</b>
<b>1 INTRODUÇÃO.....</b>	<b>20</b>
<b>2 REVISÃO DE LITERATURA.....</b>	<b>23</b>
<b>2.1 O Índice Ibovespa e sua importância no mercado financeiro.....</b>	<b>23</b>
<b>2.2 Machine Learning e a previsão de dados no mercado de ações.....</b>	<b>24</b>
<b>2.3 Aplicações de redes neurais artificiais no mercado acionário.....</b>	<b>26</b>
<b>2.4 Máquinas de Vetores de Suporte (Support Vector Machines - SVM) aplicado ao mercado de ações.....</b>	<b>29</b>
<b>3 METODOLOGIA.....</b>	<b>31</b>
<b>3.1 Natureza e fonte de dados.....</b>	<b>31</b>
<b>3.2 Análise técnica e análise fundamentalista para previsão de dados.....</b>	<b>32</b>
<i>3.2.1 Métricas de avaliação de desempenho da análise técnica na previsibilidade do Ibovespa.....</i>	<b>36</b>
<b>3.3 Modelos de Machine Learning.....</b>	<b>38</b>
<i>3.3.1 Modelo Logit.....</i>	<b>39</b>
<i>3.3.2 Regressão de Lasso.....</i>	<b>40</b>
<i>3.3.3 Modelo Random Forest.....</i>	<b>40</b>
<i>3.3.4 Modelo de Support Vector Machine (SVM).....</i>	<b>41</b>
<i>3.3.5 Modelo de Redes Neurais Artificiais.....</i>	<b>43</b>
<b>4 RESULTADOS E DISCUSSÕES.....</b>	<b>44</b>
<b>4.1 Estatística Descritiva.....</b>	<b>44</b>
<b>4.2 Série temporal: análise do Ibovespa entre 2012 e 2022.....</b>	<b>47</b>
<b>4.3 Performance dos modelos de previsão de dados.....</b>	<b>48</b>
<b>5 CONSIDERAÇÕES FINAIS.....</b>	<b>52</b>
<b>UM ESTUDO SOBRE ANOMALIAS NO MERCADO BRASILEIRO DE AÇÕES.....</b>	<b>54</b>
<b>1 INTRODUÇÃO.....</b>	<b>54</b>
<b>2 REVISÃO DE LITERATURA.....</b>	<b>57</b>
<b>2.1 Anomalias no mercado financeiro de ações.....</b>	<b>57</b>

<b>2.2</b>	<b>Tipos de anomalias no mercado de ações.....</b>	<b>60</b>
<b>2.3</b>	<b>Anomalias financeiras no Brasil: fontes e evidências.....</b>	<b>62</b>
<b>3</b>	<b>METODOLOGIA.....</b>	<b>65</b>
<b>3.1</b>	<b>Método de análise.....</b>	<b>65</b>
3.1.1	<i>Modelo Logit.....</i>	65
3.1.2	<i>Modelos ARIMA e ARIMAX.....</i>	67
<b>4</b>	<b>RESULTADOS E DISCUSSÕES.....</b>	<b>69</b>
<b>4.1</b>	<b>Anomalias de efeitos de ciclos políticos no Brasil.....</b>	<b>69</b>
4.1.2	<i>Efeitos políticos e momentos diferenciados que podem alterar o Ibovespa.....</i>	72
<b>4.2</b>	<b>Efeito Halloween.....</b>	<b>73</b>
<b>4.3</b>	<b>Efeito dia da semana.....</b>	<b>74</b>
<b>4.4</b>	<b>Efeito Janeiro.....</b>	<b>75</b>
<b>4.5</b>	<b>Resultados da análise econométrica.....</b>	<b>76</b>
4.5.1	<i>Modelo Logit para variações no índice Ibovespa (2012 – 2022).....</i>	76
4.5.2	<i>Modelo ARIMAX para análise do retorno do Índice Ibovespa (2012 – 2022).....</i>	79
<b>5</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>83</b>
<b>6</b>	<b>CONCLUSÃO.....</b>	<b>84</b>
	<b>REFERÊNCIAS.....</b>	<b>85</b>

## 1 INTRODUÇÃO GERAL

Esta dissertação é constituída por dois ensaios. De modo geral, estão relacionados ao mercado financeiro, mais especificamente ao mercado acionário do Brasil. Trata-se de sistemas dinâmicos e complexos, pois desempenham um papel extremamente importante na economia e sociedade moderna, em que a informação é um bem “inestimável” (Cavalcante *et al.* 2006).

Nesse sentido, a informação torna-se imprescindível e realizar previsões dos movimentos de mercado implica uma tarefa extremamente difícil. De acordo com a Hipótese de Mercado Eficiente (HME), os mercados são eficientes quando os preços refletem todas as informações relevantes. A hipótese de mercados eficientes, como definida por Fama (1970), advoga que o mercado seria considerado eficiente se refletisse qualquer informação disponível nos preços dos ativos, fato que inclusive impossibilita ganhos anormais. Assim, os movimentos de preços e retornos são afetados pela interação de uma infinidade de fatores que incluem não apenas indicadores financeiros e econômicos, mas também eventos políticos e expectativas dos agentes que operam nesses setores.

A direção do índice do mercado de ações refere-se ao movimento do índice de preços ou à tendência de flutuação do índice do mercado de ações em períodos futuros. Prever a direção é uma questão prática que influencia fortemente as decisões de compra ou venda dos agentes. Previsões mais acuradas podem garantir o aproveitamento de oportunidades de lucro ou a proteção contra perdas (*hedge*).

O objetivo do primeiro ensaio é aplicar técnicas de *Machine Learning* para prever movimentos do Índice Bovespa (Ibovespa), que é o mais importante indicador do desempenho médio das ações negociadas na bolsa brasileira, a B3 (Brasil, Bolsa, Balcão). Paixão *et al.* (2022) afirmam que o aprendizado de máquina é um ramo da inteligência artificial (IA) que explora o estudo e a construção de algoritmos computacionais a partir do aprendizado mediante dados.

Khaidem *et al.* (2016) elucidam que os métodos de aprendizagem em “conjunto permanecem inexplorados neste campo”. Como objetivos específicos, pretende-se aplicar diferentes métodos de previsão e realizar um comparativo do desempenho entre eles.

Nesse sentido, o presente trabalho procura integrar os vários métodos de *Machine Learning* para fazer a previsão de dados, além de integrar a análise técnica e fundamentalista para a predição do índice Ibovespa. Modelo de Lasso, modelo Logit, *Random Florest*,

*Support Vector Machine* e Redes Neurais Artificiais. Estes, ao serem empregados, tiveram boa acurácia, com destaque, de maneira geral, para a regressão lasso.

O segundo ensaio traz características das anomalias no mercado de ações que desafiam a noção de eficiência. Ahmed *et al.* (2017) confirmam que a detecção de anomalias é uma tarefa difícil e muito importante na análise de dados.

Nesse contexto, são analisadas as anomalias, no sentido de saber o que elas são e representam, os tipos no mercado de ações e as anomalias no mercado financeiro do Brasil. No tocante às anomalias mais comuns, constantemente estudadas, destacam-se aquelas que acontecem em períodos específicos de tempo, conhecidas como anomalias de sazonalidade. Quatro delas serão aqui analisadas: o efeito de ciclo político, o efeito *halloween*, o efeito janeiro e o efeito do dia da semana.

O objetivo é fazer inferências através da aplicação de modelos econométricos a respeito da significância estatística de variáveis associadas às anomalias de calendário ou sazonais e o comportamento do Ibovespa. Para tal feito, foram utilizadas variáveis *dummies* para os dias de semana e meses do ano, além das variáveis que caracterizam os movimentos do mercado. Desse modo, cabe verificar os retornos anormais no Índice Ibovespa.

Para que uma anomalia seja estatisticamente significativa, ela deve gerar retornos anormais (Jensen, 1978). Foram estimados dois tipos de modelos. O primeiro deles traduz os movimentos do Ibovespa em uma variável binária e testa a correlação desta variável com *dummies* de calendário. O segundo explora a estrutura de série temporal dos dados e estima modelos ARIMAX para os retornos do Ibovespa em função das *dummies* de calendário em duas especificações diferentes, uma sem controles e outra com controles dados por variáveis financeiras e econômicas. Os resultados da primeira análise mostram que as variáveis de calendário não possuem correlações significativas com os movimentos do mercado. Na segunda análise, os resultados variaram em razão das especificações estimadas. No modelo apenas com termos autorregressivos, de médias móveis e variáveis de calendários, estas últimas não apresentaram significância estatística. No modelo mais completo com a presença de controles, alguns coeficientes mostraram-se significantes.

# APLICAÇÃO DE MODELOS DE MACHINE LEARNING PARA PREVER MOVIMENTOS NO MERCADO FINANCEIRO

## 1 INTRODUÇÃO

Com o desenvolvimento da internet surgiram novas maneiras de se comunicar, interagir, estudar etc. Essas mudanças também se aplicam à forma como são tomadas as decisões de investimentos. De acordo com Badolia (2016), o desenvolvimento do mercado financeiro, mais especificamente o mercado de ações, foi intensificado nas últimas décadas devido à motivação constante de investidores em aumentar os seus lucros. Fato que colabora para uma maior popularidade.

Diante desse contexto, o interesse médio da população cresceu exponencialmente com a maior publicidade e com a facilidade de acesso. O número cada vez maior de acionistas e empresas demanda encontrar uma solução para prever as tendências no mercado acionário, alvo constante de analistas, pesquisadores e investidores que buscam a todo instante novas informações e modelos que subsidiem as decisões de investimento (Mokhtari *et al.* 2021).

A eficácia na tomada de decisões no mercado financeiro está intrinsecamente ligada à qualidade e rapidez na absorção de informações relevantes. A Hipótese de Mercado Eficiente (HME), proposta por Eugene Fama, postula que os preços dos ativos refletem instantaneamente todas as informações disponíveis. Nesse contexto, o preço dos ativos no mercado segue um “passeio aleatório” e mudanças futuras não podem ser previstas com informações usuais e/ou existentes (Fama, 1970).

No entanto, a constante evolução do mercado e os desafios na assimilação instantânea de informações levantam questionamentos sobre a total eficiência da HME. Além disso, o preço dos ativos não é unicamente determinado pela informação disponível, mas também é fortemente influenciado por uma complexa interação de fatores econômicos, políticos e sociais, bem como as expectativas formadas sobre eventos futuros. Realizar previsões dos movimentos de mercado torna-se uma tarefa extremamente difícil e a compreensão desses elementos é essencial para uma análise mais abrangente e informada, contribuindo para a formulação de estratégias de investimento mais sólidas em um cenário financeiro dinâmico e multifacetado.

Diante da complexidade na formação de preços dos ativos, torna-se evidente a

necessidade de os investidores buscarem não apenas compreender, mas também antecipar essas dinâmicas. A habilidade de fazer previsões sobre os movimentos do mercado é, portanto, uma peça-chave na tomada de decisões informadas. A análise preditiva, respaldada por uma compreensão abrangente dos elementos que contribuem para os movimentos do mercado, permite a adoção de medidas proativas para otimizar os resultados e minimizar riscos. Assim, ao integrar a compreensão da HME com a necessidade de antecipação diante da complexidade dos mercados, os investidores podem aprimorar a sua capacidade de navegar efetivamente em um ambiente financeiro em constante transformação.

Os avanços computacionais das últimas décadas permitiram uma maior visibilidade das técnicas de ciência de dados, dos modelos de aprendizado de máquina (*Machine Learning*) e da inteligência artificial. Nos mercados financeiros, a integração deste instrumental na “caixa de ferramentas” dos analistas e estudiosos do mercado financeiro foi um processo natural, diante das vantagens desse tipo de técnica para gerar previsões cada vez mais precisas.

Ao contrário da maioria dos métodos estatísticos mais tradicionais, os modelos de *Machine Learning* possuem a capacidade única de processar grandes volumes de dados de maneira rápida e eficiente, e sua flexibilidade permite a identificação de padrões e correlações complexas. Algoritmos de *Machine Learning* também se adequaram bem em razão de sua habilidade para lidar com a natureza dinâmica e não linear dos mercados financeiros.

Além disso, modelos de *Machine Learning* podem incorporar uma ampla variedade de variáveis, incluindo fatores econômicos, políticos e sociais, ampliando assim a abrangência da análise preditiva. Essa capacidade de considerar múltiplos inputs simultaneamente permite que uma ampla variedade de variáveis seja considerada, o que proporciona uma visão mais completa e contextualizada, possibilitando previsões mais robustas e sintonizadas com a complexidade do ambiente financeiro.

Considerando tal contexto, o presente estudo propôs aplicar técnicas de *Machine Learning* para prever movimentos do Índice Bovespa (Ibovespa), que é o mais importante indicador do desempenho médio das ações negociadas na bolsa brasileira, a B3 (Brasil, Bolsa, Balcão). Esse índice foi criado em 1968 e reúne as empresas mais importantes do cenário do país no mercado de capitais (B3, 2022).

A previsão dos movimentos do Ibovespa é uma questão prática que influencia fortemente as decisões de compra ou venda dos agentes que atuam no mercado. Previsões mais acuradas podem garantir o aproveitamento de oportunidades de lucro ou a proteção contra perdas (*hedge*).

Como objetivos específicos, pretende-se aplicar diferentes métodos de previsão e realizar um comparativo do desempenho entre eles. Estudos, pesquisas e constantes modelos e/ou sistemas são usufruídos para fazer previsão “sofisticadas” com ações (Song *et al.* 2019). Inúmeros estudiosos têm se dedicado a tentar prever e/ou gerar padrões no mercado de ações com a intenção clara de otimizar ganhos e minimizar riscos.

É por meio da pesquisa que se torna possível desenvolver, de forma crítica, teorias e/ou comparar modelos existentes a fim de obter novas evidências científicas (Tachizawa; Mendes, 2003). Tal atividade caracteriza-se como uma pesquisa quantitativa como fonte de referencial teórico principais artigos e livros da temática cujos autores de destaque para o presente estudo são: Kara *et al.* (2011); Mokhtari *et al.* (2021); Zhang *et al.* (2019) e Huang *et al.* (2005).

O período de análise foi definido entre os anos de 2012 e 2022. Nesse período, o país passou por recessões econômicas, mudanças de cenários políticos, vivenciou uma pandemia e o primeiro ano de uma guerra internacional (entre Rússia e Ucrânia). Trata-se de um recorte temporal com vários eventos que podem influenciar diretamente o Ibovespa e, por isso, pode ser também um contexto bastante interessante para a análise.

Os movimentos diários do mercado foram traduzidos em uma variável binária, de forma a categorizar altas e baixas no Ibovespa. Entre os preditores, foi considerado um conjunto amplo de variáveis que incluem métricas bastante utilizadas na análise técnica e variáveis fundamentalistas que retratam a conjuntura financeira e macroeconômica. No exercício empírico realizado, foram aplicados os modelos de classificado da classe Logit, Logit regularizado (LASSO), floresta aleatória (*Random Florest*), máquina de vetores de suporte (*Support Vector Machine*) e Redes Neurais Artificiais. Os resultados obtidos destacaram o desempenho dos modelos Logit, Lasso e do algoritmo de Redes Neurais Artificiais, que apresentaram bom desempenho com acurácias próximas de 75%. São resultados interessantes frente aos que foram obtidos na literatura consultada.

Além da introdução, o trabalho está estruturado em outras quatro seções. A segunda seção é composta por uma revisão de literatura em que são descritos artigos sobre o Ibovespa e os métodos e/ou algoritmos preditivos tais como os modelos Logit, Lasso, *Support Vector Machine* e Redes Neurais Artificiais. A terceira seção descreve os aspectos metodológicos, reportando a natureza e a fonte de dados e a descrição de cada modelo. Na quarta seção, são apresentados os resultados e a discussão com análises de estatística descritiva, da série temporal do Ibovespa e as métricas empregadas para a avaliação de desempenho dos modelos. Por último, são tecidas as considerações finais do estudo.

## 2 REVISÃO DE LITERATURA

Nesta seção, serão descritos o Índice Ibovespa e a sua importância para os investidores no mercado acionário. Posteriormente, há um comentário sobre o aprendizado de máquinas (*Machine Learning*- ML). Os tópicos seguintes retratam sobre os algoritmos de Redes Neurais e *Support Vector Machine* em previsão com aplicações de diversos autores da literatura mundial. O detalhamento de tais algoritmos e entendimento matemático será apresentado a posteriori, na metodologia.

### 2.1 O Índice Ibovespa e a sua importância no mercado financeiro

Oliveira *et al.*, (2006) define que o mercado financeiro é o conjunto de intermediários e prestadores de serviços financeiros que possibilitam a transferência de recursos dos agentes superavitários para os agentes deficitários. Nesse sentido, o Índice Bovespa é o mais importante indicador de desempenho médio na questão de cotações do mercado acionário do Brasil.

Nos estudos de Constant (2006), é um índice extremamente confiável e com uma metodologia do acompanhamento relativamente fácil no qual representa de forma fiel o comportamento médio das principais transações das ações assim como o perfil das negociações observadas nos pregões da BOVESPA. É comum na literatura fazer análises de eficiência, análises de gestão financeira, análise de risco todos estes associados ao Ibovespa.

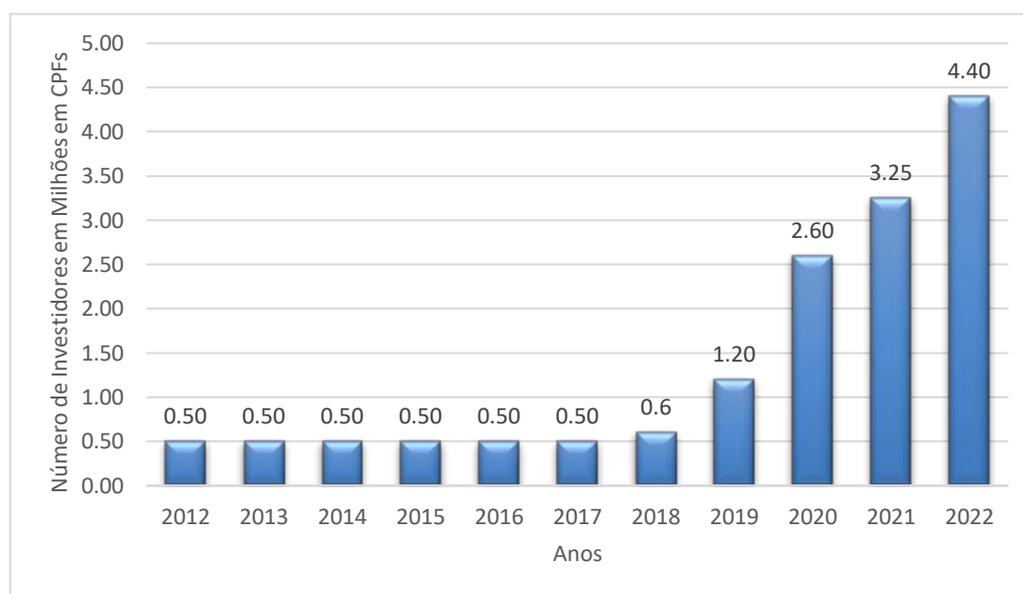
Couto *et al.* (2011) considera que a gestão financeira no período de crise econômica em 2008 das empresas que compõem o Ibovespa foi bem-sucedido, em curto prazo e que, com a rápida política de retomada de crescimento e da confiança na economia do país, foram fundamentais para o período.

Machado *et al.* (2017) verificou que a atividade econômica, a taxa de juros, o nível de exportação e importação, a oferta de moeda e a taxa de câmbio estão relacionadas em longo prazo com o índice. Silva *et al.* (2014) aplica a metodologia de modelos econométricos e verifica que as evidências positivas entre o Ibovespa a bolsa americana e o preço do petróleo é negativa entre a taxa de juros e o risco de crédito e tais variáveis serão estudadas neste ensaio.

Ainda no âmbito econômico, para estimular a economia no primeiro ano de pandemia (2020), os bancos em todo o mundo reduziram os juros e injetaram maior liquidez nos mercados. No Brasil, tal fato contribuiu para busca intensa de melhorar a diversificação de

investimentos, principalmente em ações, o que desencadeou um crescimento de 94% no número de contas na depositária de renda variável em comparação ao ano anterior, encerrando o mesmo ano com pouco mais de 2,6 milhões de investidores de pessoas físicas (B3, 2020). O Gráfico 1 evidencia o número crescente de investidores nos últimos 4 anos.

Gráfico 1- Número de investidores milhões em Cadastro de Pessoas Físicas (CPF's).



Fonte: Adaptado, Relatório sobre análise da evolução dos investidores da B3.

Evidentemente, o Índice Ibovespa é bastante importante na economia e nos diversos autores, pesquisadores, economistas que procuram fazer inferências através de modelos estatísticos e econométricos com intuito de verificar os padrões ou relacionar as principais variáveis do Índice Ibovespa. Dado a magnitude do índice é comum estudos de maior abrangência e métodos mais modernos que possam nortear investidores e analistas na tomada de decisão do mercado acionário.

## 2.2 Machine Learning e a previsão de dados no mercado de ações

Na literatura, encontramos indícios contrários à Hipótese do Mercado Eficiente (EMH), conforme analisado por Atsalakis e Valavanis (2009), bem como por Kumar, Meghwani e Thakur (2016), entre outros estudos. Contudo, a existência de um modelo preditivo capaz de

subsidiar estratégias que consistentemente resultam em retornos acima da média não apenas representa forte evidência contrárias à HME, mas também abre a possibilidade de obter lucros significativos.

Desse modo, o preço de mercado é uma referência importante no mercado de ações, pois é a única referência confiável para os investidores. Essa implicação remete a estudos de Bhalla (2017) que compreendem que todos os investidores no contexto de um mercado eficiente têm acesso a todas as informações relevantes simultaneamente e de forma simétrica. Portanto, a forma como as informações são processadas e analisadas torna-se primordial no estudo e desenvolvimento do mercado de ações em qualquer lugar do mundo.

O cenário contemporâneo de produção de informação no âmbito financeiro é profundamente moldado pelos avanços tecnológicos e computacionais. O advento de tecnologias que permitem a coleta, o armazenamento e o processamento massivo de dados desencadeou uma revolução na forma como as informações são geradas e utilizadas nos mercados financeiros.

A capacidade de armazenar e processar grandes volumes de dados alavancou o desenvolvimento de algoritmos cada vez mais poderosos algoritmos, que permitem a identificação de padrões complexos e a extração de insights valiosos. Nesse contexto, os métodos de *Machine Learning* emergiram como uma ferramenta capaz de aprimora a análise preditiva, bem como abriram a possibilidade de automação de processos, permitindo um novo contexto para a tomada de decisão e formulação de estratégias de investimento.

Em resumo, a interseção entre a produção de informação, os avanços tecnológicos, a ciência de dados e o *Machine Learning* transformou fundamentalmente a paisagem dos mercados financeiros. Essas inovações não apenas proporcionam uma compreensão mais profunda e ágil dos movimentos do mercado, mas também abrem novas oportunidades para as estratégias de investimento inovadoras e sustentáveis.

Os algoritmos de *Machine Learning* são vistos como um ramo da Inteligência Artificial (IA) que está presente em vários modelos e pode atender a diferentes demandas (Silva; Silva; Montebello, 2020). Aplicações importantes na economia estão sendo desenvolvidas como estudo de risco, energia, detecção de falhas, fraudes, inconsistências, e, claro, as aplicações preditivas no mercado de ações.

Aplicações de modelos preditivos com algoritmos de *Machine Learning* podem ter resultados bastante significativos na análise de dados financeiros. Choudhry *et al.* (2008) realizaram estudos no mercado indiano e enfatizaram comparações com o mercado

americano, no qual verificaram que há muitas diferenças entre eles. Diversos autores estudaram *Machine Learning* com aplicações no mercado financeiro. Quatro modelos/algoritmos surgem então como crescentes e predominantes que são: Redes Neurais Artificiais, máquinas de vetores de suporte, algoritmos genéticos combinados com outras técnicas e integração de métodos híbridos ou outras abordagens de inteligência artificial são as principais com aprendizagem de máquinas (Straderet *et al.* 2020)

Chhajer *et al.* (2022) concluem que a tecnologia com o aprendizado de máquinas é revolucionária e tal fato é comprovado por milhares de investidores, mesmo o mercado de ações sendo extremamente volátil. Logo, as aplicações de modelos de aprendizagem de máquina incorporando as particularidades do mercado de ações brasileiro podem ser bastante promissoras.

