



**UEPB**

**UNIVERSIDADE ESTADUAL DA PARAÍBA  
CAMPUS I - CAMPINA GRANDE  
CENTRO DE CIÊNCIAS E TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA  
CURSO DE GRADUAÇÃO EM ESTATÍSTICA**

**DAMIÃO EVERTON GOMES MAIA**

**UTILIZAÇÃO DO APRENDIZADO NÃO SUPERVISIONADO VIA *K-MEANS* PARA  
TOMADAS DE DECISÃO NO MERCADO FINANCEIRO**

**CAMPINA GRANDE - PB  
2023**

DAMIÃO EVERTON GOMES MAIA

**UTILIZAÇÃO DO APRENDIZADO NÃO SUPERVISIONADO VIA *K-MEANS* PARA  
TOMADAS DE DECISÃO NO MERCADO FINANCEIRO**

Trabalho de Conclusão de Curso (Artigo) apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

**Orientador:** Prof. Dr. Tiago Almeida de Oliveira  
**Coorientador:** Prof. Me. Ednário Barbosa de Mendonça

**CAMPINA GRANDE - PB  
2023**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

M217u Maia, Damião Everton Gomes.  
Utilização do aprendizado não supervisionado via *k-means* para tomadas de decisão no mercado financeiro [manuscrito] / Damiao Everton Gomes Maia. - 2023.  
32 p. : il. colorido.

Digitado.  
Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2023.  
"Orientação : Prof. Dr. Tiago Almeida de Oliveira, Coordenação do Curso de Estatística - CCT. "

1. K-means. 2. Aprendizado não supervisionado. 3. Bolsa de valores. 4. Mercado financeiro. I. Título

21. ed. CDD 332.642

DAMIÃO EVERTON GOMES MAIA

UTILIZAÇÃO DO APRENDIZADO NÃO SUPERVISIONADO VIA *K-MEANS* PARA  
TOMADAS DE DECISÃO NO MERCADO FINANCEIRO

Trabalho de Conclusão de Curso (Artigo) apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

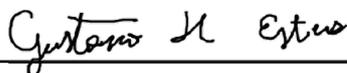
Trabalho aprovado em 13 de março de 2023.

**BANCA EXAMINADORA**



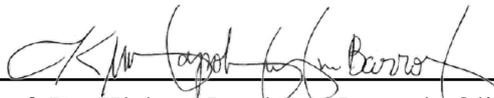
---

Prof. Dr. Tiago Almeida de Oliveira  
Universidade Estadual da Paraíba (UEPB)



---

Prof. Dr. Gustavo Henrique Esteves  
Universidade Estadual da Paraíba (UEPB)



---

Prof. Dr. Kleber Napoleão Nunes de Oliveira  
Barros  
Universidade Federal Rural de Pernambuco  
(UFRPE)

“O senhor é o meu pastor, nada me faltará.”  
(Salmos 23:1)

“Amadurecer e evoluir traz paz a alma,  
porque você para de se estressar  
com coisas desnecessárias..”  
(Marianna Moreno)

## LISTA DE ILUSTRAÇÕES

Figura 1 – Número ideal de <i>clusters</i> das ações . . . . .	21
Figura 2 – Gráfico dos retornos acumulados nas bases das ações Cielo . . . . .	24
Figura 3 – Gráfico dos retornos acumulados nas bases das ações CSN . . . . .	24
Figura 4 – Gráfico dos retornos acumulados nas bases das ações Fleury . . . . .	25
Figura 5 – Gráfico dos retornos acumulados nas bases das ações Santander . . . . .	25
Figura 6 – Gráfico dos retornos acumulados nas bases das ações Usiminas . . . . .	26

## LISTA DE TABELAS

Tabela 1	– Valores médios dos indicadores utilizados nas ações conforme os <i>clusters</i> . . . . .	22
Tabela 2	– Estatísticas utilizadas nas ações conforme os <i>clusters</i> . . . . .	23
Tabela 3	– Número de observações classificadas em cada <i>cluster</i> ao longo do tempo. . . . .	23
Tabela 4	– Retornos acumulados na base de treino . . . . .	26
Tabela 5	– Retornos acumulados na base de teste . . . . .	26

## LISTA DE ABREVIATURAS E SIGLAS

CSN	Companhia Siderúrgica Nacional
RSI	<i>Relative Strength Index</i> (Índice de Força Relativa)
MACD	<i>Moving Average Convergence Divergence</i> (Histograma Média Móvel Convergente e Divergente)
BB	<i>Bollinger Bands</i> (Bandas de Bollinger)
ROC	<i>Rate Of Change</i> (Taxa de Variação de Preço)
BPNN	<i>Back Propagation Neural Networkse</i> (Redes Neurais de Retropropagação)
SVM	<i>Support Vector Machine</i> (Máquina de Vetores de Suporte)

## SUMÁRIO

1	INTRODUÇÃO . . . . .	9
2	FUNDAMENTAÇÃO TEÓRICA . . . . .	10
2.1	Mercado Financeiro . . . . .	10
2.2	Indicadores de Análise Técnica . . . . .	12
2.3	Aprendizado de Máquinas . . . . .	13
2.4	Análise de <i>Cluster</i> . . . . .	14
2.5	Medidas de Semelhança e Distância . . . . .	15
2.6	Métodos Hierárquicos . . . . .	16
2.6.1	<i>Métodos Aglomerativos</i> . . . . .	17
2.6.2	<i>Métodos Divisivos</i> . . . . .	18
2.7	Métodos Não Hierárquicos . . . . .	18
2.7.1	<i>Método k-means</i> . . . . .	19
2.7.2	<i>Método k-medoid</i> . . . . .	19
3	MATERIAIS E MÉTODOS . . . . .	20
3.1	Coleta de dados . . . . .	20
3.2	Ações e Indicadores de Análise Técnica . . . . .	20
3.3	Método e Pacotes Utilizados . . . . .	20
4	RESULTADOS E DISCUSSÃO . . . . .	20
5	CONSIDERAÇÕES FINAIS . . . . .	27
	REFERÊNCIAS . . . . .	28

## UTILIZAÇÃO DO APRENDIZADO NÃO SUPERVISIONADO VIA *K-MEANS* PARA TOMADAS DE DECISÃO NO MERCADO FINANCEIRO

Damião Everton Gomes Maia\*  
Tiago Almeida de Oliveira†  
Ednário Barbosa de Mendonça‡

### RESUMO

Este trabalho objetivou aplicar as técnicas de aprendizado não supervisionado utilizando o método baseado no algoritmo *k-means* no mercado financeiro, visando melhorar a tomada de decisão de compra e venda de ações. Para tal, foram coletados dados referentes ao período de 04/01/2010 até 31/01/2022 das ações da Cielo, Companhia Siderúrgica Nacional (CSN), Fleury, Santander e Usiminas. Os procedimentos estatísticos, cálculos e gráficos elaborados no trabalho foram executados no *software* R 4.1.3. Para as análises foram usados variados pacotes, os quais destacam-se os seguintes: o *cluster* (2019) para análise de *cluster*, *factoextra* (2019) para visualizar a saída da análise de dados multivariados e o *PerformanceAnalytics* (2020) para realizar a avaliação de ativos financeiros. O método *k-means* foi comparado com o *buy and hold*, por meio dos resultados dos retornos acumulados nas bases de treino e teste. O resultado obtido foi que o algoritmo alcançou um desempenho superior do que o *buy and hold*. Consequentemente, é mais eficaz utilizar o método *k-means* do que a técnica *buy and hold* para a tomada de decisão a respeito da compra e venda das ações.

**Palavras-chaves:** *k-means*. aprendizado não supervisionado. bolsa de valores. mercado financeiro.

### ABSTRACT

This work aimed to apply unsupervised learning techniques using the method based on the *k-means* algorithm in the financial market, aiming to improve decision-making for buying and selling stocks. To this end, data were collected for the period from 01/04/2010 to 01/31/2022 on the shares of Cielo, Companhia Siderúrgica Nacional (CSN), Fleury, Santander and Usiminas. The statistical procedures, calculations and graphs elaborated in the work were executed in the R 4.1.3 software. For the analyses, several packages were used, of which the following stand out: the *cluster* (2019) for cluster analysis, *factoextra* (2019) to visualize the output of the multivariate data analysis and *PerformanceAnalytics* (2020) to carry out the evaluation of financial assets. The *k-means* method was compared with the *buy and hold* method, through the results of the accumulated returns on the training and test bases. The result obtained was that the algorithm achieved a superior performance than *buy and hold*. Consequently, it is more effective to use the *k-means* method than the *buy and hold* technique for making decisions regarding the purchase and sale of shares.

**Keywords:** *k-means*. unsupervised learning. stock Exchange. financial market.

---

\* Aluno Damião Everton Gomes Maia do curso de Bacharelado em Estatística, Depto de Estatística, UEPB, Campina Grande, PB, evertonmaia1234567890@gmail.com

† Prof. Dr. Tiago Almeida de Oliveira, Depto de Estatística, UEPB, Campina Grande, PB, tado-live@servidor.uepb.edu.br

‡ Prof. Me. Ednário Barbosa de Mendonça, Campina Grande, PB, ednariobm@gmail.com

## 1 INTRODUÇÃO

Um dos principais responsáveis pelo desenvolvimento do país é o mercado financeiro, uma vez que estimula a poupança e o investimento produtivo, o que é relevante para o crescimento de qualquer sociedade econômica moderna (CVM, 2014). Esse mercado tem como uma de suas divisões o mercado de capitais que é formado pelas bolsas de valores e corretoras, sendo bem eficaz em captar poupança e direcioná-la para atividades mais produtivas (BM&FBOVESPA, 2010).

No entanto, devido ao mercado de capitais ser dinâmico e os seus eventos mudarem a cada instante, o ganho pode se converter em perda rapidamente (FREITAS et al., 2018). Sendo assim, a predição do comportamento da bolsa de valores se faz essencial para os atuantes no mercado. Segundo Lima (2016), prever esse comportamento promove um diferencial estratégico que pode corresponder a ganhos significativos, por isso os profissionais da área estudam diversos tipos de indicadores que ajudam no processo de tomada de decisão e diminuição de incertezas quanto ao mercado.

