

UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ENGENHARIA DE SÃO CARLOS

IURE SILVA E MENDONÇA  
RODRIGO ANDRADE GALVANI

Aplicação de Redes Neurais e Clusterização na Análise de  
Tendências do Mercado de Capitais no Setor Elétrico Brasileiro

São Carlos  
2024



IURE SILVA E MENDONÇA  
RODRIGO ANDRADE GALVANI

**Aplicação de Redes Neurais e Clusterização na Análise de  
Tendências do Mercado de Capitais no Setor Elétrico Brasileiro**

Monografia apresentada ao Curso  
de Engenharia Elétrica - Sistemas  
de Energia e Automação, da  
Escola de Engenharia de São  
Carlos da Universidade de São  
Paulo, como parte dos requisitos  
para obtenção do Título de Enge-  
nheiro(a) Eletricista

Orientador(a): Prof. Dr. Lucas  
Gabriel Zanon

VERSÃO CORRIGIDA  
São Carlos  
2024

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

M539a

Mendonça, Iure Silva e  
Aplicação de redes neurais e clusterização na análise de  
tendências do mercado de capitais no setor elétrico brasileiro.  
/ Iure Silva e Mendonça, Rodrigo Andrade Galvani; orientador  
Lucas Gabriel Zanon -- São Carlos, 2024.

Monografia (Graduação em Engenharia Elétrica - Sistemas de Energia e Automação) -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2024.

1. K-means. 2. Redes neurais. 3. Mercado financeiro. 4. Setor elétrico. 5. Indicadores de qualidade energética. I. Mendonça, Iure Silva e. II. Galvani, Rodrigo Andrade. III. Título.

# FOLHA DE APROVAÇÃO

**Nome:** Iure Silva e Mendonça

**Título:** “Aplicação de Redes Neurais e Clusterização na Análise de Tendências do Mercado de Capitais no Setor Elétrico Brasileiro”

**Trabalho de Conclusão de Curso defendido e aprovado  
em 09/12/2024,**

**com NOTA 6,0 (seis, zero ), pela Comissão  
Julgadora:**

**Prof. Dr. Lucas Gabriel Zanon - Orientador SEP/EESC/USP**

**Dr. Rafael Ferro Munhoz Arantes - Professor Substituto  
DEP/UFSCar**

**Mestre Victor Hideki Yoshizumi - Doutorando EESC/USP**

**Coordenador da CoC-Engenharia Elétrica - EESC/USP:  
Professor Associado José Carlos de Melo Vieira Júnior**

# FOLHA DE APROVAÇÃO

Nome: Rodrigo Andrade Galvani

Título: “Aplicação de Redes Neurais e Clusterização na Análise de Tendências do Mercado de Capitais no Setor Elétrico Brasileiro”

Trabalho de Conclusão de Curso defendido e aprovado  
em 09/12/2024,

com NOTA 6,0 (seis, zero), pela Comissão  
Julgadora:

Prof. Dr. Lucas Gabriel Zanon - Orientador SEP/EESC/USP

Dr. Rafael Ferro Munhoz Arantes - Professor Substituto  
DEP/UFSCar

Mestre Victor Hideki Yoshizumi - Doutorando EESC/USP

Coordenador da CoC-Engenharia Elétrica - EESC/USP:  
Professor Associado José Carlos de Melo Vieira Júnior



## **RESUMO**

Este trabalho explora a aplicação de técnicas de análise de dados e aprendizado de máquina para analisar tendências do mercado financeiro no setor elétrico, com foco nos indicadores de qualidade energética DEC (Duração Equivalente de Interrupção por Consumidor) e FEC (Frequência Equivalente de Interrupção por Consumidor). Utilizando métodos de clustering e redes neurais, a pesquisa busca prever o comportamento de empresas do setor elétrico no mercado de capitais, considerando as flutuações dos indicadores de qualidade energética. A segmentação das empresas é realizada com base nesses indicadores, permitindo identificar padrões de comportamento e a relação entre a qualidade do fornecimento de energia e o desempenho financeiro das empresas.

**Palavras-chave:** K-means, Redes Neurais, Mercado Financeiro, Setor Elétrico, Indicadores de Qualidade Energética.



## **ABSTRACT**

**Abstract:** This study explores the application of data analysis and machine learning techniques to analyze financial market trends in the electric power sector, focusing on the energy quality indicators DEC (System Average Interruption Duration Index) and FEC (System Average Interruption Frequency Index). Using clustering methods and neural networks, the research aims to predict the behavior of electric sector companies in the capital market, considering fluctuations in energy quality indicators. Companies are segmented based on these indicators, enabling the identification of behavioral patterns and the relationship between energy supply quality and financial performance.

**Keywords:** K-means, Neural Networks, Financial Market, Electric Power Sector, Energy Quality Indicators.



## LISTA DE ILUSTRAÇÕES

Figura 1 – Geração de pontos agrupados aleatoriamente . . . . .	23
Figura 2 – Inercia para cada número de clusters . . . . .	23
Figura 3 – Clusterização passo a passo com 4 clusters . . . . .	26
Figura 4 – Neurônio Artificial . . . . .	27
Figura 5 – Topologias de RNA . . . . .	28
Figura 6 – Sequências de Etapas . . . . .	36
Figura 7 – Fluxograma de coleta e tratamento de dados . . . . .	37
Figura 8 – Formato tabulado . . . . .	38
Figura 9 – CSV aberto no LibreOffice . . . . .	39
Figura 10 – Exemplo de formatação dados ANEEL . . . . .	45
Figura 11 – Décimo Sétimo Quarter . . . . .	52
Figura 12 – Formato tabulado . . . . .	52
Figura 13 – Gráfico com valores previstos e reais . . . . .	57
Figura 14 – Gráfico de Dispersão . . . . .	58
Figura 15 – Gráfico de Resíduos . . . . .	59



## **LISTA DE TABELAS**

Tabela 1 – Relação dos Valores para Tipo de Mercado (TPMERC) . . . . .	41
Tabela 2 – Classificação dos Códigos de Ações e Direitos . . . . .	44
Tabela 3 – Formato da Tabela após tratamento de dados . . . . .	45
Tabela 4 – Tabela de preços e indicadores DEC e FEC . . . . .	49
Tabela 5 – Distribuição de tempo das empresas nos clusters. . . . .	53
Tabela 6 – Parâmetros das redes . . . . .	60



## **LISTA DE ABREVIATURAS E SIGLAS**

DEC	Duração Equivalente de Interrupção por Unidade Consumidora
FEC	Frequência Equivalente de Interrupção por Unidade Consumidora
DRP	Duração Relativa da Transgressão de Tensão Precária
DRC	Duração Relativa da Transgressão de Tensão Crítica
MAE	Mean Absolute Error (Erro Absoluto Médio)
RNA	Rede Neural Artificial
ANEEL	Agência Nacional de Energia Elétrica .
PRODIST	Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	17
<b>1.1</b>	<b>Contextualização</b>	17
1.1.1	Setor elétrico e mercado	17
1.1.2	Qualidade de Energia Elétrica	18
1.1.3	Machine Learning (aprendizado de máquina)	19
<b>1.2</b>	<b>Objetivos</b>	19
1.2.1	Objetivo Principal	19
1.2.2	Objetivos Específicos	19
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	21
<b>2.1</b>	<b>Classificação e Agrupamento</b>	21
2.1.1	O Algoritmo <i>k-means</i>	22
<b>2.2</b>	<b>Redes Neurais Artificiais</b>	27
2.2.1	Estrutura da RNA	28
2.2.2	Implementação em python	28
2.2.3	Processo de treinamento e validação	29
2.2.4	Tipologia	29
2.2.5	Análise de Resultados	30
<b>2.3</b>	<b>O mercado de renda variável</b>	31
2.3.1	As empresas do setor elétrico na B3	31
<b>2.4</b>	<b>Indicadores de Qualidade do Fornecimento de Energia Elétrica</b>	32
2.4.1	Indicadores de qualidade do serviço	32
2.4.2	Qualidade do produto	33
2.4.3	Qualidade Comercial	33
<b>3</b>	<b>MÉTODO DE PESQUISA</b>	35
<b>4</b>	<b>COLETA E TRATAMENTO DE DADOS</b>	37
<b>4.1</b>	<b>Coleta de dados</b>	37
<b>4.2</b>	<b>Dados da B3</b>	37
4.2.1	Conversão de TxT para CSV	37
4.2.2	Filtragem de Empresas	40
4.2.3	Tipo de Mercado	40
4.2.4	Ajustes Finais	41
<b>4.3</b>	<b>Dados da ANEEL</b>	45
4.3.1	Ajuste da leitura do arquivo	46
4.3.2	Filtrar interrupções desejadas	46

4.3.3	Tratamento de DRP e DRC . . . . .	47
4.3.4	Tirar colunas não usadas e agrupar valores de DEC, FEC, DRP e DRC . . . . .	47
4.3.5	Tirar DEC e FEC de linhas e por como colunas . . . . .	48
4.3.6	Montar a base de dados para a rede neural . . . . .	49
<b>5</b>	<b>MODELAGEM E RESULTADOS . . . . .</b>	<b>51</b>
5.1	Clusterização . . . . .	51
5.2	Rede Neural . . . . .	55
5.2.1	Anos de Previsão . . . . .	55
5.2.2	Indicadores . . . . .	55
5.2.3	Implementação . . . . .	55
<b>6</b>	<b>CONCLUSÕES . . . . .</b>	<b>61</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>63</b>

## 1 INTRODUÇÃO

Nessa seção serão apresentados a contextualização e os objetivos que motivam e justificam esse trabalho

### 1.1 Contextualização

#### 1.1.1 Setor elétrico e mercado

No Brasil, o setor elétrico passou por uma significativa transformação desde a década de 1990, com o processo de privatização e a subsequente listagem de empresas na B3, a bolsa de valores brasileira (GOMES et al., 2002).

Isso tem origem na crise nacional que existiu na década de 80 onde o setor público estava em grave crise econômica, notadamente o país estava com uma inflação muita alta, em estagnação econômica e grave crise fiscal. Nesse ínterim, as empresas públicas do setor elétrico, centralizadas no governo por meio da Eletrobras, foram utilizadas para conter a inflação uma vez que as tarifas da energia elétrica, unificadas nacionalmente desde 1974, tinham reajustes abaixo dos níveis de inflação (FERREIRA, 2000).

Isso acarretou, por parte das empresas, em uma diminuição nos programas de investimento que não bastou para conter as margens em um contexto que a demanda caia e os custos de empréstimo crescam a sobressaltos. Isso acarretou no aumento da tomada de empréstimos e aumento a alavancagem financeira do setor, que chegou a pontos insustentáveis de fraqueza econômica orçamentária (IPEA, 1997).

A grande reviravolta das empresas elétricas aconteceu em 1995 com a Lei Geral de Concessões, destacando-se como ponto importante de virada para o setor o capítulo IV. Tal capítulo, tratando da legislação da política tarifária, estabelece importantes pilares para a manutenção do “equilíbrio econômico-financeiro” da empresas, não ficando, com isso, mais refém de políticas anteriores que ocasionava em endividamento perigoso para o setor (BRASIL, 1995).

Nesse sentido, é necessário ressaltar que o entendimento nacional, e consequentemente a política relacionada a energia mudou de forma drástica. Mudou-se de um modelo com planejamento centralizado e com obtenção de investimento majoritariamente advindo do governo e passou-se por um afrouxamento do planejamento, uma vez que privatizando particularidades de mercado financeiro começam a ser parte constituinte da existência ou não de investimentos no setor elétrico. Assim, o setor elétrico acabou por ser submetido ao exame de análise de fluxos de caixa descontado, modelo esse primordial para definir os investimentos do mercado financeiro (ABREU, 1999).

À medida que essa dinâmica de mercado se estabelece no setor elétrico, diversas características econômicas podem ser observadas nas empresas do setor.

As privatizações permitiram, nos anos subsequentes, a comercialização das ações na bolsa de valores, reduzindo as restrições artificiais e possibilitando que mais empresas surgissem e entrassem livremente no mercado. Mais empresas listadas na bolsa garantem maior transparência e homogeneidade das informações entre os formadores de mercado.

Vale destacar que uma boa regulamentação assegura que o produto seja praticamente homogêneo. Todos esses aspectos aproximam o setor elétrico de um mercado perfeitamente competitivo, no qual os preços não são mais arbitrariamente estabelecidos, mas compostos multifatorialmente (BODIE; KANE; MARCUS, 2015).

Nesse contexto, este trabalho tem como objetivo validar se, ao longo de períodos prolongados, os indicadores de qualidade da energia são capazes de apontar tendências na precificação das empresas, refletindo na valorização das ações das companhias listadas na bolsa de valores.

### 1.1.2 Qualidade de Energia Elétrica

A Qualidade de Energia Elétrica (QEE) abrange diversos aspectos essenciais para o funcionamento eficiente e seguro das redes elétricas (ANEEL, 2024). Flutuações na tensão, como picos e afundamentos, podem danificar equipamentos elétricos, resultando em custos elevados de manutenção e substituição. Redes mal projetadas, com baixo fator de potência, reduzem a eficiência energética e aumentam os custos operacionais. A interrupção frequente no fornecimento pode parar a produção industrial, gerando prejuízos significativos, mesmo em curtos períodos de queda. Além desses desafios, a agilidade na resposta da companhia às falhas é crucial, pois tempos de resposta prolongados podem ampliar os danos e aumentar a insatisfação dos consumidores.