### **2.3 Aplicações de Redes Neurais Artificiais no mercado acionário**

O mercado de ações é um ambiente altamente complexo que compreende diversas bolsas de valores em torno do mundo. Devido à Lei da Oferta e da Procura, os preços das ações mudam a todo instante, sendo assim, difícil de fazer previsões. Diante disso, surgem novos métodos e algoritmos com a melhoria da tecnologia, e um desses algoritmos que está em evidência é o de Rede Neural Artificiais (RNAs). Trata-se de um algoritmo similar ao cérebro humano que compreende as questões complicadas e que não podem ser resolvidas por aprendizado de máquina de modo mais simples (Chhajer *et al.* 2022).

As Redes Neurais é um método computacional da área da IA, usado em ME que teve inspiração no cérebro humano.

Yu *et al.* (2020) retratam que as Redes Neurais têm grandes vantagens aos métodos convencionais estatísticos que são: capacidade de maior análise de dados “imprecisos e ruidosos”; são amplamente utilizados sem séries temporais; podem ser treinados para grandes volumes de dados e capturam relações não lineares. Ou seja, mais abstratos entre dados.

White (1988) fez a modelagem com aprendizagem de Redes Neurais para decodificar regularidades do movimento de preços de ativos e ações da *International Business Machines (IBM)*. Chenoweth (1995) fez estudos sobre os valores futuros de ações diárias relacionadas ao índice S&P 500 (*Standard and Poor's 500*), índice do mercado de ações das 500 maiores empresas de capital aberto dos Estados Unidos.

Trippi, De Sieno (1992); Bierens (1994); Mason (2014); são alguns dos pesquisadores

que aplicaram e defendem a aplicação de modelos de RNAs no mercado financeiro. Yang *et al.* (2012) explanaram que as Redes Neurais são aplicadas com sucesso para prever preço mais alto/mais baixo da Bolsa de Valores de Xangai e que o modelo pode ser importante para a precificação de derivativos e análise de riscos financeiros.

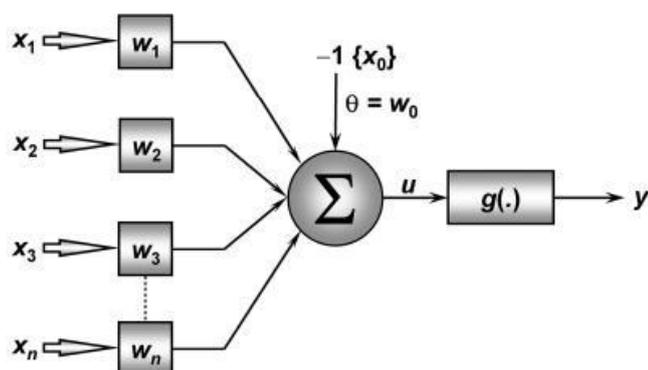
Adebvi *et al.* (2014) conduziu estudos envolvendo as redes neurais para abordar situações de processos não lineares, comparando essas situações do mundo real com o modelo ARIMA de séries temporais, que pressupõe que as séries são geradas por processos lineares. Essa análise demonstrou que os modelos baseados em Redes Neurais podem apresentar uma eficiência superior, uma vez que questões do cotidiano frequentemente envolvem os processos não lineares. As Redes Neurais alcançam os resultados superiores devido à sua capacidade de lidar com as variáveis de entrada e saída, estabelecendo as relações complexas e não lineares (Chong, 2017).

Qiu, Song e Akagi (2016) enfatizam que o uso de Redes Neurais para prever dados financeiros tornou-se um método popular e tal abordagem supera os métodos convencionais. Os mesmos autores fizeram estudos com o índice japonês Nikkei 225 e concluíram que as Redes Neurais com modelos híbridos podem ser eficazes para a predição do índice japonês, no qual o governo faz incentivos estratégicos para estudos econômicos no país.

Caramico (2012) realizou previsões utilizando Redes Neurais e concluiu que esse tipo de modelo demonstra uma notável capacidade de generalização, além de realizar previsões de forma eficaz da direção do Índice Ibovespa no Brasil. Em um estudo comparativo, Cavalheiro *et al.* (2011) investigou as relações entre o Índice Ibovespa no Brasil e o Índice de Merval na Argentina, utilizando Redes Neurais Artificiais. Os resultados indicaram que os índices, em bases mensais durante o período analisado, apresentaram cointegração em termos de causalidade fraca. Além disso, a qualidade do modelo foi mais destacada na identificação de padrões para o índice brasileiro, atribuída à menor volatilidade dos dados em comparação com o índice de Merval.

A Figura 1 representa a equivalência com Redes Neurais em que os preditores (*inputs*) formam uma camada inferior e as previsões (*output*) formam uma camada superior, podendo haver camadas intermediárias de “neurônios ocultos” (Hyndman e Athanasopoulos, 2018).

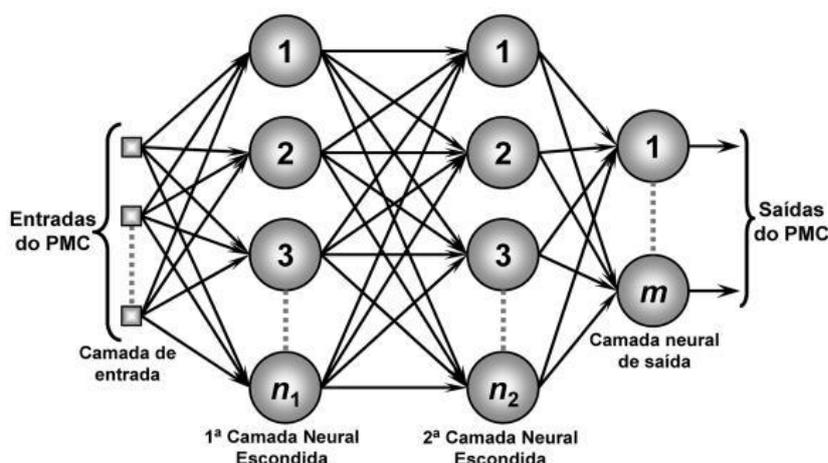
Figura 1-Rede Neural Simples com equivalência de regressão linear.



Fonte: Redes Neurais Artificiais para engenharia e ciências aplicadas, p. 58.

Ainda em Hyndman e Athanasopoulos (2018), os coeficientes relacionados aos preditores são chamados de “pesos” e mediante à combinação linear as previsões são obtidas. Na Figura 1, tem-se quatro preditores de uma Rede Neural Simples, exemplificado em que as entradas  $x_i$  representam as informações sobre o comportamento do desempenho a ser mapeado e  $w_i$  são os pesos sináptico com intuito de “quantificar a importância de cada uma frente aos objetivos funcionais atribuídos ao neurônio”,  $g(.)$  é a função de ativação da rede neural e  $u$  é o potencial de ativação (Silva *et al.* 2010). A Figura 2 representa uma Rede Neural com múltiplas camadas de neurônios, definidas como camadas ocultas.

Figura 2 - Rede Neural multicamadas com equivalência a regressão múltipla.



Fonte: Redes Neurais Artificiais para engenharia e ciências aplicadas, p. 92.

De acordo com Silva *et al.* (2010) as saídas dos neurônios da primeira camada neural

serão as entradas da segunda camada oculta. Ou seja, as saídas de cada neurônio da segunda rede, como ilustra a Figura 2, serão respectivamente as entradas dos neurônios no qual pertencem à sua camada neural de saída. As Redes Neurais são utilizadas, muitas vezes, com outros modelos, com intuito de fazer comparativos e verificar diferenças de desempenho preditivo.

#### **2.4 Máquinas de Vetores de Suporte (Support Vector Machines- SVM) aplicado ao mercado de ações**

As Máquinas de Vetores de Suporte ou *Support Vector Machines* (SVM) são uma técnica “clássica” de aprendizado de máquinas amplamente aplicadas em problemas de classificação. São várias as aplicações de estudo com SVM na literatura com mercado de capitais, principalmente em países como a Índia, China, Japão, Turquia, Cingapura e Estados Unidos da América (Suthaharan, 2016).

Manish, Thenmozhi (2007) utilizaram SVM para estudos sobre o índice S&P CNX NIFTY (índice de referência do mercado de ações da Índia com média ponderada das 50 maiores empresas da indiana) e concluíram que o modelo alcança maior precisão de previsão que outros modelos como ARIMA. Huang *et al.* (2005) investigou previsão da direção do movimento financeiro com SVM ao índice semanal NIKKEI 225 comparando o SVM as análises Discriminante Linear e Análise Discriminante Quadrática e Redes Neurais de Retropropagação de Elman e verificou que o SVM supera outros métodos de classificação e sugerem a aplicação do SVM a outras combinações de classificação.

Kara *et al.* (2011) explicam que o SVM se originou como implementação com Vapnik no livro Princípios de Minimização de Risco Estrutural (1995) que desenvolvia combinações binárias e esse método surgiu de pesquisas estatísticas sobre a estrutura de risco. Ainda conforme os autores, RNA e SVM são ferramentas úteis na previsão dos preços das ações na Bolsa de Valores de Istambul dos anos de 1997 a 2007. Previsões de movimentos de índices são complexas, porém importantes, pois podem desenvolver estratégias eficientes de negócios no mercado (Leung, Daouk, Chen, 2000).

A ideia do método SVM é construir hiperplanos separadores de decisão em que a margem de distribuição entre observações de duas classes seja máxima (Xu, Zhou, Wang, 2009). Lorena e Carvalho (2007) relatam que as SVM é uma ferramenta robusta diante de dados com grandes dimensões e se sobressaem quando comparadas a técnicas de aprendizado

comum por não gerar classificadores “super ou sub ajustados”. Uma desvantagem do modelo é a sensibilidade de escolhas de parâmetros e a interpretação dos resultados desse modelo. Chapele *et al.* (2002); Avelar *et al.* (2022); Awan *et al.* (2021) citam várias métricas usufruídas conjuntamente com SVM como: *Root Mean Square Error* (RMSE - Raiz do Erro Quadrático Médio), *Mean Absolute Error* (MAE - Erro Absoluto Médio) e *Mean Absolute Percentage Error* (MAPE - Erro Percentual Médio Absoluto), além da Accuracy (Acurácia) que é a mais utilizada para avaliar algoritmos de classificação que será abordado no presente ensaio.

### 3 METODOLOGIA

Nesta seção, são apresentadas as variáveis da base de dados e as suas respectivas fontes, assim como a descrição do tratamento desses dados, a caracterização dos indicadores de análise técnica e fundamentalista, além de uma breve descrição dos modelos aplicados.

#### 3.1 Natureza e fonte de dados

A pesquisa caracteriza-se como uma pesquisa quantitativa, adotando como principal referencial teórico e metodológico artigos e livros presentes na literatura sobre o tema. Com o objetivo de realizar as previsões dos movimentos diários de altas e baixas do Ibovespa, foram coletados dados secundários de séries históricas da B3 (Brasil, Bolsa, Balcão). O recorte temporal da análise compreendo o período entre os anos de 2012 e 2022. Os dados foram coletados da plataforma do *Yahoo Finance*, por meio da biblioteca “*yfinance*” para a linguagem *Python*.

Seguindo uma abordagem semelhante à empregada pelos trabalhos de Kara *et al.* (2011) e Huang *et al* (2005), sob posse dos dados relativos aos valores de fechamento do índice em cada dia de operação do mercado no período considerado foi construída uma série de variações ou retornos diários do índice, dados por  $r_t$ , com a aplicação da seguinte equação:

$$r_t = \text{Ln}(Y_t) - \text{Ln}(Y_{t-1}) = \text{Ln}\left(\frac{Y_t}{Y_{t-1}}\right) \quad (1)$$

Nesta expressão  $Y_t$  representa o valor de fechamento do Ibovespa na data  $t$  e  $Y_{t-1}$  o valor de fechamento na data anterior. Com essa caracterização foi definida uma variável binária para traduzir as variações do Ibovespa da seguinte forma:

$$Y = \begin{cases} 0, & \text{se } r_t \leq 0 \text{ no dia } t \\ 1, & \text{se } r_t > 0 \text{ no dia } t \end{cases} \quad (2)$$

Todo o trabalho de coleta e tratamento dos dados, de construção de gráficos e modelagem foi realizado com a aplicação da linguagem *Python*. As bibliotecas empregadas para coleta e análise foram a já mencionada *yfinance*, a *Numpy* e a *Pandas*. Os gráficos foram construídos com o uso das bibliotecas *Matplotlib* e *Seaborn*. Por sua vez, o processo de modelagem foi realizado com diversos métodos disponíveis nas bibliotecas *Scikitlearn*, *Statmodels* e *Tensorflow*.

### 3.2 Análise técnica e análise fundamentalista para previsão de dados

A formatação de preditores para os modelos foi inspirada em diversos trabalhos presentes na literatura. Na revisão realizada, percebeu-se que as variáveis são inspiradas em dois tipos de análises empregadas em mercados financeiros, que são as análises técnica e fundamentalista.

Kumbre *et al.* (2022) analisaram artigos de periódicos publicados entre 2000 a 2019 com foco em técnicas de aprendizado de máquina aplicadas ao mercado de ações e índices de ações, no qual das 2173 variáveis analisadas 1349 era de análise técnica, e com maiores números de variáveis incorporadas aos periódicos analisados existia uma tendência de utilização de variáveis macroeconômicas e de análise fundamentalista. Nti *et al.*, (2019) também realizaram uma revisão sistemática e crítica de periódicos, porém no intervalo de 2007 a 2019 com 122 trabalhos e constataram que 66% dos documentos analisados foram baseados em análises técnicas, 23% em análises fundamentalistas e 11% de análises combinadas na área de previsão do mercado de ações usando aprendizado de máquina.

Assim, muitos trabalhos consideram apenas uma dessas abordagens na formulação de preditores. No presente estudo, optou-se por combinar variáveis inspiradas em ambas as perspectivas, o que compõe apenas uma pequena parcela dos trabalhos; 11% conforme Nti *et al.* (2019). Nesse sentido, faz-se alguns comentários sobre estas abordagens de análise.

O objetivo da análise técnica é aumentar o retorno de uma determinada carteira de investimentos, compreender a interação dos vários indicadores de preços e a participação dessa carteira em um período específico (De Souza *et al.*, 2018). A análise técnica é baseada principalmente no estudo de gráficos e padrões de preços históricos para prever movimentos futuros nos preços dos ativos. Ela pressupõe que toda a informação relevante já está refletida nos preços dos ativos de forma que os padrões históricos tendem a se repetir.

No Quadro 1, são descritas as variáveis técnicas para análise do Ibovespa no presente trabalho.

Quadro 1 - Variáveis técnicas utilizadas para modelagem de previsão do Ibovespa.

VARIÁVEIS DE ANÁLISE TÉCNICA	DESCRIÇÃO	ESTUDOS FUNDAMENTADOS
Médias Móveis Simples (SMA)	Técnica utilizada para analisar determinado dados em um período. Preços de fechamento podem ser integrados dividindo por um determinado número de período.	Ayçel, Santur (2022);Nabipour (2020)
Médias Móveis Exponenciais (EMA)	A média móvel exponencial é uma técnica utilizada complementar a SMA com média ponderada que dá maior peso aos valores mais recentes.	Ayçel, Santur (2022);Nakano,Takahashi, Takahashi (2017)
Índice de Força Relativa (RSI)	É um índice de momentum e sua escala vai de 0 a 100. Oferece métrica de quanto um ativo, por exemplo, caiu ou desceu.	Kara, Boyacioglu, Baykan (2011); Hari, Dewi (2018 )
Momentum ROC	Indicador que analisa a taxa de variação dos preços de um determinado ativo.	Diler (2003);Peachavanish (2016)
Volume OBV	OBV é calculado a partir do preço de fechamento e do volume de negociação de uma ação.	Tsang, Chong (2009);Blume <i>et al.</i> (1994)
Preços Médio Ponderado pelo Volume (VWAP)	Valor médio das ações no mercado por determinado período.	Chu., <i>et al</i> (2009); Zarattini, Aziz (2023)
Volatilidade ATR	Média da variação real e mede a média da amplitude das variações nos preços dos ativos.	Deinwallner (2020);Cohen (2023)
Oscilador de Preços	Relação da diferença entre duas médias móveis de momentos diferentes.	Kara, Boyacioglu, Baykan (2011);Pandya (2021)
Moving Average Converge Divergence (MACD)	Convergência e Divergência de Médias Móveis que é a diferença entre duas médias móveis exponenciais de distintos períodos.	Aguirre <i>et al.</i> (2020);Chakrabarty <i>et al.</i> (2016)

Fonte: Resultados da pesquisa, 2023.

Por sua vez, a análise fundamentalista baseia-se em valores “intrínsecos” dos ativos com relação a fatores financeiros e econômicos. Esse tipo de análise concentra-se nos fundamentos financeiros da empresa, como receitas, lucros, dividendos, perspectivas de crescimento, e na análise da conjuntura econômica, envolvendo principalmente indicadores macroeconômicos. A análise fundamentalista lida com informações técnicas, e também com a situação econômica de um país, de forma que os investidores usam informações de uma determinada série histórica para prever retornos futuros excedentes (Fama, French, 1988).

Conjuntamente com as variáveis do Quadro 1, foram integradas à base de dados as demais variáveis que podem influenciar a direção do índice e fornecer informações sobre a

direção de alta ou baixa. O Quadro 2 explana as variáveis macroeconômicas e financeiras que foram empregadas no presente estudo. Todos esses conjuntos de variáveis foram agrupadas em um único *dataframe* compondo os preditores utilizados na modelagem dos modelos preditivos.

Quadro 2 - Variáveis macroeconômicas e financeiras para previsão do Ibovespa.

(continua)

<b>VARIÁVEIS MACROECONÔMICAS E ÍNDICE S&amp;P 500</b>	<b>FONTE</b>	<b>DESCRIÇÃO</b>
S&P 500 Fechamento	<i>Yahoo Finance.</i>	<i>Standard and Poor's 500 (S&amp;P 500) é o índice de ações das 500 maiores empresas do mundo listada nas Bolsas de Valores dos Estados Unidos da América.</i>
S&P 500 Volume	<i>Yahoo Finance</i>	É o volume de negociações que ocorre diariamente. Tem em média 2 trilhões de negociações por dia.
Poupança Mensal	Banco Central do Brasil	É uma aplicação de renda fixa em que a rentabilidade é a mesma em qualquer instituição bancária.
Taxa Selic	Banco Central do Brasil	É a taxa básica de juros no Brasil e serve para controlar a compra a compra, venda e emissão de títulos.
Indicador de Incerteza da Economia	Faculdade Getúlio Vargas (FGV)	É composto por dois indicadores: O Indicador de Incerteza na Mídia (IIE-Br-Mídia) em que aproximadamente 30 mil notícias são analisadas mensalmente e tem 80% do indicador e o Indicador de Dispersão de Expectativas (IIE-Br-Expectativa), que utiliza previsões de especialistas para a taxa de câmbio, Selic e IPCA e tem peso de 20% do indicador.

## Quadro 2 - Variáveis macroeconômicas e financeiras para previsão do Ibovespa.

(conclusão)

<b>VARIÁVEIS MACROECONÔMICAS E ÍNDICE S&amp;P 500</b>	<b>FONTE</b>	<b>DESCRIÇÃO</b>
IGP-M	Faculdade Getúlio Vargas (FGV)	Índice Geral de preços-Mercado, que foi concebido nos anos de 1940 e abrange o movimento de preços de diferentes atividades e etapas do processo produtivo.
IPCA	Instituto Brasileiro de Geografia e Estatística (IBGE)	Índice de Preços ao Consumidor Amplo é utilizado para controle da inflação por metas e é representado, de acordo com o IBGE, pelas famílias residentes em áreas urbanas com rendimento familiar entre 1 e 40 salários-mínimos por mês.
Preço do Dólar	<i>Yahoo Finance.</i>	Há uma correlação entre o Dólar e o Ibovespa. A demanda por real pode aumentar devido ao aumento no número de estrangeiros que investem. No Brasil, e com isso há uma desvalorização do dólar. O contrário também pode acontecer.
Petróleo <i>Brent</i>	<i>Yahoo Finance.</i>	Petróleo tido como referência mundial e é extraído no Mar do Norte e comercializado na Bolsa de Londres.
Minério com 62% Ferro	Yahoo Finance.	Preço do minério de ferro com 62%, conhecido como preço de referência é uma das principais commodities do mundo.

Fonte: Resultados da pesquisa, 2023.

A exemplo de Huang *et al.* (2005) que utilizaram as variáveis de taxa de juros, os índices de preços do consumidor, a taxa de juros nacional entre outras variáveis macroeconômicas, os autores incorporaram o índice S&P 500, haja visto que os Estados Unidos da América (EUA) são a principal economia do mundo e influenciam a economia do Brasil, pois é um dos maiores alvos da exportação brasileira

Outras variáveis foram integradas, como o petróleo *Brent* e o minério com 62% de ferro, que são uma das principais commodities de duas empresas destaques na composição do Ibovespa a Vale. S.A e a Petrobrás, que no início de 2022 representavam, aproximadamente, 20,65% carteira do Índice Ibovespa (B3, 2022).

A exemplo de Kara, Boyacioglu, Baykan (2011) que utilizaram a média móvel simples

de 10 dias, as médias móveis exponenciais, o indicador de força relativa e o *momentum*, o presente trabalho utiliza as variáveis de análise técnica para melhorar a previsão da direção do Índice Ibovespa. É comum a utilização de diversos indicadores técnicos no treinamento de Redes Neurais, por exemplo, ou SVM para estimar a direção de índices. Diler (2003) treinou as Redes Neurais com base em diversos indicadores técnicos para fazer previsão do índice ISE 100 (*Index of Istanbul Stock Exchange*- Índice da Bolsa de Istambul) e o mesmo procedimento será feito no presente trabalho com as variáveis técnicas já explanadas.

### 3.2.1 Métricas de avaliação de desempenho da análise técnica na previsibilidade do Ibovespa

Diversas métricas podem ser usufruídas para comparar e validar os métodos ou algoritmos utilizados em ME para fazer previsão no mercado financeiro. O Quadro 3 adiante, apresenta as métricas que vão corroborar para validação e comparação de modelos distintos.

Quadro 3 -Métricas de desempenho para validação e comparação de modelos.

(continua)

MÉTRICAS	DESCRIÇÃO	ESTUDOS FUNDAMENTADOS
Acurácia	Razão entre o número de previsões corretas e o número de todas as previsões.	Barros <i>et al.</i> (2020)
Precisão	É usada principalmente em tarefas de classificação binária. Ele concentra-se nas previsões positivas, sendo dado pela razão entre as previsões positivas corretas e todas as previsões positivas.	Sharma <i>et al.</i> (2018)
Recall/ Sensibilidade	Também é usada em tarefas de classificação binária e se concentra na classe positiva. O indicador de sensibilidade é uma taxa de verdadeiros positivos, sendo a proporção das previsões positivas corretas para todas as observações na classe positiva	Ashofter <i>et al.</i> (2021)
Especificidade	É semelhante à sensibilidade, mas focada na classe negativa. Ele mede a proporção de classe negativa que é corretamente prevista como negativa.	Sharma <i>et al.</i> (2018)

Quadro 3 - Métricas de desempenho para validação e comparação de modelos.

(conclusão)

MÉTRICAS	DESCRIÇÃO	ESTUDOS FUNDAMENTADOS
Score F1	É a média ponderada de precisão e recall. O score F1 é uma medida mais útil do que a precisão para problemas com distribuição de classe desequilibradas porque leva em consideração falsos positivos e falsos negativos.	Rahman <i>et al.</i> (2021)
Curva ROC e área sob a curva (AUC)	A curva ROC fornece uma visão geral do desempenho do modelo em diferentes valores limite. AUC é a área sob a curva ROC entre (0,0) e (1,1) que pode ser calculada usando o cálculo integral. AUC basicamente agrega o desempenho do modelo em todos os valores de limite.	Mokhtari <i>et al.</i> (2021)

Fonte: Resultados da pesquisa, 2023.