Alguns desses indicadores são: a) *Relative Strength Index* (RSI) que é utilizado para medir a magnitude das mudanças recentes de preços para avaliar as condições de sobrecompra ou sobrevenda no preço de uma ação ou outro ativo (FERNANDO, 2022); b) Histograma Média Móvel Convergente e Divergente (MACD) que proporciona a visualização da diferença entre a linha MACD e a linha de sinal, em que ele é visível em barras ou gráfico de área. Esse histograma ajuda o investidor a tomar decisões em relação ao mercado acionário, já que mostra alguns pontos de reversão das tendências de comportamento das ações e possibilita a observação dos ativos comprados e vendidos (CAPITALNOW, 2019); c) Percentual das Bandas de *Bollinger* que é usado para indicar se um ativo está sobrecomprado, estável ou sobrevendido. Esse indicador é constituído por duas médias móveis, uma superior e outra inferior que fornece a informação se o valor de uma ação está barata ou cara dentro de um determinado período de tempo (REIS, 2020); e d) *Rate of Change* (ROC) que é um indicador baseado em momento que vai medir a variação percentual no preço, entre o preço atual e o preço de um certo número de períodos anteriores (MITCHELL, 2021).

Além da utilização de indicadores, existem métodos para investimentos em ações como o *buy and hold*. Ele consiste em comprar ações e segurá-las para vender posteriormente fundamentado na concepção que a longo prazo ocorrerá a valorização dessas ações e, dessa maneira, o investidor estará lucrando mais quando decidir vendê-las. Porém, a desvantagem dessa prática é que precisa-se que se tenha paciência e foco no longo prazo o que é difícil, uma vez que os investidores, principalmente iniciantes, costumam preferir lucros imediatos (CAPITALNOW, 2020).

No que se refere a utilização de modelos estatísticos para prever o comportamento da bolsa de valores, Lima (2016) afirma que no decorrer das últimas décadas os modelos estatísticos tradicionais estão obtendo o auxílio de outras abordagens, em particular daquelas que usam inteligência artificial e suas técnicas. Conforme Sutter (2018), os métodos de agrupamento ou clusterização estão presentes no aprendizado de máquinas para finanças, pois devido ao grande volume de dados as análises de predição se tornam complicadas em função da volatilidade e valores discrepantes.

A ciência de dados é um conceito que concilia os conhecimentos de estatísticas, análise de dados e *Machine Learning*. Essa ciência é um campo multidisciplinar que usa ferramentas, processos e algoritmos para obter conhecimento e percepções de dados estruturados e não estruturados. Assim, ela consegue tratar grandes quantidades de dados, além de adicionar métodos da ciência da computação no repertório de estatísticas (FAÇANHA, 2019).

O aprendizado de máquinas é a ciência (a arte) da programação de computadores, em que

eles aprendem com os dados. Esse aprendizado pode ser dividido em aprendizado supervisionado e aprendizado não supervisionado. No aprendizado supervisionado, os dados de treinamento que você fornece ao algoritmo incluem as soluções necessárias, chamadas de rótulos. Já no aprendizado não supervisionado, os dados de treinamento não são rotulados (GÉRON, 2019).

Com relação ao aprendizado não supervisionado, uma tarefa desse aprendizado é a análise de *clusters*. A análise de *clusters* visa encontrar os agrupamentos naturais de um conjunto de padrões, pontos ou objetos. Operacionalmente define-se a clusterização como uma representação de  $n$  objetos para achar  $k$  grupos com base em uma medida de similaridade tal que as semelhanças entre objetos no mesmo grupo sejam altas, enquanto as semelhanças entre objetos em grupos diferentes sejam baixas. Os *clusters*, podem se distinguir em forma, tamanho e densidade, e a presença de ruído nos dados torna ainda mais difícil a sua detecção. Um *cluster* ideal pode ser entendido como um conjunto de pontos compacto e isolado. Constata-se, assim, que um *cluster* é uma entidade subjetiva, cujo significado e interpretação requerem conhecimento e domínio (JAIN, 2010).

O *k-Means* é um algoritmo de clusterização, com ele é possível descobrir conjuntos de dados que se assemelham, constituindo grupos identificados pela proximidade entre os seus valores, os quais por serem muito próximos, acabam por formar um *cluster* (STULP, 2014). O algoritmo *k-Means* usa o conceito da centroide. Dado um conjunto de dados, ele escolhe de maneira aleatória  $k$  registros, em que cada um vai representar um agrupamento (CAMILO; SILVA, 2009).

Logo, justifica-se o presente trabalho diante da relevância que o aprendizado de máquina vem adquirindo para a predição do comportamento da bolsa de valores. Algumas pesquisas sobre o aprendizado de máquinas foram feitas por Lima (2016), Roza (2016), Sutter (2018), Beltran (2019) e Lopes e Gosling (2021), mas existe uma escassez de estudos no que tange o aprendizado não supervisionado via *k-means* para a tomada de decisão na bolsa de valores.

Com base no que foi exposto, surge a presente questão de pesquisa: qual a utilização do aprendizado não supervisionado via *k-means* para tomada de decisão no mercado financeiro?

Portanto, a presente pesquisa tem como objetivo geral aplicar as técnicas de aprendizado não supervisionado (em especial, o algoritmo *k-means*) no mercado financeiro, visando melhorar a tomada de decisão de compra e venda de ações. Os objetivos específicos são: a) modelar, desenvolver e avaliar estratégias de compra e venda de ações; b) testar e avaliar o método *k-means* proposto com o intuito de atingir resultados próximos de testes reais; e c) pesquisar, implementar e analisar diversos indicadores relacionados ao comportamento das ações.

Este estudo pretende contribuir para: a) mostrar que o método *k-means* pode melhorar a eficiência das operações de compra e venda realizadas na bolsa de valores, reduzindo riscos e aumentando o lucro; e b) fomentar a literatura sobre aprendizado de máquinas não supervisionado via *k-means*.

O artigo está organizado em 6 seções: esta primeira é a introdução, na qual se expõe, de partida, uma visão geral do estudo; seguida da fundamentação teórica que abordou os conceitos de aprendizado de máquinas e mercado financeiro, da metodologia que orientou a pesquisa, da análise de resultados, das considerações finais e das referências.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Mercado Financeiro

O mercado financeiro é responsável pela negociação de produtos financeiros, ou seja, refere-se ao sistema por meio do qual ações, títulos públicos, câmbio e outros ativos são comercializados. Esse mercado é composto por três figuras principais que são os emissores,

intermediadores e investidores. Eles são os autores das transações de compra e venda que permite a continuidade do funcionamento do mercado financeiro (EXAME, 2021).

Na concepção de Lima (2016), o mercado financeiro determina a transação entre os agentes econômicos que possuem um excedente de capital e os que necessitam de dinheiro, esse mercado, através do intermédio de um agente financeiro, viabiliza essa negociação entre superavitários e deficitários. Essa operação é relevante para o desenvolvimento econômico, uma vez que possibilita a elevação das taxas de poupança e investimento.

A Comissão dos Valores Mobiliários (CVM, 2014) afirma que o mercado financeiro está segmentado em quatro mercados que são:

- **Mercado monetário:** Ele é usado para controlar a liquidez da economia, no qual o Banco Central tem o papel de intervir para conduzir a Política Monetária. As transferências de recursos que ocorrem nele são de curtíssimo prazo, como aquelas efetuadas entre as próprias instituições financeiras ou entre elas e o Banco Central.
- **Mercado de crédito:** Nele ocorre a captação de recursos pelas instituições financeiras dos agentes superavitários e elas emprestam esses recursos aos agentes deficitários. Assim, as instituições são as intermediárias dessas operações. Normalmente, esse mercado possui operações de curto e médio prazo que são designadas ao consumo ou capital de giro das empresas.
- **Mercado cambial:** Nesse mercado são negociadas as trocas de moedas estrangeiras por moeda nacional. Os participantes dele são todos os agentes econômicos que fazem transações com o exterior. O Banco Central é responsável por regulá-lo e fiscalizá-lo e também executa a Política Cambial.
- **Mercado de capitais:** É o mercado em que os investidores emprestam os seus recursos diretamente para as empresas e, em contrapartida adquirem títulos, que retratam as condições estabelecidas no negócio, chamados de valores mobiliários. Esses títulos podem ser títulos de dívida, que determinam uma relação de crédito entre os investidores e os emissores dos títulos, como por exemplo as debêntures, ou podem ser títulos patrimoniais, ou de capital, em que os investidores se transformam em sócios do negócio, como é o caso das ações. Nesse mercado as instituições financeiras que agem como prestadoras de serviços não possuem responsabilidade pelo cumprimento das obrigações estabelecidas e formalizadas entre as partes envolvidas nas transações.

A estrutura do mercado financeiro é bem complexa. Ela é formada por diversas instituições financeiras, como bancos, corretoras, instituições de pagamento e alguns órgãos do governo. Dentro das instituições financeiras também está incluída a Bolsa de Valores, na qual ocorre as negociações de ações de companhias de capital aberto, derivativos, opções, contratos futuros, cotas de fundos de índice e de fundos imobiliários, entre outros ativos financeiros. No Brasil somente a B3 – Brasil, Bolsa e Balcão é a bolsa de valores que se encontra em operação atualmente (PARMAIS, 2021).

Sutter (2018) aponta que as bolsas de valores oferecem sistemas para a realização de negociação de compra e venda dos ativos financeiros frequentemente presentes no mercado de capitais de maneira transparente e regulamentada. Para Pinheiro (2008), existem três tipos de pessoas que se envolvem ativamente no mercado de negociação das bolsas de valores: os especuladores, que procuram lucrar valendo-se da liquidez e da volatilidade do mercado de ações; os gestores financeiros, que são encarregados por gerir as empresas, captando recursos a baixo

custo e executando investimentos sem riscos; e os investidores, os quais usam o mercado buscando lucro a longo prazo, por meio das transações de compra e venda de títulos ou dividendos.

No que tange as ações, são um título de renda variável que equivale a menor fração do capital de uma entidade e possui potencial de constituir retorno financeiro para seus investidores. Quando um investidor adquire uma ação, ele se torna um coproprietário da empresa e possivelmente é remunerado, dependendo da habilidade da empresa em gerar lucro. O retorno financeiro aos investidores pode acontecer no médio e longo prazo, por meio dos dividendos de uma empresa, ou pela venda de ações a preços mais altos do que foram compradas (SUTTER, 2018).

## 2.2 Indicadores de Análise Técnica

A tomada de decisão na bolsa de valores deve ser fundamentada, por isso é essencial analisar os ativos ou derivativos. Assim, para realizar essa avaliação corretamente, os indicadores de análise técnica são especialmente relevantes. Esses indicadores são ferramentas usadas para avaliar o comportamento dos ativos no intuito de coletar e contextualizar as informações dos padrões gráficos e auxiliam na tomada de decisão. Desta forma, os indicadores técnicos são essenciais na hora de realizar operações de *day trade*, *swing trade* ou de outros períodos. Além disso, o ideal é usar diversos indicadores para ter informações mais íntegras e confirmar movimentos e tendências (BLOG BRA, 2021).

A seguir, descreve-se brevemente cada indicador técnico utilizado neste trabalho.