Dessa forma, a QEE pode ser avaliada por três principais aspectos: a qualidade do “produto” em si, ou seja, a energia elétrica fornecida; a qualidade da continuidade do serviço, que se refere à frequência e duração das interrupções; e a qualidade do atendimento ao consumidor, conforme especificado no Módulo 8 do (ANEEL, 2024). Além disso, é fundamental considerar os indicadores de segurança do trabalho, que refletem o comprometimento das distribuidoras com a segurança de seus colaboradores e o impacto sobre a segurança da população em geral.

Como o setor elétrico desempenha um papel fundamental no desenvolvimento social e econômico no mundo moderno, é imprescindível que os padrões de qualidade energética sejam rigorosamente cumpridos. O cumprimento desses padrões não só melhora a satisfação dos consumidores, como também fortalece a confiança nas empresas fornecedoras de energia. Clientes satisfeitos tendem a manter essa confiança, o que pode se refletir positivamente no desempenho financeiro dessas empresas no mercado (DROSOS et al., 2020).

### 1.1.3 Machine Learning (aprendizado de máquina)

O *machine learning* é uma subárea da inteligência artificial que permite que sistemas computacionais aprendam padrões a partir de dados históricos e façam previsões ou decisões sem a necessidade de programação explícita. Seu uso tem crescido exponencialmente em diversas áreas, devido à capacidade de lidar com grandes volumes de dados e identificar correlações complexas (RUSSELL; NORVIG, 2020).

Com base na perspectiva apresentada no item anterior, de que a qualidade da energia pode influenciar positivamente o desempenho financeiro, o uso de técnicas de *machine learning*, mais especificamente utilizando clusterização por meio do método K-means, e redes neurais artificiais com o método perceptron multicamadas, surge como uma abordagem promissora para prever mudanças no mercado a partir de indicadores de qualidade energética. Essa estratégia se destaca pela possibilidade de aproveitar o grande volume de dados disponíveis sobre interrupções no fornecimento de energia e cotações da bolsa, permitindo alimentar modelos capazes de lidar com a complexidade e possivelmente identificar padrões.

## 1.2 Objetivos

### 1.2.1 Objetivo Principal

Este trabalho tem como objetivo analisar o comportamento de empresas do setor elétrico no mercado de capitais (com enfoque na distribuição), utilizando indicadores de qualidade de energia como parâmetros centrais. Busca-se compreender como esses indicadores, que impactam diretamente os consumidores e são destacados nos relatórios financeiros das empresas, podem influenciar sua valorização. Para alcançar esse objetivo, será empregada uma abordagem que combina redes neurais e técnicas de clusterização. As redes neurais serão utilizadas para investigar tendências implícitas nos indicadores e explorar uma possível relação causal com a valorização das empresas. Já as técnicas de clusterização serão aplicadas para identificar trajetórias e padrões organizacionais, revelando estratégias de mercado associadas ao desempenho nos indicadores.

### 1.2.2 Objetivos Específicos

- a) Revisar literatura referente a clusterização e sua aplicação.
- b) Revisar literatura referente ao funcionamento das redes neurais e suas aplicações.
- c) Desenvolver os métodos para tratar os dados de qualidade energética
- d) Avaliar se empresas com melhores índices tendem a valorizar mais.
- e) Verificar se os indicadores podem ajudar a explicar trajetória das empresas



## 2 REVISÃO BIBLIOGRÁFICA

Nesse capítulo, abordam-se bibliografias que apresentam o panorama de diversos autores sobre métodos de clusterização, o mercado de renda variável e análise preditiva.

### 2.1 Classificação e Agrupamento

A classificação é uma disciplina humana antiga, com registros de sua prática desde os gregos, destacando-se Aristóteles, que já no século IV a.C. realizou uma ampla classificação de plantas e animais, um marco de início para a classificação (ARISTOTELES, 350 a.C.), ainda hoje observada pelo seu valor histórico. Como mencionado por (HARTIGAN, 1975), a classificação é uma forma de analisar os objetos, e não um estudo dos objetos em si, permitindo sua aplicação nas mais diversas áreas do conhecimento.

Com o avanço da capacidade computacional, das técnicas de processamento e do crescimento exponencial na geração e disponibilidade de dados, impulsionado por áreas como inteligência artificial e aprendizado de máquina, a disciplina de classificação expandiu seus escopos e aplicações. No entanto, embora os grandes volumes de dados modernos permitam análises mais robustas, a clusterização de um número limitado de amostras de um conjunto de dados extenso pode introduzir vieses nos resultados (HAN; KAMBER; PEI, 2011). Nesse contexto, o cenário elétrico brasileiro se destaca positivamente, pois os dados são regulados e disponibilizados de forma pública e acessível pelos agentes reguladores, permitindo análises mais consistentes e confiáveis.

Dentro dessa disciplina, destaca-se o *clustering*, ou agrupamento. A partir de parâmetros previamente selecionados e tratados, realizamos a ingestão de dados e, utilizando algoritmos específicos, processamos e definimos o número de grupos, a seleção dos parâmetros que determinam a inclusão de um objeto em um grupo, e, finalmente, a classificação dos objetos em cada grupo. Assim, cada *cluster* passa a reunir características ou propriedades próprias, de modo que os dados dentro de um *cluster* podem ser claramente diferenciados dos dados fora dele (LINDEN, 2009).

Os algoritmos de *clustering* podem ser divididos em quatro grandes grupos (HAN; KAMBER; PEI, 2011): *clustering* hierárquico, particional, baseado em densidade e baseado em grade.

Na primeira subdivisão, o *clustering* hierárquico, parte-se de um agrupamento previamente estabelecido e cria-se novos subgrupos a partir deste original, onde cada objeto contido em um subgrupo está correlacionado. Esse método é caracterizado por uma decomposição hierárquica, mas não tem a capacidade de corrigir fusões ou divisões errôneas, embora possa incorporar outras técnicas, como *microclustering*, ou considerar “ligações” entre objetos.

Na subdivisão particional, após a definição inicial do número de grupos (o que também

pode ser realizado por algoritmos), são feitas realocações iterativas dos objetos até que a convergência entre os grupos seja atingida. Esse método busca *clusters* mutuamente exclusivos de forma esférica, com base em distância, podendo usar a média ou medóide como centro do *cluster*. É eficaz para conjuntos de dados de pequeno a médio porte.

O *clustering* baseado em densidade é projetado para identificar *clusters* com formas arbitrárias, em regiões densas de objetos no espaço, separados por regiões de baixa densidade. Cada ponto dentro de um *cluster* deve ter um número mínimo de pontos em sua "vizinhança". Este método também pode ser usado para filtrar *outliers*, identificando regiões de alta densidade.

Finalmente, o *clustering* baseado em grade utiliza uma estrutura de dados de grade multiresolução, o que resulta em tempos de processamento rápidos, geralmente independentes do número de objetos, mas dependentes do tamanho da grade.

Neste trabalho, serão utilizados algoritmos contidos na segunda subdivisão, destacando-se os algoritmos *k-means* e *k-medoids*.

### 2.1.1 O Algoritmo *k-means*

O algoritmo *k-means* a ser utilizado será implementado por meio da biblioteca *Scikit-learn*, que oferece ferramentas para aprendizado de máquina, mineração de dados, entre outras. Concebido por MacQueen em 1967, o *k-means* tem se consolidado como uma das principais ferramentas de agrupamento (MACQUEEN, 1967).

O *k-means* tenta separar as amostras em um número  $n$ , fornecido previamente, de igual variância, minimizando o critério de inércia ao mínimo possível. Dessa forma, o algoritmo divide um número  $N$  de amostras  $X$  em  $K$  *clusters* disjuntos  $C$ , cada um descrito por um centroide  $u_j$ , que é a média das amostras contidas no *cluster*.

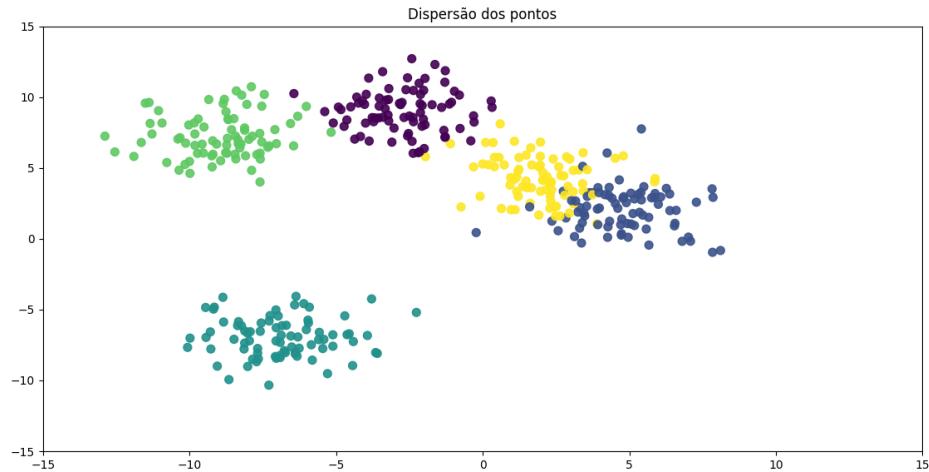
$$\sum_{i=0}^n \min_{u_j \in C} (||x_i - u_j||^2) \quad (2.1)$$

O algoritmo pode ser basicamente entendido em três passos. O primeiro é a inicialização, na qual são escolhidos os centróides, sendo que, nos casos mais comuns, a escolha do centróide é feita basicamente escolhendo  $k$  amostras da base de dados  $X$ . O segundo passo consiste em atribuir cada amostra ao centróide mais próximo dela. O terceiro passo é a criação de novos centróides por meio da média de todas as amostras marcadas previamente para cada centróide. Esses dois últimos passos são feitos em *loop* até que os centróides não mudem mais significativamente.

É necessário ressaltar que o número de *clusters* a ser separado precisa ser previamente escolhido. Para isso, será utilizado o método *elbow*, no qual basicamente é rodado o algoritmo para diversos números de *clusters* e isso tende a gerar uma saída exponencial decrescente para o valor da inércia. Dessa forma, escolhe-se o menor número de *clusters* para o qual a inércia já estiver no entorno do seu valor final.

Para demonstrar esse processo de forma mais clara, recorre-se ao seguinte exemplo: primeiro são gerados alguns pontos aleatórios que ficam agrupados em formatos de bolhas, como pode ser observado na imagem 1.

Figura 1 – Geração de pontos agrupados aleatoriamente

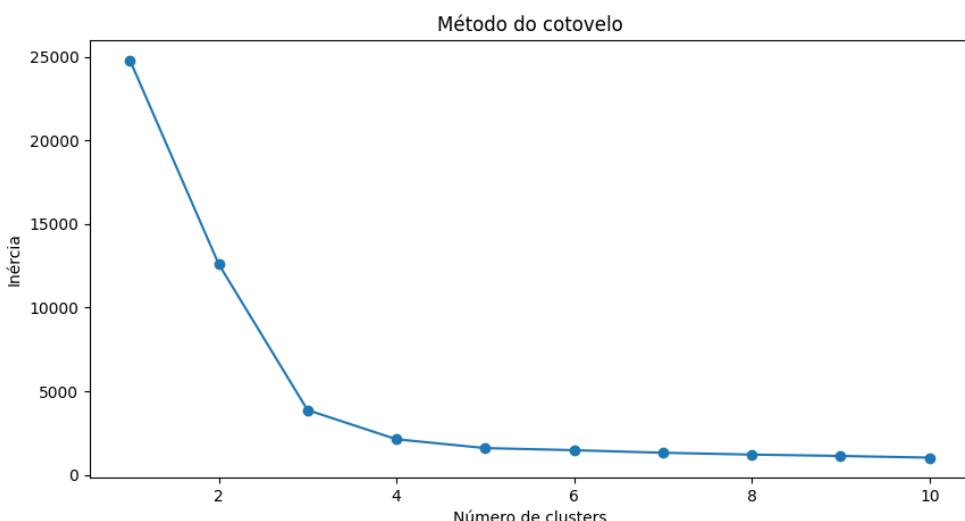


Fonte: Elaborado pelos autores.

É de se ressaltar que as cores atribuídas demonstram como foram geradas as amostras aleatórias.

Após isso, busca-se determinar o número de *clusters* ideal através do método *elbow*, executando o algoritmo *k-means* para um número de *clusters* de 1 a 11. Encontra-se que as inéncias chegam aos mínimos valores, como mostrado na imagem 2.

Figura 2 – Inercia para cada número de clusters



Fonte: Elaborado pelos autores.

A partir de 3 *clusters*, a inércia começa a apresentar um comportamento próximo ao linear. Dessa forma, esse ponto é considerado o “cotovelo” e, portanto, o número ideal de *clusters* para o algoritmo.

Dessa forma, o comportamento do algoritmo será explicado ao longo das iterações mostradas na figura 3. Na primeira iteração, temos a inicialização, onde são escolhidos 4 centróides em posições dispersas ao longo das amostras, e as amostras são marcadas com cores de acordo com a proximidade ao centróide.

Na segunda iteração, é calculada a posição média de todas as amostras previamente atribuídas, definindo, assim, o novo centróide.

O processo de marcação é então refeito, e os demais passos de cálculo da posição e redefinição dos centróides são realizados até que a posição dos centróides não se altere significativamente. Dessa forma, o algoritmo para e os *clusters* são definidos.

O algoritmo em Python, sem a geração das imagens iteração a iteração, pode ser observado a seguir.