Tais métricas são bastante utilizadas na literatura e não podem ser vistas de forma isolada, pois elas se complementam no efeito de previsão. Todas essas medidas são baseadas na Matriz de Confusão, que é uma abordagem bastante popular para avaliar a performance de modelos de classificação. Uma matriz de confusão é simplesmente uma matriz que compara as classes observadas de uma variável categórica com as classes previstas pelo modelo. Para a presente aplicação, em que a variável alvo apresenta duas classes, a matriz de confusão é uma matriz 2x2, como na Figura 3 a seguir, onde as linhas representaram os valores reais e as colunas os valores preditos.

Ao prever a classe alvo (denominada como positivo) corretamente, tem-se um verdadeiro positivo. No entanto, ao prever erroneamente essa classe, tem-se um falso positivo (um erro estatístico do tipo I). Quando se prevê a classe alternativa (denominada como negativo) e a classe alvo é a classe correta, chama-se de falso negativo (um erro do tipo II). Quando se prevê a classe alternativa e a previsão é correta, tem-se um verdadeiro negativo. Com base na matriz de confusão, é possível calcular as métricas mencionadas acima.

Figura 3 -Matriz de Confusão

		Classe Observada	
		Positivo	Negativo
Classe Prevista	Positivo	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Negativo	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Fonte: Resultados da pesquisa, 2023.

A Acurácia é uma métrica de performance geral do modelo. Ela indica as classificações corretas do modelo dentre todas as classificações realizadas ( $Total = VP + FP + VN + FN$ ):

$$acurácia = \frac{VP + VN}{Total} \quad (3)$$

A Precisão é definida como a proporção de predições corretas de uma categoria em relação a todas as previsões feitas dessa categoria. Nesse caso, o número de verdadeiros positivos entre o total de positivos previstos ( $VP + FP$ ):

$$precisão = \frac{VP}{VP + FP} \quad (4)$$

A Sensibilidade (também conhecida como recall) é definida como a proporção de previsões corretas da categoria alvo. Ou seja, uma taxa de verdadeiros positivos.

$$sensibilidade = \frac{VP}{VP + FN} \quad (5)$$

A Especificidade, por sua vez, mede a proporção de previsões negativas realizadas corretamente:

$$especificidade = \frac{VN}{VN + FP} \quad (6)$$

### 3.3 Modelos de Machine Learning

A análise empírica propõe aplicar diferentes modelos de classificação, como os modelos Logit, Lasso, SVM e modelos de Redes Neurais Artificiais (RNA), e fazer comparativos em termos do desempenho preditivo em relação aos movimentos do Índice Ibovespa. Contudo, foram feitos testes iniciais com as variáveis descritas adiante com a modelagem de Lasso para os dados de treino e teste, assim como a verificação da acurácia do modelo.

### 3.3.1 Modelo Logit

A regressão logística é um tipo de regressão formulada para fazer previsão de uma variável categórica binária em dois grupos e não uma medida dependente métrica (Hair, 2009). Lindner (2016) “a Regressão Logística se diferencia da Regressão Linear, pois seus cálculos envolvem variáveis dependentes qualitativas, ao invés de quantitativas” representado por meio de:

$$Y_i = X_i \cdot \beta_i + \varepsilon; i = 1, \dots, k \quad (7)$$

Onde: em que  $Y_i$  variável resposta;  $\beta_i$  é o intercepto do modelo,  $X_i$  variável explicativa e  $\varepsilon$  o erro associado às observações. Na regressão, a probabilidade de ocorrência de um certo evento pode ser estimada e assim sendo, a variável dependente  $Y$  assume apenas dois valores possíveis (0 ou 1) e há um conjunto de  $p$  variáveis independentes  $X$ . Ainda em Hair (2009), a opção pela regressão logística tem vantagens como: Não é necessário supor a normalidade multivariada. É uma técnica genérica e, ao mesmo tempo, robusta. Tem grandes semelhanças com a Regressão Linear múltipla.

A função de Regressão Logit é descrita através de:

$$P(X) = \frac{e^{\beta_0 + \beta_1 \cdot X}}{1 + e^{\beta_0 + \beta_1 \cdot X}} \quad (8)$$

Expressão que pode ser reescrita como:

$$\log \left[ \frac{P(X)}{1 - P(X)} \right] = \beta_0 + \beta_1 \cdot X \quad (9)$$

A expressão entre colchetes é conhecida como razão de chances. Com várias variáveis preditoras, tem-se a equação, a seguir:

$$P(X) = \frac{\exp(\beta_0 + \sum_{j=1}^k \beta_j \cdot X_j)}{1 + \exp(\beta_0 + \sum_{j=1}^k \beta_j \cdot X_j)} \quad (10)$$

O método adota o critério de máxima verossimilhança para estimar os parâmetros. Uma das vantagens desse método é que os estimadores de máxima verossimilhança têm várias propriedades teóricas interessantes que podem ser exploradas para derivar intervalos de confiança para os coeficientes de regressão, realizar testes de hipótese, entre outros testes e estatísticas.

No presente trabalho, tem-se a variável de resposta  $Y$  é categórica, de modo que há um

problema de classificação. O objetivo da construção de um modelo é prever a categoria de  $Y$  com base nos dados  $X$ . De modelo simples, tem-se a formulação:

$$E(Y|X = x) = P(Y = 1|X = x) \quad (11)$$

O que significa que o modelo de regressão linear está tentando prever a probabilidade de ocorrência de  $Y = 1$ .

### 3.3.2 Regressão de Lasso

A estimação de Lasso é uma regressão cujo a resposta do problema é dada da seguinte maneira:

$$\hat{\beta}_L = \underset{\beta}{\arg \min} = \left( \sum_{i=1}^n (Y_i - X_i\beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right), \lambda \geq 0 \quad (12)$$

Onde:  $\beta$  é o parâmetro;  $Y_i$  é a variável de resposta;  $X_i$  é a variável explicativa;  $\lambda$  parâmetro do Lasso.

A regressão de Lasso é um problema de minimização dos erros quadráticos adicionado de uma penalização  $L_1$ . Géron (2019) cita que uma característica dessa regressão é que ela tende a eliminar pesos que são menos importantes na regressão. Ou seja, executa de forma automática a seleção de “recursos no qual gera um modelo esparso”.

Klosterman (2019) chama a atenção de que quando um modelo está sendo ajustado, e está “aprendendo” a relação entre características e variável de resposta, então o modelo pode começar a perceber o ruído em um conjunto de dados, de modo que esse modelo pode provocar o ajuste e generalizar tal feito para novos dados o que é prejudicial na qualidade preditiva, daí a importância da regressão de Lasso que impõe penalidades e tenta “erradicar esses espúrios” da modelagem.

### 3.3.3 Modelo Random Forest

*Random Forest* é um algoritmo de classificação e regressão que utiliza o método de árvores de decisão criado por Breiman (2001) e tem alta popularidade, pois, por ser mais simples, implantar do que impulsiona (Lohrmann, Luukka, 2019). A classificação é obtida por meio da construção de um determinado conjunto de árvores aleatórias de classificação e

regressão (*CART-Classification and Regression Trees*) (Yeh, Chi, Lin, 2014). As principais vantagens do CART incluem simplicidade do modelo, o que facilita a interpretabilidade; o resultado do modelo pode ser facilmente explicado pelas regras de decisão e o modelo pode ser visualizado de forma intuitiva. Tudo isso exige uma preparação mínima de dados. Enquanto alguns algoritmos podem exigir a normalização dos dados, a criação de variáveis *dummies*, a contabilização de valores ausentes (em alguns casos), dentre outros procedimentos, essas etapas não são necessárias para as árvores de decisão.

As principais desvantagens incluem a instabilidade devido a pequenas variações no conjunto de dados e dados ligeiramente diferentes, que podem levar à geração de árvores completamente diferentes.

Sendo assim, as árvores de decisão tendem ao *overfitting* em situações em que o conjunto de dados possui um número muito grande de preditores (recursos). Truncar o número de recursos, aplicando um algoritmo de redução de dimensionalidade pode melhorar o desempenho.

De acordo com Jiang *et al.*, (2020) os subconjuntos de dados originais são gerados por um método chamado *bootstrap* (reamostragem) e impureza de Gino no qual a divisão ideal é feita sob critérios específicos e a árvore é gerada através da pesquisa de subconjuntos de recursos de candidatos.

### 3.3.4 Modelo de Support Vector Machine (SVM)

Strader *et al.* (2020) utilizaram *Support Vector Machine* (SVM) para fazer previsões do mercado de ações. Os autores relatam que uma máquina de vetores de suporte faz representações com exemplos como pontos em um espaço, visando criar um gap entre as categorias o mais amplo possível. SVM é uma família de algoritmos implementados em regressão e classificação de estudos em séries temporais (Kara, Boyacioglu, Baykan, 2011).

O algoritmo de máquinas de vetores de suporte realiza mapeamento de vetores de dimensão  $n$ ,  $x \in \mathbb{R}^{n^1}$  de uma dimensão alta que se caracteriza então como o espaço “H” e se constrói um hiperplano de separação neste espaço, no qual dados de treinamento são separáveis, o hiperplano que maximiza a distância no espaço H entre o hiperplano e a imagem mais próxima  $\Phi(x_i)$  do veto  $x_i$ , sendo esse o hiperplano “ideal” (Chapelle *et al.* 2002).

Ainda consoante os autores, a suposição da distância máxima seja igual a  $Y$  em que as

imagens  $\Phi(x_1), \dots, \Phi(x_l)$  e os vetores de treinamento  $x_1 \dots x_l$  esteja, dentro da esfera de raio  $R$  vale o Teorema 1.

Teorema 1: Dado um conjunto de treinamento  $Z = \{(x_1, y_1), \dots, (x_l, y_l)\}$  e um hiperplano  $(w, b)$  a margem  $Y(w, b, Z)$  e o raio  $R(Z)$  definidos em (8) a margem e (9) o raio.

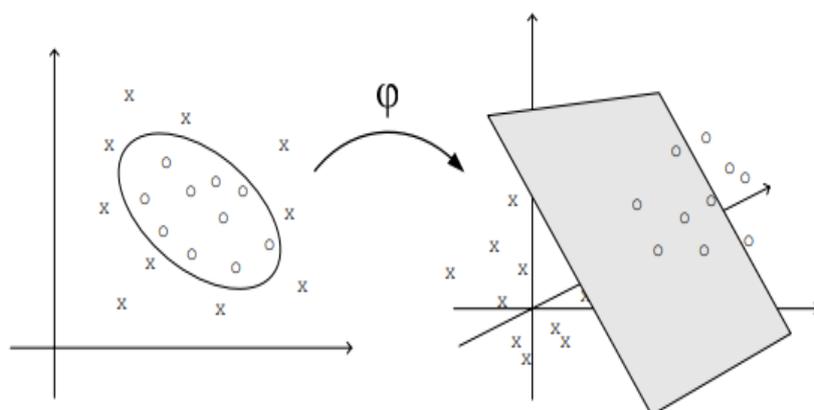
$$Y(w, b, Z) = \min_{x_i, y_i \in Z} \frac{y_i(w \cdot \Phi(x_i) + b)}{\|w\|} \quad (13)$$

$$R(Z) = \min_{a, x_i} \|\Phi(x_i) + a\| \quad (14)$$

O algoritmo de margem máxima  $L_l: (\mathcal{X} \times \mathcal{Y})^l \rightarrow \mathbb{H} \times \mathbb{R}$  toma como entrada a completude de tamanho de treinamento e retorna a um hiperplano no espaço de tal forma que os recursos a margem  $Y(w, b, Z)$  é maximizada. Ou seja, quanto maior a margem, melhor será o desempenho do hiperplano (Chapelle *et al.* 2002). Karatzoglou, Meyer, Hornik (2006) evidenciam que no SVM, para classificar, fazer uso de regressão ou detectar algo, envolve resolver problemas quadráticos de otimização. Na literatura, é comum a expressão “truque do kernel” que usa transformações variáveis por kernels.

O “truque de Kernel” possibilita que um determinado espaço original seja mapeado em um espaço de produto escalar com característica de alta dimensão no qual os dados podem, então ser, linearmente, separáveis (Bonesso, 2016). Na Figura 4, tem-se uma representação de um mapeamento de espaço bidimensional para um espaço tridimensional.

Figura 4 - Mapeamento de um espaço bidimensional para espaço tridimensional de dados.



Fonte: Adaptado de (PRESS *et al.* 2007).

No espaço de dimensão mais alta, é possível separar melhor os dados com linearização da superfície (Bonesso, 2016). Na prática computacional, são utilizados espaços de características de dimensões maiores.

### 3.3.5 Modelo de Redes Neurais Artificiais

De Pauli *et al.* (2020) cita que as observações das variáveis de resposta denotadas por  $y_i = f_o(B_j, x_{ij})$  e as variáveis independentes;  $i = 1, \dots, n, x_{ij}$ , em que  $j = 1, \dots, p$ , de tal modo que  $p \times 1$  é um vetor de variáveis explicativas de entrada, da mesma forma,  $\beta$  é um vetor de coeficientes ( $p \times 1$ ).

Aqui, seguem os passos de Chandra *et al.* (2016) no algoritmo de coevolução cooperativa que “determina como os pesos são codificados nos subcomponentes implementados como subpopulações”. Passo 1: Decompor o problema em *Neuron Level* (Nível de Neurônio-NL) ou *Synapse Level* (nível de sinapse -SL); Passo 2: Inicializa e avaliar de modo cooperativo cada subpopulação em cada ciclo até o fim para cada subpopulação.

No Algoritmo, a Rede Neural é decomposta usando o método NL ou SL em  $k$  subcomponentes e, assim, na fase de inicialização, todas as subpopulações são preenchidas com valores aleatórios e de tal modo que são também avaliadas. O objetivo desse processo evolutivo é aumentar a aptidão que tende a diminuir o erro da Rede Neural, no qual é definido como a raiz do erro quadrático médio de modo específico para séries temporais (Chandra 2016). A avaliação cooperativa é feita de um indivíduo  $x$  em uma subpopulação  $y$  (em particular) com a concatenação de indivíduos mais aptos do restante da população. (Potter, De Jong 1994).

## 4 RESULTADOS E DISCUSSÕES

Nesta seção, são apresentados a estatística descritiva do Índice Ibovespa, a série temporal compreendida do período de análise e a performance dos modelos de previsão de dados com as métricas de desempenho apresentadas na seção anterior.

### 4.1 Estatísticas Descritivas

Foi feito a Estatística Descritiva do Índice Ibovespa na análise temporal já descrita na Tabela 1.

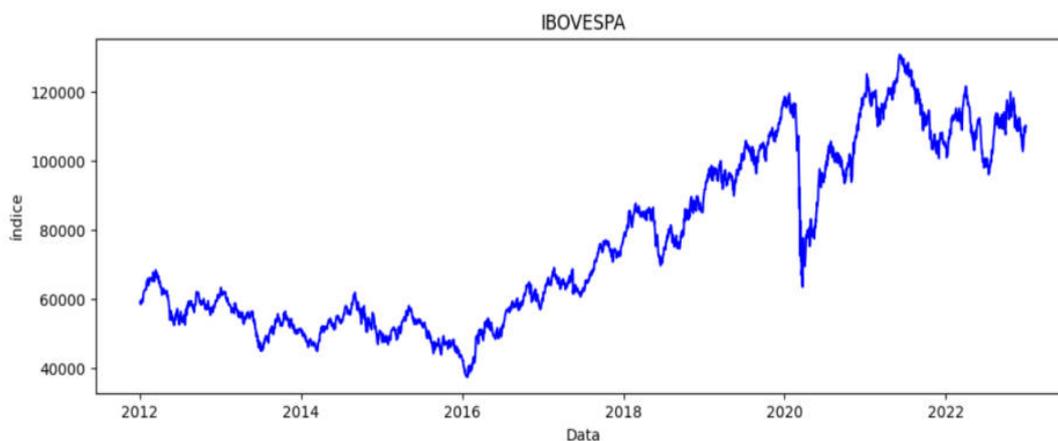
Tabela 1-Estatística descritiva do preço do Ibovespa entre 2012 e 2022.

índice	Abertura	Alta	Baixa	Fechamento	Volume	Retorno
<b>média</b>	76656,25	77401,31	75919,27	76673,56	5.682.767.000.000	0,00023
<b>desvio padrão</b>	25195,41	25385,57	24990,32	25201,54	3.740.803.000.000	0,01577
<b>mínimo</b>	37501,0	38031,0	37046,0	37497,0	0	-0,15993
<b>25%</b>	54442,0	54914,0	53963,0	54448,0	3.195.800,0	-0,00820
<b>50%</b>	66504,0	66968,0	65860,0	66662,0	4.103.000,0	0,000296
<b>75%</b>	101209,0	102218,0	100098,0	101216,0	7.488.400,0	0,00900
<b>máximo</b>	130776,0	131190,0	129526,0	130776,0	26.029.300.000.000	0,13022

Fonte: Resultados da pesquisa, 2023.

Com um período relativamente grande há grandes disparidades entre valores máximos e mínimos, por exemplo. O mínimo da série 37.501 aconteceu no ano de 2016 e o valor máximo 120.776 no ano de 2021. A média e o desvio padrão na Tabela 1 são relativamente altos como já mencionado ao longo período de análise e valores de 25%, 50% e 75% foram integrados na complementação da descrição dos dados. Os Gráficos 2 e 3 adiante, complementam a Estatística Descritiva com a variação do índice de 2012 e 2022.

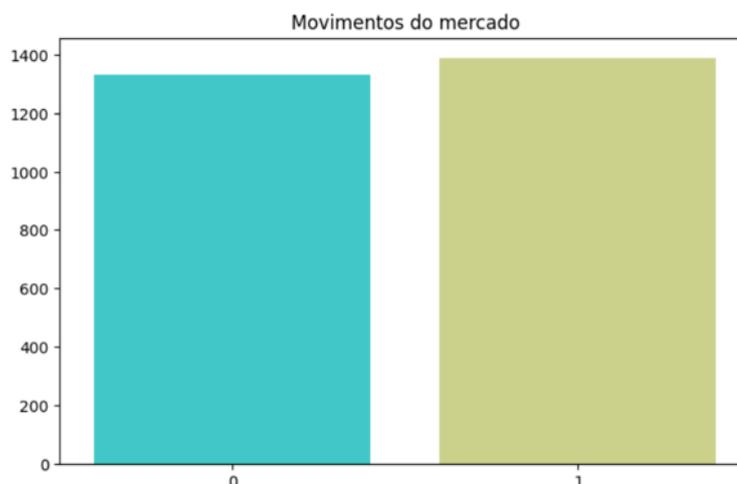
Gráfico 2- Série do Índice Ibovespa de 2012 e 2022.



Fonte: Resultados da pesquisa, 2023.

Visualiza-se que o índice sofre grandes quedas nos anos de 2015, 2016 e no ano de 2020. Ainda em 2020, o índice tem elevação e continua nos anos seguintes. Sem dúvida, fazer análises desse índice é muito importante. O resultado de práticas comerciais, os corretores de ações, os serviços atrelados a investimentos, estão conectados tanto aos investidores quanto aos especuladores em ações ordinárias que dedicam atenção às previsões do mercado (Graham, 2015). O Gráfico 3 representa os movimentos do Mercado do período analisado.

Gráfico 3-Movimentos do mercado de altas e baixas do Ibovespa entre 2012 e 2022.

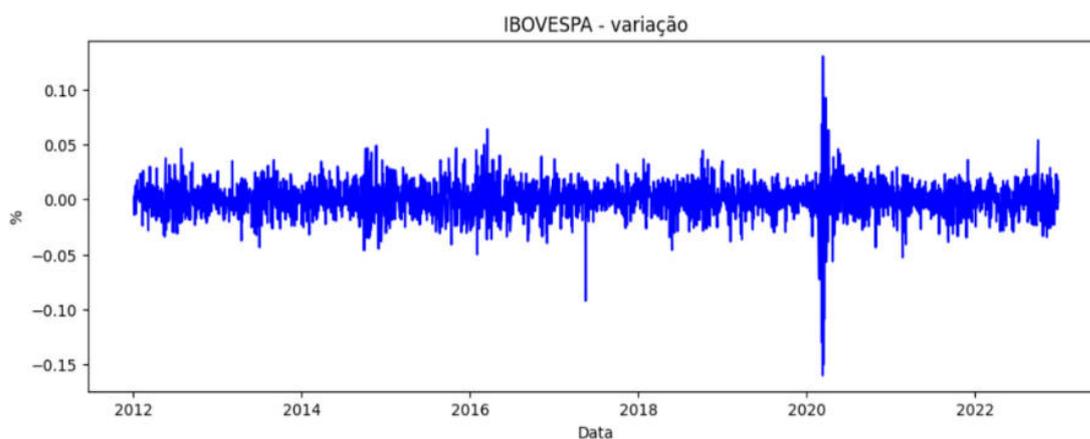


Fonte: Resultados da pesquisa, 2023.

No período analisado, houve 1397 altas e 1346 baixas referentes ao dia anterior, e apenas um único dia que não houve variação do índice, porém esse dia foi contabilizado como baixa, pois não houve variação já descrita na metodologia. Percebe-se que não houve discrepância entre altas e baixas e nos anos de estudo.

Em resultados de estudos de mercado chinês e americano de ações, os simples movimentos intradiários de subida e descida em todas as negociações anteriores podem afetar o sentimento de investidores. Isso resulta em movimentos de mercado na mesma direção durante o dia de negociação (Zhang *et al.* 2016). Apesar de não haver uma grande discrepância, como já dito, no quantitativo de altas e baixas, a oscilação da variação durante os meses é grande. Assim como apresentado na equação (1), segue o retorno de forma gráfica do Índice Ibovespa no Gráfico 4.

Gráfico 4 -Retorno do Ibovespa entre 2012 e 2022.



Fonte: Resultados da pesquisa, 2023.

Percebem-se grandes variações dos retornos para os últimos anos da série devido a vários momentos diferenciados. A variação no mês de março de 2020 foi de -29,90% e teve variação positiva máxima de 15,90% no mês de novembro.

## 4.2 Série temporal: análise do Ibovespa entre 2012 e 2022

Séries temporal é uma metodologia com aplicação em estudo e análise de dados com objetivo de modelar padrões e autocorrelação e, dessa forma, possibilita fazer previsões (Cintra, Melo, Bueno Filho, 2019).

Sumamente, Keiel e Bender (2018) a análise de séries temporais financeiras avalia preços e/ou retornos de ativos ao longo do tempo, proporcionando o entendimento do dinamismo e construção de modelos para previsão de valores futuros. Para continuação de testes econométricos e estatísticos, feitos os testes Dickey-Fuller e Teste de Estacionariedade Kwiatkowski-Phillips Schmidt-Shin (KPSS) no presente trabalho com a série de retornos criada no *python* do Ibovespa. De acordo com Mushtaq (2011), uma tarefa importante em econometria é visualizar de forma apropriada a tendência em uma série, pois muitas séries financeiras apresentam comportamento de tendência ou não estacionariedade na média. A Tabela 2 exibe o teste de estacionariedade Dickey-Fuller e (KPSS).

Tabela 2-Teste de Estacionariedade Dickey-Fuller e Teste de Estacionariedade Kwiatkowski-Phillips Schmidt-Shin (KPSS) na série do retorno do Ibovespa entre 2012 e 2022.

	<b>ADF</b>	<b>KPSS</b>
<b>Estatística de teste</b>	-18,84	0,072
<b>Valor-P</b>	0.0000	0,01
<b>Lags Usados</b>	6	7
<b>Número de observações usadas</b>	2733	2733
<b>Valores Críticos (1%)</b>	-3,43	0,739
<b>Valores Críticos (5%)</b>	-2,86	0,463
<b>Valores Críticos (10%)</b>	-2,57	0,347

Fonte: Resultados da pesquisa, 2023.