- **Relative Strength Index (RSI):** O RSI mede a velocidade e mudança do movimento de preço. Calcula-se utilizando o valor da variação dos preços de alta e queda no decorrer de um período de tempo especificado (SILVA, 2015). A fórmula do RSI é:

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

$$RS = \frac{AG}{AL} \quad (2)$$

Em que RS é a média dos períodos de alta (AG - *Average Gain*) dividido pela média dos períodos de baixa (AL - *Average Loss*). Um período de alta ou baixa é definido quando o preço de um ativo passa por uma variação positiva (SILVA, 2015).

- **Histograma Média Móvel Convergente e Divergente (MACD):** MACD é um indicador específico de oscilação no preço. Ele é utilizado nos preços de fechamento de uma ação para identificar tendência, demonstrando a relação entre duas médias móveis (SILVA, 2015). A Média móvel pode ser compreendida como um indicador muito utilizado por quem faz trade, sendo uma relevante forma de fazer análise técnica. Por meio dela, percebe-se o equilíbrio do preço de um ativo no mercado, identificando se há uma tendência de alta, de baixa ou até mesmo do valor do ativo manter-se neutro. Assim, entende-se que o termo “móvel” é dado devido ao fato dos dados estarem em constante atualização conforme o tempo passa (EXAME, 2022).

Em suma, é composto de dois elementos: a linha MACD e a linha de sinal. A linha MACD é constituída pela diferença entre duas médias móveis exponenciais (EMA), uma de curto e outra de longo prazo, frequentemente sendo computadas com 12 e 26 períodos, respectivamente (SILVA, 2015).

$$MACD = EMA[12] - EMA[26] \quad (3)$$

- **Percentual das Bandas de Bollinger:** Percentual das Bandas de *Bollinger* são bandas de volatilidade colocadas acima e abaixo de uma média móvel. Esse indicador tem uma forte ligação com a volatilidade. Logo, quanto maior é a volatilidade de um ativo, maior é seu desvio padrão. As bandas são formadas por um conjunto de três curvas calculadas em relação aos preços. Elas são traçadas a partir de uma determinada distância de uma média móvel (SILVA, 2015). A fórmula das bandas de *bollinger* é a seguinte:

$$UpperBand = SMA[20] + 2 \times SD[20] \quad (4)$$

$$MiddleBand = SMA[20] \quad (5)$$

$$LowerBand = SMA[20] - 2 \times SD[20] \quad (6)$$

No qual SMA é a média móvel simples de 20 períodos, e SD é o desvio padrão (SILVA, 2015).

- **Rate of Change (ROC):** ROC é um indicador que mostra a dinâmica de um ativo como uma porcentagem. Ele é calculado subtraindo-se o preço de um número de períodos anteriores, do preço atual, dividindo-se pelo preço de um número de períodos anteriores, e depois multiplicado por 100 para adquirir uma porcentagem. Esse indicador apresenta a diferença percentual entre o preço atual e o preço de fechamento de  $N$  períodos anteriores. ROC é categorizado como um indicador de dinâmica de preços ou um indicador de velocidade, uma vez que mede a taxa de mudança ou a força de impulso de mudança (SILVA, 2015).

$$ROC = \frac{CP - CPA}{CPA} \times 100. \quad (7)$$

Em que o CP é o preço de fechamento e CPA é o preço de fechamento de  $N$  períodos anteriores (SILVA, 2015).

## 2.3 Aprendizado de Máquinas

O processo de aprendizagem é inerente ao ser humano desde os seus primeiros momentos de vida, o qual ocorre por meio de assimilações dos mais diversos estímulos externos. O aprendizado de máquinas é baseado por essa imitação, só que ele acelera o processo de aprendizagem por repetição no intuito de encontrar padrões de comportamento no que está sendo observado. Quando a máquina obtém uma boa quantidade de informações sobre determinado assunto é razoável que ela seja capaz de reproduzir formas de inteligência (SUTTER, 2018).

Com o avanço da tecnologia houve um aumento do uso dos meios eletrônicos pela sociedade, gerando um crescimento exponencial nas produções de dados. Logo, acontece uma gradativa evolução na ciência de dados e na computação visando armazenar e processar maiores quantidades de dados, e ainda na inferência e em ter algoritmos competentes para solucionar problemas de otimização (SOUSA, 2019).

Diante disso, Sousa (2019) afirma que os pesquisadores de Inteligência Artificial, cientistas de dados e da computação, e outros, empenharam-se para desenvolver uma área atualmente chamada de Aprendizado de Máquinas, que é um sistema habilitado em reconhecer *e-mails* de *spam*, prever quem são os compradores viáveis de um produto específico, resolver problemas de

visão e reconhecimento de fala.

Os filtros de *spam* são exemplos de aprendizado de máquinas, nele pode-se assinalar *e-mails* como spam e como regulares. Assim, os exemplos que o sistema usa para o aprendizado são chamados de conjunto de treinamento e cada exemplo de treinamento é chamado de instância de treinamento ou amostra (GERÓN, 2019).

O Aprendizado de Máquinas pode ser supervisionado ou não supervisionado. O supervisionado possui uma entrada e saída bem definidas que se relacionam entre si, ou seja, existem os dados para analisar e procurar restando a tarefa de aprender o mapeamento desse percurso. Para isso, há a fase de treinamento, em que alguns dados são postos do sistema com a intenção de achar parâmetros ajustáveis aos dados desconhecidos do conjunto de testes, a fim de prever o resultado dado um conjunto de amostras de treinamento, junto com seus rótulos de treinamento. Por sua vez, no aprendizado não supervisionado apenas os dados de entrada são definidos. O seu objetivo é descobrir os padrões ou estruturas existentes no conjunto inicial de dados, ou seja, na entrada. Desse jeito, a análise advém a partir do funcionamento do programa, o que ocorre e não ocorre com os dados. Nesse aprendizado não se possui ideia (ou pouca) do resultado, já que toda a tarefa é elaborada conforme as semelhanças compreendidas pela máquina (SOUSA, 2019).

## 2.4 Análise de *Cluster*

Corrar, Paulo e Dias Filho (2007), definem a análise de *cluster* como um método estatístico que possibilita agrupar elementos, indivíduos, produtos e até mesmo comportamentos de elementos de uma amostra, tendo como base as similaridades e diferenças das características que estes itens possuem. Mingoti (2005) complementa que esse método é exploratório no qual divide-se em grupos uma população (ou amostra) que em muitos casos o número de grupos não é conhecido à priori, porém precisa ser estimado por meio dos dados amostrais observados. Os grupos são estabelecidos de forma a obter-se homogeneidade dentro dos grupos e heterogeneidade entre eles.

No aprendizado não supervisionado existem determinados padrões nos dados de entrada, alguns são mais comuns que outros. A máquina faz o agrupamento desses dados que aparentemente apresentam algo em comum, essa tarefa significa reunir em um *cluster* dados semelhantes entre si, porém cada *cluster* deve ser diferente um do outro. Os seus algoritmos são usados para explorar padrões ou criar grupos. A clusterização ou *clustering* é a análise dos *clusters* que analisa o grau de semelhanças ou diferenças entre eles e seus elementos, utiliza-se uma função de distância, na qual quanto menor a distância mais semelhantes são os dados analisados (SOUSA, 2019). Ou seja, a clusterização é uma tentativa de formar agrupamentos ou associações entre os dados de entrada (SUTTER, 2018).

Ketchen e Shook (1996) concordam que a análise de *clusters* é uma técnica estatística que classifica os dados em conjuntos ou grupos semelhantes. Eles ainda declaram que o uso dessa análise representa um desafio complexo, uma vez que solicita várias escolhas metodológicas que definem a qualidade de uma solução de *cluster*. Desde o final da década de 1970, ela vem sendo utilizada nas pesquisas de gestão estratégica para fazer investigações de questões de relevância central.

Nos últimos anos, essa análise tem incentivado os pesquisadores a explorar novas formas para compreender o comportamento dos dados. Isso é explicado de certa maneira por causa da facilidade computacional desse método e sua habilidade de criar resultados consistentes, mesmo em bases de dados pequenas. Apesar disso, os pesquisadores costumam se equivocar quando sustentam que a clusterização é um território no qual pode tudo, já que a literatura mostra que eles precisam ter cuidado, em particular com o efeito dos *outliers* na formação dos *clusters* (LOPES; GOSLING, 2021).

Os principais propósitos da utilização da análise de *clusters* de dados são: a) estrutura subjacente, ou seja, para conseguir informações sobre os dados, gerar hipóteses, detectar anomalias e identificar características salientes; b) classificação natural por meio do reconhecimento do grau de similaridade entre formas ou organismos (relação filogenética); e c) compressão que é um método que organiza os dados visando resumi-los através de protótipos de *cluster* (JAIN, 2010).

Conforme Reis (2001), a análise de *cluster* pode ter as seguintes fases: selecionar os indivíduos ou uma amostra de indivíduos que serão agrupados, definir um conjunto de variáveis das quais vai se obter a informação necessária ao agrupamento dos indivíduos, estabelecer uma medida de similaridade ou dissimilaridade entre cada dois indivíduos, escolher um critério de agregação ou desagregação dos indivíduos, ou seja, determinar um algoritmo de partição/classificação e validar os resultados obtidos.

## 2.5 Medidas de Semelhança e Distância

O primeiro passo no desenvolvimento de uma análise de *cluster* envolve a definição da medida de distância (dissimilaridade) ou de semelhança (similaridade), que servirão de base para atribuir cada observação a um grupo específico. Quando uma variável do banco de dados é essencialmente uma métrica, geralmente é usada uma métrica de distância, pois quanto maior a diferença entre os valores das duas observações específicas, menor a semelhança entre elas, ou seja, maior a diferença. Já quando a variável é binária, geralmente é usada a medida de semelhança. O importante é a frequência da resposta de convergência de duas observações específicas de 1-1 ou 0-0 pares. Nesse caso, quanto maior a frequência do par convergente, maior a similaridade entre as observações (FÁVERO; BELFIORE, 2017).

### *Medidas de Distância*

Segundo Doni (2004), em geral os métodos de análise de *cluster* pedem uma medida de dissimilaridade entre os elementos que vão ser agrupados, frequentemente expressa por uma função de distância ou métrica. Do ponto de vista de Neto e Moita (1998) nessa análise, a similaridade entre duas amostras pode ser retratada como uma função da distância entre os dois pontos que representam estas amostras no espaço  $n$ -dimensional.

Em seguida são apresentadas algumas distâncias enquadradas na definição de medidas de dissimilaridade.