### Script 2.1 – Algoritmo K-means

```

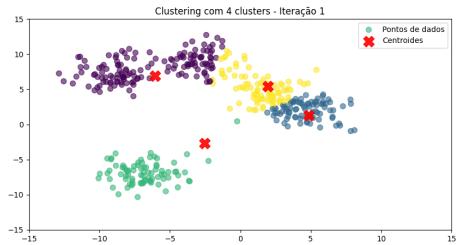
1 import sklearn.datasets as skl
2 from sklearn.cluster import KMeans
3 import numpy as np
4 import matplotlib.pyplot as plt
5
6 # Configurando a semente para reproduzibilidade
7 np.random.seed(42)
8
9 # Figura com a posição inicial dos pontos aleatórios
10 plt.figure(figsize=(10, 5))
11
12 # Parâmetros dos clusters
13 k = 5 # número de clusters
14 n = 400 # número de pontos
15
16 # Gerando os dados com make_blobs (distribuição em forma de bolhas)
17 data = skl.make_blobs(n_samples=n, n_features=2,
18                       centers=k, cluster_std=1.5, random_state=42)
19
20 # X contém as coordenadas dos pontos; labels contém os rótulos originais
21 X, labels = data
22
23 # Exibindo os dados iniciais com rótulos originais
24 plt.scatter(X[:, 0], X[:, 1], c=labels, cmap='viridis',
25             s=50, alpha=0.9)
26 plt.xlim(-15, 15)
27 plt.ylim(-15, 15)
28 plt.title('Dispersão dos pontos')

```

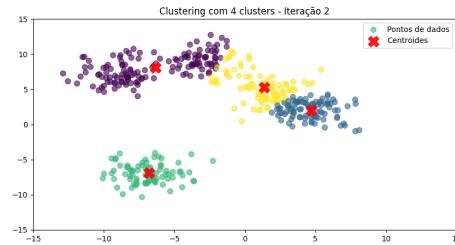
```
29 plt.xlabel('Coordenada_X')
30 plt.ylabel('Coordenada_Y')
31 plt.show()
32
33 # Aplicando o metodo do cotovelo
34 inertias = []
35
36 # Loop para calcular a inercia para diferentes numeros de clusters
37 for i in range(1, 11):
38     kmeans = KMeans(n_clusters=i, random_state=42)
39     kmeans.fit(X)
40     inertias.append(kmeans.inertia_)
41
42 # Exibindo o grafico do metodo do cotovelo
43 plt.figure(figsize=(10, 5))
44 plt.plot(range(1, 11), inertias, marker='o', label='Inercia')
45 plt.title('Metodo do Cotovel')
46 plt.xlabel('Numero de clusters')
47 plt.ylabel('Inercia')
48 plt.legend()
49 plt.show()
50
51 # Configurando o KMeans com o numero de clusters definido (3)
52 kmeans = KMeans(n_clusters=3, random_state=42)
53 kmeans.fit(X)
54
55 # Exibindo o resultado do clustering
56 plt.figure(figsize=(10, 5))
57 plt.scatter(X[:, 0], X[:, 1], c=kmeans.labels_, cmap='viridis',
58             s=50, alpha=0.9, label='Pontos')
59 plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1],
60             marker='x', s=200, c='red', label='Centroides')
61 plt.xlim(-15, 15)
62 plt.ylim(-15, 15)
63 plt.title('Clustering com 3 clusters')
64 plt.xlabel('Coordenada_X')
65 plt.ylabel('Coordenada_Y')
66 plt.legend()
67 plt.show()
```

Figura 3 – Clusterização passo a passo com 4 clusters

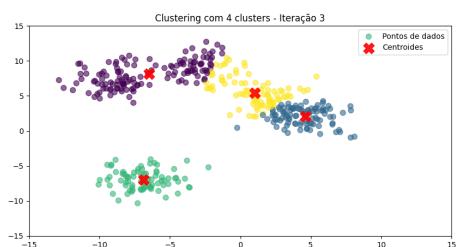
(a) Iteração 1



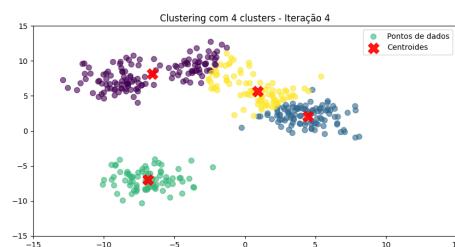
(b) Iteração 2



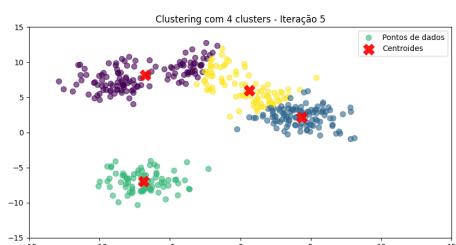
(c) Iteração 3



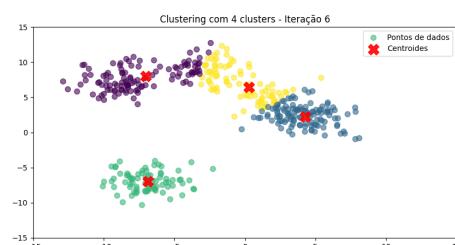
(d) Iteração 4



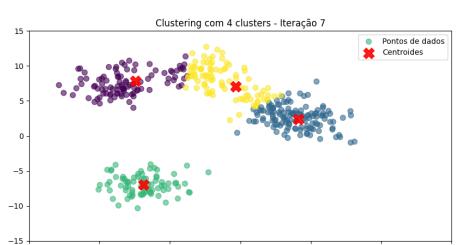
(e) Iteração 5



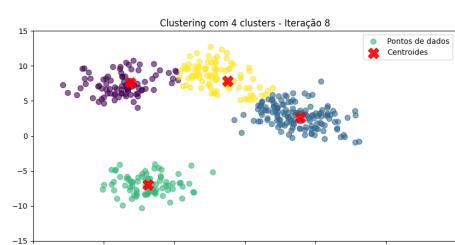
(f) Iteração 6



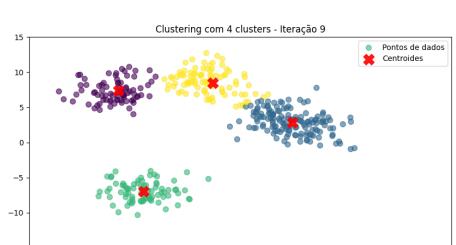
(g) Iteração 7



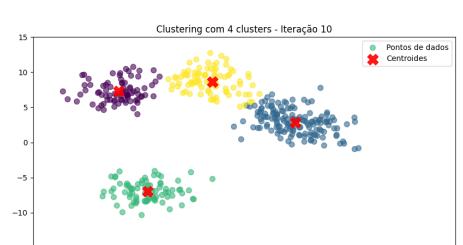
(h) Iteração 8



(i) Iteração 9



(j) Iteração 10



Fonte: Elaborado pelos autores.

## 2.2 Redes Neurais Artificiais

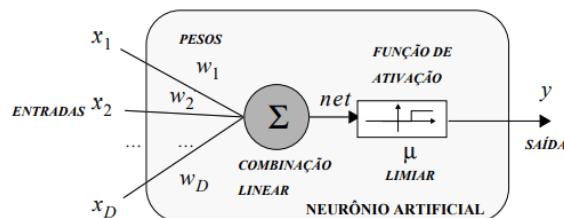
As redes neurais artificiais (RNAs) têm suas raízes na tentativa de simular o comportamento do cérebro humano, que é composto por bilhões de neurônios interconectados (Haykin 1998). A ideia inicial remonta aos anos 1940, com o modelo do neurônio de McCulloch e Pitts, que introduziu o conceito de representar funções lógicas através de redes neurais simples. O avanço significativo veio com a introdução do algoritmo de retropropagação nos anos 1980, que permitiu treinar redes com múltiplas camadas ocultas de maneira eficiente (Goodfellow, Bengio e Courville 2016).

Matematicamente, cada neurônio recebe entradas, que são ponderadas por pesos e somadas junto a um termo de viés. O resultado passa por uma função de ativação, produzindo a saída. Isso pode ser expresso pela equação:

$$y = A(x) \sum_{j=1}^D w_j x_j - \mu$$

Em que  $A(x)$  representa a função de *Heaveside* (função de escada),  $A(x) = 1$  para  $x$  maior ou igual a zero e  $A(x) = 0$  caso o contrário,  $x$  apresenta as entradas  $w$  o peso de cada entrada e  $\mu$  uma função de ativação linear. Dessa forma, o neurônio artificial pode ser representado pela Figura 4:

Figura 4 – Neurônio Artificial



Fonte: (RAUBER; RUEDEL; PADILHA, 2024)

As funções de ativação introduzem não-linearidade no modelo, o que permite às redes aprenderem padrões complexos (GOODFELLOW; BENGIO; COURVILLE, 2016):

- **ReLU (Rectified Linear Unit):** É uma função de ativação que retorna o valor da entrada quando positivo e zero caso contrário. É amplamente utilizada devido à sua eficiência em mitigar problemas de *vanishing gradient*, além de ser computacionalmente simples.
- **Sigmoides:** É uma função de ativação que transforma a entrada em valores no intervalo de 0 a 1. É especialmente útil em problemas de classificação binária, onde a saída representa uma probabilidade.

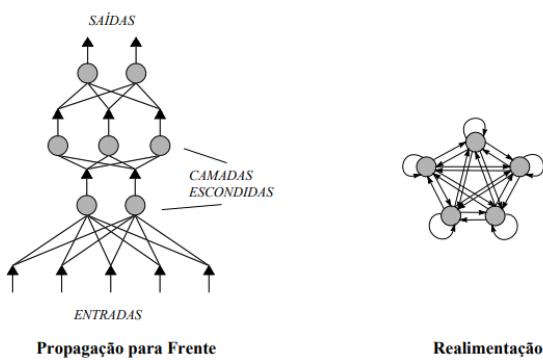
### 2.2.1 Estrutura da RNA

As RNAs são compostas por três tipos de camadas ([CHOLLET, 2017](#)):

- **Camada de entrada:** Recebe as variáveis independentes do modelo.
- **Camadas ocultas:** Executam a maior parte do processamento e aprendem representações internas dos dados.
- **Camada de saída:** Gera as previsões do modelo.

A arquitetura de uma rede neural pode variar em profundidade (número de camadas) e largura (número de neurônios por camada). A escolha de arquitetura depende da complexidade do problema a ser resolvido. Sendo assim o potencial e flexibilidade do cálculo baseado em redes neurais vêm da criação de conjuntos de neurônios que estão interligados entre si ([RAUBER; RUEDEL; PADILHA, 2024](#)). A topologia da rede pode ser observada na Figura 5:

Figura 5 – Topologias de RNA



Fonte: ([RAUBER; RUEDEL; PADILHA, 2024](#))

Para que a rede entenda como transformar as entradas em saída de forma coerente é aplicado um algoritmo de retropropagação (*backpropagation*), o qual é um método de otimização usado para ajustar os pesos da rede, minimizando a função de custo através do gradiente descendente ([HAYKIN, 1998](#)). O processo envolve: a entrada ser propagada através da rede para calcular a saída, o cálculo da perda comparando a função de custo e a saída da rede com os valores reais, cálculo dos gradientes e ajustes dos pesos de acordo com os gradientes

### 2.2.2 Implementação em python

Python é uma das linguagens preferidas para desenvolvimento de redes neurais devido à sua vasta gama de bibliotecas e ferramentas ([CHOLLET, 2017](#)), como NumPy, Usada

para operações matriciais e vetoriais de forma eficiente, Pandas, usada para trabalhar com *dataframes* e TensorFlow e Keras as quais simplificam a criação e treinamento de redes neurais. Keras, em particular, oferece uma API intuitiva para modelagem (ABADI et al., 2016).

Um código exemplo possível é mostrado no script 2.2.

Script 2.2 – Rede Exemplo

```

1
2 from keras.models import Sequential
3 from keras.layers import Dense
4
5 # Criando o modelo
6 # Camada de entrada com 3 features
7 # Camada oculta
8 # Camada de saída para regressão
9 modelo = Sequential()
10 modelo.add(Dense(16, input_dim=3, activation='relu'))
11 modelo.add(Dense(8, activation='relu'))
12 modelo.add(Dense(1, activation='sigmoid'))
13
14 # Compilando o modelo
15 modelo.compile(loss='mean_absolute_error', optimizer='adam',
16 metrics=['mae'])

```

Nesse código tem-se o modelo com as camadas de entrada, oculta (com 8 neurônios) e saída. Na linha final tem-se a métrica MAE que será explicada a seguir.

### 2.2.3 Processo de treinamento e validação

O treinamento de redes neurais envolve dividir os dados em conjuntos de treinamento e validação. O desempenho do modelo é medido por métricas como *Mean Absolute Error* (MAE), que indica a média das diferenças absolutas entre valores previstos e reais (Chollet 2017). Durante o treinamento, é importante monitorar o *overfitting*, que ocorre quando o modelo se ajusta excessivamente aos dados de treinamento e não generaliza bem para novos dados.

### 2.2.4 Tipologia

No presente trabalho, utilizou-se a rede neural do tipo Perceptron Multicamadas (MLP - *Multi-Layer Perceptron*), amplamente reconhecida na literatura como uma das arquiteturas mais utilizadas em problemas de aprendizado supervisionado, especialmente para dados tabulares e regressão (TAUD; MAS, 2018). A escolha dessa tipologia deve-se à sua capacidade de modelar relações não lineares entre variáveis de entrada e saída, o que é

essencial em problemas como o presente estudo, que analisa a relação entre os indicadores de qualidade de energia e os preços médios ponderados das ações no mercado financeiro.