Para visualizar se a série é estacionária ou não, verifica-se o valor crítico do teste Dickey Fuller para avaliação, que tem como hipótese nula ( $H_0$ ). A série possui raiz unitária, e não é estacionária. Logo, rejeita-se a hipótese nula aos níveis usuais de significância: 1%, 5% e 10%. Silva (2019) elucida que uma série temporal tem características como: média, variância e autocorrelação constantes ao longo do tempo e assim a série é estacionária. Complementando o primeiro teste, tem-se o Kwiatkowski-Phillips Schmidt-Shin (KPSS) para verificar a estacionariedade da série.

Marques (2013) elucida que o KPSS é importante na variação na argumentação das hipóteses de testes cuja hipótese nula (a série é estacionária) é que o  $\rho$  valor seja menor que 1

e hipótese alternativa (a série não é estacionária) que  $\rho = 1$ . Verifica-se, então, conclusivamente, que a série é estacionária para os critérios de defasagem 6 no teste Dickey Fuller e 7 no critério KPSS. Vários testes dependem da estacionariedade da série que foi verificada com o teste citado.

Modelos de previsão como *Naive Bayes* ou *Random Forest* (RF) podem atingir precisão acima de 95% em um conjunto de dados estacionário, e a única maneira de obter resultados bem-sucedidos é que as séries tenham estacionariedade, ou ainda que modelos sejam projetados e/ou evoluam, especificamente, para essas situações de estacionariedade (Dixit, Jain, 2021). Sendo assim, constatada a série estacionária nos retornos do Índice Ibovespa serão feitos os testes de previsão adiante.

### 4.3 Performance dos modelos de previsão de dados

Com o conjunto de dados das variáveis descritas no tópico anterior, os dados são integrados em um único *dataframe* no qual são processados para treinamento e o restante é dividido para dados em testes e validação do modelo. É comum que pesquisadores e/ou acadêmicos utilizem 80% de dados para treino e 20% para testes, ou ainda, 70% de treino e 30% para testes.

Nti *et al.*, (2020); Su *et al.*, (2022); Singh (2022); Bhandari *et al.*, (2022); Demirel *et al.*, (2020); Raso *et al.*, (2019); Wanjawa *et al.* (2014) com estudos predição de dados no mercado de ações aplicaram em suas bases de dados 80% para dados de treino e 20% para dados de testes. Aqui, adotou-se o mais comum na literatura, 80% para treinamento e 20% para testes.

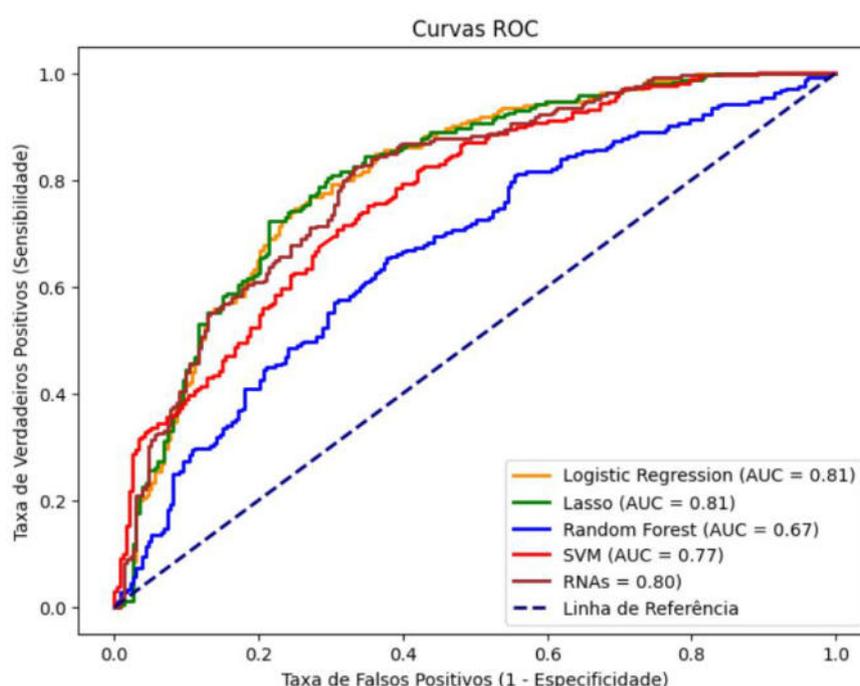
Em seguida, o modelo avalia com as métricas de desempenho da previsão em relação aos dados de validação. Tal validação é importante para previsão do alvo “imprevisto” de dados de teste e comparação com o alvo verdadeiro (Mokhtari *et al.* 2021). Segue na Tabela 3, a performance de desempenho dos diferentes algoritmos empregados na pesquisa, seguido pela Gráfico 5, Curva ROC dos modelos preditivos para o Índice Ibovespa na análise de 2012 e 2022.

Tabela 3-Performance de modelos para análise de desempenho.

Modelos	Acurácia	Precisão	Sensibilidade	Especificidade	Score F1	ROC
Modelo Logit	74%	77%	71%	77%	74%	81%
Modelo Lasso	75%	77%	72%	77%	75%	81%
Random Florest	63%	65%	60%	66%	62%	67%
Support Vector Machine	69%	70%	71%	68%	70%	77%
Redes Neurais Artificiais	74%	75%	73%	74%	74%	80%

Fonte: Resultados da pesquisa, 2023.

Gráfico 5-Curva ROC dos modelos preditivos para o Índice Ibovespa na análise de 2012 e 2022.



Fonte: Resultados da pesquisa, 2023.

Pode-se, então, concluir uma superioridade preditiva do modelo Lasso e RNAs, seguido do modelo Logit. Os resultados adequam-se ao trabalho de Kara *et al.* (2011), pois utilizaram indicadores técnicos para as previsões do índice turco, que considera um intervalo de 10 anos para os seus estudos no qual se aproxima bastante deste trabalho no âmbito de tal quesito. Para efeitos comparativos, ambos os trabalhos obtiveram como desempenho preditivo com RNAs 75% e 74% e para o modelo SVM os autores turcos obtiveram 71,52% enquanto o presente trabalho obteve 69%.

Com análises do trabalho de Kraus, Feuerriegel (2017) no mercado de ações na

Alemanha com os modelos Lasso, *Random forest*, SVM e obtiveram acurácia máxima de 55,20% com melhoras superiores ao utilizar *Deep learning* (aprendizagem profunda) ainda assim acurácia máxima de 60,1% em que a presente pesquisa obteve com mínimo de acurácia 63% o que indica bons resultados para as técnicas de previsão apresentadas.

Em comparativas com Su *et al.*, (2022), o presente trabalho teve melhor acurácia, com superioridade em 15,57%, e melhor predição com o modelo SVM em 13,86%. Por isso, os resultados contradizem aqueles que foram encontrados por Meher *et al.* (2024), que apontam que modelos logísticos têm previsão entre 55-60%, enquanto o ensacamento de árvores aleatórias tem predição de 85-90%. Na presente pesquisa, foi obtido para modelo Logit acurácia superior e para modelagem *Random forest* acurácia inferior descritos pelos autores.

Diler (2003) projetou um modelo de RNA para o índice da Bolsa de Valores de Istambul e obteve 60,81% como resultado para estimação do índice para o dia seguinte, enquanto a presente pesquisa resultou, como já dito, 74%. Huang *et al.* (2005) obtiveram a performance de 73% no modelo SVM muito próximo do resultado deste trabalho, que teve 69% de acurácia.

Para Braga *et al.* (2022), o classificador de SVM foi importante para previsão de dados históricos cinco ações (Petrobrás, Itaú, Bradesco, Vale e Ambev) no Brasil e ainda segundo os autores a regressão linear pode ser utilizada para auxiliar na tomada de decisão ou ainda para se ter um “alvo” de compra ou venda de ações com os dados obtidos no estudo dos autores. Corroborando resultados de Strader *et al.* (2020), Redes Neurais são melhores para prever os números e valores de índices do mercado de ações e SVM pode se adaptar melhor a problemas de classificação para a direção do índice.

Cao *et al.* (2005) elucidam que as Redes Neurais superam modelos lineares e que RNAs é uma ferramenta útil para previsão de ações no mercado emergentes da China. Nabipour *et al.* (2020) estudaram a previsão de quatro grupos de bolsas de valores diversificados e empregaram nove modelos de aprendizado de máquinas com superioridade de Redes Neurais e modelos *long short-term memory* (LSTM- Memória de Longo Prazo). Ainda em Nabipour *et al.* (2020) utilizaram dados binário e esse artifício também é utilizado no presente trabalho para prever a direção do Índice Ibovespa em alta e baixas diárias e mensais

Destaca-se o modelo de Lasso que teve a melhor acurácia e a curva ROC (*Receiving Operator Characteristics*). Geralmente, tal modelo não é aproveitado em outros periódicos da temática, com *Machine Learning*, sendo este um dos diferenciais da pesquisa, e que teve bom desempenho para efeitos preditivos. Com relação ao Gráfico 5, Oliveira Ritta *et al.* (2015)

ênfatizam que quanto mais distante a curva ROC, se distância da diagonal principal ( $x = y$ ) então, melhor será a classificação do modelo. Martinez (2003) ênfatiza que a curva ROC possibilita fazer análises empíricas eficazes de modelos diferenciando classes em amostragem. Destacam-se com mesma porcentagem a regressão de Lasso e o modelo Logit com 81% e a modelagem RNAs com 80%.

Para efeitos comparativos do modelo Logit com Mokhtari *et al.*, (2018), o atual trabalho teve melhores desempenhos em todas as performances, com destaque para a curva ROC que teve desempenho de 81% enquanto os autores citados obtiveram, apenas 73%.

Com relação a modelagem RF, Lohrman e Luukka (2019) utilizaram tal modelo para a predição de dados do S&P 500 com acurácia entre 44% e 47,78% (com ajustes de parâmetros), também utilizando variáveis técnicas. O presente trabalho teve resultados melhores de acurácia do que os autores, mas os mesmos, ênfatizam que com outras variáveis técnicas o modelo de *Random Florest* pode ter desempenhos superiores.

Como resultado, pode-se afirmar fortificados por Mokhtari *et al.* (2018) que modelos híbridos de análise técnica e análise fundamental podem compensar desvantagens de algoritmos individuais e tais análises podem ser complementares, assim como as métricas de desempenho que contribuíram para este trabalho.

## 5 CONSIDERAÇÕES FINAIS

O aumento de investidores em ações é crescente no Brasil. No ano de 2020, somente com pessoas físicas, houve um aumento de mais de 2,6 milhões de investidores em relação ao ano anterior. O mercado de ações fornece um meio para proteção dos investidores de riscos do mercado financeiro e cria oportunidades ímpares para a especulação e arbitragem, com a expectativa também de ganhos em índices futuros.

Desse modo, o presente estudo tentou nortear investidores, analistas e pesquisadores que estudam o mercado financeiro, em particular, o mercado de ações do Brasil. Para isso, duas categorias principais são constantemente citadas na literatura, que são as análises técnica e fundamentalista, as quais foram utilizadas neste ensaio. Algoritmos de *Machine Learning* foram operacionalizados e, como resultado, com base nas métricas de desempenho elencadas, anteriormente, modelos de Lasso e Redes Neurais foram os algoritmos de melhor desempenho para previsibilidade do Índice Ibovespa. Tal resultado diferencia-se dos trabalhos da literatura, pois a pesquisa utiliza o modelo de Lasso, no qual teve boa acurácia e vai de encontro com dezenas de trabalhos da literatura internacional com bom desempenho preditivo da modelagem de RNAs que teve boa performance de desempenho.

No tocante à predição do Índice Ibovespa, diversos fatores podem influenciar a queda ou aumento. Das várias variáveis dispostas no trabalho, o IGP-M, o IPCA, Indicador de Incerteza da Economia, Taxa Selic e o índice S&P 500 foram bastante significativos para a pesquisa. O preço do petróleo *Brent* e o minério de ferro refinado 62% cotado na Bolsa de Valores contribuíram também para prever o Índice Ibovespa, pois tais commodities são os dois principais ativos das empresas: Petrobrás e a Vale S.A. A composição das duas empresas ao índice equivale a 20,65%, no início de 2022, e a 22,44%, no fechamento do pregão seguinte, vigorando de maio a setembro.

Incorporar as variáveis técnicas, as macroeconômicas, o índice S&P 500, os preços de *comodities*, e as variáveis técnicas já descritas para fazer melhores previsões da direção do Ibovespa podem aumentar a precisão, no qual implica uma questão importante para futuros estudos.

Conclui-se que os métodos de *Machine Learning* são bons classificadores e atingiram um bom desempenho de acurácia. Fatores como a escolha do horizonte de tempo, os indicadores para estudo de predição e a utilização de mais de um modelo para fins

comparativos podem ser determinantes para o estudo de prenúncios de índice de ações, que no presente trabalho foi o Ibovespa.

Fazer o estudo de direção do Ibovespa é importante para orientar os investidores no mercado de ações e oportunizar ganhos com maior potencial com contratos futuros da bolsa de valores como, por exemplo, o Ibovespa Futuro (INDFUT) um derivativo da B3 negociado por meio de especulação e desempenho, vinculados ao Índice Ibovespa, e, portanto, exige muito conhecimento e alto risco.

## UM ESTUDO SOBRE ANOMALIAS NO MERCADO BRASILEIRO DE AÇÕES

### 1 INTRODUÇÃO

São crescentes as transformações digitais interligadas a serviços e aos negócios nas vidas das pessoas. O fato ficou bastante evidente durante a pandemia da COVID-19, iniciada no ano de 2020, que transformou inúmeros serviços e deu oportunidade para o desenvolvimento de diversas estratégias de empresas em todo o mundo (Shaik *et al*, 2021).

Big data, blockchain, computação em nuvem, internet das coisas, criptomoedas, inteligência artificial são algumas das tecnologias emergentes que passaram a se desenvolver de forma mais acelerada e transformam o cotidiano de grande parte das pessoas (Hanafizadeh; Kim, 2020). Nessa perspectiva, a velocidade em que os dados são gerados e a forma como são processados também modificou o mercado financeiro. A informação sempre foi essencial para a tomada de decisão de investidores e para o desenvolvimento econômico, procurando prevenir e detectar anomalias financeiras (Ahmed *et al*, 2017; Bollaert *et al*, 2021).

No mercado de ações, a ideia de "mudanças de preços aleatórias e imprevisíveis" foi inicialmente introduzida por Bachelier (1900) e posteriormente aprimorada por Fama (1965), desenvolvendo-se para o conceito de eficiência de mercado em seu trabalho de 1970. Este último apresentou a Hipótese do Mercado Eficiente (HME), na qual toda a informação relevante e disponível tem reflexo nos preços das ações, tornando impraticável obter retornos inconsistentes. Em outras palavras, as mudanças de preços são consideradas imprevisíveis, sendo diretamente afetadas por notícias que chegam de maneira aparentemente aleatória (Yavrumyan, 2015).

Algumas teorias contemporâneas apontam para a existência de anomalias que desviam dessa expectativa postulada pela HME. Avramov *et al*. (2013) apresentam teorias de precificação que indicam uma relação entre o nível de risco e os retornos no mercado de ações. Segundo essas teorias, ativos mais arriscados deveriam, logicamente, gerar retornos mais elevados. Essa perspectiva, porém, colide com a visão da HME, que sustenta a eficiência do mercado em incorporar rapidamente todas as informações disponíveis.

As anomalias, como são conhecidas, desencadeiam questionamentos fundamentais sobre a eficiência plena do mercado. Os seus ativos mais arriscados podem, de fato, gerar

retornos superiores. Por isso, surge a necessidade de reavaliar a eficácia do mecanismo de precificação e incorporação de informações nos preços das ações. Dessa forma, o estudo das anomalias não apenas desafia teoricamente a HME, mas também enriquece o entendimento do mercado financeiro, destacando a complexidade inerente e a possibilidade de que, em certos cenários, as premissas tradicionais possam não capturar totalmente a dinâmica do comportamento do mercado.

As anomalias de mercado podem se manifestar de diversas formas e desafiam a noção de eficiência de mercado ao apresentarem comportamentos sistemáticos que não podem ser explicados exclusivamente por informações disponíveis. Uma das mais notáveis são as chamadas anomalias de calendário que denotam padrões irregulares associados a determinados períodos do tempo. Um exemplo clássico de anomalia de calendário é o "efeito janeiro", que sugere que os retornos médios de ações em janeiro tendem a ser superiores em comparação com outros meses. Outras anomalias, como o "efeito segunda-feira", destacam os padrões específicos de comportamento de preços em determinados dias da semana.

Se padrões sazonais podem ser identificados e explorados, isso levanta a questão de até que ponto os preços das ações incorporam efetivamente todas as informações relevantes. A investigação dessas anomalias não apenas desafia a teoria tradicional, mas também acrescenta camadas de complexidade ao entendimento do mercado financeiro, reconhecendo que fatores temporais podem influenciar substancialmente o comportamento dos investidores e a dinâmica de preços.

A importância desse tema foi enfatizada por Ahmed *et al.* (2017). O autor destaca que o desconhecimento do efeito das anomalias pode conduzir a decisões equivocadas e resultar em conclusões desastrosas, seja por estudiosos do tema ou analistas. Portanto, pesquisas nesse sentido são essenciais para fornecer evidências e contribuir para a formação de conhecimento no campo.

Analisar anomalias específicas no mercado financeiro brasileiro envolve a consideração de fatores que podem ser distintos de outros mercados e que estão presentes na literatura internacional. Algumas anomalias podem estar relacionadas a características únicas da economia, regulação e comportamento dos investidores. Também é importante destacar que a observação de anomalias pode variar ao longo do tempo e estar sujeita a mudanças nas condições econômicas.

Nesse sentido, o presente estudo dedica-se a fazer uma revisão da literatura a respeito de anomalias no mercado financeiro, com maior direcionamento ao mercado brasileiro. De forma

adicional, com o propósito de contribuir com a geração de evidências sobre o tema, em um segundo momento, o estudo procura aplicar modelos econométricos com o objetivo de realizar inferências a respeito da significância estatística de variáveis associadas às anomalias de calendário ou sazonais e o comportamento do Ibovespa. De forma específica, isso foi feito ao testar hipóteses de associações significativas entre variáveis *dummies* para dias de semana e meses do ano e variáveis, que caracterizam os movimentos do mercado e os retornos observados no índice.

Os resultados obtidos mostram evidências de que variáveis de calendário não apresentam efeitos significantes sobre os movimentos diários do índice. Por sua vez, diferentes versões dos modelos estimados para avaliar a correlação destas variáveis sobre os retornos mostraram resultados diferenciados. Modelos mais simples, sem a presença de variáveis de controle relacionadas ao ambiente econômico mostraram a ausência de correlações significativas. Por sua vez, modelos mais completos, que incorporam controles para variáveis macroeconômicas e financeiras, indicaram coeficientes significantes para alguns meses do ano e dias da semana.

O trabalho contém quatro partes, além desta introdução. A primeira é a revisão de literatura em que são referenciados autores que estudaram anomalias no mercado financeiro, os principais tipos de anomalias estudadas no contexto mundial, bem como as evidências sobre anomalias no mercado de ações no Brasil. A segunda parte descreve a metodologia aplicada que inclui o modelo Logit e os modelos de séries temporais ARIMA e ARIMAX. Na terceira parte, são apresentadas as análises com base nos resultados obtidos. Por último, as considerações finais do estudo.

## 2 REVISÃO DE LITERATURA

O objetivo desta seção é fazer uma descrição do que pode ser classificado como anomalia no mercado financeiro, com ênfase no mercado de ações. Além disso, há o registro de alguns tipos de anomalias mundialmente conhecidas e citadas na literatura, bem como evidências sobre as anomalias financeiras no Brasil.

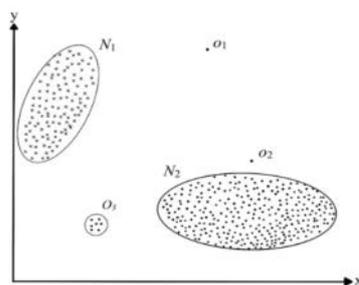
### 2.1 Anomalias no mercado financeiro de ações

Conforme o dicionário Oxford, anomalia significa algo ou alguma coisa, situação que é diferente, que é anormal ou inesperado. Esse termo está presente em diversas áreas como Genética, Biologia, Anatomia, Meteorologia, Geografia entre outros. Frankfurter *et al.* (2001) também cita uma edição especial do *Journal of Financial Economics*, de 1978, como sendo a primeira publicação a associar o termo “anomalias” para fazer referência a contradições à HME.

Em Frankfurter *et al.* (2001), o termo anomalia financeiro faz referência à irregularidade, um desvio na ordem comum, natural e/ou fora do padrão ou regular. Com essa definição, surgem outras perguntas: o que é regular? O que é algo natural? Ou fora do padrão? Perguntas como estas são feitas há algum tempo no estudo de mercado financeiro. Em sentido semelhante, Aggarwal (2013) define que anomalias podem transmitir informações e/ou características “anormais de um sistema” que têm um impacto amplo.

A Figura 5, a seguir, busca apresentar de forma didática, por meio de um diagrama bidimensional simples que faz referência a uma dispersão de pontos de dados, elementos que constituem regiões “normais”  $N_1$  e  $N_2$ , em que estão concentradas a grande parte dos pontos de dados e podem ser observados elementos mais distantes, que estão fora das regiões rotuladas como normais. Seja de forma individual ou em pequenos agrupamentos apresentadas em  $O_1, O_2$  e  $O_3$ , tais pontos podem ser classificados como anomalias (Hilal *et al.*, 2022).

Figura 5—Anomalia em forma bidimensional simples.



Fonte: Hilal *et al.* 2022.

Para Kuhn (1977), nem todas as anomalias precipitam-se em crises financeiras, porém as anomalias exigem que alguma ação seja tomada no sentido de resolvê-la. Por sua vez, Jensen (1978) expõe que, para que uma anomalia seja significativa estatisticamente, ela deve gerar retornos excessivos.

De acordo com Saffer *et al.* (2014) a Hipótese do Mercado Eficiente (HME) elucida que investidores que estão bem-informados, sob todas as informações disponíveis, agem de modo racional e em “conformidade”. Portanto, nenhum investidor pode vencer o mercado gerando retornos anormais. Contudo, verifica-se em muitas bolsas de valores do mundo que estes mercados não seguem a HME. Ou seja, há uma divergência entre a hipótese e o que acontece na realidade.

Há inúmeras evidências de anomalias financeiras no mercado de ações. Farmer *et al.* (2004) analisam as diferenças de flutuações na Bolsa de Valores em Londres. Caporale *et al.* (2016) confirmaram em seus estudos as evidências de anomalias na diferença de preços no mercado da Bolsa de Valores da França e no mercado norte-americano. Plastun *et al.* (2019) manifestam que existem sazonalidades de anomalias do *gap* de preços e efeito momentum no mercado de ações.

Su *et al.* (2022) enfatizam que, no mercado de ações chinês, é relevante compreender as leis das mudanças desse mercado. Portanto, fazer previsão com anomalias da diferença de preços é algo significativo. David *et al.* (2020) trazem hipóteses diferenciadas de ativos e a sensibilidade ao humor de investidores que podem explicar a sazonalidade transversal dos retornos das ações.

Sharma, Kumar e Vaish (2022) concluíram que as anomalias de mercado indiano surgem pela ineficiência do mercado de ações e investigações sobre causas, e aponta de tempos em tempos para o comportamento de investidores. Cengiz *et al.* (2017) investigaram o

Índice Borsa Istanbul (BITS) e verificaram que o mercado de ações na Turquia apresenta anomalias de mercado, além de que o BIST não faz parte de um mercado eficiente. Ainda nesse tocante surge as finanças comportamentais que se baseia na premissa de que os investidores, ou uma parcela pequena, porém significativa, estão sujeitos a “preconceitos comportamentais” e as suas decisões sobre finanças estão menos atreladas a racionalidades. Ou seja, não são totalmente racionais (Woo *et al.* 2020). Evidentemente, estudos sobre as finanças comportamentais crescem ao longo dos anos, envolvendo a área de psicologia e mercado financeiro.