### *Distância euclidiana*

De acordo com Neto e Moita (1998), a maneira mais comum de calcular a distância entre dois pontos  $p$  e  $q$  no espaço  $n$ -dimensional é a distância euclidiana. Essa distância para Fávero e Belfiore (2017) é definida pela seguinte expressão:

$$d_{pq} = \sqrt{(X_{1p} - X_{1q})^2 + (X_{2p} - X_{2q})^2 + \dots + (X_{kp} - X_{kq})^2} = \sqrt{\sum_{j=1}^k (X_{jp} - X_{jq})^2}. \quad (8)$$

### *Distância quadrática euclidiana*

Para Fávero e Belfiore (2017) distância quadrática euclidiana é outra opção de distância euclidiana que pode ser utilizada quando a variável apresenta uma pequena dispersão de seus

valores, facilitando o uso da distância euclidiana ao quadrado para interpretar os *outputs* da análise e a alocação das observações no grupo. Sua expressão é dada por:

$$d_{qe} = (X_{1q} - X_{1e})^2 + (X_{2q} - X_{2e})^2 + \dots + (X_{kq} - X_{ke})^2 = \sum_{j=1}^k (X_{jq} - X_{je})^2. \quad (9)$$

### ***Distância de Minkowski***

A distância de Minkowski é a expressão mais geral de dissimilaridade que outros derivaram dela. É dado por (FÁVERO; BELFIORE, 2017):

$$d_{fg} = \left[ \sum_{j=1}^k (|X_{jf} - X_{jg}|)^m \right]^{\frac{1}{m}}. \quad (10)$$

Em que  $m$  é um número inteiro e um valor positivo ( $m=1,2,\dots$ ). Quando  $m=2$ , pode-se verificar que a distância euclidiana é um caso especial da distância de Minkowski.

### ***Distância de Manhattan***

A distância de Manhattan demonstra resultados semelhantes ao da distância Euclidiana. No entanto, ela minimiza o efeito de uma grande diferença entre uma das dimensões de um elemento, uma vez que a mesma não é elevada ao quadrado. Essa medida é definida pela expressão (DONI, 2004):

$$d_{xy} = \sum_{i=1}^p |X_i - Y_i|. \quad (11)$$

### ***Distância de Chebyshev***

Conforme Doni (2004), a distância de Chebychev é usada no caso em que se objetiva definir dois elementos como diferentes, se somente umas das dimensões é diferente. Sendo definida por:

$$d_{ch} = \text{máx}(|C_1 - H_1|, |C_2 - H_2|, |C_p - H_p|). \quad (12)$$

### ***Distância de Canberra***

Segundo Fávero e Belfiore (2017), a distância de Canberra é usado quando a variável tem apenas valores positivos, assume que o valor está entre 0 e  $j$  (número de variáveis). Sua expressão é dada por:

$$d_{ca} = \sum_{j=1}^k \frac{|X_{jc} - X_{ja}|}{(X_{jc} + X_{ja})}. \quad (13)$$

## **2.6 Métodos Hierárquicos**

Os métodos hierárquicos de agrupamento correspondem a uma série de sucessivos agrupamentos ou sucessivas divisões de elementos, na qual os elementos são agregados ou desagregados

(DONI, 2004). Os esquemas de aglomeração hierárquicos podem ser aglomerativos ou divisivos, dependendo de como o processo é iniciado. Se todas as observações forem consideradas independentes, e de acordo com sua distância (ou semelhanças), um grupo é formado até que seja realizado apenas um agrupamento até o estágio final, então esse processo é denominado aglomerativo. Por outro lado, se todas as observações forem consideradas agrupadas e, etapa após etapa, for separado cada observação para formar grupos menores, até que essas subdivisões produzam grupos individuais (ou seja, observações completamente separadas), então é um processo divisivo (FÁVERO; BELFIORE, 2017).

### 2.6.1 Métodos Aglomerativos

Segundo Doni (2004), no método aglomerativo cada elemento começa representando um grupo, e a cada etapa, um grupo ou elemento é ligado a outro conforme a sua similaridade, até a última etapa, em que é gerado um grupo único com todos os elementos.

A seguir são mostrados os variados métodos de agrupamento que constituem o método aglomerativo:

**Método de Encadeamento Único (*Nearest Neighbor ou Single Linkage*):** Esse método de encadeamento favorece as menores distâncias (por isso a nomenclatura *nearest neighbor*) para formar um novo agrupamento em cada estágio de aglomeração, mesclando observações ou grupos. Nesse sentido, recomenda-se aplicá-lo a observações relativamente distantes, ou seja, situações diferentes, e esperar formar agrupamentos considerando um mínimo de homogeneidade. Ele é expresso pela seguinte função (FÁVERO; BELFIORE, 2017):

$$d_{(MN)W} = \min \{d_{MW}; d_{NW}\}. \quad (14)$$

em que  $d_{MW}$  e  $d_{NW}$  são as distâncias entre as observações mais próximas dos *clusters*  $M$  e  $W$  e dos *clusters*  $N$  e  $W$ , respectivamente.

**Método de Encadeamento Completo (*Furthest Neighbor ou Complete Linkage*):** O método de encadeamento completo é o oposto do método de encadeamento único, ou seja, prioriza as maiores distâncias entre as observações ou grupos para formar um novo agrupamento (daí, a nomenclatura *furthest neighbor*), portanto, recomenda-se utilizá-lo quando não houver distância considerável entre as observações, e o pesquisador exige a determinação da heterogeneidade entre elas. A sua função é a seguinte (FÁVERO; BELFIORE, 2017):

$$d_{(MN)W} = \max \{d_{MW}; d_{NW}\}. \quad (15)$$

em que  $d_{MW}$  e  $d_{NW}$  são as distâncias entre as observações mais distantes dos *clusters*  $M$  e  $W$  e dos *clusters*  $N$  e  $W$ , respectivamente.

**Método de Encadeamento Médio (*Between Groups ou Average Linkage*):** No método de encadeamento médio, dois grupos são fundidos com base na distância média entre todos os pares de observações pertencentes a esses grupos (daí, a nomenclatura *average linkage*). Portanto, embora o cálculo da medição da distância entre os agrupamentos tenha mudado, o método de encadeamento médio mantém a solução de ordenar as observações em cada grupo fornecida pelo método de encadeamento único no caso de haver uma distância considerável entre as observações. Se as observações forem muito próximas, a solução de classificação fornecida pelo método de encadeamento completo também estará. A sua expressão é dada por (FÁVERO; BELFIORE, 2017):

$$d_{(MN)W} = \frac{\sum_{p=1}^{m+n} \sum_{q=1}^w d_{pq}}{(m+n) \times (w)}. \quad (16)$$

em que  $d_{pq}$  representa a distância entre qualquer observação  $p$  do *cluster*  $MN$  e qualquer observação  $q$  do *cluster*  $W$ , e  $m + n$  e  $w$  representam, respectivamente, a quantidade de observações nos *clusters*  $MN$  e  $W$ .

**Método da Centroide:** Na visão de Doni (2004), o centroide é o ponto médio no espaço multidimensional e retrata o centro de gravidade do referente grupo. As características desse método são a robustez à presença de ruídos e o fenômeno da reversão. A função da distância é estipulada por:

$$d_{(UV)W} = \frac{N_U \times d_{UW} + N_V \times d_{VW}}{N_U + N_V} - \frac{N_U \times N_V \times d_{UV}}{(N_U + N_V)^2}. \quad (17)$$

onde  $N_U$  e  $N_V$  são os números de elementos no grupo  $U$  e  $V$ , respectivamente;  $d_{UW}$  e  $d_{VW}$  são as distâncias entre os elementos  $UW$  e  $VW$ , respectivamente.

**Método de Ward (Método da soma de erros quadráticos ou variância):** Esse método é baseado na variação dentro do grupo em vez de ligação (ANDERBERG, 1973). Algumas das suas características são: **a)** apresentar resultados satisfatórios para as distâncias euclidianas como para outras distâncias; **b)** poder mostrar resultados insatisfatórios quando o número de elementos em cada grupo for quase igual; **c)** ter tendência a combinar grupos com poucos elementos; e **d)** ser sensível à presença de *outliers* (DONI, 2004). A equação é dada por:

$$d_{(UV)W} = \frac{((N_W + N_U) \times d_{(UW)} + (N_W + N_V) \times d_{(VW)} - N_W \times d_{(UV)})}{N_W + N_U + N_V}. \quad (18)$$

onde  $N_U$ ,  $N_V$  e  $N_W$  são os números de elementos no grupo  $U$ ,  $V$  e  $W$ , respectivamente;  $d_{(UW)}$ ,  $d_{(VW)}$  e  $d_{(UV)}$  são as distâncias entre os elementos  $UW$ ,  $VW$  e  $UV$ , respectivamente.

### 2.6.2 Métodos Divisivos

Os métodos divisivos atuam de uma forma distinta dos métodos aglomerativos, isto é, um grupo inicial com todas as observações é dividido em dois subgrupos, de tal maneira que as observações de um subgrupo estejam distantes das observações do outro subgrupo. Assim, esses subgrupos são divididos em subgrupos dissimilares e o processo continua até cada observação formar um grupo (DONI, 2004). Para Kaufman e Rousseeuw (2009), a literatura cita pouco os métodos divisivos, já que requerem uma maior capacidade computacional que os métodos aglomerativos.

## 2.7 Métodos Não Hierárquicos

Os métodos não hierárquicos utilizam algoritmos para maximizar a homogeneidade dentro de cada agrupamento sem um processo hierárquico (FÁVERO; BELFIORE, 2017). Na concepção de Doni (2004), os métodos não-hierárquicos ou por particionamento foram elaborados na intenção de agrupar elementos em  $K$  grupos, em que o  $K$  é a quantidade de grupos estabelecida preliminarmente.

Diferente do esquema hierárquico, em que o pesquisador pode estudar as várias possibilidades de distribuição das observações, e até definir o número de *clusters* interessantes de acordo com cada estágio de agrupamento, o esquema de agrupamento não hierárquico deve ser estipulado *a priori*, do número de *clusters* define-se o centro de aglomeração e atribui observações a partir dele. É por esta razão que quando o número de *clusters* que podem ser formados não podem ser razoavelmente estimados com base nas variáveis em estudo e observações do banco de dados, é recomendado desenvolver um esquema de agrupamento hierárquico antes do esquema não hierárquico (FÁVERO; BELFIORE, 2017).

Os métodos não hierárquicos mais conhecidos são o método *k-means* (k-médias) e o método *k-medoid* (k-medóides), que são descritos a seguir.

### 2.7.1 Método *k-means*

O *k-Means* é um algoritmo de aprendizagem não supervisionado, usado para particionar dados em *k* agrupamentos diferentes. Ele agrupa dados que possuem características relevantes e parecidas. Empiricamente, uma boa solução para o processo de clusterização é a que os elementos do grupo sejam mais semelhantes entre si, do que comparados com outro grupo (SOUSA, 2019).