Na suas camadas de entrada serão usadas 2 entradas, valores de DEC e FEC. As camadas ocultas, as quais ficam entre a entrada e a saída, processam os dados por meio de combinações lineares ponderadas e funções de ativação, isso permite que a rede aprenda padrões complexos e não lineares, o que é o ideal para um problema abordado pelo estudo, já que é difícil de visualizar uma função que relate a entrada com a saída. Por fim a ultima camada da rede gera o resultado final. Aqui, foi configurada com um único neurônio para fornecer um valor contínuo, representando a previsão do preço médio ponderado

Dentro dessa tipologia, também é possível aplicar técnicas para melhorar a precisão ou mitigar efeitos indesejados, como o overfitting. O overfitting ocorre quando a rede neural se ajusta excessivamente aos dados de treinamento, perdendo a capacidade de generalizar para novos dados. Para evitar esse problema e melhorar o desempenho geral, este trabalho empregou algumas estratégias, como o aumento do número de camadas, a normalização por lotes (Batch Normalization) e o uso de Dropout.

O aumento do número de camadas permite que a rede aprenda representações mais complexas dos dados, capturando relações sutis entre as variáveis. Por outro lado, a normalização por lotes (Batch Normalization) reduz a instabilidade durante o treinamento ao normalizar as ativações dentro de cada lote, acelerando a convergência e melhorando a robustez do modelo. Já o Dropout é uma técnica que introduz aleatoriedade no treinamento ao "desligar" uma fração dos neurônios em cada iteração, forçando a rede a aprender padrões mais gerais e reduzindo o risco de overfitting

## 2.2.5 Análise de Resultados

Para avaliar a performance das previsões da Rede Neural, serão utilizados diversos métodos e métricas de avaliação, que permitem analisar tanto a acurácia do modelo quanto a sua capacidade de generalização. Além de MAE o qual já foi explicado, as principais métricas empregadas são as seguintes:

- **Erro Quadrático Médio (MSE):** Calcula a média dos quadrados das diferenças entre os valores previstos e os reais. O MSE penaliza grandes erros de forma mais significativa do que o MAE, tornando-o sensível a outliers.
- **Raiz do Erro Quadrático Médio (RMSE):** Representa a raiz quadrada do MSE, oferecendo uma medida da magnitude do erro, mas na mesma unidade das variáveis de entrada. Assim como o MSE, o RMSE é sensível a grandes erros e outliers.
- **Coeficiente de Determinação ( $R^2$ ):** Métrica que indica a proporção da variabilidade dos dados que é explicada pelo modelo. Um valor de  $R^2$  próximo de 1 indica que o

modelo explica bem a variabilidade dos dados, enquanto um valor próximo de 0 indica que o modelo não consegue explicar a variabilidade.

## 2.3 O mercado de renda variável

Os autores (BODIE; KANE; MARCUS, 2015) estabelecem uma distinção entre ativos reais e financeiros. Enquanto os primeiros constituem a base da riqueza material de uma sociedade, determinada por sua capacidade produtiva e composta por terrenos, edificações, equipamentos e maquinários, além do conhecimento empregado na produção de bens e serviços, os ativos financeiros, por sua vez, existem em diversos formatos, como documentos físicos ou números em sistemas computacionais, e não contribuem diretamente para a capacidade produtiva da economia.

Ainda assim, os ativos financeiros desempenham uma função crucial: servem como instrumentos para a alocação de renda ou riqueza entre os investidores. É nessa característica que se fundamenta o interesse deste trabalho. No mercado acionário atual, a informação é amplamente disseminada e, legalmente, acessível; além disso, as taxas estão cada vez mais reduzidas, incentivando uma alocação mais eficiente. Assim, diversos participantes do mercado são estimulados a buscar os ativos mais bem precificados.

Dessa forma, o preço das ações é um reflexo da avaliação coletiva dos agentes e refletem, da melhor forma quantos as informações existem e permitem, a melhor alocação de recursos possível na economia dos países atualmente. Mesmo que esse direcionamento de investimento possa ser ineficiente em momentos particulares de irracionalidade ou iliquidez de mercado.

Quando olhamos para as empresas do setor elétrico, e mais propriamente para os indicadores de negócio ligados ao setor busca-se entender como cada um desses indicadores é precificado e, portanto, qual é a sua importância para a economia na qual estão inseridos.

### 2.3.1 As empresas do setor elétrico na B3

A B3 classifica as ações em setores e subsetores. As empresas do setor elétrico, foco deste estudo, estão inseridas no setor de utilidade pública, no subsetor de energia elétrica, com o segmento homônimo. Esse grupo reúne 77 empresas que atuam em diferentes áreas, como geração, transmissão e distribuição de energia (B3, s.d.). A análise da interseção entre as empresas listadas na B3 e aquelas com dados de indicadores de qualidade identificou 22 empresas com, no mínimo, 10 anos de dados disponíveis. Sob o viés financeiro, destacam-se o preço médio diário e a quantidade de papéis negociados, que constituem os principais objetos de análise deste trabalho.

## 2.4 Indicadores de Qualidade do Fornecimento de Energia Elétrica

A qualidade do fornecimento de energia elétrica pelas empresas distribuidoras é avaliada com base em três grandes aspectos: qualidade de serviço, qualidade do produto e qualidade comercial. Cada um desses aspectos possui indicadores específicos, que permitem mensurar e regular o desempenho das empresas.

O descumprimento dos limites regulatórios estabelecidos para esses indicadores pode acarretar uma série de sanções, incluindo a obrigatoriedade de elaboração de um plano de resultados, restrições na distribuição de proventos aos acionistas, abertura de processo de caducidade da concessão e a compensação financeira automática às unidades consumidoras afetadas. Essas penalidades criam um incentivo econômico claro e mensurável para que as distribuidoras assegurem o fornecimento de energia dentro dos limites estabelecidos. Esse mecanismo não apenas protege os interesses financeiros das empresas, mas também promove benefícios sociais, garantindo que a população seja atendida com um serviço de qualidade e em conformidade com padrões adequados.

Os três aspectos, seus respectivos indicadores e detalhes adicionais são apresentados nas subseções a seguir.

### 2.4.1 Indicadores de qualidade do serviço

Os Indicadores de Qualidade do Serviço buscam avaliar a continuidade do fornecimento de modo que são avaliados as interrupções do fornecimento, que afetam diretamente a satisfação do consumidor. São aqueles que, na maioria das vezes estão também mais próximos da realidade do consumidor, uma vez que os indicadores individuais podem constar na conta de luz.

Os indicadores de qualidade que foram considerados para esse estudo foram:

- **DEC ou Duração Equivalente de Interrupção por Unidade Consumidora:** é um indicador que representa o tempo médio, em horas, que um conjunto de unidades consumidoras permanece sem fornecimento de energia elétrica (ANEEL, 2024).
- **FEC ou Frequência Equivalente de Interrupção por Unidade Consumidora:** indica a quantidade média de vezes em que o fornecimento foi interrompido nas unidades consumidoras, para cada conjunto de unidades consumidoras (ANEEL, 2024).

Outros indicadores:

- **DIC ou Duração de Interrupção Individual por Unidade Consumidora:** assim como o DEC representa o tempo médio, em horas, que se permanece sem fornecimento de energia, porém em vez de ser a média para unidades consumidores é o valor individual para cada unidade (ANEEL, 2024).

- **FIC Frequência de Interrupção Individual por Unidade Consumidora:** Frequência de interrupções (como o FEC) porém para cada unidade consumidora (ANEEL, 2024).

É possível perceber que a partir do FIC e DIC pode-se fazer uma média para se obter DEC e FEC por isso serão considerados apenas DEC e FEC para esse estudo.

#### 2.4.2 Qualidade do produto

A qualidade do produto diz respeito a qualidade de energia, ou seja, a conformidade de tensão no regime permanente. Existem diversos fatores que são considerados para se avaliar essa qualidade, a amplitude da tensão, sua frequência, a existência de harmônicos, dentre outros. Sendo assim existem índices para quantificar essa qualidade.

Os utilizados nesse estudo são:

- **DRP ou Duração Relativa da Transgressão de Tensão Precária:** percentual do tempo total da medição no qual a unidade consumidora permaneceu com tensão na faixa precária, ou seja tempo que amplitude da tensão ficou abaixo do nível mínimo aceitável (ANEEL, 2024).
- **DRC ou Duração Relativa da Transgressão de Tensão Crítica:** percentual do tempo total da medição no qual a unidade consumidora permaneceu com tensão na faixa crítica, ou seja tempo que amplitude da tensão ficou acima do nível mínimo aceitável (ANEEL, 2024).

Existem também outros indicadores acerca do comportamento da tensão, porém eles são mais específicos para estudos que envolvem análise do comportamento dessas ondas.

#### 2.4.3 Qualidade Comercial

Esse aspecto de qualidade é associado ao cumprimento dos prazos para realização dos serviços comerciais solicitados. De mês em mês as empresas distribuidoras devem encaminhar as demandas registradas pelos usuários para assim apurar os seguintes índices:

- **DER ou Duração Equivalente de Reclamação:** prazo médio que a empresa leva a resolver as reclamações procedentes. (ANEEL, 2024).
- **FER ou Frequência Equivalente de Reclamação a cada 1000 unidades consumidoras:** quantidade de reclamações a cada 1000 unidades consumidoras (ANEEL, 2024).

Como o desejo é analisar se é possível observar uma tendência a partir de dados de interrupção esses índices não serão usados.



### 3 MÉTODO DE PESQUISA

O presente estudo utiliza uma metodologia quantitativa baseada em dados, com foco no setor de energia, especificamente na distribuição elétrica, para avaliar se empresas com menor número de interrupções elétricas são mais valorizadas no mercado de ações ao longo do tempo. Conforme destacado por (BERTRAND; FRANSOO, 2010), um modelo quantitativo fundamenta-se em um conjunto de variáveis que se modificam dentro de um domínio, permitindo o estabelecimento de relações quantitativas e causais. Essa análise vai além da observação dos valores absolutos de mercado, buscando identificar se, a longo prazo, essas empresas apresentam um desempenho de valorização superior em termos percentuais.

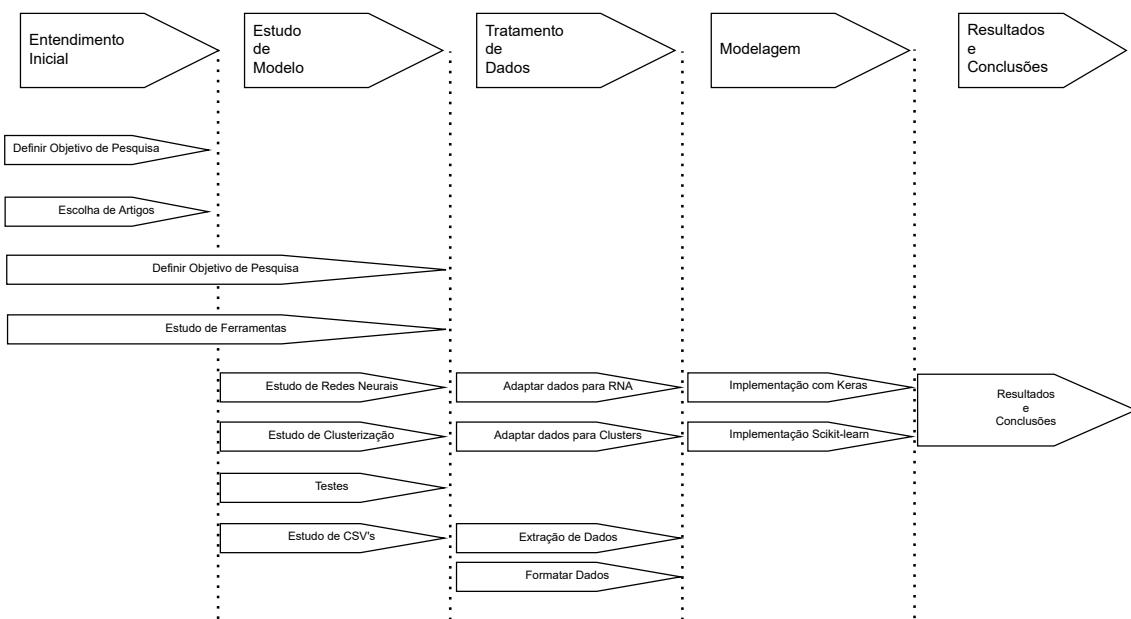
Para tanto, o estudo combina a análise de tendências financeiras com técnicas de aprendizado de máquina e *clustering*, utilizando dados históricos do mercado e dos indicadores de qualidade de serviço elétrico (DEC, FEC, DRP e DRC). A pesquisa é dividida em cinco etapas principais:

1. **Entendimento inicial:** Nessa etapa foram definidos os objetivos que se desejam alcançar com esta pesquisa,
2. **Estudo do modelo:** Foram analisados todos os aspectos necessários para a aquisição, tratamento (quantitativo e qualitativo) e implementação dos dados.
3. **Coleta e Tratamento de Dados:** Foram compiladas informações de empresas do setor de distribuição de energia elétrica listadas na B3, incluindo dados financeiros e indicadores de qualidade de serviço. Optou-se por utilizar as cotações históricas da B3, por ser a fonte oficial para dados de cotações, garantindo confiabilidade, abrangência e um histórico completo para análise. Além disso, foram utilizados dados de interrupções fornecidos pela ANEEL, que é a agência reguladora oficial do setor elétrico no Brasil. Esses dados são indispensáveis para estudos relacionados à qualidade do serviço, pois representam a fonte mais confiável e abrangente sobre indicadores como DEC e FEC. Não há fontes alternativas com a mesma precisão e padronização, o que reforça sua importância para análises consistentes no contexto do setor elétrico. Após a coleta será feita a filtragem e tratamento desses dados. Esses processos serão mais explorados no capítulo 4 deste trabalho.
4. **Modelagem:** Desenvolvimento de uma rede neural que utiliza os índices DEC e FEC como variáveis de entrada para prever a tendência de valorização das ações. Após isso utilização de uma rede neural para prever a valorização das ações com base nos índices de DEC e FEC, oferecendo uma abordagem preditiva e matemática para analisar a evolução do mercado.