Diversos autores procuram estudar as causas e consequências de anomalias financeiras. Zhang *et al.* (2008) contemplam que, em mercados emergentes, as anomalias podem ser causadas por diferentes razões, e nas análises dos autores, considerações políticas do governo podem intensificar anomalias mensais no mercado financeiro. No contexto chinês, as anormalidades dos retornos são explicadas por manobras políticas com objetivo de criar uma aparência de um mercado de ações “estável e próspero” no qual conduz para prevenção de ressentimento social em momentos mais delicados. Quando há evidências sobre anomalias, os estudiosos confirmam inicialmente a sua existência e, só então, procuram fazer estimativas, avaliações ou previsões para explicar suas origens, utilizando, para isso, análise quantitativa, a modelagem ou ainda constroem novas teorias (Woo *et al.* 2020).

As anomalias de preços interessam a acadêmicos, aos investidores e aos gestores políticos que podem fazer decisões para prevenção de risco no contexto de mercados emergentes (Zada, *et al.* 2021). Vidal-Garcia *et al.*, (2023) estudaram a evolução dos mercados financeiros nas últimas décadas e constataram a “elevada instabilidade e volatilidade” e que, no estudo da Bolsa de Londres de 2000 a 2002, houve detecção de comportamentos de estruturas diferenciadas, o que permitir concluir que fazer estudos da evolução de ações e antecipar padrões futuros pode ser importante no desenvolvimento de estratégias eficientes.

Plastun *et al.* (2020), no estudo de anomalias na diferença de preços no mercado de ações nos Estados Unidos da América, apresentaram conclusões semelhantes a outros estudos, pois as anomalias evoluem e são importantes aos profissionais da área e acadêmicos. Sem dúvida, cabe verificar o comportamento do mercado financeiro brasileiro, verificando como este evolui ao longo dos anos. Tal análise é de interesse de muitos atores do mercado.

## 2.2 Tipos de anomalias no mercado de ações

Guerino *et al.* (2020) e Rtayli *et al.* (2020) detectaram as anomalias no sistema financeiro explorando *outliers* em sistemas de lavagem de dinheiro, sistemas de pagamento e transações bancárias. No mercado de ações, as anomalias são distorções que contradizem a Hipótese do Mercado Eficiente e assim criam oportunidades de retornos diferenciados ou anormais aos investidores e os preços de mercado muitas vezes desviam-se do seu valor intrínseco (Degutis, Novickyté, 2014).

Ahmed *et al.* (2017) relatam que, desde 2007, com a crise financeira em nível mundial, os profissionais de risco no mercado financeiro estão mais preocupados em fazer combinações de teoria com as situações práticas e, portanto, fazer a precificação é uma ação cada vez mais importante. Para Yavrumyan (2015), existem três categorias diferentes de anomalias que são: anomalias fundamentais, anomalias técnicas e anomalias de calendário. Já Shulmerich *et al.* (2015) fazem estudos das anomalias no mercado financeiro tais como: efeito fim de semana, efeito janeiro, efeito virada do mês e /ou feriado, viés doméstico, efeito S&P 500 e o enigma de Value Line, no qual os autores apontam que dada persistência de ocorrência não há oportunidades de arbitragem.

Do ponto de vista das finanças comportamentais, Abdulrasool e Othman (2012) relatam que a área dos investidores, atrelada ao viés do conservadorismo, está relacionada com as anomalias. No Quadro 4 adiante, estão elencados alguns tipos de anomalias relacionadas ao mercado de ações reconhecidos na literatura internacional.

Quadro 4-Tipos de anomalias no mercado de ações.

<b>Efeito</b>	<b>Descoberta</b>	<b>Autor</b>
<b>Anomalias de calendário ou sazonalidade</b>		
<b>Efeito Janeiro</b>	Nas primeiras semanas do ano, os preços das ações são maiores em janeiro do que nas últimas semanas de dezembro.	Rozeff e Kinney (1976)
<b>Efeito Halloween</b>	Retornos no mercado de ações tendem a ser maiores no período de novembro a abril do que entre maio a outubro.	Bouman e Jacobsen (2002)
<b>Efeito Virada do ano</b>	Com taxações e impostos há mais vendas em dezembro e compras de ações em janeiro. Perca de ações são maiores em dezembro devido a maiores negociações.	Guin (2005); Dyl (1977); Givoly e Ovadia (1983)
<b>Efeito Semana do mês</b>	Os retornos são maiores na primeira semana do mês do que na última semana do mês.	Linn e Lockwood (1988); Hensel and Ziemba (1996)
<b>Efeito Fim de semana Ou Efeito dia da semana</b>	O preço de fechamento da segunda-feira é inferior ao preço de fechamento da sexta-feira. O Efeito fim semana ocorre quando empresas e governos percebem más notícias no fim de semana.	Cross (1973); French (1980); Guin (2005)
<b>Efeito Feriado</b>	Retorno anormal do mercado de ações antes do início de feriados.	Lakonishok e Smidt (1988); Petengill (1989)
<b>Efeito Ciclo Político</b>	Retornos mais elevados durante o primeiro e o último ano de eleições para presidente.	Santa e Valkanov (2003)
<b>Efeito do Fim do dia</b>	Há um aumento de negociações e dos valores das ações nos últimos 15 minutos do dia.	Guin (2005)
<b>Efeito Dia da Semana</b>	Os retornos das ações não são constantes e diferem, geralmente a evidência mais notável é para a segunda-feira em que os retornos são incomumente baixos ou negativos	Gibbons, Hesse (1981)
<b>Efeito Ramadã</b>	Em países islâmicos os retornos tendem a ser menores no Ramadão do que no restante dos meses.	Jedrzej e Bialkowski (2010)
<b>Anomalias Fundamentais</b>		
<b>Anomalias de Tamanho</b>	Instituições de pequena capitalização obtêm retornos maiores do que instituições de maior capitalização.	Banz (1981)
<b>Anomalias de Início de Oferta ao Público</b>	Empresas que entram pela primeira vez no mercado são mal avaliadas em contrapartida seus retornos são maiores em comparação ao mercado no curto prazo.	Loughran e Ritter (2004)
<b>Efeito de Valor</b>	Instituições com menor multiplicador de rentabilidade tem maiores retornos frente a instituições com alto multiplicador de rentabilidade.	Goodman e Peavy (1983)
<b>Outras Anomalias</b>		
<b>Efeito <i>Momentum</i></b>	Ações com maiores valores superam o restante das outras ações, e ações com menores valores atuam com o mesmo “desempenho”	Egadeesh e Titman (1993)
<b>Reversão e Momento</b>	No intervalo de três a cinco anos ações com retornos mais baixos tendem a produzir maiores retornos nos próximos três a cinco anos.	Debondt e Thaler (1985)
<b>Quebra de Intervalo em negociação</b>	O sinal para venda de ações é gerado quando as taxas excedem o nível de suporte, ou seja, quando os preços sobrem além do último pico de venda e os preços caem abaixo do último vale.	Brock (1992); Josef (1992); Lakonishok et al (1992)
<b>Médias em movimento</b>	Ações que são comparadas como médias em curto prazo aumentam acima das médias de longo prazo.	Brock (1992); Josef (1992); Lakonishok et al. (1992)

Fonte: Adaptado de Abdulrasool e Othman (2012).

Muitas outras questões estão relacionadas ao mercado de ações que podem influenciar o surgimento de algumas anomalias. Por exemplo, o comportamento da atividade econômica, da produção e do consumo, pode afetar bastante o mercado financeiro. Ariadi *et al.*, (2015) explicam que o conhecimento financeiro está ligado a ações de poupar e consumir. Este mesmo conhecimento evita consumos excessivos. É cabível fazer a reflexão de que o mercado financeiro é composto por investidores, seres humanos que tomam decisões eficientes, racionais, mas que estão sujeitos a errar.

Godoy (2022) relata que, além da racionalidade, os agentes são regidos por instituições, sentimentos e emoções. Não obstante, grande parte dos seres humanos, em muitos momentos de ira, replicam comportamentos nocivos de forma consciente e proposital.

Assim, as novas teorias e debates, sob a abordagem da Economia Comportamental, ganham relevância na análise dos agentes econômicos que atuam no mercado de ações. Racionalidade e irracionalidade são inerentes aos investidores de forma que esses componentes precisam ser estudados para uma compreensão mais profunda sobre a tomada de decisões em mercados financeiros.

### **2.3 Anomalias financeiras no Brasil: fontes e evidências**

É progressivo o aumento do número de investidores no Brasil nos últimos anos. Esse resultado pode ser confirmado por vários fatores. Dentre eles, uma maior propagação de informação nos últimos anos e o desenvolvimento de uma cultura educacional financeira. Além disso, há a entrada crescente dos investidores estrangeiros e a abertura de capital é intensificada pelas empresas, principalmente, no setor privado, o que justifica a necessidade de compreender os diferentes mecanismos que guiam a dinâmica do mercado e a compreensão de novas informações sobre os preços das ações (Gartner, 2011).

Lemos e Costa Jr (1995) estudaram o Modelo de Apreçamento de Ativos de Capital, do inglês *Capital Asset Pricing Model* (CAPM), utilizando as cotações mensais no mercado brasileiro de capitais, entre os anos de 1974 e 1993. Os autores constataram que investidores tendem a superestimar as informações atuais e a subestimar os dados anteriores. Também verificaram que existe um “comportamento anômalo semelhante ao encontrado no mercado norte-americano”. Para Saturnino *et al.* (2017), o CAPM é o principal e talvez o mais tradicional modelo para estimar retornos esperados das ações, de tal forma que é a base para

os vários modelos que surgiram, posteriormente, com a incorporação de variáveis adicionais.

Leite Filho e Colares (2014) analisaram o comportamento do Ibovespa no período entre 1990 e 2012 e obtiveram como resultado a não existência de anomalias de calendário. O resultado levou os autores a concluir que o mercado brasileiro opera de forma que impossibilita ganhos extraordinários relacionados com esse tipo de evento, e reforça que esta é uma evidência da validade da HME. Os autores também estudaram o efeito Eleições Presidenciais e constataram que as eleições interferem direta e indiretamente, principalmente por agentes especulativos no mercado.

Ainda sobre as anomalias de calendário no mercado brasileiro, uma das mais estudadas é o chamado efeito dia-da-semana ou efeito segunda-feira, no qual os retornos nesse dia são menores que os retornos nos outros dias da semana. Lemgruber *et al.* (1988) encontraram indícios da existência de anomalias de calendário, mais especificamente, o efeito dia da semana para o Índice Ibovespa na década de 80. Já o artigo de Santos *et al.* (2007) apresenta os testes estatísticos sobre os retornos do Ibovespa, no período de 1995 a 2006 e confirmou tal efeito.

Almeida *et al.* (2017) relatam o efeito *Halloween* nos retornos do mercado acionário na análise do período de 1994 a 2014 empregando a análise de regressão com variáveis *dummy*. Os autores relatam uma tendência de crescimento no período de novembro a abril em relação ao que ocorre no período entre os meses de maio e outubro. Nesse sentido, estratégias de investimento baseado no efeito Halloween permitiriam retornos estatisticamente significativos e superiores a uma estratégia de investimento *buy-and-hold* (comprar e segurar).

Sousa e Lucena (2021) estudaram informativos de notícias sobre o mercado e concluíram que os ativos negociados podem estar erroneamente avaliados, de modo que os autores detectaram bolhas de mercado. Os autores estudaram cinco janelas anômalas entre 1993 e 2020 e notaram que, após os períodos de anomalias, há sempre um grande volume de notícias negativas, ou ainda uma igualdade de notícias positivas e negativas.

O estudo de Neto e Saito (2003) dedica-se à análise dos pagamentos de dividendos e concluí que existem retornos anormais positivos para ações de maior *dividend yield* (rendimento de dividendo), aspecto que confronta a validade da HME.

Fama, Cioffi e Coelho (2008) elucidam que o mercado de ações no Brasil apresenta anomalias, mas não descartam totalmente a HME. Para os autores, as análises de correlação indicam que os mercados estão mais correlacionados e que o ambiente dos negócios está mais propício a reações exacerbadas, o que contribui para elevar a volatilidade.

Afonso *et al.*, (2015) estudou a relação entre o anúncio da meta da taxa Selic na BMF & Bovespa e retornos anormais de ações listadas na Bolsa. Com dados entre os anos de 2009 e 2013, os autores concluíram que o mercado de capitais brasileiro possui uma gama de variáveis que podem “enviesar os retornos das ações das empresas”.

### 3 METODOLOGIA

Nesta seção, são apresentados os aspectos metodológicos empregados no estudo. Foram aplicadas duas abordagens diferentes para fazer as inferências a respeito da significância estatística de variáveis associadas às anomalias de calendário. A primeira abordagem traduz os movimentos de alta e baixa diárias no Ibovespa em uma variável binária. Para modelar a correlação das variáveis de calendário com esses movimentos foi empregado um modelo de classe Logit.

A segunda abordagem utiliza a estrutura de série temporal do Ibovespa, e modela os retornos calculados com base no índice com o uso da metodologia Box-Jenkins. Em específico, a aplicação de modelos ARIMAX, incorporando as variáveis de calendário.

Por sua vez, as variáveis de calendário foram incorporadas aos modelos por meio de variáveis binárias (*dummies*) criadas para os dias de semana e meses do ano. A presente análise assume a hipótese de que as variáveis são adequadas para captar possíveis efeito de calendário conforme descritos na literatura.

#### 3.1 Método de análise

Nesta seção, são apresentados os modelos econométricos para a investigação das anomalias sazonais no mercado financeiro entre 2012 e 2022.

##### 3.1.1 Modelo Logit

O modelo Logit é uma ferramenta estatística robusta para lidar com as variáveis dependentes que assumem os valores discretos e dicotômicos. Bastante empregado em estudos de econometria aplicada, desenvolve-se a partir da distribuição logística e corresponde a uma forma de regressão que permite a estimação de probabilidades condicionais.

No contexto do presente estudo, os movimentos do Ibovespa (*ibov*) foram traduzidos na forma de uma variável binária  $y$  da seguinte forma:

$$Y = \begin{cases} 0, & \text{se } ibov \leq 0 \text{ no dia } t \\ 1, & \text{se } ibov > 0 \text{ no dia } t \end{cases} \quad (15)$$

Definido a variável explicada dessa forma, temos um modelo Logit binário, em que a probabilidade condicional de um determinado evento ocorrer em função de um conjunto de variáveis explicativas  $X$  é descrito por meio de uma função logística da seguinte forma:

$$P(y = 1|X) = P(X) = \frac{e^{X\beta}}{1 + e^{X\beta}} \quad (16)$$

Se essa expressão representa a probabilidade de observar o resultado de interesse, a probabilidade do resultado complementar é dada por:

$$1 - P(X) = \frac{1}{1 + e^{X\beta}} \quad (17)$$

Com um pouco de manipulação, escrevemos:

$$\frac{P(X)}{1 - P(X)} = e^{X\beta} \quad (18)$$

A quociente do lado esquerdo da expressão acima é conhecido como razão de chance (*odds ratio*) em favor do resultado de interesse. Tomando o logaritmo natural da razão de chance obtemos a seguinte expressão:

$$\ln\left(\frac{P(X)}{1 - P(X)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (19)$$

Essa expressão apresenta o logaritmo da razão de chance, também conhecido como Logit. Por isso, tal classe de modelos recebe essa denominação - como uma função linear do conjunto de variáveis explicativas.

Embora as probabilidades fiquem (por necessidade) entre 0 e 1, os Logits não são limitados. Se o Logit for positivo, significa que, quando o valor de uma variável explicativa  $X$  aumenta, as chances de a variável de resposta ser igual a 1 (indicando que algum evento de interesse acontece) aumentam. Se o Logit for negativo, as chances de a variável de resposta ser igual a 1 diminuem à medida que o valor de  $X$  aumenta. Em outras palavras, o Logit torna-se negativo e cada vez maior à medida que a razão de chances diminui de 1 para 0 e torna-se cada vez maior e positivo quando as chances aumentam infinitamente, a partir de 1.

A estimação dos parâmetros/coeficientes do modelo de regressão logística é usualmente realizada por meio da técnica de máxima verossimilhança. Essa abordagem busca encontrar os valores dos parâmetros que maximizam a probabilidade de observar os dados disponíveis na amostra. No contexto do modelo Logit, a função de verossimilhança é construída com base na distribuição de probabilidade condicional dos dados dado os parâmetros estimados. Nesse

caso, a função de verossimilhança é expressa como o produto das probabilidades condicionais de observar os resultados da variável dependente para cada caso na amostra.

Cada coeficiente estimado nessa equação é um coeficiente angular parcial e mede a variação no Logit estimado para uma variação unitária do valor da variável explicativa dada (mantendo-se tudo o mais constante). Conforme destaca Gujarati (2011), na estimação deste tipo modelo o que importa, a princípio, são os sinais esperados dos coeficientes de regressão e a sua significância estatística e/ou prática.

### 3.1.2 Modelos ARIMA e ARIMAX

O modelo ARIMAX é frequentemente visto como um derivado do modelo ARIMA e, a diferença entre eles é que este modelo utiliza além de parâmetros auto-regressivos e médias móveis, como o ARIMA, inclui outras variáveis exógenas (Bennett *et al.* 2014). Lydia *et al.* (2016) concluíram que uma grande vantagem do ARIMAX é admitir a entrada de variáveis que podem estar relacionadas com a variável dependente, que no presente trabalho é dada pelos retornos do Ibovespa.

De acordo com Siami Namin *et al.* (2018), o modelo ARIMA ( $p, d, q$ ) indica um modelo com componentes autorregressivo (AR) de ordem  $p$ ; integrado (I) de forma que a série seja estacionária após  $d$  diferenciações; e com componentes de médias móveis (MA) de ordem  $q$ . Os processos ARIMA representam uma classe de processos estocásticos utilizados para a análise de séries temporais, compondo a abordagem conhecida como metodologia Box-Jenkins (Contreras, 2003). De acordo com Mondal *et al.* (2014), o modelo ARIMA é baseado no ARMA, no qual a diferença é que o ARIMA converte dados não estacionários e estacionários e que esse modelo muitas vezes é aplicado em preços de ações e, portanto, não são valores gerados de forma aleatória, e são tratados como série temporal.

Contudo, o modelo mais apropriado deve ser identificado pelo estudioso e/ou pesquisador no seu conjunto de dados. Yuhong (2005) analisa que, para a quantificação de qualidade de ajuste do modelo, os Critérios de Informação Akaike (AIC) são bastante utilizados na modelagem estatística. Na comparação de modelos, o AIC mais baixo é geralmente considerado mais próximo dos dados.

AIC tem a vantagem de permitir a parcimônia na escolha da modelagem de modo que o modelo com maior número de coeficientes nem sempre é a melhor escolha (Burnham,

Anderson, 2004). Conforme Akaike (1974), os menores valores de AIC exprimem melhores ajustes.

$$AIC = -2 \cdot \log \text{like} + 2p \quad (20)$$

em que  $p$  é o número de coeficientes e  $\log \text{like}$  é o logaritmo da função de verossimilhança no qual considera as estimativas dos coeficientes.

O outro critério também utilizado na seleção de modelos está relacionado aos Critérios de Informação Bayesiano (BIC), também chamado de critério de Schwars, proposto por Schwars em 1978 (Emiliano, 2009).

Konish e Kitagawa (2008) definem o critério BIC matematicamente da seguinte forma:

$$-2 \log p_i(x_n) = -2 \log \int f_i((x_n | \theta_i) \pi_i \theta_i d\theta_i \quad (21)$$

Sendo,  $\hat{\theta}_i$  o estimador de máxima verossimilhança e  $f_i((x_n | \theta_i)$  o modelo.

Ainda em Emiliano (2009), os critérios AIC e BIC fundamentam-se na verossimilhança, o que difere os dois são os critérios de penalizações e ambos servem para comparar modelos encaixados, mas também modelos não encaixados.

Por sua vez, o *ARIMAX* ( $p, d, q, r$ ) a combinação de modelos AR modelos Auto-Regressivo ( $p$ ), Integrado ( $d$ ), Média Móvel ( $q$ ) e Exógeno ( $r$ ). Camelo (2019) representa de forma simplificada o modelo matematicamente da forma:

$$y_t = \rho + \sum_{i=1}^p \beta_i y_{t-i} + \sum_{j=1}^r \omega_j x_j + \sum_{j=1}^q (\theta_j \varepsilon_{t-j}) + \varepsilon_t \quad (22)$$

Nesta expressão,  $y_t$  é a variável explicada no tempo  $t$ ;  $\rho$  é a constante;  $y_{t-i}$  é a variável dependente defasada  $i$  períodos;  $\beta_i$  é o coeficiente de  $y_{t-i}$ , cada termo  $x_j$  representa as variáveis exógenas. O índice  $p$  é o número máximo de intervalos de tempo,  $r$  é o número máximo de variáveis exógenas,  $\omega_j$  são os coeficientes das variáveis exógenas;  $\theta_j$  é o coeficiente do termo  $\varepsilon_{t-j}$  no qual representa o erro no tempo  $t$  defasado de  $j$  períodos,  $\varepsilon_t$  é o erro do modelo (ruído branco).

Lee e Hamzah (2010) citam que a variação de calendários com series temporais tem sido investigada desde a década de 1980 por muitos pesquisadores e enfatiza as aplicações do método ARIMAX em estudos de previsão de vendas com efeito Ramadã.

## 4 RESULTADOS E DISCUSSÕES

Nesta seção, são apresentadas as inferências sobre as anomalias de efeitos políticos, efeito *Halloween*, efeito dia da semana e efeito Janeiro, seguido pelos resultados econométricos do modelo *Logit* e os modelos ARIMA e ARIMAX.

### 4.1 Anomalias de efeitos de ciclos políticos no Brasil

É comum na literatura o termo anomalias de efeito político, em que há retornos mais elevados durante o primeiro e último ano de mandatos de presidentes. Wong e McAleer (2009) concluíram em sua pesquisa que há uma tendência cíclica para as últimas dez administrações presidenciais nos Estados Unidos da América. Verificaram que durante a primeira metade do mandato há um valor mínimo para o preço das ações, e em seguida, durante o terceiro e quarto ano há um “pico”, o que contraria os resultados mais comuns das anomalias de efeito político.

Jensen e Schmith (2005) ilustram a utilidade de usar desempenho do mercado acionário em avaliação de impactos de partidos políticos específicos na economia e aclararam com relação às eleições de 2002 que, no Brasil, os agentes do mercado financeiro pareciam bastantes incertos sobre a eleição do presidente Luíz Inácio Lula da Silva. Esse fato aumentou a incerteza sobre a situação econômica do país em períodos futuros, porém não sinalizou, claramente, um declínio.

Assim sendo, serão feitas análises importantes sobre os efeitos de ciclo político de eleições para presidentes, como mostra a Tabela 4. Por não contemplarem o ciclo eleitoral completo, foram excluídos os anos do período de 2012 a 2014.

Tabela 4- Altas e baixas totalizadas por ano de 2014 e 2022.

<b>Ano</b>	<b>Altas</b>	<b>Baixas</b>
2014	116	133
2015	112	134
2016	139	110
2017	136	109
2018	129	116
2019	135	113
2020	130	118
2021	128	119
2022	127	123

Fonte: Resultados da pesquisa, 2023.

No primeiro ano do segundo mandato da presidente Dilma Vana Rousseff, em 2014, não houve retornos mais elevados. Na tabela acima, podemos visualizar uma maior quantidade de baixas para o Ibovespa do que altas. Esse fato repete-se para o ano seguinte, em que houve muitas tensões políticas e econômicas. Com o *impeachment* da presidente em 2016, surge o novo presidente Michel Miguel Elias Temer Lulia e, de 2016 a 2018, o número de altas do Índice Ibovespa supera o número de baixas em 29, 27 e 13, respectivamente. Nesse período, não se pode afirmar que há retornos mais elevados durante o primeiro e último ano de mandatos de presidentes, pois houve muitas mudanças no cenário nacional e os maiores aumentos em números de altas compreende os anos de 2016 e 2017, o que contraria o efeito de ciclo político de que há retornos mais elevados durante o primeiro e o último ano de presidentes eleitos.