Dentre os métodos não hierárquico, o *k-means* é o mais utilizado por pesquisadores de diversas áreas do conhecimento. Como o número de *clusters* é inicialmente definido pelo pesquisador, quando o número de *clusters* que podem ser formados não é conhecido, o processo pode ser explicado em detalhes após a aplicação do esquema hierárquico aglomerativo, neste caso, o *output* obtido por este pode servir de *input* para o não hierárquico.

De acordo com Doni (2004), algumas das características desse método são: **a)** ser sensível a ruídos, pois um elemento com um valor bastante alto pode deturpar a distribuição dos dados; **b)** ter tendência a criar grupos esféricos; **c)** o número de grupos é o mesmo no decorrer do processo; e **d)** ser inadequado para encontrar grupos com formas não convexas ou de tamanhos muito distintos.

Um ponto interessante a ser destacado é que o algoritmo *k-means* suporta espaços multi-dimensionais, o que possibilita a sua utilização em problemas de agrupamento com observação de grandes dimensões. A principal barreira desse algoritmo encontra-se em sua inicialização, quando define-se antecipadamente a quantidade de *clusters* e suas respectivas localizações, assim, é possível que o algoritmo convirja para um ótimo local (SUTTER, 2018).

### 2.7.2 Método *k-medoid*

O *k-Medoid* é uma variação do *k-means*. Nesse método ao invés de determinar o centro do agrupamento e utilizá-lo como referência, emprega-se o conceito do objeto mais central do agrupamento. As suas variações mais conhecidas são os algoritmos PAM (*Partitioning Around Medoids*) e CLARA (*Clustering LARge Applications*) (CAMILO; SILVA, 2009).

Segundo Doni (2004), o algoritmo *k-medoid* usa o valor médio dos elementos em um grupo como um ponto referência que é denominado de medóide. Esse é o elemento mais central localizado em um grupo. O *k-medoid* quer basicamente encontrar *K* grupos em *N* elementos e, aleatoriamente, achar um elemento representativo (medóide) para cada grupo. Cada elemento remanescente é agrupado com o medóide ao qual ele é mais similar. Dessa maneira, repetidamente, troca um dos medóides por um dos não medóides ao passo que a qualidade do agrupamento resultante é melhorada.

As características desse método são: **a)** ter os mesmos resultados, independente da ordem; **b)** ter tendência a achar grupos esféricos; **c)** seu processamento ser mais custoso que o *k-means*; **d)** ser não aplicável à grandes bases de dados, uma vez que o custo do seu processamento é alto; e **e)** ser mais robusto do que o *k-means* na presença de ruídos, pois o medóide é menos influenciado pelos ruídos do que a média (DONI, 2004).

### 3 MATERIAIS E MÉTODOS

#### 3.1 Coleta de dados

Os dados utilizados neste trabalho são dados reais da Bolsa de Valores de São Paulo (BM&BOVESPA), a oitava maior bolsa do mundo. A base de dados foi obtida através de uma função do *software* Excel, o período de coleta foi de 04/01/2010 a 31/01/2022 e esses dados foram observados e analisados diariamente, e nesses dados contêm as informações sobre os valores das ações nos momentos de:

- **Abertura:** O preço da primeira negociação no período;
- **Alto:** O maior preço de todas as negociações no período;
- **Baixo:** O menor preço de todas as negociações no período;
- **Fechamento:** O preço da última negociação no período;

#### 3.2 Ações e Indicadores de Análise Técnica

As ações foram escolhidas de forma aleatória, sendo elas as ações da Cielo, Companhia Siderúrgica Nacional (CSN), Fleury, Santander e Usiminas e os indicadores técnicos utilizados foram o *Relative Strength Index* (RSI), Histograma Média Móvel Convergente e Divergente (MACD), Percentual das Bandas de *Bollinger* e o *Rate of Change* (ROC).

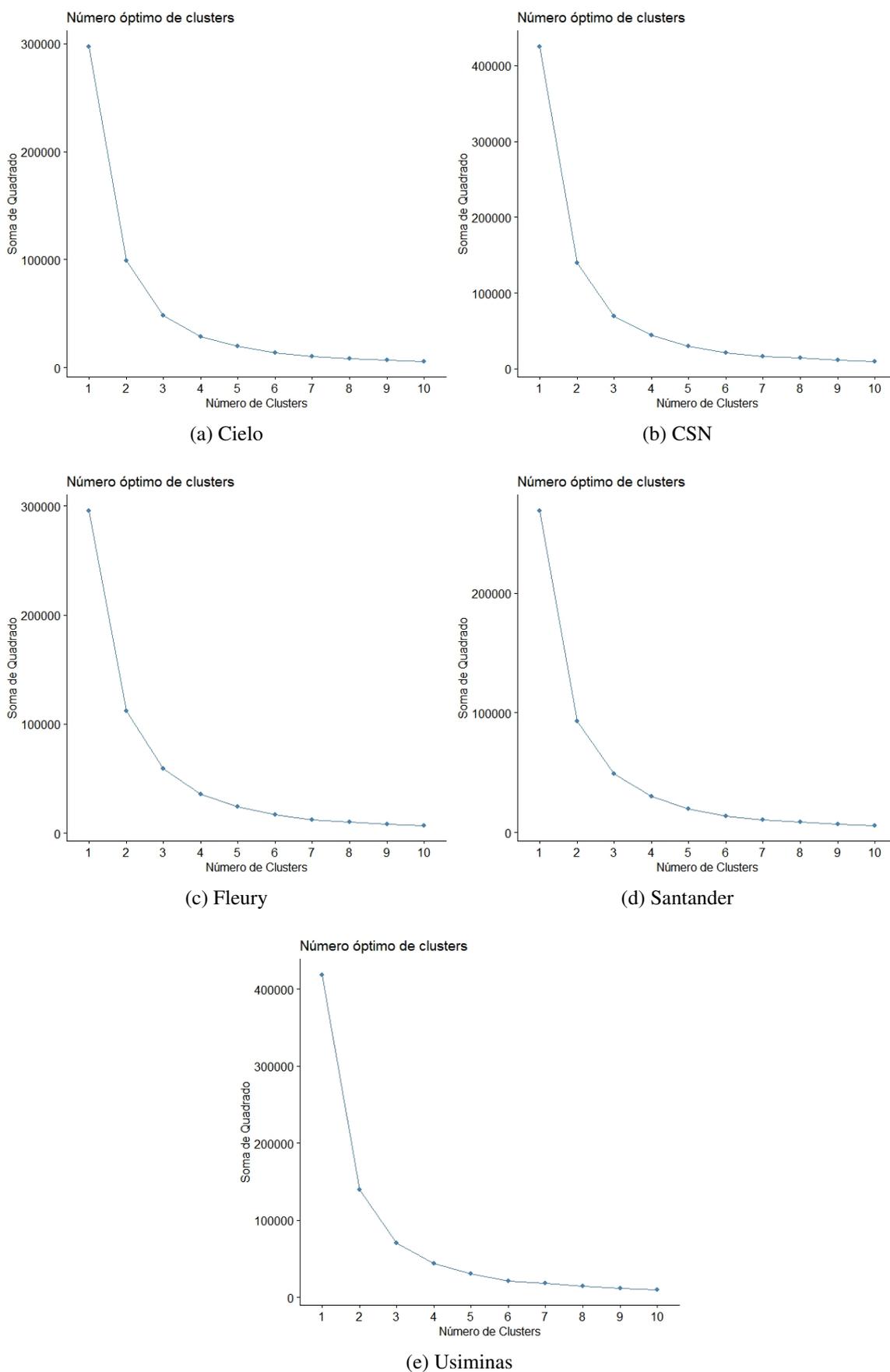
#### 3.3 Método e Pacotes Utilizados

Após coletar os dados, foi realizada uma observação detalhada para a verificação de dados faltantes. Como em todo trabalho, deve-se observar o banco de dados antes de iniciar qualquer análise. Assim, utilizando o método de *Elbow* foram criados os gráficos (*scree plot*), observa-se a curva (cotovelo) no gráfico, sendo considerado para as ações o número ideal de *clusters*, dando início a análise usando uma abordagem não hierárquica, com o método *k-means*, após a formação dos *clusters*, uma análise descritiva foi iniciada para observar os valores das ações em cada um dos *clusters* por meio dos indicadores.

Os procedimentos estatísticos, cálculos e gráficos realizados neste trabalho foram executados no *software* R 4.1.3, uma das ferramentas mais utilizadas atualmente. Esse *software* pode ser obtido gratuitamente em <https://www.r-project.org/>. Para as análises foram utilizados diversos pacotes, podendo destacar os seguintes pacotes: *cluster* (2021) para análise de *cluster*, *factoextra* (2020) para visualizar a saída da análise de dados multivariados e o *PerformanceAnalytics* (2020) para realizar a avaliação de ativos financeiros, entre outros. Em seguida foi apresentado as estatísticas utilizadas nas ações por *clusters* e finalmente os resultados dos retornos acumulados nas bases de treino e teste, comparando o método *k-means* com o *buy and hold*.

### 4 RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados das análises estatísticas fundamentadas pelas técnicas multivariadas em especial o método baseado no algoritmo *k-means*. Com todos os dados coletados, foi realizada uma análise descritiva para verificar as ações em cada *cluster* no *software* R 4.1.3. Para facilitar a visualização das informações obtidas pelas estatísticas, foram criados gráficos e tabelas.

Figura 1 – Número ideal de *clusters* das ações

Fonte: Produzido pelos autores, 2023.

Na Figura 1 foi apresentada a curva (cotovelo) que demonstra a quantidade ideal de *clusters* a serem utilizados na análise. Conforme essa figura o número ideal de *clusters* a serem considerados é igual a 4 para cada uma das ações, uma vez que, a partir de 4 *clusters*, não se obtém um decaimento significativo nos valores da soma de quadrados.

A Tabela 1 demonstra os valores médios de cada indicador que foi usado nas ações de acordo com os *clusters*. Tendo como base os resultados da Tabela 1, para as ações obteve-se as estatísticas descritivas das ações que formam cada *cluster*. Isso poderá ser visualizado na Tabela 2.

Tabela 1 – Valores médios dos indicadores utilizados nas ações conforme os *clusters*

Ações	Variáveis	Clusters			
		1	2	3	4
Cielo	RSI	46,6998	69,1456	34,7564	57,0537
	MACD	-0,1906	0,4497	-0,4707	0,1645
	BB	0,3576	0,9371	0,0856	0,7122
	ROC	-0,0013	0,0085	-0,0101	0,0031
CSN	RSI	70,2478	55,5094	29,8017	43,0369
	MACD	1,2156	0,3352	-0,9357	-0,2785
	BB	0,9421	0,6836	0,0792	0,3021
	ROC	0,0173	0,0044	-0,0152	-0,0038
Fleury	RSI	57,4013	71,7588	46,8945	35,0043
	MACD	0,1522	0,5600	-0,0727	-0,5765
	BB	0,7351	0,9465	0,3771	0,0581
	ROC	0,0043	0,0106	-0,0018	-0,0094
Santander	RSI	45,1725	55,8432	67,1188	33,0441
	MACD	-0,1395	0,2097	0,4903	-0,7101
	BB	0,3314	0,7123	0,9416	0,0743
	ROC	-0,0028	0,0029	0,0107	-0,0114
Usiminas	RSI	57,0170	44,2949	71,2339	30,4743
	MACD	0,3900	-0,3014	1,2360	-0,8416
	BB	0,7207	0,3309	0,9655	0,0896
	ROC	0,0048	-0,0039	0,0216	-0,0160

Fonte: Produzido pelos autores, 2023.