**5. Resultados e Conclusões:** Agrupamento das empresas com base nos resultados previstos e análise comparativa dos *clusters* identificados para entender o impacto da qualidade do serviço elétrico na valorização de mercado.

Esse método pode ser resumido pela figura 6

Figura 6 – Sequências de Etapas



Fonte: elaborado pelos autores

Essa figura apresenta uma ordem lógica das etapas realizadas no estudo. No início do diagrama, são mostrados os processos iniciais, como a definição de objetivos e a escolha de artigos. Algumas etapas se iniciam em um estágio e se estendem para outro, como o processo de 'Definir Objetivo de Pesquisa', que se inicia no entendimento inicial e segue até o estudo de modelo. Isso ocorre porque os modelos utilizados e os objetivos do estudo estão interligados, influenciando-se mutuamente. Dessa forma, a figura fornece uma visão geral da sequência das etapas, permitindo compreender como o estudo foi conduzido ao longo do tempo.

## 4 COLETA E TRATAMENTO DE DADOS

### 4.1 Coleta de dados

Para criar os *clusters* e treinar as redes neurais, é fundamental realizar a coleta de dados antes de iniciar o processo de tratamento.

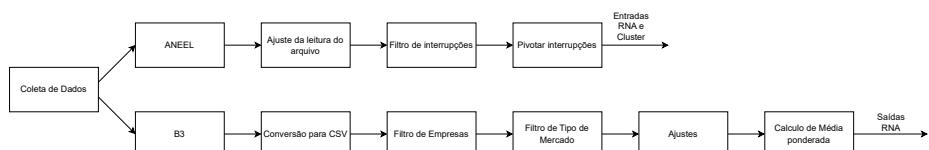
Optou-se pelo uso do formato CSV para armazenar esses dados devido às suas diversas vantagens. Este formato é de fácil leitura e manipulação, além de ser amplamente suportado por linguagens de programação, como Python. Os arquivos CSV são compactos, ocupando menos espaço de armazenamento, e permitem integração simples com *scripts*, o que facilita tanto o pré-processamento quanto a análise.

Os dados financeiros podem ser baixados no site da B3 (B3, 2024) porém ainda é necessário fazer algumas mudanças para converter no formato CSV. Os dados são divididos por ano e contém as cotações de todas as empresas da B3.

Os dados de interrupção são disponibilizados no site da ANEEL divididos por ano. Eles contém todos os tipos de interrupções previstas no módulo 8 do PRODIST ((ANEEL), 2024) referente a cada dia e unidade distribuidora de energia.

O tratamento de dados é essencial para assegurar que as informações utilizadas em uma análise sejam precisas, consistentes e relevantes, aumentando a confiabilidade dos resultados obtidos. Neste trabalho, o tratamento será aplicado aos dados de interrupções e indicadores de continuidade fornecidos pela ANEEL, bem como às cotações históricas disponibilizadas pela B3. O processo geral é representado pelo fluxograma da Figura 7.

Figura 7 – Fluxograma de coleta e tratamento de dados



Fonte: elaborado pelos autores

### 4.2 Dados da B3

#### 4.2.1 Conversão de TxT para CSV

As cotações históricas são disponibilizadas no site da B3 (B3, 2024) em formato TXT, utilizando uma tabulação que dificulta significativamente a interpretação dos dados, como ilustrado na Figura 8.

Para interpretar esse arquivo, é necessário utilizar o guia disponibilizado pela B3 (B3, 2024), no qual é possível identificar o que pode ser descartado, bem como os espaçamentos e a separação de cada parte do TXT para convertê-lo em CSV. Para isso, foi utilizado o *script* 4.1 em Python elaborado pelos autores:

Figura 8 – Formato tabulado

Fonte: Arquivo de demonstração B3

Script 4.1 – Conversão para CSV

```
1 import pandas as pd
2
3
4 arquivo_bovespa = 'Arquivo_Tabulado.txt'
5
6 tamanho_campos
7     =[2,8,2,12,3,12,10,3,4,13,13,13,13,13,13,13,5,18,18,18,13,1,8,7
8 ,13,12,3]
9
10 dados_acoes = pd.read_fwf(arquivo_bovespa, widths=tamanho_campos,
11 encoding='Windows-1252')
12 dados_salvar = 'Arquivo_csv.csv'
13
14 ## Nomear as colunas
15
16 dados_acoes.columns = [
17 "tipo_registro",
18 "data_pregao",
19 "cod_bdi",
20 "cod_negociacao",
21 "tipo_mercado",
22 "noma_empresa",
23 "especificacao_papel",
24 "prazo_dias_merc_termo",
25 "moeda_referencia",
26 "preco_abertura",
27 "preco_maximo",
28 "preco_minimo",
29 "preco_medio",
30 "preco_ultimo_negocio",
31 "preco_melhor_oferta_compra",
32 "preco_melhor_oferta_venda",
33 "numero_negocios",
34 "quantidade_papeis_negociados",
35 "volume_total_negociado",
36 "preco_exercicio",
37 "ndicador_correcao_precos",
38 "data_vencimento",
39 "fator_cotacao",
40 "preco_exercicioPontos",
```

```

40 "codigo_isin",
41 "num_distribuicao_papel"]
42
43 # Eliminar a ultima linha
44 linha=len(dados_acoes["data_pregao"])
45 dados_acoes=dados_acoes.drop(linha-1)
46
47 # Ajustar valores com virgula (dividir os valores dessas colunas por
48 # 100)
49 listaVirgula=[
50 "preco_abertura",
51 "preco_maximo",
52 "preco_minimo",
53 "preco_medio",
54 "preco_ultimo_negocio",
55 "preco_melhor_oferta_compra",
56 "preco_melhor_oferta_venda",
57 "volume_total_negociado",
58 "preco_exercicio",
59 "preco_exercicioPontos"
60 ]
61 for coluna in listaVirgula:
62     dados_acoes[coluna]=[i/100. for i in dados_acoes[coluna]]
63
64 dados_acoes.head()
65
66 dados_acoes.to_csv(dados_salvar, index=False)

```

Após essa etapa, obtém-se um arquivo CSV, que apresenta uma estrutura mais acessível para análise e pode ser facilmente aberto em softwares de planilha, como o *LibreOffice*, conforme ilustrado na Figura 9.

Figura 9 – CSV aberto no LibreOffice

tipo_registro	data_pregao	cod_bdi	cod_negociacao	tipo_mercado	nome_empresa	especificacao_papel	prazo_dias_merc_termo	moeda_referencia	preco_abertura	preco_maximo	preco_minimo	preco_medio	preco_ultimo_negocio
1	200701022.0	CBEE3		10AMPLA ENERG ON *				R\$	0.88	0.89	0.85	0.87	0.86
1	2007010296.0	CEE3F		20COELBA	ON *			R\$	220.0	220.0	220.0	220.0	220.0
1	200701022.0	CEPE5		10CELPE	PNA*EJ			R\$	21.0	21.0	21.0	21.0	21.0
1	2007010296.0	CLSC5F		20CELESC	PNB N2			R\$	33.0	33.0	33.0	33.0	33.0
1	200701022.0	CLSC6		10CELESC	PNB N2			R\$	33.71	34.55	33.71	34.37	34.5
1	2007010296.0	CLSC6F		20CELESC	PNB N2			R\$	34.2	34.5	34.0	34.27	34.4
1	200701022.0	CMIG3		10CEMIG	ON * N1			R\$	91.2	92.0	91.04	91.47	91.82
1	2007010296.0	CMIG3F		20CEMIG	ON * N1			R\$	91.0	91.3	90.26	91.0	91.27
1	200701022.0	CMIG4		10CEMIG	PN * N1			R\$	105.0	105.9	103.7	104.65	105.9
1	2007010296.0	CMIG4F		20CEMIG	PN * N1			R\$	106.0	106.0	103.4	104.58	105.35
1	2007010262.0	CMIG4T		30CEMIG	PN * N1	30.0		R\$	105.9	105.91	105.9	105.9	105.91
1	2007010278.0	CMIGB16		70CMIG	PN * N1	0.0		R\$	3.62	3.62	3.62	3.62	3.62
1	200701022.0	COCE5		10COELCE	PNA*			R\$	11.57	11.89	11.5	11.65	11.65
1	2007010296.0	COCE5F		20COELCE	PNA*			R\$	11.8	11.9	11.51	11.79	11.51
1	200701022.0	CPFE3		10CPFL ENERGIA ON	NM			R\$	30.05	30.7	30.0	30.35	30.09
1	2007010296.0	CPFE3F		20CPFL ENERGIA ON	NM			R\$	29.9	30.52	29.9	30.26	30.2
1	2007010262.0	CPFE3T		30CPFL ENERGIA ON	NM	30.0		R\$	30.72	31.04	30.72	30.85	31.04
1	2007010262.0	CPFE3T		30CPFL ENERGIA ON	NM	50.0		R\$	30.79	30.8	30.68	30.69	30.69
1	2007010262.0	CPFE3T		30CPFL ENERGIA ON	NM	90.0		R\$	31.0	31.08	31.0	31.05	31.08
1	200701022.0	CPL3		10COPEL	ON *			R\$	21.85	22.0	21.71	21.92	21.99
1	2007010296.0	CPL3F		20COPEL	ON *			R\$	21.85	21.85	21.85	21.85	21.85
1	2007010262.0	CPL3T		30COPEL	ON *	30.0		R\$	22.19	22.2	22.18	22.19	22.19

Fonte: Arquivo de demonstração B3

#### 4.2.2 Filtragem de Empresas

O arquivo transformado em CSV contém todas as empresas listadas na Bolsa de Valores Brasileira, com foco específico nas associadas ao mercado de energia elétrica (B3, 2024). Foi criado um CSV contendo o nome de todas as empresas que atuam no mercado em 2024. A partir desse primeiro CSV, realiza-se a filtragem da base de dados, comparando as *strings* de cada cédula. Para essa tarefa, utiliza-se o *script* 4.2 em Python:

Script 4.2 – Filtragem do CSV

```

1 import pandas as pd
2
3
4 # Carregar o primeiro CSV (com os valores da bolsa)
5 df_valores = pd.read_csv('Arquivo_csv.csv', sep=',')
6
7 #Carregar o segundo CSV (com os codigos das empresas elericas)
8 df_empresas = pd.read_csv('Empresas_eletricas.csv', sep=',')
9
10 #Coluna com o nome das Empresas
11 lista_codigos = df_empresas['Codigo'].tolist()
12
13 #Funcao para verificar se o codigo da bolsa comeca com um codigo da
14 #lista
15 def filtrar_codigos(codigo):
16     return any(codigo.startswith(c) for c in lista_codigos)
17
18 #Filtrar o primeiro DataFrame usando a coluna correta para comparacao
19 df_filtrado = df_valores[df_valores['cod_negociacao'].apply(
20     filtrar_codigos)]
21
22 df_filtrado.to_csv('Arquivo_filtrado.csv', index=False)

```

#### 4.2.3 Tipo de Mercado

O autores (BODIE; KANE; MARCUS, 2015) classificam os ativos financeiros em três grandes grupos: títulos de renda fixa, ações e derivativos. O primeiro grupo representa essencialmente dívidas, nas quais o tomador do empréstimo se compromete a pagar uma quantia fixa ou vinculada a um índice.

O segundo grupo, as ações, corresponde a uma participação na propriedade da corporação. Nesse caso, não há promessa de pagamento ao comprador, e a remuneração está condicionada ao aumento do valor patrimonial da empresa.

Por sua vez, os derivativos, que englobam principalmente o mercado de opções e contratos futuros, oferecem retornos baseados no preço de outros ativos. Eles são amplamente utilizados para proporcionar proteção ou transferir riscos, a um custo, para outras partes.

Como o objetivo deste trabalho é avaliar como os indicadores de qualidade de energia podem interferir na especificação das empresas, utilizaremos apenas a classe de ações, uma vez que a cotação das empresas reflete, em alguma medida, a qualidade do serviço entregue pelas distribuidoras de energia.

Ao avaliarmos o mercado de renda variável no Brasil, recorremos à bolsa de valores B3, onde os tipos de mercado são identificados por códigos que podem ser observados na Tabela 1, conforme informações provenientes de (B3, 2024).