Para análise do período de 2019 a 2022, durante o mandato do presidente Jair Messias Bolsonaro, há maiores altas do que baixas para todos os anos. O número de altas é maior do que as baixas do índice em todos os anos, com destaque para os dois primeiros anos em que a amplitude é maior. Sendo assim, não se pode considerar que há o efeito político para os períodos analisados. Para a verificação de tal efeito, seria necessária uma análise temporal mais vasta. Outro fato que pode mudar bastante a direção do índice é a questão da instabilidade política que ocorreu nos anos de 2015 e 2016, no Brasil. Ainda no tocante às eleições, a Tabela 5 relaciona o ano de 2022 no mês de outubro, no período de eleição para presidente do Brasil entre os candidatos Luiz Inácio Lula da Silva e Jair Messias Bolsonaro, e como a especulação financeira reage a momentos que antecedem a eleição.

Tabela 5-Análise do Índice Ibovespa para período de eleição para presidente do Brasil em 2022.

<b>Data</b>	<b>Fechamento Ajustado</b>	<b>Variação Percentual</b>
<b>30/09/2022</b>	110037.0	2,20%
<b>03/10/2022</b>	116130.0	5,54%
<b>04/10/2022</b>	116230.0	0,08%
<b>05/10/2022</b>	117198.0	0,83%
<b>06/10/2022</b>	117561.0	0,31%
<b>07/10/2022</b>	116375.0	-1,01%
<b>10/10/2022</b>	115941.0	-0,37%
<b>11/10/2022</b>	114827.0	-0,96%
<b>13/10/2022</b>	114300.0	-0,46%
<b>14/10/2022</b>	112072.0	-1,96%
<b>17/10/2022</b>	113624.0	1,38%
<b>18/10/2022</b>	115743.0	1,87%
<b>19/10/2022</b>	116274.0	0,46%
<b>20/10/2022</b>	117171.0	0,77%
<b>21/10/2022</b>	119929.0	2,35%
<b>24/10/2022</b>	116013.0	-3,27%
<b>25/10/2022</b>	114626.0	-1,20%
<b>26/10/2022</b>	112764.0	-1,62
<b>27/10/2022</b>	114641.0	1,66%
<b>28/10/2022</b>	114539.0	-0,09%
<b>31/10/2022</b>	116037.0	1,31%
<b>01/11/2022</b>	116929.0	0,77%

Fonte: Resultados da pesquisa, 2023.

Os resultados da eleição do primeiro turno foram divulgados em 02/10/2022, no dia seguinte a variação do índice foi 5,54% em relação ao último dia (na sexta anterior). Já no dia 31/10/2022, um dia após a eleição do segundo turno, a variação foi menor, 1,31%. O mercado financeiro tende a absorver questões políticas como a escolha do presidente, muitas vezes, de forma intensificada. A especulação no Brasil pode ser verificada em momentos de eleição, crises em empresas estatais de grande porte como a Petrobrás ou ainda em momentos que antecederam o impeachment, como descrito no tópico adiante.

#### 4.1.2 Efeitos políticos e momentos diferenciados que podem alterar o Ibovespa

No ano de 2016, houve o *impeachment* da presidente Dilma Vana Rousseff. Dentre muitas tratativas, no dia 10 de agosto de 2016, o Plenário decidiu que a presidente iria a julgamento e ela foi afastada. No dia 31/8/2016, a presidente sofre o *impeachment*. Contudo, permaneceu com direitos políticos. Então, os anos de 2015 e 2016 podem ter gerado expectativas em investidores no qual a instabilidade política e econômica afetou negativamente os investimentos, de maneira geral, o Brasil.

Há momentos peculiares no cenário brasileiro. No dia 17 de maio de 2017, foi divulgado pelo empresário Joesley Batista, dono da JBS, áudios do presidente Michel Temer, dando o aval para a compra de silêncio de Eduardo Cunha (ex-presidente da Câmara dos deputados) no qual estava preso. O impacto no mercado de ações foi imenso, e assim houve uma desvalorização do Ibovespa de -8,80% em um único dia. Foi, então, acionado o *circuit breaker*, um mecanismo de suspensão temporário da Bolsa de Valores da B3, pois os investidores vendiam as ações de modo desenfreado. O referido dia foi tão significativo no mercado de ações que, por isso, foi denominado como “Joesley day”.

Outra consequência desse episódio foi a valorização do dólar, que naquele mesmo dia teve a maior alta em 18 anos. Episódios assim podem criar alterações no Ibovespa e podem gerar gaps de preços no mercado acionário. A Tabela 6 retrata o fechamento do índice e a variação percentual dos momentos elencados.

Tabela 6-Efeitos políticos na variação percentual ao Ibovespa.

<b>Data</b>	<b>Fechamento</b>	<b>Variação</b>	<b>Evento</b>
11/08/2016	58.300	2,42%	Afastamento da presidente
31/08/2016	57.901	-1,15%	Impeachment da presidente
18/05/2017	61.597	-8,80%	Joesley day

Fonte: Resultados da pesquisa, 2023.

Momentos de tensão que antecederam a guerra entre Ucrânia e Rússia também alteraram significativamente o Ibovespa. No dia 17/2/2022, é noticiado o confronto entre “pró-russo” e o exército ucraniano, e, no 24/2/2022, é declarada a guerra entre os dois países.

A variação dos índices para o dia 17 e 24 do mês referido são -1,435% e 1,389% respectivamente.

Notícias sobre a pandemia ocasionada pelo vírus Covid-19 também impactaram enormemente a Bolsa de Valores. Quando foi noticiado o primeiro caso de Covid-19 no Brasil, no dia 26/2/2020, o Ibovespa teve uma queda de 7,0%. No dia 11 de março de 2020, a Organização Mundial da Saúde (OMS) caracteriza a Covid-19 como uma pandemia com consequência no dia seguinte, de modo que o Ibovespa teve variação de 14,78% em queda, seguido por altas e baixas anômalas no mesmo mês.

Fazer previsão sobre as questões políticas, os momentos que antecedem uma guerra ou mesmo os eventos pandêmicos são extremamente difíceis em relação ao Índice Bovespa. O que se percebe é que, em momentos de tensão dos investidores, o Ibovespa tende a subir, porém, no acontecimento de eventos descritos, o índice tende a descer, como mostra a variação negativa da Tabela 6. Sem dúvida, retornos anormais surgem em momentos diferenciados e prever esses momentos, como descritos acima, são muito difíceis, algo que não ocorre somente com o Ibovespa, mas com outros índices do mercado acionário mundial.

#### **4.2 Efeito Halloween**

De acordo com Bouman e Jacobsen (2002), existe um ditado na Europa que sugere “venda em maio e vá embora, mas lembre-se de voltar em setembro”, de forma a ser possível pensar que o mês de maio é tido como o início do mercado em baixa. Assim, os investidores optam por vender as suas ações nos meses de novembro a abril, pois proporcionam retornos mais elevados do que nos outros meses do ano. Para concretização do efeito *Halloween*, os retornos de ações são maiores no período de novembro a abril do que entre maio e outubro. Bouman e Jacobsen (2002) identificaram esse efeito no Brasil, com observações de processo estocástico do preço do ativo o qual tem como limite final de tempo o ano de 1994.

Almeida *et al.*, (2017) fizeram análises com efeito *Halloween* no Brasil, entre 1994 e 2014, com regressões de variáveis *dummies* e ações das empresas listadas na Bolsa de Valores e constataram o retorno médio dos meses de maio a outubro é 1,2% ao mês, enquanto nos meses de novembro a abril é 3,3%.

Em estudos de Lucey e Zhao (2008), o efeito Halloween pode ser um reflexo do efeito Janeiro no qual é ele próprio diminuído em significância e termos absolutos. Para Haggard *et al.* (2010), o efeito *Halloween* é significativo no EUA entre 1954 e 2008, além de considerar

que geralmente as anomalias estão mais presentes em dados mais antigos quando investidores com experiência e perspicácia poderiam tirar proveito com fins lucrativos.

No Brasil, tal efeito foi analisado em quantidades de altas em relação a baixas para o Índice Ibovespa no período de 2012 a 2022, como mostra a Tabela 7. Entre 2012 e 2015, há quantidade de baixas maiores do que altas. Entre 2015 e 2016, 2016 e 2017 e 2021 e 2022, o mesmo número de altas e baixas para o Ibovespa. Entre 2018 e 2021, há maiores altas em relação a baixas para o índice. Portanto, não há evidências, quantitativas do efeito descrito para o índice em relação à quantidade entre altas e baixas nos períodos de novembro a abril.

Tabela 7- Efeito Halloween para o período de 2012 e 2022 do índice Ibovespa.

Anos	Quantidades de Altas de novembro a abril	Quantidades de Baixas de novembro a abril
2012 e 2013	54	64
2013 e 2014	57	62
2014 e 2015	55	66
2015 e 2016	61	61
2017 e 2018	66	53
2018 e 2019	62	56
2019 e 2020	60	60
2020 e 2021	69	50
2021 e 2022	61	61

Fonte: Resultados da pesquisa, 2023.

### 4.3 Efeito dia da semana

O fenômeno efeito dia da semana é um tipo de anomalia em que o retorno médio diário do mercado é diferente em alguns dias da semana e argumenta-se que essa anomalia faz parte da teoria do mercado de capitais eficiente e tal fenômeno foi constatado em mercado desenvolvidos (Reino Unido, França, Japão e Canadá), mas também em mercado emergentes (Hong Kong, Malásia e Turquia) (Drogalas *et al.*, 2007).

Gibbons e Hess (1981) relatam que várias teorias são investigadas, porém “ninguém provou de forma satisfatória” e surge uma investigação aprofundada do equilíbrio de retorno ao longo do tempo no mercado acionário. Stosic *et al.* (2022) discorrem que as anomalias do efeito dia da semana são bem conhecidas na literatura financeira, mas os fenômenos associados a ela precisam ser explorados na econofísica, assim como comentam o retorno de

segunda-feira, pois tende a exibir comportamentos estruturais e persistentes multifractais mais ricas do que os retornos em outros dias. Bayar e Kan (1999) relataram um estudo do efeito do dia da semana em dezenove países, entre 1993 e 1998, e constataram que taxas médias de retornos mais elevadas nas terças-feiras e quartas-feiras, enquanto nas quintas-feiras e sextas-feiras as taxas de retorno médios foram menores.

No Brasil, Costa Jr (1990) fez análises sobre o efeito dia da semana com o Ibovespa no período de 1986 a 1989 e constatou que esse fenômeno é semelhante ao do mercado americano, em que o retorno é menor na segunda-feira e maior na sexta-feira. Isso sugere que esse fenômeno é também conhecido como efeito fim de semana e que não existe uma explicação razoável para esse efeito. Caretta e Costa Jr apresentaram estudos do efeito dia da semana em seis países da América Latina entre 1994 e 1999, e somente no Peru e Venezuela, tal efeito foi significativo. No Brasil, não houve distinção significativa entre os diferentes dias da semana com estudo de regressão e teste F.

#### **4.4 Efeito Janeiro**

Para Famá *et al.* (2008), os retornos no mês de janeiro são em média maiores em comparação aos outros meses do ano. Retornos diferenciados no mercado de ações foram observados primeiramente por Wachtel (1942) e Rozeff e Kinney (1976), no qual os retornos são, significativamente, maiores em janeiro (Almeida *et al.* 2017).

Avdalovic e Milenkovic (2017) analisaram com dados em painel o mercado de ações de países emergentes da Bósnia e Herzegovina, Macedônia, Montenegro, Croácia, Romênia e Bulgária e Sérvia entre 2008 e 2014. Tal feito indicou fraca sustentabilidade da (EHM) dos países elencados e tal efeito está presente de forma significativa somente na Macedônia. Sahin *et al.*, (2017) examinaram os mercados emergentes da Bolsa de Istanbul e Bucareste, entre 2000 e 2014, utilizando preços de fechamento diário com abordagem de relação de potência no que resultou que o efeito de Janeiro é persistente para ambas as bolsas.

Com a utilização da metodologia padrão de estudo de eventos ajustados ao risco Klock (2014) sugere que ações com baixo desempenho apresentam risco negativo anormal com retornos ajustados de até 60 dias antes do último dia de negociação do ano, com retornos seguintes, significativos positivos durante o mês de janeiro com início 30 dias antes do final do ano com amostras do mercado mundial de 2010 e 2012.

Costa Jr *et al.* (1990) estudou as cotações mensais da série do Ibovespa entre 1969 e

1988 com regressão múltipla. Desse modo, constataram tanto para a série deflacionada quanto para a série não deflacionada resultados no teste F não significativos. Os autores explicam que, na época, exista uma taxaço sobre os ganhos de capital em que os investidores que possuíam ações e haviam sofrido perdas de valores no decorrer do ano eram motivados a vende-las no mês de dezembro para abatimento de impostos. Portanto, no mês de janeiro, o mercado voltava com “equilíbrio”, ou seja, as ações que tiveram um declínio artificial no final do ano sofriam retornos “acima do normal” no período seguinte a dezembro.

Na questão tributária, Potin *et al.*, (2015) analisaram as ações de empresas brasileiras entre 1996 e 2013 com subamostras de grupos de ações com ou sem ADR (*American Depositary Receipts*) no que resultou em evidências do efeito Janeiro nas ações de empresas brasileiras do Ibovespa com retornos negativo em dezembro e explicado pela hipótese de gestão tributária dos ganhos de capital.

Leito Filho *et al.* (2014) analisaram o retorno do Ibovespa entre 1990 e 2012 e constataram que os meses de janeiro possuíam retornos médio bem próximos aos outros meses do ano e com análises de outros testes, como Kolmogorov-Smirnov e Mann-Whitney, indicando a inexistência para o período de destaque o efeito janeiro no mercado de ações no Brasil.

#### **4.5 Resultados da análise econométrica**

Com o objetivo de testar a significância estatística de variáveis de calendário, que podem representar efeitos de anomalias, foram estimados modelos da classe Logit para as variações diárias do Ibovespa e modelos de regressão para os retornos diários calculados com base no índice.

As variáveis de calendário foram incorporadas aos modelos por meio de variáveis binárias (*dummies*) criadas para os dias de semana e meses do ano. A presente análise assume a hipótese de que tais variáveis são adequadas para captar o possível efeito de calendário, conforme descrito na literatura.

##### *4.5.1 Modelo Logit para variações no Índice Ibovespa (2012 – 2022)*

Definindo as variações do Ibovespa como uma variável dependente binária, com altas sinalizadas pelo valor 1 e baixas (ou estabilidade) por zero, aplicou-se um modelo Logit. Com

a proposta de realizar inferências sobre o efeito das variáveis de calendário sobre os movimentos do Ibovespa, a presente análise adota o conjunto de *dummies* para os dias de semana e meses.

A estimação do modelo Logit é realizada com a aplicação do método de Máxima Verossimilhança. A Tabela 8 traz os valores dos coeficientes estimados para as duas especificações do modelo: a primeira delas inclui apenas *dummies* de calendário; e a segunda incorpora algumas variáveis de controle que incluem médias móveis simples e exponenciais do próprio Ibovespa e variáveis relacionadas com o ambiente econômico e financeiro brasileiro.

De acordo com os resultados, as *dummies* de calendário não se mostram estatisticamente significantes para explicar variações do Ibovespa. Nesse sentido, os movimentos do mercado não seriam correlacionados com dias ou meses específicos. A ausência de significância estatística em variáveis relacionadas ao calendário pode ser consistente com a ideia de que as informações contidas nessas variáveis já estão incorporadas nos preços do mercado. Este resultado é coerente com as evidências encontradas por Leite Filho e Colares (2014).

Tabela 8 –Resultado do Modelo Logit para variações do Ibovespa.

	Logit com apenas dummies de calendário		Logit com dummies de calendário e controles	
	Coefficiente	Erro Padrão	Coefficiente	Erro Padrão
<b>Intercepto</b>	-0.0091	0.1761	1.0023	0.8287
<b>Terça</b>	0.1651	0.1320	0.1382	0.1425
<b>Quarta</b>	0.0955	0.1311	0.0319	0.1414
<b>Quinta</b>	0.0972	0.1321	0.0170	0.1435
<b>Sexta</b>	0.0079	0.1318	-0.0678	0.1470
<b>Fevereiro</b>	-0.1085	0.2101	0.1240	0.2469
<b>Março</b>	-0.1676	0.2047	0.0290	0.2342
<b>Abril</b>	-0.0656	0.2067	0.1303	0.2468
<b>Mai</b>	-0.1749	0.2074	0.0365	0.2377
<b>Junho</b>	-0.1116	0.2055	0.0690	0.2347
<b>Julho</b>	0.2380	0.2087	0.1852	0.2352
<b>Agosto</b>	-0.0370	0.2033	0.0501	0.2261
<b>Setembro</b>	-0.1490	0.2081	-0.0414	0.2276
<b>Outubro</b>	0.1002	0.2062	0.1132	0.2282
<b>Novembro</b>	-0.1199	0.2133	-0.0618	0.2372
<b>Dezembro</b>	0.1319	0.2092	0.2282	0.2318
<b>brent</b>			-0.0032	0.0037
<b>sp500</b>			-0.0001	0.0002
<b>dolar</b>			0.2239	0.1753
<b>iron</b>			0.0019	0.0023
<b>poup</b>			-0.1075	1.5155
<b>selic</b>			-1.0439	1.6565
<b>cdi</b>			1.0341	1.6505
<b>IIE-Br</b>			-0.0095***	0.0055
<b>inpc</b>			-0.6414	0.5666
<b>ipca</b>			0.4521	0.6255
<b>ibov_SMA_5</b>			-0.0026***	0.0002
<b>ibov_SMA_20</b>			-0.0002***	0.0001
<b>ibov_EMA_5</b>			0.0028***	0.0002
<b>ibov_EMA_20</b>			0.0001	0.0001

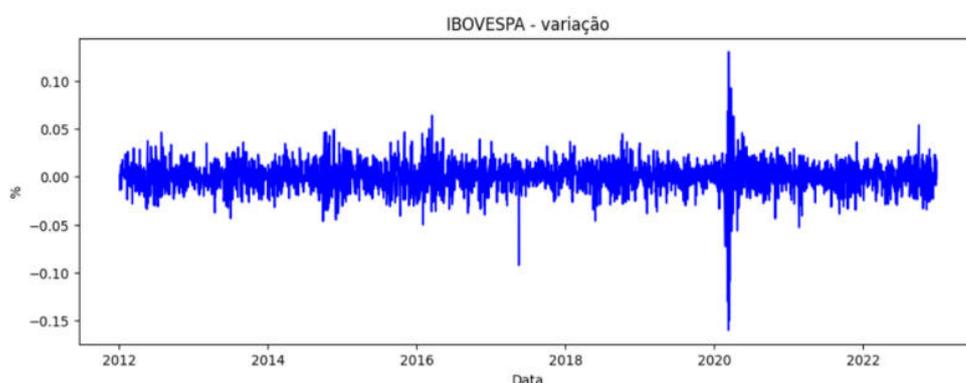
Fonte: Resultados da pesquisa, 2023. Significância estatística: \*  $p < 10\%$ , \*\*  $p < 5\%$ , \*\*\*  $p < 1\%$ .

#### 4.5.2 Modelo ARIMAX para análise do retorno do Índice Ibovespa (2012 – 2022)

Alternativamente ao modelo anterior, o presente estudo investiga se as variáveis de calendário poderiam exercer efeitos significantes sobre os retornos do Ibovespa. Para realizar essa estimação, optou-se por aplicar uma modelagem de séries temporais com modelos ARIMAX.

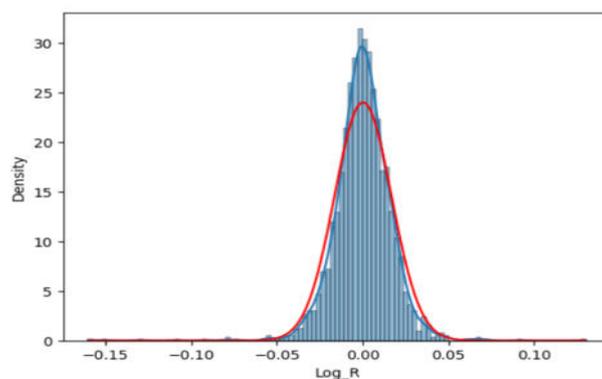
As Figuras 6 e 7 apresentam a série da distribuição de retornos do Ibovespa e sua distribuição. A análise destes gráficos indica uma série sem tendências e com uma distribuição muito próxima de uma distribuição normal com média zero. Verifica-se alguns momentos de maior volatilidade, seguindo cenários econômicos e políticos que foram comentados anteriormente.

Figura 6– Série temporal de retornos do Ibovespa (2012-2022).



Fonte: Resultados da pesquisa, 2023

Figura 7– Histograma da distribuição dos retornos (2012-2022).



Fonte: Resultados da pesquisa, 2023

Assim como qualquer modelagem de séries temporais com a abordagem Box-Jenkins, verificou-se a estacionariedade da série de retornos do Ibovespa. Foram aplicados os testes de Dickey Fuller Aumentado e o Teste de Estacionariedade Kwiatkowski-Phillips Schmidt-Shin (KPSS). A Tabela 9 exibe os resultados da aplicação desses testes.

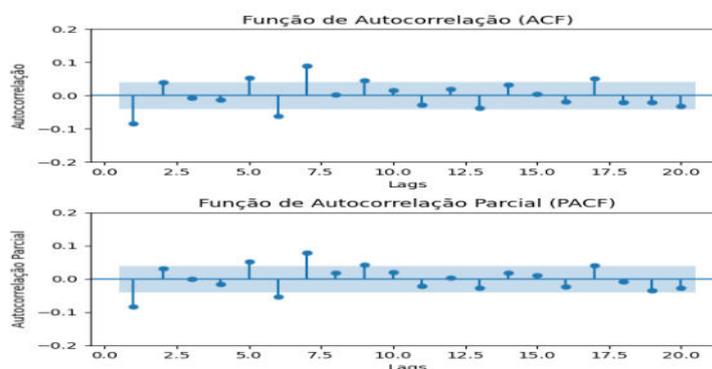
Tabela 9 - Teste de Estacionariedade Dickey-Fuller e Teste de Estacionariedade Kwiatkowski-Phillips Schmidt-Shin (KPSS) na série do retorno do Ibovespa entre 2012 e 2022.

	<b>ADF</b>	<b>KPSS</b>
<b>Estatística de teste</b>	-18,84	0,072
<b>Valor-P</b>	0.0000	0,01
<b>Lags Usados</b>	6	7
<b>Número de observações usadas</b>	2733	2733
<b>Valores Críticos (1%)</b>	-3,43	0,739
<b>Valores Críticos (5%)</b>	-2,86	0,463
<b>Valores Críticos (10%)</b>	-2,57	0,347

Fonte: Resultados da pesquisa, 2023.

Os resultados dos testes ADF e KPSS são consistentes, pois ambos indicam a rejeição da hipótese nula de não estacionariedade. Na sequência, foram realizados testes para verificar a ordem adequada dos termos autorregressivos e de média móvel dos modelos a serem estimados. Além da avaliação gráfica das funções de autocorrelação e autocorrelação parcial, foram estimadas diversas especificações do modelo em diferentes ordens e os critérios de informação de Akaike (AIC) e Bayesiano (BIC) foram empregados para a escolha da ordem do modelo mais adequada.

Figura 8– Funções de autocorrelação e autocorrelação parcial.



Fonte: Resultados da pesquisa, 2023.

Na presente análise, os testes baseados em critérios de informação indicaram que os modelos mais adequados são os modelos de ordem 4, tanto para termos autorregressivos, quanto para termos de médias móveis. O menor valor de AIC obtido foi de -12.823,9310 e o menor valor de BIC foi de -12.788,4042.

A Tabela 10 retrata os coeficientes estimados de três especificações com base no modelo ARIMA. A primeira delas caracteriza um ARIMA (4, 1, 4). A segunda retrata um ARIMAX (4,1,4) contando apenas com *dummies* de calendário como variáveis exógenas. A terceira representa um ARIMAX (4,1,4) contando com *dummies* de calendário e um conjunto de variáveis financeiras e econômicas como variáveis exógenas.

Tabela 10- Resultados da estimação dos modelos ARIMA e ARIMAX para os retornos do Ibovespa.