Na Tabela 2 são apresentadas as estatísticas descritivas das ações em cada *cluster*. Ao utilizar essa tabela para tomar a decisão de investir ou não se leva em consideração a média e mediana do *cluster* observado. Quando a média e mediana forem positivas é nesse *cluster* que se deve investir, caso algumas das medidas seja igual a zero e a outra positiva, deve-se analisar e decidir se venderá ou se manterá a ação e quando as duas medidas forem negativas não se deve investir nesse *cluster*.

Portanto, nas ações da Cielo os *clusters* mais rentáveis foram os *clusters* 1 e 4, na CSN os *clusters* 1 e 2. As ações da Fleury, Santander e Usiminas apresentam apenas um *cluster* rentável, sendo eles os *clusters* 4, 2 e 1, respectivamente, porém deve-se analisar os *clusters* 2 da Fleury, o 4 da Santander e o 3 da Usiminas, pois apesar da mediana possuir valor 0, as médias das ações da Fleury e da Santander possuem valores próximos aos dos *clusters* rentáveis, já nas ações da Usiminas a média é ainda maior do que a do *cluster* rentável, talvez assim possam ser considerados rentáveis.

Tabela 2 – Estatísticas utilizadas nas ações conforme os *clusters*

Ações	Estatísticas	Clusters			
		1	2	3	4
Cielo	Mín	-0,0645	-0,1389	-0,0790	-0,0632
	Máx	0,0758	0,0624	0,0944	0,1119
	Média	0,0010	-0,0003	0,0001	0,0011
	Mediana	0,0008	0,0009	-0,0012	0,0009
	DP	0,0180	0,0170	0,0224	0,0171
CSN	Mín	-0,1361	-0,2051	-0,1166	-0,1390
	Máx	0,2062	0,1818	0,1493	0,1659
	Média	0,0045	0,0020	-0,0022	-0,0022
	Mediana	0,0030	0,0005	-0,0050	-0,0031
	DP	0,0386	0,0349	0,0331	0,0317
Fleury	Mín	-0,0571	-0,0699	-0,0897	-0,0758
	Máx	0,0953	0,1111	0,0894	0,0677
	Média	0,0003	0,0022	-0,0004	0,0028
	Mediana	0,0000	0,0000	0,0000	0,0018
	DP	0,0174	0,0209	0,0179	0,0201
Santander	Mín	-0,0972	-0,1115	-0,0890	-0,0829
	Máx	0,0703	0,1578	0,0842	0,1432
	Média	0,0004	0,0010	-0,0012	0,0007
	Mediana	0,0000	0,0007	-0,0020	0,0000
	DP	0,0200	0,0214	0,0202	0,0239
Usiminas	Mín	-0,1614	-0,1300	-0,1568	-0,1237
	Máx	0,1797	0,3511	0,2387	0,1527
	Média	0,0026	-0,0009	0,0034	-0,0035
	Mediana	0,0012	-0,0025	0,0000	-0,0053
	DP	0,0346	0,0357	0,0410	0,0350

Fonte: Produzido pelos autores, 2023.

De acordo com a Tabela 3, percebe-se que para as ações da Cielo os *clusters* 1 e 4 foram os que tiveram maior número de entradas, para as ações da CSN foram os *clusters* 2 e 4 com o maior número de entradas, para as ações da Fleury o maior número de entradas foram nos *clusters* 1 e 3, já nas ações da Santander e Usiminas os *clusters* com os maiores números de entradas foram nos *clusters* 1 e 2.

Tabela 3 – Número de observações classificadas em cada *cluster* ao longo do tempo.

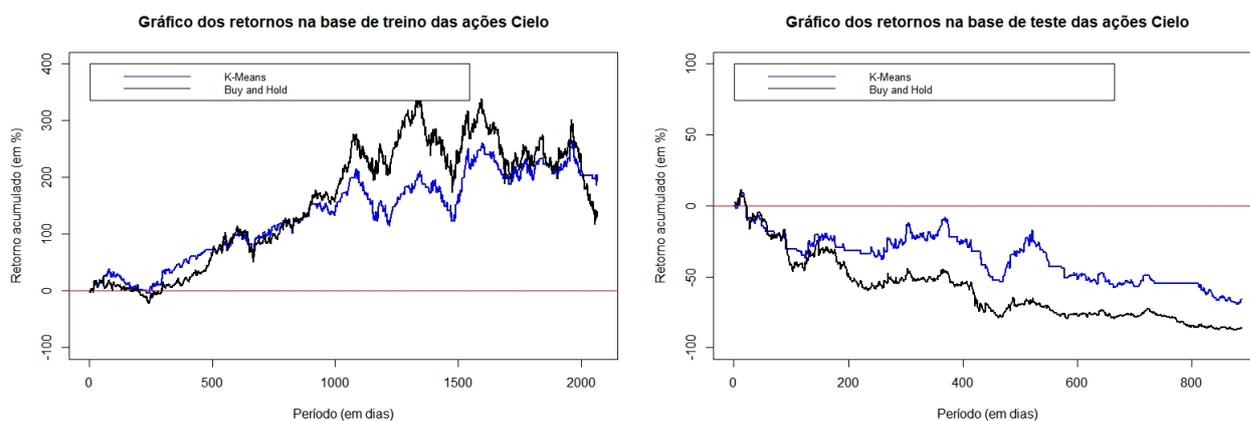
Ações	Clusters			
	1	2	3	4
Cielo	634	418	374	641
CSN	357	554	468	688
Fleury	655	297	751	364
Santander	691	687	362	327
Usiminas	555	708	312	491

Fonte: Produzido pelos autores, 2023.

Na Figura 2 observa-se que na base de treino os retornos acumulados foram bem su-

periores que na base de teste, porém o que se deve analisar é na base de teste. Dessa maneira, percebe-se que apesar dos retornos acumulados terem sido negativos, o método do algoritmo *k-means* apresentou um desempenho melhor do que o método *buy and hold*.

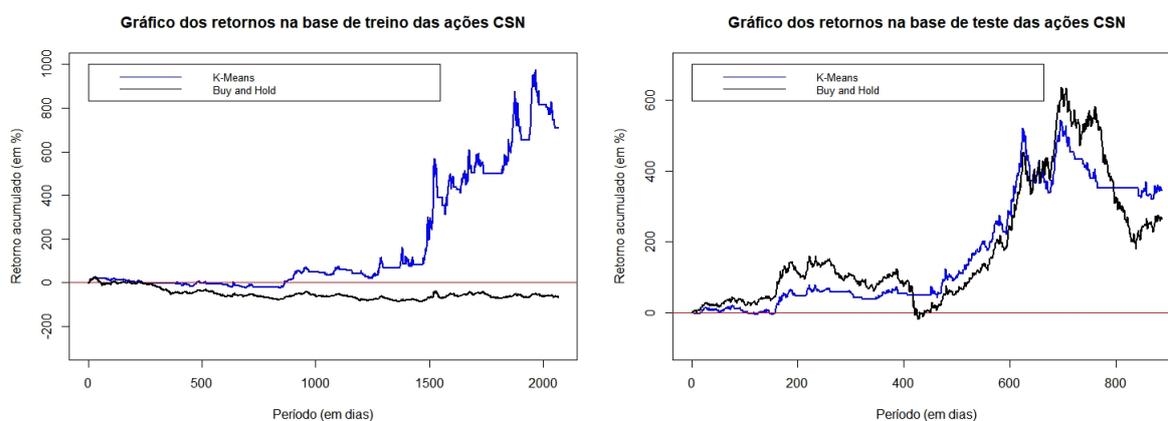
Figura 2 – Gráfico dos retornos acumulados nas bases das ações Cielo



Fonte: Produzido pelos autores, 2023.

Conforme a Figura 3, percebe-se que os retornos acumulados das ações da CSN foram bem altos, dando ênfase para o algoritmo *k-means* que mais uma vez superou o método *buy and hold*, pode-se notar que quando os valores das ações estavam caindo o algoritmo *k-means* indicou que não comprasse as ações (no gráfico nota-se um comportamento contínuo no *k-means*), mas quando começou a subir, logo ele indicou a compra das ações, sendo assim obteve um desempenho superior ao do método *buy and hold*.

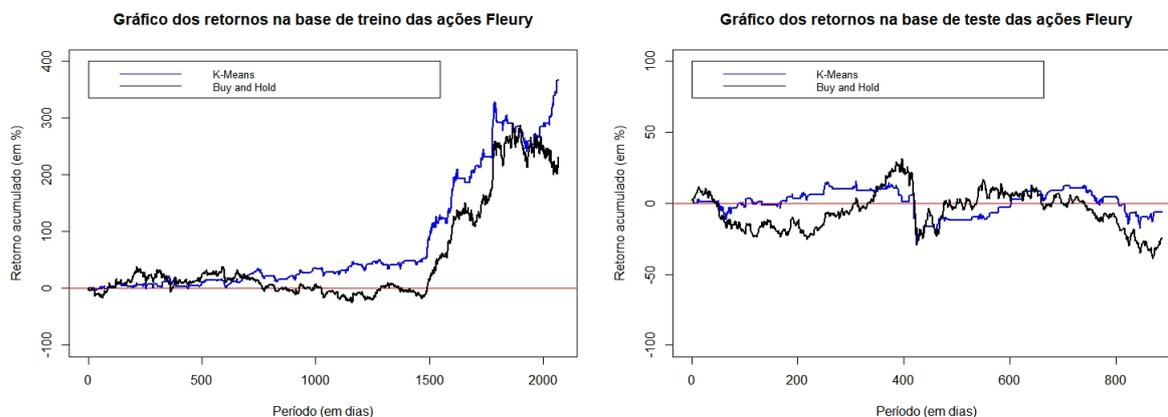
Figura 3 – Gráfico dos retornos acumulados nas bases das ações CSN



Fonte: Produzido pelos autores, 2023.

Na Figura 4 percebe-se que os retornos acumulados foram negativos, mas o método do algoritmo *k-means* foi melhor do que o método *buy and hold*, apresentando valores melhores, fazendo com que os investidores possam minimizar seus prejuízos.