Tabela 1 – Relação dos Valores para Tipo de Mercado (TPMERC)

Código	Descrição
010	Vista
012	Exercício de Opções de Compra
013	Exercício de Opções de Venda
017	Leilão
020	Fracionário
030	Termo
050	Futuro com Retenção de Ganho
060	Futuro com Movimentação Contínua
070	Opções de Compra
080	Opções de Venda

Fonte: Elaborado pelos autores

Analizando esses dados, verificamos que os códigos 010, 017 e 020 correspondem ao mercado acionário, enquanto os demais códigos estão relacionados ao mercado de derivativos, que não faz parte do escopo deste trabalho.

Dentre os códigos analisados, o mercado identificado pelo código 010, o mercado à vista, é o principal, representando mais de 80% das negociações.

Por outro lado, o mercado fracionário (código 020) ainda apresenta taxas consideráveis e é predominantemente utilizado por pequenos investidores, o que dificulta a proposição de aproximação a um mercado eficiente. Já o mercado de leilão (código 017) é utilizado para a entrada de empresas na bolsa, mas suas informações históricas não estão plenamente disponíveis. Por essa razão, esses mercados não serão considerados nesta análise, a fim de minimizar possíveis distorções.

#### 4.2.4 Ajustes Finais

Os ajustes finais são realizados para aprimorar a visualização dos dados e torná-los mais adequados para uso nas análises. Esses ajustes incluem: consolidar dados de diversos anos provenientes dos arquivos CSV, remover colunas irrelevantes, separar corretamente as datas no formato desejado e agrupar as cotações, que serão utilizadas como saídas da rede neural. As funções podem ser observadas no Script 4.3.

### Script 4.3 – Funções Úteis

```

1
2 def juntar_csvs(caminho_arquivos, caminho_saida):
3     # Lista explicita de arquivos CSV
4     ano = 2010
5     arquivos_csv = [
6         f'Cota_{ano}_tipo10.csv',
7         f'Cota_{ano+1}_tipo10.csv',
8         f'Cota_{ano+2}_tipo10.csv',
9         f'Cota_{ano+3}_tipo10.csv',
10        f'Cota_{ano+4}_tipo10.csv',
11        f'Cota_{ano+5}_tipo10.csv',
12        f'Cota_{ano+6}_tipo10.csv',
13        f'Cota_{ano+7}_tipo10.csv',
14        f'Cota_{ano+8}_tipo10.csv',
15        f'Cota_{ano+9}_tipo10.csv',
16    ]
17
18    lista_dataframes = []
19
20    for arquivo in arquivos_csv:
21        # Cria o caminho completo para cada arquivo
22        caminho_completo = f'{caminho_arquivos}\\{arquivo}'
23
24        # Le o CSV e adiciona a coluna 'Ano' com o nome do arquivo
25        ano = arquivo.split('.')[0]
26        df = pd.read_csv(caminho_completo)
27        df['Ano'] = ano # Adiciona a coluna do ano
28        lista_dataframes.append(df)
29
30    # Concatena todos os DataFrames em um unico DataFrame
31    df_final = pd.concat(lista_dataframes, ignore_index=True)
32
33    # Salva o DataFrame final em um novo arquivo CSV
34    df_final.to_csv(caminho_saida, index=False)
35
36    print("CSV combinado criado com sucesso!")

```

Separar valores de data pela coluna data\_pregao como por exemplo, 20210102 que representa a data 02/01/2021 (em dia, mês e ano) no Script 4.4

### Script 4.4 – Separar datas

```

1
2 def processar_data_pregao(caminho_entrada, caminho_saida):
3     # Carrega o CSV em um DataFrame
4     df = pd.read_csv(caminho_entrada)
5

```

```

6   # Verifica se a coluna 'data_pregao' existe no DataFrame
7   if 'data_pregao' not in df.columns:
8       print("A coluna 'data_pregao' não foi encontrada no arquivo.")
9       return
10
11  # Converte a coluna 'data_pregao' para string (caso não seja)
12  df['data_pregao'] = df['data_pregao'].astype(str)
13
14  # Cria as colunas 'Ano' e 'Mes'
15  df['Ano'] = df['data_pregao'].str[:4].astype(int)
16  df['Mes'] = df['data_pregao'].str[4:6].astype(int)
17
18  # Salva o DataFrame modificado em um novo CSV
19  df.to_csv(caminho_saida, index=False)

```

Serão utilizados os preços médios de cada transação, considerando que uma mesma empresa pode ser negociada de várias formas diferentes, cada uma com um código de negociação distinto. Em cada uma dessas negociações, os valores das transações e a quantidade de papéis negociados variam. Assim, ao calcular essas médias, é necessário ponderar os valores pelos respectivos números de papéis negociados, garantindo que os resultados refletem proporcionalmente o volume de cada transação.

A forma que a empresa é negociada é indicada por um código de negociação, cuja descrição pode ser consultada na Tabela 2, busca-se calcular o preço médio mensal proporcional dessa empresa. Para isso, utiliza-se a média ponderada entre os preços médios de cada código de negociação e o número de papéis negociados por cada código ao longo do mês.

Com essa abordagem, consolida-se toda a informação de preço da empresa, considerando todos os mercados nos quais ela é negociada. Essa padronização permite igualar o espectro temporal aos valores regulatórios dos indicadores de qualidade de energia, que também são reportados em escala mensal.

Tabela 2 – Classificação dos Códigos de Ações e Direitos

Código	Descrição
1	Direito de Subscrição – Ação Ordinária
2	Direito de Subscrição – Ação Preferencial
3	Ações Ordinárias
4	Ações Preferenciais
5	Ações Preferenciais Classe A
6	Ações Preferenciais Classe B
7	Ações Preferenciais Classe C
8	Ações Preferenciais Classe D
9	Recibo de Subscrição – Ação Ordinária
10	Recibo de Subscrição – Ação Preferencial
11	BDRs e Units

Fonte: Elaborado pelos autores

#### Script 4.5 – Agrupar e calcular média ponderada

```

1
2 def calcular_media_ponderada(caminho_csv, caminho_saida):
3     # Carregar o CSV
4     df = pd.read_csv(caminho_csv)
5
6     # Agrupar os dados por Ano, Mes e nome_empresa
7     grouped = df.groupby(['Ano', 'Mes', 'nome_empresa'])
8
9     # Calcular a media ponderada de 'preco_medio' usando
10    'papeis_neg' como peso
11    def media_ponderada(grupo):
12        return (grupo['preco_medio'] * grupo['papeis_neg']).sum() /
13            grupo['papeis_neg'].sum()
14
15    # Aplica a fun ao de media ponderada para cada grupo
16    df_ponderada = grouped.apply(m_ponderada).reset_index(name='
17        v_ponderado')
18
18    # Salvar o resultado em um novo arquivo CSV
19    df_media_ponderada.to_csv(caminho_saida, index=False)

```

Tabela 3 – Formato da Tabela após tratamento de dados

Ano	Mes	noma_empresa	preco_medio ponderado
2010	1	AMPLA ENERG	1.6983301727770754
2010	1	CEB	26.7136065557377048
2010	1	CEE-E-D	3.7253333333333334
2010	1	CEE-GT	2.4622608695652173
2010	1	CELESC	36.43879749041478
2010	1	CELPE	45.55333333333335
2010	1	CEMIG	30.159496116951072
2010	1	COELBA	31.736808510638298
2010	1	COELCE	30.85893876706661
2010	1	COPEL	38.04660248997145
2010	1	COSERN	9.736476190476191
2010	1	CPFL ENERGIA	36.53481566484034
2010	1	ELEKTRO	21.1922
2010	1	ELETROBRAS	36.734611458078305

Fonte: Elaborado pelos autores

### 4.3 Dados da ANEEL

A ANEEL disponibiliza os dados em formato de CSV no seu site oficial ([ANEEL, 2024](#)), os dados vêm formatados como mostrado pela figura 10

Figura 10 – Exemplo de formatação dados ANEEL

DatGeracaoConjuntoDados	SigAgente	NumCNPJ	IndeConjUndConsumidoras	DscConjUndConsumidoras	SigIndicador	AnoIndice	NumPeriodoIndice	VlrIndiceEnviado
2024-11-05 EAC		4065033000170	12595 Cruzeiro do Sul		FECIP	2010	4	1
2024-11-05 EAC		4065033000170	12599 Porto Walter		DECIPC	2010	12	0
2024-11-05 EAC		4065033000170	12590 TAQUARI		DECXNC	2010	3	0
2024-11-05 CERR		5938444000196	12315 CANTA		FEC	2010	2	0
2024-11-05 CERR		5938444000196	12323 SAO LUIS DO ANUA		DECXP	2010	10	0
2024-11-05 CERR		5938444000196	12323 SAO LUIS DO ANUA		DEC	2010	9	16 77
2024-11-05 CEA		5965546000109	7322 MAZAGAO		DECIPC	2010	11	0
2024-11-05 CEA		5965546000109	7325 SERRA DO NAVIO		DECIPC	2010	10	0
2024-11-05 CEA		5965546000109	7318 PORTO GRANDE		FECXN	2010	7	0
2024-11-05 ETQ		25086034000171	12154 Grande Ponte Alta do TO		DECINO	2010	4	0
2024-11-05 ETQ		25086034000171	12159 Grande Natividade		DECIND	2010	4	3 43
2024-11-05 ETQ		25086034000171	12156 Grande Guara n o Urbano		FECIND	2010	2	4 8
2024-11-05 ETQ		25086034000171	12151 Grande Axa do TO		DECXP	2010	7	0
2024-11-05 ETQ		25086034000171	11738 Formoso		FECIND	2010	11	5 41
2024-11-05 ETQ		25086034000171	11763 Porto Nacional Urbano		DECIND	2010	9	2 7
2024-11-05 ETQ		25086034000171	12134 Grande Presidente Kennedy		FEC	2010	1	5 41
2024-11-05 ETQ		25086034000171	6433 DJANPOLIS		FECIND	2010	10	1 96
2024-11-05 ETQ		25086034000171	11731 Colinas n o urbano		FECXN	2010	3	0
2024-11-05 ETQ		25086034000171	12144 Grande F lima		FECXN	2010	11	0
2024-11-05 ETQ		25086034000171	12156 Grande Guara n o Urbano		FECING	2010	1	59
2024-11-05 ETQ		25086034000171	11731 Colinas n o urbano		DECXNC	2010	1	0
2024-11-05 ETQ		25086034000171	6494 MIRANTE		DECXP	2010	8	0
2024-11-05 ETQ		25086034000171	11731 Colinas n o urbano		FECXPC	2010	6	0
2024-11-05 ETQ		25086034000171	12129 Grande Palmeira polis		FECXNC	2010	9	0
2024-11-05 ETQ		25086034000171	11763 Porto Nacional Urbano		DECINC	2010	7	0

Fonte: ([ANEEL, 2024](#))

É evidente que os dados ainda precisam passar por diversos processos de filtragem e formatação. Existem caracteres que não são identificados, múltiplos tipos de interrupções que não serão utilizados e uma coluna sem índice, que representa os decimais da coluna VlrIndiceEnviado. O arquivo que contém as explicações de todas as colunas está disponível no site da ANEEL ([ANEEL, 2024](#)).

Outro desafio é o número de linhas do arquivo CSV. Como o arquivo contém uma grande quantidade de dados, nem todas as funções descritas podem ser utilizadas a qualquer momento do processo. Isso ocorre porque arquivos extensos podem demandar uma quantidade significativa de memória RAM do computador usado para o tratamento de dados. Portanto, é essencial considerar a ordem em que esse tratamento é realizado.

#### 4.3.1 Ajuste da leitura do arquivo

O primeiro passo é fazer com que o arquivo seja lido a partir de agora com as separações usando ponto e vírgula e as casas decimais separadas por vírgula o que é feito de forma simples com a função no script 4.6:

Script 4.6 – Separar por ponto e vírgula

```

1
2 def processar_csv(caminho_entrada, caminho_saida):
3
4     # Carregar o CSV com o delimitador correto e o encoding adequado
5     df = pd.read_csv(caminho_entrada, encoding='latin1', sep=';')
6
7     # Salvar o CSV processado no caminho de saída
8     df.to_csv(caminho_saida, index=False)

```

Fonte: elaborado pelos autores

#### 4.3.2 Filtrar interrupções desejadas

Para a rede neural, são utilizados os dados de DEC e FEC, pois esses indicadores são amplamente reconhecidos como representativos da qualidade do fornecimento de energia elétrica. Eles fornecem uma medida objetiva da confiabilidade e resiliência da rede elétrica, influenciando diretamente o desempenho e a análise do setor como um todo. Com isso o Script 4.7 tira as linhas que não são referentes a esses indicadores.

Script 4.7 – Filtrar interrupções

```

1 def Apenas_DEC_FEC(caminho_entrada, caminho_saida):
2     df = pd.read_csv(caminho_entrada)
3
4     # Filtrar as linhas onde a coluna 'SigIndicador'
5     df_filtrado = df[df['SigIndicador'].isin(['DEC', 'FEC'])]
6
7     # Salvar o DataFrame filtrado em um novo arquivo CSV
8     df_filtrado.to_csv(caminho_saida, index=False)

```

Fonte: elaborado pelos autores

### 4.3.3 Tratamento de DRP e DRC

Os dados de DRP e DRC são utilizados no processo de clusterização, mas estão estruturados de forma diferente na base de dados. Esses parâmetros são apresentados por trimestre e por unidade consumidora, ao contrário de DEC e FEC, que estão agrupados pelo conjunto dessas unidades. Dessa forma, o filtro para DRP e DRC segue uma lógica semelhante ao utilizado para DEC e FEC, mas com algumas adaptações. É necessário agrregar os valores de DRP e DRC por unidade consumidora em um único conjunto (calculando a média). Paralelamente, os valores de DEC e FEC precisam ser reorganizados em trimestres para que a base de dados mantenha consistência e coerência.