					(continua)	
	Coeficiente	Erro-padrão	Coeficiente	Erro-padrão	Coeficiente	Erro-padrão
<b>ar.L1</b>	-0.0311	0.173	-0.0823	0.424	-0.9165***	0.006
<b>ar.L2</b>	1.0285***	0.121	0.8047**	0.331	-0.4065***	0.009
<b>ar.L3</b>	0.2534*	0.138	0.1224	0.389	-0.0828***	0.009
<b>ar.L4</b>	-0.6106***	0.129	-0.4808*	0.282	-0.1949***	0.010
<b>ma.L1</b>	-0.0562	0.170	-0.0170	0.419	0.5779***	0.009
<b>ma.L2</b>	-1.0026***	0.106	-0.7779***	0.296	-0.0847***	0.011
<b>ma.L3</b>	-0.1687	0.137	-0.0474	0.379	-0.3163***	0.010
<b>ma.L4</b>	0.5979***	0.112	0.4463*	0.240	0.0927***	0.006
<b>sigma2</b>	0.0003***	3.63e-06	0.0003***	3.76e-06	0.0002***	3.24e-06
<b>Terça</b>			0.0017	0.001	0.0018*	0.001
<b>Quarta</b>			0.0012	0.001	0.0007	0.001
<b>Quinta</b>			0.0008	0.001	0.0004	0.001
<b>Sexta</b>			0.0003	0.001	0.0002	0.001
<b>Fevereiro</b>			-0.0010	0.001	0.0017**	0.001
<b>Março</b>			-0.0014	0.001	0.0016***	0.001
<b>Abril</b>			0.0004	0.001	0.0024***	0.001
<b>Mai</b>			-0.0018	0.001	0.0005	0.001
<b>Junho</b>			-0.0020	0.001	0.0002	0.001
<b>Julho</b>			0.0018	0.001	0.0020***	0.001
<b>Agosto</b>			-0.0004	0.001	0.0009	0.001
<b>Setembro</b>			-0.0013	0.001	0.0003	0.001
<b>Outubro</b>			0.0002	0.001	0.0004	0.001
<b>Novembro</b>			-0.0008	0.001	0.0001	0.001
<b>Dezembro</b>			-0.0003	0.001	0.0007	0.001
<b>brent</b>					2.825e-05***	1.26e-05
<b>sp500</b>					-1.484e-06**	7.6e-07

	Coeficiente	Erro-padrão	Coeficiente	Erro-padrão	Coeficiente	Erro-padrão
<b>dolar</b>					0.0027***	0.000
<b>iron</b>					-7.409e-06	8.56e-06
<b>poup</b>					0.0018	0.005
<b>selic</b>					-0.0032	0.006
<b>cdi</b>					0.0031	0.006
<b>IIE-Br</b>					-4.153e-05***	1.31e-05
<b>inpc</b>					-0.0026	0.002
<b>ipca</b>					0.0024	0.002
<b>ibov_SMA_5</b>					-2.284e-05***	4.5e-07
<b>ibov_SMA_20</b>					-1.625e-06***	2.58e-07
<b>ibov_EMA_5</b>					2.432e-05***	4.69e-07
<b>ibov_EMA_20</b>					9.89e-08	3.36e-07

Fonte: Resultados da pesquisa, 2023. Significância estatística: \*  $p < 10\%$ , \*\*  $p < 5\%$ , \*\*\*  $p < 1\%$ .

No caso das duas primeiras especificações, os coeficientes estimados indicam que nenhuma das variáveis de calendário se mostraram estatisticamente significantes. Considerando um nível de significância de 10%, para especificação mais completa, temos que foram significativos os coeficientes associados às *dummies* para a terça-feira e para os meses de fevereiro, março, abril e julho.

Sobre a anomalia do efeito Halloween não houve retornos positivos para os meses de fevereiro, e março apesar de estatisticamente significantes, e para os demais meses não houve diferenciação no quesito retornos positivos. Em comparativas com Potin *et al.*, (2015) não foram constatados retornos significativos no mês de janeiro, porém houve retornos negativos em dezembro não significantes.

Também se observou como significantes os coeficientes relacionados a algumas variáveis macroeconômicas e financeiras, a exemplo de preços do petróleo, o índice S&P500, a cotação do dólar, o índice de incerteza econômica e médias móveis do próprio Ibovespa. A significância de retornos mais elevados nas terças-feiras já havia sido reportada nos resultados de Bone e Ribeiro (2002) e Leite Filho e Colares (2014). Estes autores denominam tal efeito como “efeito Brasília”, uma vez que nas terças-feiras haveria uma variação diferenciada no retorno das ações brasileiras em função dos resultados das reuniões da câmara de deputados federais e senadores da república às segundas feiras, e pelo movimento parlamentar mais elevado no dia.

## 5 CONSIDERAÇÕES FINAIS

Com a combinação de um grande conjunto de informações, o presente estudo tentou o entendimento do comportamento, de anomalias sazonais que podem ocorrer no Brasil e/ou já ocorreram. Dessa forma, a pesquisa tentou contribuir com os analistas, os investidores e as empresas que lidam com o mercado de ações para que possam tomar as melhores decisões sobre os investimentos. O conhecimento antecipado dos investidores sobre o comportamento sazonal de variação no preço do Ibovespa pode ajudar a fomentar as estratégias de negociação que podem gerar retornos mais elevados.

Conclui-se que os modelos Logit, ARIMA E ARIMAX desempenharam, através dos testes econométricos e estatísticos, boas inferências para o Índice Ibovespa com as variáveis utilizadas na pesquisa. Constatou-se que, para o período analisado, não há como prever mudanças da direção de preços em momentos diferenciados como: pandemia, guerra ou mesmo *impeachment*, haja vista que momentos assim são bastante diferenciados e causam muitas mudanças para o índice.

Com análises quantitativas e econométricas, não foram constatadas anomalias do efeito *Halloween*. Sobre o efeito de ciclos políticos, momentos como *impeachment*, escândalos políticos e crises econômicas influenciam fortemente a variação do índice e impactam a confiança e/ou a desconfiança de investidores, sendo eles nacionais ou estrangeiros. Para o efeito dia da semana, foi constatado retornos diferenciais para a terça-feira a um nível de significância de 10% e para os demais dias não há retornos significativos. Para os meses, fevereiro, março, abril e julho, há retornos diferenciados do Ibovespa.

Tem-se como sugestão a continuidade de estudos do Ibovespa com a inclusão de outras variáveis técnicas ou fundamentalistas, assim como a utilização de outros modelos econométricos e estatísticos para verificação de mudanças anormais no Índice Ibovespa. Sem dúvida, anomalias no mercado financeiro já são estudadas há algum tempo, mas precisam de melhores explicações e análises assertivas para o Brasil, que tem muitos investidores estrangeiros e cresce o número de investidores nacionais.

## 6 CONCLUSÃO

Com os resultados apresentados, torna-se possível concluir que, especificamente, no primeiro ensaio o aprendizado de máquinas (*Machine Learning*) foi importante na previsão do Índice Ibovespa com destaque para os modelos Logit, Regressão de Lasso e Redes Neurais, no qual tiveram as maiores métricas de desempenho como acurácia, precisão, especificidade e curva ROC. Isso vai ao encontro com dezenas de trabalhos na literatura internacional e nacional. Diferentemente de outros trabalhos analisados, a Regressão de Lasso foi utilizada nessa pesquisa, no qual obtiveram-se bons resultados de previsão.

No que tange ao segundo ensaio, foram identificadas anomalias financeiras de ordem mundial na revisão de literatura, e mediante de análises econométricas e testes estatísticas. Desse modo, foram analisadas anomalias do tipo sazonais no Brasil que foram: anomalia de efeitos de ciclo políticos, anomalia de efeito *Halloween*, efeito dia da semana, e efeito janeiro no qual foram constatados apenas o efeito dia da semana para a terça-feira com retornos diferenciados e o efeito político no Brasil, que pode ser intensificado em momentos de crises econômicas, guerras, pandemias, entre outros momentos diferenciados. Anomalias do efeito *Halloween* não foram constatadas, como sugerem alguns autores citados na presente pesquisa, tão pouco o efeito janeiro no período de análise.

Em síntese, pela importância e complexibilidade do assunto, sugere-se o aprimoramento das técnicas empregadas a outras variáveis e/ou que outros modelos sejam incorporados, pois novas modelagens são empregadas com o desenvolvimento de *Machine Learning* para fazer previsão do Ibovespa com intuito de nortear centenas de investidores no mercado acionário, a fim de que possam fazer as melhores escolhas.

## REFERÊNCIAS

- ABDULRASOOL, H. D.; OTHMAN, R. Analysing Global Research on Stock Market Anomalies: A Behavioural Finance Perspective. **International Journal of Academic Research in Business and Social**, Al Diwaniya, v, 12, n. 6, p. 1891-1912, ago. 2012.
- ADEBVI, A. A.; ADEWUMI, A. O; AYO, C.K. Comparison of ARIMA and artificial neurais networks models for stock price prediction. **Journal of Applied Mathematics**, Cairo, v. 2014,[s.n], set.2014.
- AFONSO, A. S; SILVA, S. F; BORTOLON, P. M; MACEDO, M. A.S. Política Monetária e Mercados de Capitais: Análise do Impacto da Divulgação da Taxa Selic nos retornos das ações empresas listadas na BMF&BOVESPA. *In*: CONGRESSO ANPCONT, 2015, Curitiba. Anais [...].São Paulo: Anpcont, 2015, p.1-17.
- AGGARWAL, C. C. An Introduction to Outlier Analysis. **Outlier analysis**, Berlim, v.1, [s.n], p. 1–34, dez. 2016.
- AGUIRRE, A. A. A; MEDINA, R. A. R; MÉNDES, N. D. D. Machine learning applied in the stock market through the Moving Average Convergence Divergence (MACD) indicator. **Investment Management & Financial Innovations**, Sumy, v. 17, n. 4, p. 44, nov.2020.
- AHMED, M.; CHOUDHURY, N.; UDDIN, S. Anomaly detection on big data in financial markets. *In*: **Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017**, Sydney,p. 998-1001, ago.2017.
- AKAIKE, H. A new look at the statistical model identification. **IEEE Transaction on Atomic Control**, Alfenas, v. 9, n. 6, p. 716-723,ago. 1974.
- ALMEIDA, J. R.; ALMEIDA, G. R.; BERGMANN, D. R. O Efeito Halloween no Mercado Acionário Brasileiro. **Revista Brasileira de Finanças**, Rio de Janeiro, v. 14, n. 4, p. 1-34, out. 2017.
- ARIADI, R; MALELAK, M, Ing; ASTUTI, Dewi. Analisa hubungan financial literacy dan demografi dengan investasi, saving dan konsumsi. **Finesta**, Surabaya, v. 3, n. 1, p. 7-12, 2015.
- ASHOFTEH, A.; BRAVO, J. M. A conservative approach for online credit scoring. **Expert Systems with Applications**, Amsterdam, v. 176,[s.n], p. 114835,ago.2021.
- ATSALAKIS, G. S.; VALAVANIS, K. P. Suveryingsotck market forecasting techniques-Part II:Sooft computing methods.**Expert Sytems with Applications**, Amsterdam, v. 36, n. 3. p-5932-5941, abr. 2009.
- AVDALOVIC, S. M; MILENKOVIC, I. January effect on stock returns: Evidence from emerging Balkan equity markets. **Industrija**, Belgrado, v. 45, n. 4, p. 7–21,dez.2017.
- AVELAR, E. A.; OREFICI, P. J.; BORGES, L. S.Inteligência artificial e previsão de preços de ativos financeiros: uma revisão sistemática. **Sistemas & Gestão**, Belgradov. 17, n. 3, p-271-285, dez. 2022.

AVRAMOV, D.; CHORDIA, T.; JOSTOVA, G.; PHILIPPOV, A. Anomalies and financial distress. **Journal of Financial Economics**, Amsterdam, v. 108, n. 1, p. 139–159, abr. 2013.

AWAN, M. J.; RAHIM, M. S. M.; NOBANEE, H.; MUNAWAR, A.; YASIN, A.; ZAIN, A. M. Social Media and Stock Market Prediction: A Big Data Approach. **Computers, Materials & Continua**, Amsterdam, v. 67, n. 2, p. 2569–2583, abr. 2021.

AYCEL, Ü.; SANTUR, Y. A new moving average approach to predict the direction of stock movements in algorithmic trading. **Journal of New Results in Science**, Tokat, v. 11, n. 1, p. 12-25, abr. 2022.

B3. **Índice Ibovespa (Ibovespa B3)**. São Paulo, 2019. Disponível em: [https://www.b3.com.br/pt\\_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/perfil-pessoas-fisicas/perfil-pessoa-fisica/#:~:text=Em%202018%2C%20o%20n%C3%BAmero%20de](https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/perfil-pessoas-fisicas/perfil-pessoa-fisica/#:~:text=Em%202018%2C%20o%20n%C3%BAmero%20de). Acesso em: 19 dez. 2023.

B3. **Índice Ibovespa (Ibovespa B3)**. São Paulo, 2021. Disponível em: [https://www.b3.com.br/pt\\_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm](https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm). Acesso em: 06 dez. 2023.

BADOLIA, L. **How can I get started investing in the stock market**. New Delhi, Educreation Publishing, 2016.

BARROS, M. A.; MENDES, M. S.; LEITE, P. A. M. A relação entre o sentimento do investidor e acurácia na previsão dos analistas. *In*: CONGRESSO ANPCONT, 14, 2020, Foz do Iguaçu. Anais[...]. São Paulo: Anpcont, 2020, p.1-15.

BAYAR, A.; KAN, O. B. Day of the Week Effects: Recent Evidence from Nineteen Stock Markets. **Central Bank Review**, Ankara, v. 2, n. 2, p. 77–90, fev. 2002.

BENNETT, C.; STEWART, R. A.; LU, J. Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks. **Energies**, Basel, v. 7, n. 5, p. 2938-2960, abr. 2014.

BHALLA, V. K. **Investment management: Security analysis and portfolio management**. New Delhi: S. Chad and Company Ltd, 2017.

BHANDARI, H. N.; RIMAL, B.; POKHREL, R. N.; RIMAL, R.; DAHAL, R. K.; KHATRI, C. K. R. Predicting stock market index using LSTM. **Machine Learning with Applications**, Amsterdam, v. 9, n. 100320, p. 100320, set. 2022.

BIERENS, H. J. Comment on artificial neural networks: an econometric perspective. **Econometric Reviews**, Londres, v. 13, n. 1, p. 93-97, mar. 1994.

BLUME, L.; EASLEY, D.; O'HARA, M. Market Statistics and Technical Analysis: The Role of Volume. **Journal of Finance**, Pensilvânia, v. 49, n. 1, p. 153–81, mar. 1994.

BOLLAERT, H.; LOPEZ, S. F.; SCHWIENBACHER, A. Fintech and Access to Finance. **Journal of Corporate Finance**, Amsterdam, v. 68, p. 101941, jun. 2021.

- BONE, R. B.; RIBEIRO, E. P. Eficiência fraca, efeito dia-da-semana e efeito feriado no mercado acionário brasileiro: uma análise empírica sistemática e robusta. **Revista de Administração Contemporânea**, São Paulo, v. 6, n.1, p. 19-37, abr.2002.
- BONESSO, D. **Estimação dos Parâmetros do Kernel em um Classificador SVM na Classificação de Imagens Hiperespectrais em uma Abordagem Multiclasse**. 2013.Dissertação (Mestrado em Sensoriamento Remoto) – Programa de Pós-Graduação em Sensoriamento Remoto, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.
- BOUMAN, S.; JACOBSEN, B. The Halloween Indicator, Sell in May and Go Away: Another Puzzle. **The American economic review**, Pittsburch, v. 92, p. 1618-1635, dez.2002.
- BRAGA, B.M.; SISCOUTO, R.A.; ALMEIDA, L.; SILVA, F. A. Machine learning aplicado em ações no mercado financeiro B3. **Colloquium Exactarum**, São Paulo, v. 14, n. 1, p. 57-66, jul. 2022.
- BURNHAM, K. P.; ANDERSON, D. R. Multimodel Inference. **Sociological Methods & Research**, Londres, v. 33, n. 2, p. 261-304, nov.2004.
- CAMELO, H. N. **Modelos híbridos estocástico - matemático para previsão de velocidade do vento**. 2019.Tese (Doutorado em Ciências Climáticas) – Programa de Pós-Graduação em Ciências Climáticas, Universidade Federal do Rio Grande do Norte, Natal, 2019.
- CAO, Q.; LEGGIO, K. B.; SCHNIEDERJANS, M. J. A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market. **Computers & Operations Research**, Sheffield, v. 32, n. 10, p. 2499–2512, out.2005.
- CAPORALE, G. M.; GIL-ALANA, L.; PLASTUN, A.; MAKARENKO, I. Intraday Anomalies and Market Efficiency: A Trading Robot Analysis. **Comput Econ**, Berlim, v. 47, n. 2, p. 275–295, jan.2016.
- CARAMICO, A. Proposição de um modelo preditivo do Ibovespa utilizando redes neurais artificiais. **Revista CAD**. São Paulo, v 7, n. 1. p. 21-42, dez.2012.
- CAVALCANTE, R. C.; BRASILEIRO, R. C.; SOUZA, V. L.F; NOBREGA, J. P. Computational Intelligence and Financial Markets: A Survey and Future Directions. **Expert Systems with Applications**, Amsterdam, v. 55, p. 194–211, ago. 2016.
- CAVALHEIRO, E. A.; VIEIRA, K. M.; CERETTA, P. S.; CORREA, J. C. S.; CUNHA, C. F.O. Aplicação de Redes Neurais Polinomiais na Previsão do Ibovespa e Merval. **Desenvolvimento em Questão**, São Paulo, v. 9, n.18. p. 196-224, set. 2011.
- CENGIZ, H.; BILEN, O.; BUYUKLU, A. H.; DAMGACI, G. Stock market anomalies: the day of the week effects, evidence from Borsa Istanbul. **J Globentreppres**, Heidelberg, v. 7, n. 4, p. 1-11, dez.2017.
- CERETTA, P. S.; COSTA JÚNIOR, N. C. A. Efeito dia da semana: evidência na América Latina. **Teoria e Evidência Econômica**, Passo Fundo, v. 8, n. 14, p. 27-35, mai.2000.

- CHAKRABARTY, A.; DUBEY, R. A flexible approach towards multi-frequency reengineering of the Moving average convergence divergence indicator. **Global Journal of Flexible Systems Management**, Toronto, v. 15, [s.n], p. 219-234, mai.2014.
- CHANDRA, R.; CHAND, S. Evaluation of co-evolutionary neural network architectures for time series prediction with mobile application in finance. **Applied Soft Computing**, Amsterdam, v. 49, [s.n], p. 462–473, set. 2016.
- CHAPELE, O.; VAPNIK, V.; BOUSQUET, O.; MUKHERJEE, S. Mukherjee. Choosing multiple parameters for support vector machines. **Machine Learning**, Alphen an den Rijnv. 46, n. 1-3, p. 131-159, jan.2002.
- CHHAJER, P.; SHAH, M.; KSHIRSAGAR, A. The applications of artificial neural networks, support vector machines, and long-short term memory for stock market prediction. **Decision Analytics Journal**, Amsterdam, v.5, [s.n], p. 100015, nov.2022.
- CHONG, E.; HAN, C.; PARK, F. C. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. **Expert Systems with Applications**, Amsterdam, v. 83, [s. n], p. 187-205, out.2017.
- CHOUDHRY, R; GARG, K. A Hybrid Machine Learning System for Stock Market Forecasting. World Academy of Science, **Engineering and Technology**, Trieste, v. 2, n. 3, p. 689-692, jan.2008.
- CHU, H. H.; CHEN, T. L.; CHENG, C. H.; HUANG, C. C. Fuzzy dual-factor time-series for stock index forecasting. **Expert Systems with Applications**, Amsterdam, v. 36, n. 1, p. 165-171, jan.2009.
- CINTRA, R. A.; MELO, M. I. P.; BUENO FILHO, J. S. S. Modelos de séries temporais para a previsão da temperatura média mensal de Lavras, MG. **Sigmae**, Alfenas, v. 8, n. 2, p. 596-605, jun. 2019.
- COHEN, G. Trading crypto currencies using algorithmic average true range systems. **Journal of Forecasting**, Amsterdam, v. 42, n. 2, p. 212-222, ago.2023.
- CONSTANT, P. A. de. **Previsão do Ibovespa utilizando modelos híbridos**. 2006. Dissertação (Mestrado em Engenharia Elétrica) –Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006. Disponível em: <http://pee.ufrj.br/teses/textocompleto/2006042001.pdf>. Acesso em: 6 dez. 2023.
- CONTRERAS, J.; ESPÍNOLA, R.; NOGALES, J. F.; CONEJO, J. A. ARIMA models to predict next-day electricity prices. **IEEE Transactions on Power Systems**, Nova Jersey, v. 18, n. 3, p. 1014–1020, ago. 2003.
- COSTA JÚNIOR, N.C.A. Sazonalidades do Ibovespa. São Paulo. **Revista de Administração de Empresas**, São Paulo, v.30, n.3, p.79-84, set. 1990.
- COUTO, V. D; FABIANO, D; RIBEIRO, K. C. S. Gestão Financeira de Curto prazo: Uma Análise do Comportamento das Empresas do Ibovespa à Crise Econômica. **Revista de Contabilidade do Mestrado em Ciências Contábeis da UERJ**, Rio de Janeiro, v. 16, n. 3, p. 46 -64, jun. 2011.

- DAVID, H.; DANLING, J. D.; MENG, D. Y. Mood beta and seasonalities in stock returns. **Journal of Financial Economics**, Amsterdam, v. 137, [s. n], p. 272-295, jul. 2020.
- DEGUTIS, A.; NOVICKYTĖ, L. The Efficient market hypothesis: A critical review of literature and methodology. **Ekonomika**, Lithuania v. 93, n. 2, p. 7-23, jan.2014.
- DEINWALLNER, R. U. Average: True Range: High Volatility as a Success Factor for trading. **International Journal of Recent Scientific Research**, Hanôver, v. 11, n. 1, p. 36805-36812, out. 2020.
- DEMİREL, U.; ÇAM, H.; ÜNLÜ, R. Predicting stock prices using machine learning methods and deep learning algorithms: The sample of the Istanbul Stock Exchange. **Gazi University journal of science**, New York, v.1, n.1, p. 63-88, jul. 2020.
- DICHTL, H.; DROBETZ, W. Are stock markets really so inefficient? The case of the Halloween Indicator. **Finance Research Letters**, Amsterdam v. 11, [s.n], p. 112-121, jan.2014.
- DILER, A, I. Predicting direction of ISE national-100 index with back propagation trained neural network. **Journal of Istanbul stock Exchange**, Orebro, v. 7, n.25-26, p. 65-82, jun.2003.
- DIXIT, A.; JAIN, S. Effect of stationarity on traditional machine learning models: Time series analysis. **Thirteenth international conference on contemporary computing**, v. 16, n. 2, p. 303-308, ago.2021.
- DROGALAS, G.; ATHIANOS, S.; BAKAS, G.; GEORGE, E. Seasonalities in Stock Markets: The Day of the Week Effect. **SSRN Electronic Journal**, v.1, [s.n], out. 2007.
- EMILIANO, P. **Fundamentos dos Critérios de Informação: Akaike e Bayesiano**. 2009. Dissertação (Mestrado em Agropecuária) – Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Universidade Federal de Lavras, Lavras, 2009.
- FAMA, E. Efficient capital markets: A review of Theory and Empirical Work. **The Journal of finance**, Amsterdam, v. 25, n. 2, p. 383-417, mai.1970.
- FAMA, F. E.; FRENCH, K. Dividend yields and expected stock returns. **Journal of Financial Economics**, New York, v. 25, n. 2, p. 3-25, out. 1988.
- FAMA, R.; CIOFFI, P. L. M. de; COELHO, P. A. R. Contexto das finanças comportamentais: anomalias e eficiência do mercado de capitais brasileiro. **Revista de Gestão USP**, São Paulo, v. 15, n. abr., p. 65-78, jun. 2008.
- FARMER, J. D.; GILLERMOT, L.; LILLO, F.; SZABOLCS, M.; SEM, A. What really causes large price changes? **Quantitative Finance**, Nova York, v. 4, n. 4, p. 383-397, abr.2004.
- FRANKFURTER, M. F.; MCGOUN, E. G. Anomalies in finance What are they and what are they good for? *International Review of Financial Analysis*. **Department of Management**, Bucknell University, Lewisburg, v.17, n.10, p.407- 429, dez.2001.