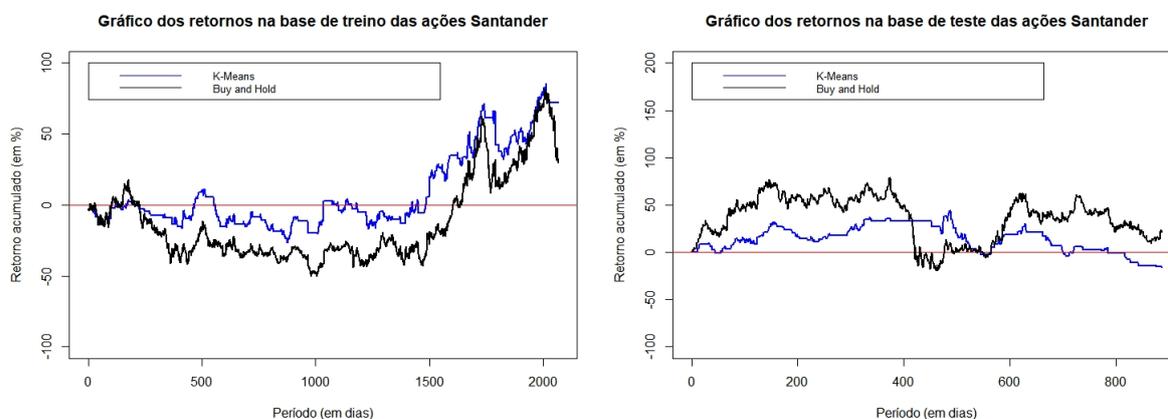
Figura 4 – Gráfico dos retornos acumulados nas bases das ações Fleury



Fonte: Produzido pelos autores, 2023.

Na Figura 5 observa-se que os retornos acumulados foram negativos, mas nesse caso das ações da Santander o método *buy and hold*, foi melhor do que o método do algoritmo *k-means* apresentando valores melhores e positivos, desse modo nota-se que o algoritmo *k-means* nem sempre vai funcionar.

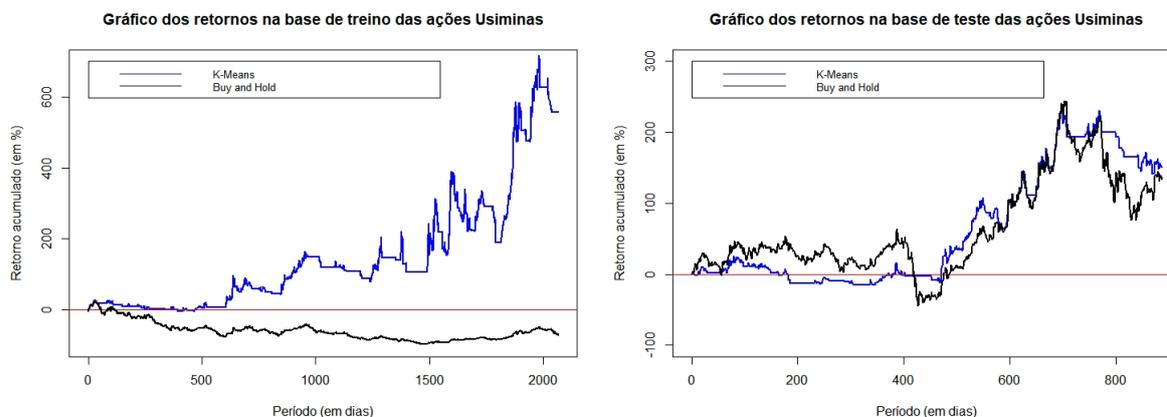
Figura 5 – Gráfico dos retornos acumulados nas bases das ações Santander



Fonte: Produzido pelos autores, 2023.

Na figura 6 os retornos acumulados de ambos os métodos são bem próximos, mas ainda assim o método do algoritmo *k-means* apresentou melhores resultados para as ações da Usiminas.

Figura 6 – Gráfico dos retornos acumulados nas bases das ações Usiminas



Fonte: Produzido pelos autores, 2023.

Na Tabela 4 os retornos acumulados na base de treino com o método *k-means* foi muito superior ao método *buy and hold* em todas as ações, porém o que realmente importa são os dados de teste.

Tabela 4 – Retornos acumulados na base de treino

Ações	Retornos acumulados	
	<i>K-Means</i>	<i>Buy and Hold</i>
<b>Cielo</b>	204,7434	139,3324
<b>CSN</b>	709,3632	-66,0919
<b>Fleury</b>	366,9663	225,3559
<b>Santander</b>	72,1068	29,7949
<b>Usiminas</b>	557,1056	-71,0976

Fonte: Produzido pelos autores, 2023.

Na tabela 5 os retornos acumulados na base de teste com o método *k-means* foi superior ao método *buy and hold* em 4 das 5 ações, apresentando melhores resultados. A tabela demonstra em números absolutos que o algoritmo *k-means* foi melhor na maioria das vezes. Nas ações da Cielo o algoritmo teve um retorno menos desfavorável. As ações da CSN o algoritmo foi superior ao *buy and hold* em 80,5% e no caso das ações da Santander o algoritmo não funcionou, obtendo retorno negativo de aproximadamente 16%.

Tabela 5 – Retornos acumulados na base de teste

Ações	Retornos acumulados	
	<i>K-Means</i>	<i>Buy and Hold</i>
<b>Cielo</b>	-65,7831	-85,6105
<b>CSN</b>	344,4206	263,9231
<b>Fleury</b>	-5,9991	-24,2010
<b>Santander</b>	-15,8163	20,7964
<b>Usiminas</b>	150,5792	134,0422

Fonte: Produzido pelos autores, 2023.

Assim, os resultados encontrados por este estudo são semelhantes aos de Vo, Luo e Vo (2016). O presente artigo obteve resultados positivos com relação a eficácia no auxílio para tomada de decisões referentes a compra e venda de ações. O estudo de Vo, Luo e Vo (2016) verificou o problema de análise de tendências de séries temporais com máquinas de aprendizado supervisionadas e não supervisionadas. Aplicando essas técnicas, propuseram um método com a combinação do algoritmo *k-means* de agrupamento e algoritmo de treinamento SVM para o problema de previsão de tendências. Esse método usou *k-means* para agrupar os dados de entrada e, a partir de cada agrupamento, treinou a classificação SVM para prever o resultado de saída de uma tendência de série temporal, como ascendente, descendente ou sem tendência. Os resultados desse experimento mostraram que o método de combinação proposto tem uma precisão maior do que o *Back Propagation Neural Networks* (BPNN) ou o *Support Vector Machine* (SVM) tradicional. Além disso, para implementar esse método de combinação, foram usados os dados de séries temporais financeiras obtidos no site *Yahoo Finance* e os resultados experimentais também validaram a eficácia do método.

## 5 CONSIDERAÇÕES FINAIS

O objetivo geral deste estudo foi aplicar as técnicas de aprendizado não supervisionado utilizando o método baseado no algoritmo *k-means* no mercado financeiro, visando melhorar a tomada de decisão de compra e venda de ações. Para analisar se a aplicação do algoritmo *k-means* causa impacto positivo nessa tomada de decisão apresentou-se as estatísticas usadas nas ações por *clusters* e, por fim os resultados dos retornos acumulados na base de teste, comparando o método *k-means* com o *buy and hold*.

Este trabalho teve como resultado que na grande maioria das vezes o algoritmo superou o *buy and hold*. Portanto, é mais eficaz utilizar o método *k-means* do que a técnica *buy and hold* para tomar decisões referentes a compra e venda das ações. No entanto, existem momentos que o algoritmo não possui um desempenho eficaz, mas ainda nesses momentos o algoritmo fornece ao investidor perdas inferiores do que o *buy and hold*. Sendo assim, o algoritmo consegue reduzir os riscos e aumentar os lucros de quem investe na bolsa de ações.

É relevante destacar as limitações deste estudo, visto que os resultados encontrados estão limitados à sua amostra. Assim, a sugestão para pesquisas futuras seria examinar a aplicação das técnicas de aprendizado supervisionado no mercado financeiro, tendo em vista que essas técnicas poderiam obter resultados mais favoráveis para a tomada de decisão de compra e venda de ações.

## REFERÊNCIAS

- ANDERBERG, M. R. *Cluster analysis for applications*. New York: Acafenic press, 1973.
- BELTRAN, R. D. **Detecção de fraudes bancárias utilizando métodos de clustering**. 2019. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Universidade Federal do Pampa, Alegrete, 2019. Disponível em: <<https://dspace.unipampa.edu.br/bitstream/riu/4173/1/Rafael%20Duarte%20Beltran%20-%202019.pdf>>. Acesso em: 25 mar. 2022.
- BM&FBOVESPA. **Introdução ao mercado de capitais**. São Paulo: BM&FBovespa, 2010. Disponível em: <<https://centraldefavoritos.com.br/wp-content/uploads/2016/09/mercado-de-capitais.pdf>>. Acesso em: 15 fev. 2022.
- BLOG BRA. **Post: 7 Indicadores de análise técnica para suas operações na bolsa**. Ago. 2021. Disponível em: <<https://sejabra.com.br/2021/09/30/7-indicadores-de-analise-tecnica-para-suas-operacoes-na-bolsa/>>. Acesso em: 08 mar. 2022.
- CAMILO, C. O.; SILVA, J. C. **Mineração de dados: conceitos, tarefas, métodos e ferramentas**. 2009. Relatório Técnico - Universidade Federal de Goiás, Goiânia, 2009. Disponível em: <[https://rozero.webcindario.com/disciplinas/fbm/g/dm/RT-INF\\_001-09.pdf](https://rozero.webcindario.com/disciplinas/fbm/g/dm/RT-INF_001-09.pdf)>. Acesso em: 10 mai. 2022.
- CAPITALNOW. **Indicador MACD: o que é e como interpretar?** Out. 2019. Disponível em: <<https://capitalresearch.com.br/blog/indicador-macd/>>. Acesso em: 03 mar. 2022.
- \_\_\_\_\_. **Você sabe o que é buy and hold? Aprenda já**. Fev. 2020. Disponível em: <<https://capitalresearch.com.br/blog/buy-and-hold/>>. Acesso em: 03 mar. 2022.
- CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. **Análise multivariada: para os cursos de administração, ciências contábeis e economia**. São Paulo: Atlas, 2007.
- CVM. **O mercado de valores mobiliários brasileiro**. 3. ed. Rio de Janeiro: Comissão de Valores Mobiliários, 2014.
- DONI, M. V. **Análise de cluster: métodos hierárquicos e de particionamento**. 2004. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) - Faculdade de Computação e Informática, Universidade Presbiteriana Mackenzie, São Paulo, 2004. Disponível em: <<http://meusite.mackenzie.com.br/rogerio/tgi/2004Cluster.PDF>>. Acesso em: 30 mai. 2022.
- EXAME. **Mercado financeiro: o que é, como funciona e por que acompanhar**. Mai. 2021. Disponível em: <<https://exame.com/invest/academy/mercado-financeiro-como-funciona/>>. Acesso em: 12 mai. 2022.
- \_\_\_\_\_. **O que é média móvel, o indicador usado por traders**. Dez. 2022. Disponível em: <<https://exame.com/invest/guia/o-que-e-media-movel-o-indicador-usado-por-traders/>>. Acesso em: 16 fev. 2023.
- FAÇANHA, S. S. **Uso de técnicas e ferramentas de ciência de dados para realização de análises preditivas sobre medições de vazão, temperatura e pressão de eteno na**

**indústria de petróleo e gás.** 2019. Trabalho de Conclusão de Curso (Bacharelado em Engenharia Metalúrgica) - Universidade Federal do Ceará, Fortaleza, 2019. Disponível em: <[https://repositorio.ufc.br/bitstream/riufc/49254/3/2019\\_tcc\\_ssfa%20c3%27anha.pdf](https://repositorio.ufc.br/bitstream/riufc/49254/3/2019_tcc_ssfa%20c3%27anha.pdf)>. Acesso em: 03 abr. 2022.