Portanto a base de dados será trimestral para os *clusters* e será mensal para a rede neural.

### 4.3.4 Tirar colunas não usadas e agrupar valores de DEC, FEC, DRP e DRC

Para tirar as colunas foi utilizada uma função muito parecida com a que removia colunas indesejadas da seção anterior. O agrupamento será feito pegando todos os valores de mesmo CNPJ, mês, ano e nome da empresa, fazendo a média dos valores de DEC e FEC das diversas unidades consumidores que tem o mesmo CNPJ, tem-se o DEC e o FEC para esse CNPJ como todo.

Esse código pode ser visto no script 4.8:

Script 4.8 – Retirar colunas

```

1
2 df = pd.read_csv(caminho_entrada, low_memory=False)
3
4 # Substituir virgulas por pontos
5 df['VlrIndiceEnviado'] =
6 df['VlrIndiceEnviado'].str.replace(',', '.', regex=False)
7
8 # Converter para numerico
9 df['VlrIndiceEnviado'] =
10 df['VlrIndiceEnviado'].astype(float)
11
12 #Verificar valores
13 if df['VlrIndiceEnviado'].isna().sum() > 0:
14     print(
15         "Aviso: não pode ser convertido"
16
17 # Agrupar pelas colunas:
18 # Calcular a media de 'VlrIndiceEnviado' para cada grupo
19 df_agrupado =
20 df.groupby(['SigAgente', 'NumCNPJ',
21 'SigIndicador', 'AnoIndice'],

```

```

22     'NumPeriodoIndice'], as_index=False)[
23         'VlrIndiceEnviado'].mean()
24
25     # Renomear as colunas para algo mais amigavel
26     df_agrupado.columns = ['SigAgente', 'NumCNPJ', 'SigIndicador',
27     'AnoIndice', 'NumPeriodoIndice',
28     'VlrIndiceEnviado']
29
30     # Salvar o DataFrame resultante em um novo arquivo CSV
31     df_agrupado.to_csv(caminho_saida, index=False)

```

Fonte: elaborado pelos autores

#### 4.3.5 Tirar DEC e FEC de linhas e por como colunas

Por fim, foi feita a pivotagem dos dados de DEC e FEC para facilitar como descrever a entrada da rede neural, essa pode ser observada no script 4.9.

Script 4.9 – Pivotagem

```

1
2 # Carregar o arquivo CSV
3 df_entradas = pd.read_csv(caminho_entrada)
4
5     # Criar os DataFrames para DEC e FEC
6     df_dec =
7     df_entradas[df_entradas['SigIndicador'] ==
8     'DEC'].copy() # Filtrar para DEC
9     df_fec =
10    df_entradas[df_entradas['SigIndicador'] ==
11    'FEC'].copy() # Filtrar para FEC
12
13    # Renomear a coluna 'VlrIndiceEnviado' para 'DEC' ou 'FEC'
14    df_dec = df_dec.rename(columns={'VlrIndiceEnviado': 'DEC'})
15    df_fec = df_fec.rename(columns={'VlrIndiceEnviado': 'FEC'})
16
17    # Mesclar os dois DataFrames:
18    df_completo = pd.merge(
19        df_dec[['AnoIndice', 'NumPeriodoIndice',
20        'SigAgente', 'NumCNPJ', 'DEC']],
21        df_fec[['AnoIndice', 'NumPeriodoIndice',
22        'SigAgente', 'NumCNPJ', 'FEC']],
23        on=['AnoIndice', 'NumPeriodoIndice', 'SigAgente', 'NumCNPJ'],
24        how='outer'
25    )
26
27    # Substituir os valores ausentes com 0
28    df_completo['DEC'] = df_completo['DEC'].fillna(0)

```

```

29 df_completo['FEC'] = df_completo['FEC'].fillna(0)
30
31 # Salvar o DataFrame resultante em um novo arquivo CSV
32 df_completo.to_csv(caminho_saida, index=False)

```

Fonte: elaborado pelos autores

#### 4.3.6 Montar a base de dados para a rede neural

Para montar a base de dados que será utilizada na RNA é necessário filtrar o CSV com DEC e FEC alinhado com apenas as 22 empresas citadas. E depois cruzar os dados finais da B3 também filtrados com essas 22 empresas. Uma parte desse CSV pode ser observado na tabela 4

Tabela 4 – Tabela de preços e indicadores DEC e FEC

Ano	CNPJ	Preço médio ponderado	DEC	FEC
2010	8467115000100	3.725	5,73	1.506
2010	15139629000194	31.737	3.447	1.123
2010	8324196000181	9.736	2.247	1.060
2010	2328280000197	21.192	1.686	0.921
2010	61695227000193	35.284	3.026	1.094

Fonte: Elaborado pelos autores

A entrada da rede será os valores de DEC e FEC e a saída o preco\_medio ponderado.



## 5 MODELAGEM E RESULTADOS

### 5.1 Clusterização

Foi realizada a clusterização trimestral (por *quarter*) das 22 empresas do setor elétrico no período de 2010 a 2019, utilizando quatro dimensões: os indicadores DEC, FEC, DRP e DRC. A análise revelou três tendências principais na distribuição das empresas, que foram observadas por meio de *pairplots*.

A primeira tendência identificada corresponde à maioria das empresas, que se concentram nos extremos inferiores das dimensões, geralmente próximas a zero. Esse grupo representa, em um dado *quarter*, empresas com pouquíssimos casos de interrupção e alta conformidade de tensão em regime permanente, se aproximando do ideal. Esse *cluster* foi denominado como *cluster 0*.

O segundo grupo apresentava valores intermediários nas dimensões, indicando a ocorrência de possíveis incidentes. No entanto, a maioria das empresas desse grupo manteve seus indicadores dentro dos limites regulatórios estabelecidos. Esse *cluster* foi denominado como *cluster 2*.

Por fim, foi identificado um terceiro grupo composto por uma minoria de *outliers*, caracterizados por parâmetros significativamente inadequados, com valores elevados e fora dos padrões esperados. Esse *cluster* foi denominado como *cluster 1*.

Com base nesses padrões, foi utilizado o método elbow para determinar o número ideal de *clusters*, resultando em três agrupamentos. Foram definidos três centróides iniciais: dois posicionados próximos aos extremos superiores e inferiores das dimensões e um no ponto médio entre esses extremos, representando os máximos e mínimos locais e os valores intermediários observados.

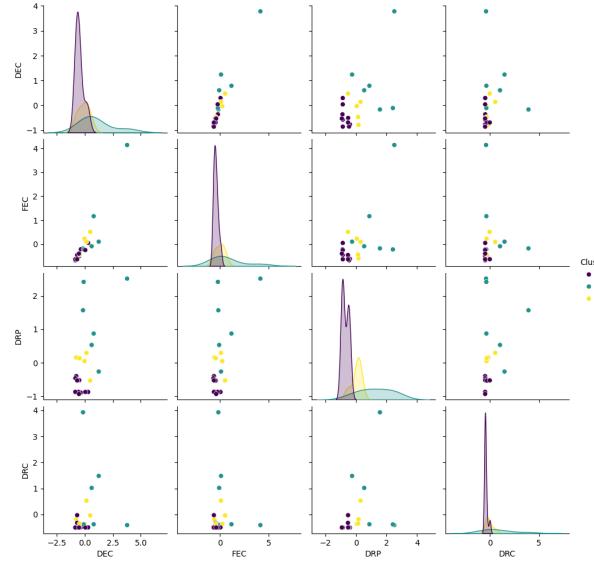
O resultado obtido ao longo dos quartos foi sistematizado no gráfico 12 abaixo inspirado em um mapa de calor, foi somado o número correspondente a nota de cada *cluster* de forma a ordenar da empresa com a maior posição, e que em média entregou os melhores resultados até a menor.

A tabela apresenta um mapeamento das empresas classificadas como "boas", "medianas" ou "ruins" com base em seus índices de interrupções. As empresas "boas", são as que se apresentam mais no topo da tabela, com a maioria dos trimestres destacadas em verde (ou representadas pelo número 2), possuindo os menores índices de interrupção, enquanto as "ruins" ficam embaixo em vermelho, com os trimestres com os piores índices de interrupção. As empresas "medianas" são as do meio do eixo y, e contêm majoritariamente trimestres em amarelo.

Foi analisado também o percentual de tempo que cada empresa fica em cada *cluster*. Os resultados podem ser analisado na tabela 5

A tabela 5 representa a porcentagem de trimestres a empresa permaneceu com indica-

Figura 11 – Décimo Sétimo Quarter



Fonte: Elaborado pelos autores

Figura 12 – Formato tabulado

CNPJ	Distribuição de Clusters por Empresa (2010 a 2019) - Ordenado por Desempenho																																												
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39						
	Quarter																																												
CEMIG DISTRIBUICAO	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
COPEL DISTRIBUICAO	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
LIGHT SERVICOS DE ELETRICIDADE	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
COSERN	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
AMPLA ENERGIA E SERVICOS	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
COMPANHIA PAULISTA DE FORCA E LUZ	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
ELEKTRO REDES S.A.0	2	2	2	2	2	2	1	1	1	1	1	2	1	1	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
ENEL CEARA	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
ENERGISA MINAS RIO	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
CPF1 PIRATININGA	2	2	2	2	2	2	2	2	2	2	2	1	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
ELETROPAULO	1	2	1	1	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
EDP ESPIRITO SANTO	1	2	1	1	2	1	2	1	2	1	1	1	2	0	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1		
EDP SAO PAULO	2	2	1	1	2	1	2	1	2	1	1	1	2	1	2	0	1	2	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
ENERGISA SERGIPE	2	1	1	1	2	1	1	2	1	2	2	1	0	0	0	1	1	2	1	0	1	1	1	1	2	2	2	1	2	2	1	2	0	2	0	2	0	1	2	2	0	2			
COELBA	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	0	1	2	1	2	2	2	1	2	2	2	2	0	1	0	0	1	0	0	0	0	0	0			
ENERGISA MATO GROSSO DO SUL	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	2	1	2	2	2	1	2	2	2	2	2	2	0	0	1	0	0	0	0	0	0	0	
Companhia Energética de Pernambuco	1	1	1	2	1	1	2	2	2	2	1	2	2	1	1	0	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
CEE-E-D	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	1	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
EQUATORIAL MARANHAO	1	0	0	0	1	0	0	0	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
ENERGISA MATO GROSSO	2	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
RGE SUL	1	1	1	1	0	0	0	0	1	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
EQUATORIAL PARA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fonte: Elaborado pelos autores

dores bons ruins ou médios.

Analizando os 858 desempenhos da figura 12, junto com as porcentagens da tabela 5, observa-se que 49,65% estão classificados como bons, 32,40% como medianos e 15,95% como ruins. Apenas cinco empresas apresentam desempenho bom em mais de 80% do tempo. Esse panorama pode indicar que o setor elétrico nacional ainda enfrenta desafios significativos, especialmente no que diz respeito aos investimentos em infraestrutura, que são necessários para garantir um regime de abastecimento elétrico estável e permanente. As empresas localizadas na parte inferior da figura 12 passaram, em sua maioria, por

Tabela 5 – Distribuição de tempo das empresas nos clusters.

<b>Empresa</b>	<b>%bom</b>	<b>%ruim</b>	<b>%médio</b>
CEMIG DISTRIBUICAO S.A.	97,44%	0,00%	2,56%
COPEL DISTRIBUIÇÃO S.A.	94,87%	0,00%	5,13%
LIGHT SERVICOS DE ELETRICIDADE S.A.	92,31%	0,00%	7,69%
CIA ENERGETICA DO RIO GDE NORTE - COSERN	87,18%	0,00%	12,82%
AMPLA ENERGIA E SERVICOS S.A.	84,62%	5,13%	10,26%
CIA PAULISTA DE FORCA E LUZ	69,23%	0,00%	30,77%
ELEKTRO REDES S.A.	66,67%	2,56%	30,77%
CIA ENERGETICA DO CEARA - COELCE	61,54%	2,56%	35,90%
ENERGISA MINAS RIO	58,97%	0,00%	41,03%
CIA PIRATININGA DE FORCA E LUZ	48,72%	0,00%	51,28%
CIA ESTADUAL DE DISTRIB ENER ELET-CEEE-D	46,15%	41,03%	12,82%
ELETROPAULO METROP. ELET. SAO PAULO S.A.	46,15%	0,00%	53,85%
EDP ESPIRITO SANTO	43,59%	2,56%	53,85%
ENERGISA SERGIPE	41,03%	12,82%	46,15%
CIA ELETRICIDADE EST. DA BAHIA - COELBA	35,90%	12,82%	51,28%
ENERGISA MATO GROSSO DO SUL	33,33%	23,08%	43,59%
EDP SÃO PAULO	30,77%	2,56%	66,67%
CIA ENERGETICA DE PERNAMBUCO - CELPE	28,21%	23,08%	48,72%
ENERGISA MATO GROSSO	12,82%	74,36%	12,82%
EQUATORIAL MARANHÃO	12,82%	25,64%	61,54%
RGE SUL	0,00%	66,67%	33,33%
EQUATORIAL PARA	0,00%	100,00%	0,00%

Fonte: Elaborado pelos autores

processos de fusão e aquisição. Essa dinâmica reflete uma reestruturação recente no setor, onde grandes conglomerados têm incorporado operações regionais para maximizar a eficiência. Entretanto, será necessário realizar estudos mais aprofundados para avaliar a eficácia dessa estratégia a longo prazo.