GARTNER, A. Análise da reação das ações do setor financeiro brasileiro as divulgações da Taxa Selic ocorridos entre 2004 e 2011. **Revista Finanças Aplicadas**, São Paulo, v. 1, n. 1, p. 1-12, dez.2011.

GIBBONS, M. R; HESS, P. Day of the Week Effects and Asset Returns. **The Journal of Business**, Chicago, v. 54, n. 4, p. 579–596, out.1981.

GODOY, T. Mercado de Ações: passado, futuro e função. **O Eco da graduação**, Brasília, v. 7, n. 1, p. 1-16, out.2022.

GRAHAM, B. **O Investidor inteligente**. Harper Collins. Rio de Janeiro,2015.

GUERINO, A.P. F.; MICHELE, G. Cash payment anomalies and money laundering: An econometric analysis of Italian municipalities. **International Review of Law and Economics**, Amsterdam, v. 56, p. 105–121,dez. 2018.

GUJARATI, D. N.; PORTER, D. C. **Econometria básica**. 5. ed. Porto Alegre: AMGH, 2011.

HAGGARD, K. S; WITTE, H. D. The Halloween effect: trick or treat? *International Review of Financial Analysis*, Amsterdam, v. 19, n. 5, p. 379-387, dez. 2010.

HAIR, F. J.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados**. Tradução: Adonai Schlup, 6.ed. Porto Alegre: Bookman 2009.

HANAFIZADEH, P; KIM, S. Digital Business: A new forum for discussion and debate on digital business model and digital transformation. **Digital Business**, Amsterdam, v. 1, n. 1, p. 100006, set. 2020.

HARI, Y.; DEWI, P. L. Forecasting System Approach for Stock Trading with Relative Strength Index and Moving Average Indicator. **Journal of Telecommunication, Electronic and Computer Engineering**, Batu Pahat, v. 10, n. 2-3, p. 1-5,mai. 2018.

HENRIQUE, B. M; SOBREIRO, V. A; KIMURA, H. Literature review: Machine learning techniques applied to financial market prediction, **Expert Systems with Applications**, Amsterdam, v. 124, [s. n], p. 226-251,jun.2019.

HILAL, W.; ANDREW G. S.; YAWNEY, J. A Review of Anomaly Detection Techniques and Applications in Financial Fraud. **Expert Systems with Applications**, Amsterdam, v. 193, [s. n], p. 116429, mai.2022.

HYNDMAN, R. J; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2.ed.São Paulo: Os textos, 2018.

JENSEN, M. Some anomalous evidence regarding market efficiency. **Journal of Financial Economics**, Amsterdam, v. 6, n. 2-3, p. 95-101,fev.1978.

JENSEN, N.; SCHMITH, S. Market Responses to Politics Rise of Lula and the Decline of the Brazilian Stock Market. **Comparative Political Studies**, California, v. 38, n. 10, p. 1245-1270,dez. 2005.

JIANG, M.; LIU, J.; ZHANG, L.; LIU, C. An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. **Physica A: Statistical Mechanics and its applications**, Amsterdam, v. 5, n.41, p. 122272,mar. 2020.

- KARA, Y.; BOYACIOGLU, M. A.; BAYKAN, M. K. O. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. **Expert Systems with Applications**, Amsterdam, v. 38, n. 5. p. 5311-5319, mai.2011.
- KARATZOGLOU, A.; MEYER, D; HORNIK, K. Support vector machines in R. **Journal of Statistical Software**, Zurique, v. 15, p. 1-28, abr.2006.
- KEIEL, G.; BENDER, F. A. Modelagem de séries temporais financeiras: uma abordagem estatística para a identificação de modelos de média condicional. **Scientia Cum Industria**, Caxias do Sul, v. 6, n. 1, p. 22-28, mai. 2018.
- KHAIDEM, L.; SAHA, S.; DEY, S.R. R. Predicting the direction of stock market prices using random forest. **Arxiv Preprint Arxiv**: Ithaca, v. 0, n. 0. p. 1-20, abr. 2016.
- KLOCK, S. The January Effect: A Test of Market Efficiency. Longwood University, **Theses & Honors Papers**, 1 jan. 2014.
- KLOSTERMAN, S. Projeto de Ciência de Dados com Pyhton. **Abordagem de estudo de caso para criação de projetos de ciência de dados bem-sucedidos usando python, pandas e scikit-learn**. Tradução: Silva, C. C. A. da. São Paulo: Novatec, 2019.
- KONISH, S.; KITAGAWA, G. **Information Criteria and Statical Modeling**. New York: Springer, 2008.
- KRAUS, M.; FEUERRIEGEL, S. Decision support from financial disclosures with deep neural networks and transfer learning. **Decision Support Systems**, Amsterdam, v. 104, [s. n], p. 38–48, dez. 2017.
- KUMAR, D.; MEGHWANI, S. S.; THAKUR, M. Proximal support vector machine based hybrid prediction models for trend fore casting in financial markets. **Journal of Computacional Science**, Amsterdam, v. 17, n. 1, p.1-13, nov. 2016.
- KUMBURE, M. M.; LOHRMANN, C.; LUUKKA, P.; PORRAS, J. Machine learning techniques and data for stock market forecasting: A literature review. **Expert Systems with Applications**, Amsterdam, v. 197, n 8, p. 16659, jul. 2022.
- LEITE, G.A.; COLARES, A. F.V. Estudo das Anomalias de Calendário nos Retornos do Ibovespa no Período de 1990 a 2012. *In*: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 24, 2014. Anais[...].Curitiba: ABEPRO, 2014.
- LEMGRUBER, E. F; BECKER, J. L; CHAVES, T. B. S. O efeito fim de semana no comportamento dos retornos diários de índices de ações. *In*: ENANPAD, 12, 1988. Anais[...]. Natal: ANPAD, 1988, p. 873-878.
- LEMOS, M. O.; COSTA, N. C. A. O efeito de sobre-reação no curto prazo no mercado de capitais brasileiro. *In*: ENANPAD, 19, 1995. Anais. João Pessoa: ANPAD, 1995, p.293-309.
- LEUNG, M. T.; DAOUK, H.; CHEN, A. S. Forecasting stock indices: A comparison of classification and level estimation models. **International Journal of Forecasting**, Amsterdam, v. 16, n.2, p. 173-190, jun.2000.
- LINDNER, A.; PITOMBO, C. S. Modelo logit binomial com componentes principais para estimação de preferência por modo de transporte motorizado. **Journal of Transport Literature**, Amsterdam, v.10, n.3, p.5-9, abr.2016.

LOHRMANN, C.; LUUKKA, P. Classification of intraday S&P500 returns with a Random Forest. **International Journal of Forecasting**, Amsterdam, v. 35, n. 1, p. 390-407, mar. 2019.

LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma Introdução as Support Vector Machines. **Revista de Informática Teórica e Aplicada**, Porto Alegre, v. 14, n. 2, p. 43-67, dez. 2007.

LUCEY, B. M.; ZHAO, S. Halloween or January? Yet another puzzle. **International Review of Financial Analysis**, Amsterdam, v. 17, n. 5, p. 1055-1069, dez. 2008.

LUCHESE, A. L. B.; ARAÚJO, M. P.; WALTER, S. A. Aplicação do modelo ARIMA à previsão da soja como ferramenta para tomada de decisão. **Revista Contemporânea**, Rio de Janeiro, v. 3, n. 10, p. 18181-18202, out. 2023.

LYAKINA, M.; KOYUNDZHIYSKA, B.; POPP, J. Technical analysis and its theoretical basis for trading activity management, **Ekonomicko-Manazers Kespektrum**, Zilina, v. 15, n. 2, p. 52-64, dez. 2021.

LYDIA, M.; KUMAR, S. S.; SELVAKUMAR, A. I; KUMAR, G. E. P. Linear and nonlinear autoregressive models for short-term wind speed forecasting. **Energy Conversion and Management**, Amsterdam, v. 112, p. 115-124, mar. 2016.

MACHADO, M. R. R.; GARTNER, I. R; MACHADO, L. S. Relação entre Ibovespa e Variáveis Macroeconômicas: evidências a partir de um Modelo Markov-Switching. **Revista Brasileira de Finanças**, Rio de Janeiro. v. 15, n. 3, p. 435-468, out. 2017.

MANISH, K.; THENMOZHI, M. Support Vector Machines Approach to Predict the S&P CNX NIFTY Index Returns. 10th Capital Markets Conference, **Indian Institute of Capital Markets Paper**, Available at, 2007.

MARQUES, A. F. C. **Desafios na previsão de séries temporais financeiras: O caso da taxa de câmbio EUR/USD**. 2013. Dissertação (Mestrado em Ciências do Trabalho) – Programa de Pós – Graduação em Ciências do Trabalho e Empresa, Universidade de Lisboa, 2013.

Disponível em:

[https://repositorio.ul.pt/bitstream/10451/30341/1/ulfc1211569\\_tm\\_Andr%C3%A9\\_Marques.pdf](https://repositorio.ul.pt/bitstream/10451/30341/1/ulfc1211569_tm_Andr%C3%A9_Marques.pdf). Acesso em: 23 dez. 2023.

MARTINEZ, E. Z.; FRANCISCO, L. N.; PEREIRA, B. B. de. A Curva ROC para testes diagnósticos. **Cadernos Saúde Coletiva**, Rio de Janeiro, v. 11, n. 3, p. 7-31, jun. 2003.

MASOD, N. Predicting Direction of Stock Prices Index Movement Using Artificial Neural Networks: The Case of Libyan Financial Market. **British Journal of Economics, management & trade**, Novi Sad, v. 4, n. 4, p. 597-619, jan. 2014.

MEHER, B.K.; SINGH, M.; BIRAU, R.; ANAND, A. Forecasting stock prices of fintech companies of India using random forest with high-frequency data. **Journal of Open innovation: technology, market, and complexity**, Amsterdam, v. 10, n. 1, p. 100180, mar. 2024.

MOKHTARI, S.; YEN, K. K.; LIU, J. Effectiveness of Artificial Intelligence in Stock Market Prediction based on Machine Learning. **International Journal of Computer applications**, Ithaca, v. 183, n. 7, p. 1-8, jun. 2021.

MONDAL, P.; SHIT, L.; GOSWAMI, S. Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. **International Journal of Computer Science, Engineering and Applications**, Iowa, v. 4, n. 2, p. 13, abr. 2014.

MORAIS, I.; STONA, F.; SCHUCK, G. **Econometria aplicada no EViews**. Porto Alegre: FEE, 2016.

MUSHTAQ, R. Augmented dickey-fuller test. **Econometrics: Mathematical Methods & programming e journal**, Nova York, v. 1, [s. n], ago. 2011.

NABIPOUR, M.; NAVYERI, P.; JABANI, H.; SHAHAB, S.; MOSAVI, A. Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis. **IEEE**, Nova York, v. 8, [s. n]., p- 150199-150212, ago. 2020.

NAKANO, M.; TAKAHASHI, A.; TAKAHASHI, S. Generalized exponential Moving average (EMA) model with particle filtering and anomaly detection. **Expert Sytems with Applications**, Nova York, v. 73, n. 1, p. 187-200, mai. 2017.

NETO, J. A. N.; SAIT, R. Pagamentos de dividendos e persistência de retornos anormais das ações: evidência no mercado brasileiro. **Revista de Administração de Empresas**, Rio de Janeiro, v. 38, n. 2, p. 135-143, jun. 2003.

NTI, I. K.; ADEKOYA, A. F.; WEYORI, B. A. A systematic review of fundamental and technical analysis of stock market predictions. **Artificial Intelligence Review**, Berlim, v. 53, n. 1, p. 3007–3057, ago. 2019.

NTI, I. K; ADEKOYA, A. F.; WEYORI, B. A. A comprehensive evaluation of ensemble learning for stock-market prediction. **Journal of Big data**, Berlim, v. 7, n. 1, mar. 2020.

OLIVEIRA, I. L. A. Computational Intelligence and Financial Markets: A Survey and Future Directions. **Expert Systems With Applications**, Amsterdam, v. 55. n. 1, p. 194-211, ago. 2006.

OLIVEIRA, R. C.; GORLA, M. C.; HEIN, N. Modelo de regressão logística para análise de risco de crédito em uma instituição de microcrédito produtivo orientado. **Iberoamerican Journal of Industrial Engineering**, Florianópolis, v. 7, n. 13, p. 103-122, set. 2015.

PAIXÃO, G. M. M.; SANTOS, B. C.; ARAUJO, R. M.; RIBEIRO, M. H.; MORAES, J. L.; RIBEIRO, A. L. Machine Learning na Medicina. **Revisão e Aplicabilidade. arquivos brasileiros de cardiologia**, Belo Horizonte, v. 118, [s.n], p. 95-102, jan. 2022.

PALIWAL, M.; KUMAR, U. A. A Neural networks and statistical techniques: a review of applications. **Expert systems with Applications**. Amsterdam, v. 36, [s. n], p. 2-17, jan.2009.

PANDYA, B, J.; JALIYA, K. U. Opinion and Technical indicator based optimized deep learning for prediction of stock market. **Indian Journal of Computer Science and Engineering (IJCSE)**, Chennai, v. 12, n. 6, p. 1860-1874,dez. 2021.

PEACHAVANISH, R. Stock selection and trading based on cluster analysis of trend and momentum indicators. **In proceedings of the international multi conference of engineers and computer scientists**, Hong Kong, v. 1, p. 317-321, mar.2016.

PLASTUN, A.; MAKARENKO, I.; KHOMUTENKO, L.; SHCHERBAK, S.; TRYFONOVA, O. Exploring price gap anomaly in the Ukrainian stock market. Invest. Manag. **Financ. Innov**, Sumy, v. 16, n.2, p. 150-158, jun.2019.

PLASTUN, A; SIBANDE, X; GUPTA, R; WO HAR, E. M. Price Gap Anomaly in the US Stock Market: The Whole Story. **The North American journal of economics and finance**, Amsterdam, v. 52, p. 1-17,abr. 2020.

POTIN, S. A.; POTIN, S.; CUNHA, P. M. C.; BORTOLON, M. P. Efeito Janeiro nas ações e adrs de empresas brasileiras após o início da tributação de ganhos de capital. **Revista Eletrônica de Administração**, Porto Alegre, v. 21, n. 2, p. 320–347, ago.2015.

POTTER, M. A.; JONG, K. A cooperative coevolutionary approach to function optimization, Of Lecture Notes in Computer Science, **Springer**, Berlin, v. 866,[s.n], p. 249–257,out. 1994.

PRESS, W.H.; TEVKOLSKY, S.A.; VETTERLING, W.T.; FLANNERY, B. P. **Numerical recipes: the art of Scientific Computing**. 3.ed. New York: Cambridge University Press, 2007.

QIU, M.; SONG, Y.; AKAGI, F. Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. **Chaos, solitons & fractals**, Amsterdam, v. 85, p-1-7,abr.2016.

RAHMAN, M.; SA, L. C.; MASUD, K. Predicting Firms' Financial Distress: Na Empirical Analysis Using the F-Score Model. **Journal of Risk and Financial Management**, Basel, v. 14, n. 5, mai.2021.

RAMOS, A. L. *et al.* Evaluation of an iron ore price forecast using a geometric Brownian motion model. **REM. International Engineering Journal**, Ouro Preto, v. 72, n. 1, p. 9-15, mar. 2019.

RAŞO, H.; DEMIRCI, M. Predicting the Turkish Stock Market BIST 30 Index using Deep Learning. **Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi**, Ankara, v. 11, n. 1, p. 253–265,jan. 2019.

RTAYLI, N.; ENNEYA, N. Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization. **Journal of Information security and applications**, Amsterdam, v. 55, [s.n], p. 102596,dez. 2020.

SAFFER, M.; KEVIN, S. A study on market anomalies in Indian stock market. **Int. J. Bus. Admin. Res.** Bangalore, v. 1, n. 3, p. 128-137, jan. 2014.

SAHIN, S.; TOPALOGLU, E. E.; EGE, I. January Effect Revisited: Evidence from Borsa Istanbul and Bucharest Stock Exchange. **International Journal of economics and finance**, Ontário, v. 10, n. 1, p. 159, dez. 2017.

SANTOS, O. J; MUSSA. A; TROVÃO. R; SILVA, C. R. O. R. Anomalias do mercado acionário: a verificação do efeito segunda-feira no IBOVESPA, no período de 1995 a 2006. *In: ENCONTRO BRASILEIRO DE FINANÇAS*, 2007. São Paulo. **Anais [...]**. São Paulo: Sociedade Brasileira de Finanças, 2007, p.45-54.

SATURNINO, O.; LUCENA, P.; SATURNINO, V. Liquidez e valor no mercado de ações brasileiro: Modelo de cinco fatores. **Revista Eletrônica de Administração**, Porto Alegre, v. 23, n. 2, p. 191–224, ago.2017.

SCHULMERICH, M.; LEPRCHER, Y.M.; EU, C. K. **Stock market anomalies, in applied Asset and risk management**. Berlim: Springer,2015.

SHAIKH, A. A.; SHARMA, R.; KARJALUOTO, H. Digital innovation & enterprise in the sharing economy: An action research agenda. **Digital business**, Amsterdam, v. 1, n. 1, p. 100002, set.2021.

SHARMA, A.; KUMAR, A.; VAISH, K, A. Market anomalies and investor behaviour. **Afro-Asian journal of finance and accounting**, Genebra, v. 12, n. 1, p. 62-81, mar.2022.

SHARMA, M. S. S.; SINGH, G. Performance Analysis of Statistical and Supervised Learning Techniques in Stock Data Mining. **MDPI and ACS Style**, Basel, v. 3, n. 4, nov. 2018.

SIAMI NAMIN, S.; SIAMI NAMIN, A. Forecasting Economics and Financial Time Series. **ARIMA vs. LSTM**. Cornell University, v.1, [s. n], p. 1-19, mar.2018.

SILVA, E. J.; SILVA, S. L.; MONTEBELLO, L. I. M. Investimentos em tempos de pandemia: uma abordagem utilizando Machine Learning na construção de portfólios. *In: SEMINÁRIOS EM ADMINISTRAÇÃO*, 23, 2020, São Paulo. Anais[...], São Paulo: Semead, 2020, p.45-54.

SILVA, I. N.; SPATTI, H. D.; FLAUZINO, A. R. **Redes neurais artificiais: para engenharia e ciências aplicadas. Fundamentos teóricos e práticos**. São Paulo: Artliber, 2010.

SILVA, M. G. M. **Investigação de Memória a longo prazo no mercado acionário brasileiro**. 2019. Dissertação (Mestrado em Modelagem Matemática e Computacional) – Centro de Educação Tecnológica, Universidade Federal de Minas Gerais, Belo Horizonte, 2019.

SILVA, R.; PEREIRA, R. Mercado de ações brasileiro: uma investigação empírica sobre suas relações de longo prazo e de precedência temporal pré-crise de 2008. **Nova Economia**, Belo Horizonte, v. 24, n. 2, p. 1-120, ago. 2014.

SINGH, G. Machine Learning Models in Stock Market Prediction. **International Journal of Innovative Technology and Exploring Engineering**, Bhopal, v. 11, n. 3, p. 18–28, fev.2022.

SONG, Y.; LEE, J. W.; LEE, J. A. Study on novel filtering and relationship between input-features and target-vectors in a deep learning model for stock price prediction. **Applied Intelligence**, Berlim, v. 49, n. 3, p. 897-911, mar. 2019.

STOSIC, D.; STOSIC, D; VODENSKA, I.; STANLEY, E. H.; STOSIC, T. A New Look at Calendar Anomalies: Multifractality and Day-of-the-Week Effect. **Entropy**, Basel, v. 24, n. 4, p. 562, abr.2022.

STRADER, J. T.; ROZYCKI, J. J.; ROOT, H. T.; HUANG, J. Y. H. Machine Learning Stock Market Prediction Studies. Review and Research Directions. **Journal of International Technology and Information Management**, California, v. 28, n. 4, p. 1-22, dez. 2020.

SU, Z.; BAO, H.; CUI, X. B.; CUI, X. The prediction of price gap anomaly in Chinese stock market: Evidence from the dependent functional logit model Finance. **Research Letter**, Amsterdam, v. 47, n. B, p. 1-7, jun. 2022.

SUTHAHARAN, S. Support Vector Machine. In: Machine Learning Models and Algorithms for Big Data Classification. **Integrated Series in Information Systems Springer**. Boston, v. 36, n.6 p. 207-235, ago. 2016.

TACHIZAWA, T; MENDES, G. **Como fazer monografia na prática**. Rio de Janeiro: FGV, 2003.

TRIPPI, R. R.; DESIENO, D. Trading equity index futures with a neural network. **Journal of Portfolio Mngement**, Nova York, v. 19, n. 1, p. 27-33, out.1992.

TSANG, H. W. W.; CHONG, L. T. T. Profitability of the On-Balance Volume Indicator. **Economics Bulletin**, Orebro, v. 28, n. 3, p-1-8, set.2009.

- VIDAL-GARCÍA, J.; VIDAL, M. Financial Market Anomalies. **SSRN**, Nova York, v.17, p. 1-17, mai.2023.
- WAMKAYA, B. W.; LAWRENCE, M. ANN model to predict stock prices at stock exchange markets. **Arxiv preprint arxiv**, Chicago,v.1, [s.n], fev. 2015.
- WHITE, H. Economic prediction using neural networks: The case of IBM daily stock returns. **IEEE International Conference on Neural Networks**, San Diego, v. 2, p.451-458,jul. 1988.
- WONG, W. K; MCALEER, M. Mapping the Presidential Election Cycle in US stock markets. **Mathematics and Computers in Simulation**, Amsterdam, v. 79, n. 11, p. 3267-3277,jul. 2009.
- WOO, K.Y; MAI, C; MCALEER, M; WONG, W. K. Review on Efficiency and Anomalies in Stock Markets. **Economies**, Basel, v. 8, n. 1, p. 1-51, mar.2020.
- XU, X.; ZHOU, C.; WANG, Z. Credits coring algorithm based on link analysis ranking with support vector machine. **Expert Systems with Applications**, Amsterdam, v. 36, n. 2, p. 2625-2632, mar.2009.
- YANG, B.; JIANKUN, H.; SICHANG, Z. Stock Market Prediction Using Artificial Neural Networks. **Advanced Engineering Forum**, Bach, v. 6-7, p 1055-1060, set.2012.
- YEH, C. C.; CHI, D. J.; LIN, Y. R. Going-concern prediction using hybrid random forests and rough set approach. **Information sciences**, Amsterdam, v. 254, n.6, p. 98-110, jan.2014.
- YU, P, YAN, X. Stock price prediction based on deep neural networks. **Neural Comput & Applic**, Londres, v. 32, [s.n], p-1609-1628, abr.2020.
- YUHONG, Y. Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. **Biometrika**, Oxford, v. 92, n. 4, p. 937-950,dez. 2005.
- ZADA, H.; HASSAN, A.; WONG, W. K. Do jumps matter in both equity market returns and integrated volatility: A comparison of asian developed and emerging markets. **Economies**, Basel, v. 9, n. 2, jun.2021.
- ZHANG, W.; LIN, S.; ZHANG, Y. Intraday Market-Wide Ups/Downs and Returns. **Journal of Management Science and Engineering**, Novi Sad, v. 1, n. 1, p. 28-57, dez. 2016.
- ZHANG, Y.J.; WANG, J. L. Do high-frequency stock market data help forecast crude oil prices? Evidence from the MIDAS models. **Energy Economics**, Amsterdam, v. 78, p. 192-201,fev. 2019.
- ZHANG, Z.; SUN, W.; WANG, H. A new perspective on financial anomalies in emerging markets: the case of China. **Applied financial economics**, Londres, v. 18, n. 21, p. 1681-1695, nov.2008.