FÁVERO, L. P. L.; BELFIORE, P. **Manual de análise de dados: estatística e modelagem multivariada com Excel, SPSS e Stata.** 1. ed. Rio de Janeiro: Elsevier, 2017.

FERNANDO, J. **Relative Strength Index (RSI).** Fev. 2022. Disponível em: <[https://www.investopedia.com/terms/r/rsi.asp#:~:text=The%20relative%20strength%20index%20\(RSI\)%20is%20a%20momentum%20indicator%20used,a%20stock%20or%20other%20asset.&text=An%20RSI%20reading%20of%2030,an%20oversold%20or%20undervalued%20condition](https://www.investopedia.com/terms/r/rsi.asp#:~:text=The%20relative%20strength%20index%20(RSI)%20is%20a%20momentum%20indicator%20used,a%20stock%20or%20other%20asset.&text=An%20RSI%20reading%20of%2030,an%20oversold%20or%20undervalued%20condition)>. Acesso em: 03 mar. 2022.

FREITAS, T. A. et al. Risco de mercado: a importância do gerenciamento para mensurar o risco de uma carteira de investimento. In: Simpósio de excelência em gestão e tecnologia (SEGeT), 15, 2018, Resende. **Anais...** Resende, 2018. Disponível em: <<https://www.aedb.br/seget/arquivos/artigos18/10326142.pdf>>. Acesso em: 20 mar. 2022.

GÉRON, A. **Mãos à obra: aprendizado de máquina com scikit-learn & tensorflow: conceitos, ferramentas e técnicas para a construção de sistemas inteligentes.** Rafael Contatori. Rio de Janeiro: Alta Books, 2019.

JAIN, A. K. *Data clustering: 50 years beyond K-means.* **Pattern Recognition Letters**, [S. l.], v. 31, ed. 8, p. 651–666, jun. 2010. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0167865509002323>>. Acesso em: 05 mai. 2022.

KAUFMAN, L.; ROUSSEEUW, P. J. **Finding Groups in Data: An Introduction to Cluster Analysis.** New York: John Wiley & Sons, 2009.

KASSAMBARA, A.; MUNDT, F. (2020). **Factoextra: Extract and Visualize the Results of Multivariate Data Analyses.** R package version 1.0.7. 2020.

KETCHEN, D. J.; SHOOK, C. L. *The application of cluster analysis in strategic management research: an analysis and critique.* **Strategic Management Journal**, [S. l.], v. 17, ed. 6, p. 441-458, jun. 1996. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0266%28199606%2917%3A6%3C441%3A%3AAID-SMJ819%3E3.0.CO%3B2-G>>. Acesso em: 30 mai. 2022.

LIMA, M. L. **Um modelo para predição de bolsa de valores baseado em mineração de opinião.** 2016. Dissertação (Mestrado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia de Eletricidade, Universidade Federal do Maranhão, São Luís, 2016. Disponível em: <[https://tedebc.ufma.br/jspui/bitstream/tede/297/1/Dissertacao\\_MilsonLouseiroLima.pdf](https://tedebc.ufma.br/jspui/bitstream/tede/297/1/Dissertacao_MilsonLouseiroLima.pdf)>. Acesso em: 21 fev. 2022.

LOPES, H. E. G.; GOSLING, M. S. Análise de *clusters* na prática: lidando com *outliers* na pesquisa gerencial. **Revista de Administração Contemporânea**, v. 25, n. 1, p. e200081,

jan./fev. 2021. Disponível em:

<<https://rac.anpad.org.br/index.php/rac/article/view/1425/1523>>. Acesso em: 21 fev. 2022.

MAECHLER, M., ROUSSEEUW, P., STRUYF, A., HUBERT, M., HORNIK, K. *Cluster: Cluster Analysis Basics and Extensions. R package version 2.1.2.* 2021.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada.** 1. ed. Belo Horizonte: Editora UFMG, 2005.

MITCHELL, C. *Price Rate Of Change Indicator (ROC).* Ago. 2021. Acesso em:

<[https://www.investopedia.com/terms/p/pricerateofchange.asp#:~:text=The%20Price%20Rate%20of%20Change%20\(ROC\)%20is%20a%20momentum%2D,certain%20number%20of%20periods%20ago](https://www.investopedia.com/terms/p/pricerateofchange.asp#:~:text=The%20Price%20Rate%20of%20Change%20(ROC)%20is%20a%20momentum%2D,certain%20number%20of%20periods%20ago)>. Acesso em: 03 mar. 2022.

NETO, J. M. M.; MOITA, G. C. Uma introdução à análise exploratória de dados multivariados. *Química Nova*, [S. l.], v. 21, n. 4, p. 467-469, 1998. Disponível em:

<<https://doi.org/10.1590/S0100-40421998000400016>>. Acesso em: 10 fev. 2022.

PARMAIS. **Mercado financeiro – o que é e como funciona?** Dez. 2021. Disponível em:

<<https://www.parmais.com.br/blog/o-que-e-mercado-financeiro/>>. Acesso em: 12 mai. 2022.

PETERSON, B. G.; CARL, P. (2020). *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis. R package version 2.0.4.* 2020.

PINHEIRO, J. L. **Mercado de capitais: fundamentos e técnicas.** 4. ed. São Paulo: Atlas, 2008.

R CORE TEAM. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Áustria. 2022.

REIS, E. **Estatística multivariada aplicada.** 2. ed. Edições Sílabo, 2001.

REIS, T. **Bandas de Bollinger: saiba como funciona essa ferramenta de análise.** Dez.

2020. Disponível em: <<https://www.sun0.com.br/artigos/bandas-de-bollinger/>>. Acesso em: 03 mar. 2022.

ROZA, F. S. **Aprendizagem de máquina para apoio à tomada de decisão em vendas do varejo utilizando registros de vendas.** 2016. Monografia (Bacharelado em Engenharia de Controle e Automação) - Universidade Federal de Santa Catarina, Florianópolis, 2016.

Disponível em: <[https://repositorio.ufsc.br/bitstream/handle/123456789/171569/PFC\\_2016-1%20Felippe\\_Roza.pdf?sequence=1&isAllowed=y](https://repositorio.ufsc.br/bitstream/handle/123456789/171569/PFC_2016-1%20Felippe_Roza.pdf?sequence=1&isAllowed=y)>. Acesso em: 05 mai. 2022.

SILVA, E. J. **Modelagem e aplicação de técnicas de aprendizado de máquina para negociação em alta frequência em bolsa de valores.** 2015. Dissertação (Mestrado em

Ciência da Computação) - Programa de Pós-Graduação em Ciência da Computação,

Universidade Federal de Minas Gerais, Belo Horizonte, 2015. Disponível em:<

<https://repositorio.ufmg.br/bitstream/1843/ESBF-9XYGE7/1/evertonsilva.pdf>>. Acesso em: 13 fev. 2022.

SOUSA, M. C. C. **Uma análise do algoritmo *k-means* como introdução ao aprendizado de máquinas**. 2019. Monografia (Licenciatura em Matemática) - Universidade Federal do Tocantins, Araguaína, 2019. Disponível em: <<http://repositorio.uft.edu.br/handle/11612/1764>>. Acesso em: 30 mai. 2022.

SUTTER, L. F. M. **Aplicação de técnicas de aprendizado de máquina na predição de tendências das ações na bolsa de valores**. 2018. Trabalho de Conclusão Curso (Bacharelado em Engenharia Computacional) - Universidade Federal de Juiz de Fora, Juiz de Fora, 2018. Disponível em: <[https://www.ufjf.br/engcomputacional/files/2018/02/Monografia\\_LorranFerreira.pdf](https://www.ufjf.br/engcomputacional/files/2018/02/Monografia_LorranFerreira.pdf)>. Acesso em: 18 fev. 2022.

STULP, G. T. **Aplicação e avaliação de algoritmos de mineração de dados para localizar padrões em bancos de dados**. 2014. Monografia (Bacharelado em Ciência da Computação) – Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Ijuí, 2014. Disponível em: <<https://bibliodigital.unijui.edu.br:8443/xmlui/bitstream/handle/123456789/3008/TCC%20-%20Guilherme%20Toldo%20Stulp.pdf?sequence=1&isAllowed=y>>. Acesso em: 24 fev. 2022.

VO, V.; LUO, J.; VO, B. Time series trend analysis based on k-means and support vector machine. **Computing and Informatics**, [S.l.], v. 35, n. 1, p. 111–127, 2016. Disponível em: <<https://www.cai.sk/ojs/index.php/cai/article/view/1445>>. Acesso em: 20 nov. 2022.

## AGRADECIMENTOS

Primeiramente agradeço a Deus por ter me dado saúde e forças para continuar alcançando os meus objetivos.

Agradeço em especial a minha noiva e futura esposa Larissa Scarlet, que sem Deus e ela, nada disso teria acontecido, que sempre me apoiou e incentivou em todos os momentos.

Gostaria de agradecer aos meus pais que me deram total apoio nessa caminhada. Quero agradecer aos colegas e amigos que a universidade juntou, alguns deles são a Débora dos Santos Farias e Sóstenes Jerônimo, que praticamente em todo o decorrer do curso estudamos juntos, compartilhando saberes e conhecimentos.

E por fim, e não menos importante, quero agradecer ao meu orientador, Tiago Almeida de Oliveira, e ao meu coorientador, Ednário Barbosa de Mendonça, pela constante ajuda e encorajamento, compartilhando seus conhecimentos para que fosse possível a realização desse trabalho. Quero agradecer também a todos os professores que me ajudaram de alguma forma a chegar até aqui.