Ao analisar a empresa Equatorial Energia, sua estratégia torna-se evidente. O processo teve início com a aquisição da antiga CEMAR, que, no estudo de clusterização, aparece como "Equatorial Maranhão". O perfil da companhia é descrito como tendo uma "experiência diferenciada em reestruturação operacional e financeira de companhias no setor elétrico brasileiro" (Equatorial Energia S.A., 2013).

A estratégia de aquisições continuou em 2012, com a compra da CELPA, considerada a empresa com a pior classificação no *clustering*. Com isso, a Equatorial passou a deter duas das cinco distribuidoras classificadas negativamente no estudo.

Nos anos seguintes, a empresa ampliou sua atuação com novas aquisições. Em 2018, adquiriu as distribuidoras de Alagoas e Piauí, que não entraram nos critérios de clusterização. Em 2021, a companhia deu sequência à estratégia, comprando a CEEE-D, também mal classificada, e a CEA.

Dessa forma, a estratégia da Equatorial Energia está fundamentada na aquisição de empresas mal estruturadas, como evidenciado pelos indicadores analisados. A companhia demonstra foco em realizar “*turnarounds*” e reverter a situação dessas distribuidoras (Equatorial Energia S.A., 2022).

A empresa Neoenergia pode ser observada nos *clusters* por meio de suas distribuidoras Elektro, COSERN, CELPE e COELBA. Algumas dessas empresas, como a COSERN e a Elektro, apresentam boas classificações, refletindo uma estratégia mais voltada para a consolidação no mercado. A Neoenergia destaca compromissos como “disciplina de custos” “continua melhoria operacional assegurando o atendimento das exigências regulatórias” e “firme plano de recuperação e combate ao furto de energia” (Neoenergia Distribuição Brasília S.A., 2020).

Embora reconheça os desafios apresentados pelos indicadores de qualidade de energia, a empresa, com distribuidoras consolidadas há mais de 20 anos, prioriza manter-se bem estabelecida no mercado, em vez de focar em uma expansão agressiva. A COSERN, a COELBA e a CELPE já fazem parte do grupo desde 2000, enquanto a Elektro foi incorporada apenas em 2011. Mais recentemente, o grupo adquiriu a CEB, reforçando seu perfil consolidado.

Dessa forma, a Neoenergia apresenta uma estratégia voltada mais para a gestão e eficiência, em contraste com o foco em crescimento acelerado.

O grupo Energisa, representado no estudo pelas empresas Energisa Mato Grosso, Energisa Mato Grosso do Sul, Energisa Sergipe e Energisa Minas Rio, apresenta a maior parte de suas distribuidoras posicionadas na parte inferior da tabela de desempenho. Relatórios destinados a investidores frequentemente apontam que algumas dessas empresas operam fora dos limites regulatórios, o que impacta negativamente os resultados financeiros do grupo. Além disso, mencionam a ocorrência de eventos sistêmicos que afetam o desempenho.

Esses desafios refletem a característica interligada do sistema de distribuição com os sistemas de transmissão e geração, onde regiões afastadas do Centro-Sul enfrentam sérios problemas devido à falta de infraestrutura elétrica adequada. Essa realidade é evidenciada pelo fato de a empresa com melhor desempenho do grupo ser a que atua na região Sudeste, onde as condições estruturais são mais favoráveis (S.A., 2019).

Por outro lado, as empresas na parte superior da figura parecem mais consolidadas. Apesar de suas rentabilidades nem sempre superarem as das empresas anteriormente mencionadas, elas passaram por menos processos de fusão e aquisição. Muitas dessas empresas também se destacam pela alta distribuição de dividendos, uma estratégia que aparentemente visa atrair investidores de longo prazo.

Dessa forma, o país passa por um processo de consolidação das empresas de distribuição, com os grupos Equatorial, Energisa, CPFL e Neoenergia atendendo cerca de 93 milhões de clientes, de forma que mais de 15 empresas se tornaram 4.

## 5.2 Rede Neural

### 5.2.1 Anos de Previsão

Para a realização da análise, os dados do período de 2010 a 2019 foram utilizados como base para o treinamento da rede neural, enquanto os anos de 2020 a 2023 foram reservados para a etapa de previsão. A entrada da rede neural consiste exclusivamente nos indicadores DEC e FEC, associados a cada mês, ano e CNPJ das empresas.

### 5.2.2 Indicadores

Como se deseja se prever o valor médio de cada mês para uma determinada empresa, os indicadores DRC e DRP não serão utilizados como entrada da rede neural, uma vez que até 2020 eles se apresentam aglomerados em trimestres.

### 5.2.3 Implementação

Vale destacar que, após a aplicação dos filtros necessários e a verificação de quais empresas estavam ativas e possuíam dados adequados para a análise, foi possível selecionar nove empresas para a previsão dos resultados entre os anos de 2020 e 2024.

O processo para definir a quantidade de camadas e neurônios da rede neural envolve o tipo de função que se espera que associe a entrada com a saída, como não há como saber qual é esse tipo de função, foi desenvolvido um processo de tentativa e erro visando minimizar o MAE apresentado pela rede. Ao considerar esses fatores o Script 5.1 foi elaborado

Script 5.1 – RNA elaborada

```

1
2 import pandas as pd
3 from sklearn.model_selection import train_test_split
4 from tensorflow import keras
5 from tensorflow.keras import layers
6
7 # Carregar os dados do CSV de entrada
8 entrada = 'Base.csv'
9 df = pd.read_csv(entrada, sep=';')
10
11 # Garantir que as colunas numericas estejam no formato correto
12 df['DEC'] = df['DEC'].astype(float)
13 df['FEC'] = df['FEC'].astype(float)
14 df['preco_medio_ponderado'] = df['preco_medio_ponderado'].astype(float)
15
16 # Codificar o CNPJ
17 df['cod_bdi'] = pd.factorize(df['cod_bdi'])[0]
18

```

```

19 # Definir as entradas e a saída
20 X = df[['DEC', 'FEC', 'cod_bdi']]
21 y = df['preco_medio_ponderado']
22
23 # Dividir em conjunto de treino e teste
24 X_train, X_test, y_train, y_test =
25 train_test_split(X, y, test_size=0.2, random_state=42)
26
27 # Criar a rede neural
28 model = keras.Sequential([
29     layers.Dense(64, activation='relu', input_shape=
30     (X_train.shape[1],)),
31     layers.Dense(32, activation='relu'),
32     layers.Dense(1)
33 ])
34
35 # Compilar o modelo
36 model.compile(optimizer='adam', loss=
37 'mean_squared_error', metrics=['mae'])
38
39 # Treinar o modelo
40 history = model.fit(X_train, y_train, epochs=2000,
41 validation_split=0.2, verbose=1)
42
43 # Avaliar o modelo
44 loss, mae = model.evaluate(X_test, y_test)
45 print(f'MAE: {mae}')
46
47 # Salvar o modelo treinado
48 model.save('modelo_CNPJ.keras')

```

Fonte: elaborado pelos autores

Esse código irá salvar o modelo e também irá calcular o MAE baseado nas previsões usando pontos que já foram dados ao sistema comparando com os valores reais.

No estudo, foram testados diferentes tipos de redes MLP com o objetivo de reduzir a MAE. Durante os testes, variaram-se o número de camadas ocultas e a quantidade de neurônios em cada camada. O menor valor obtido para a MAE foi 7,88, um valor considerado elevado em relação à magnitude dos dados analisados.

As previsões foram feitas a partir do Script [5.2](#)

### Script 5.2 – Código para previsões

```

1
2 import pandas as pd
3 from tensorflow.keras.models import load_model
4

```

```

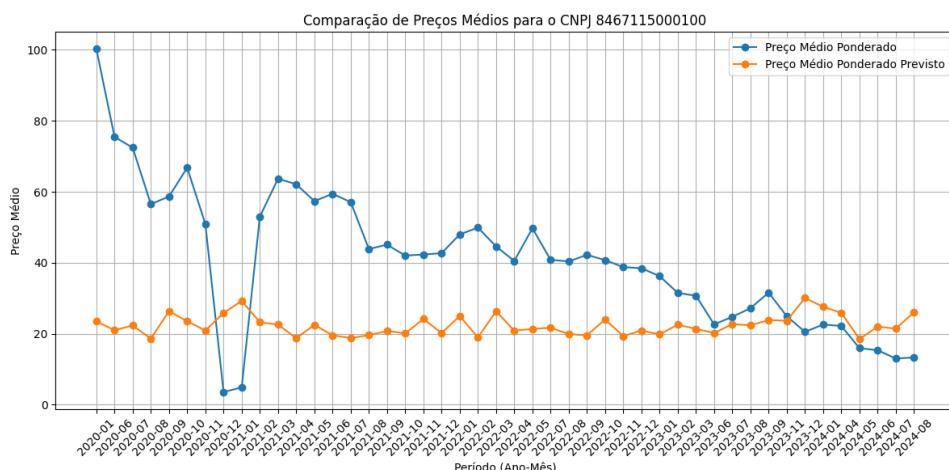
5 # Carregar o novo arquivo CSV de entrada (com dados de DEC e FEC)
6 df_novo = pd.read_csv('Entrada.csv', sep=';')
7
8 # Garantir que as colunas numericas estejam no formato correto
9 df_novo['DEC'] = df_novo['DEC'].astype(float)
10 df_novo['FEC'] = df_novo['FEC'].astype(float)
11
12 # Selecionar as colunas de entrada (DEC e FEC)
13 X_novo = df_novo[['DEC', 'FEC']]
14
15 # Carregar o modelo treinado
16 model = load_model('modelo_teste.keras')
17
18 # Fazer previsões com o modelo
19 predicoes = model.predict(X_novo)
20
21 # Adicionar as previsões ao DataFrame original para análise
22 df_novo['preco_medio_ponderado_previsto'] = predicoes
23
24 # Exibir as primeiras previsões
25 print(df_novo[['DEC', 'FEC', 'preco_medio_ponderado_previsto']].head())
26
27
28 df_novo.to_csv('Previsões.csv', index=False)

```

Fonte: elaborado pelos autores

As previsões são salvas em formato CSV, o que permite gerar gráficos que comparam, ao longo do tempo, o valor médio previsto pela rede neural com o valor médio real das empresas, conforme ilustrado na Figura 13. Nesse gráfico, a linha amarela representa a previsão, enquanto a linha azul representa o valor real.

Figura 13 – Gráfico com valores previstos e reais



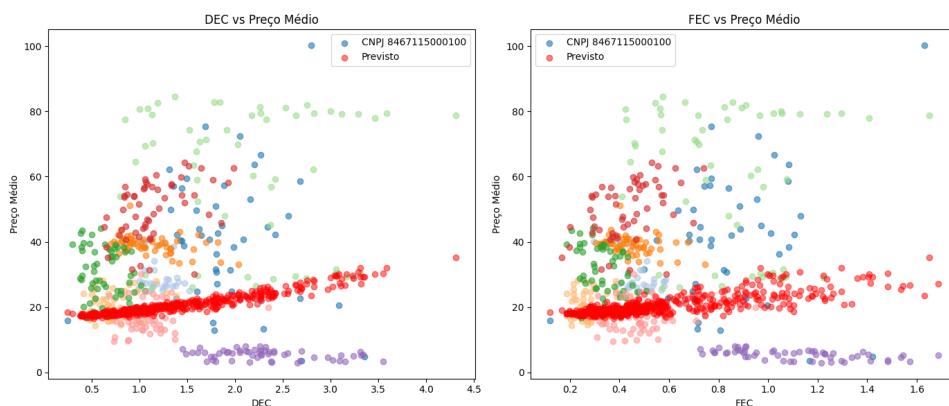
Fonte: Elaborado pelos autores

Observa-se que, na figura, os valores previstos pela rede neural concentram-se em torno de 20, sugerindo uma possível convergência das previsões para esse valor. Em contrapartida, os valores reais, representados em azul, apresentam uma variação significativamente maior. Isso indica que a rede não conseguiu prever com precisão os valores de mercado com base nos indicadores DEC e FEC das empresas. Pensando na rede neural esse comportamento pode vir de diversos fatores:

- **Estrutura da Rede Neural:** A estrutura da rede, incluindo o número de camadas e unidades de processamento, pode ser inadequada, o que pode levar a um aprendizado insuficiente ou até à incapacidade de aprender as relações entre os dados (GOODFELLOW; BENGIO; COURVILLE, 2016).
- **Overfitting:** Se a rede foi treinada por muitas épocas (iterações) ou com um modelo muito complexo, ela pode ter “decorado” os dados de treinamento, não generalizando bem para novos dados. Isso resulta em previsões imprecisas em dados de teste ou em produção.
- **Características Relevantes:** Se os indicadores como DEC e FEC não são suficientemente explicativos para o comportamento do mercado, a rede pode não conseguir identificar uma relação significativa entre essas variáveis e o valor de mercado. Uma análise mais profunda dos dados e a possível inclusão de outras variáveis preditivas podem ser necessárias.

Com isso os resultados obtidos se afastam significativamente dos valores reais. Na Figura 14, é apresentado um gráfico de dispersão que ilustra como cada empresa se comporta com base nos valores de DEC e FEC, reais e previstos.

Figura 14 – Gráfico de Dispersão



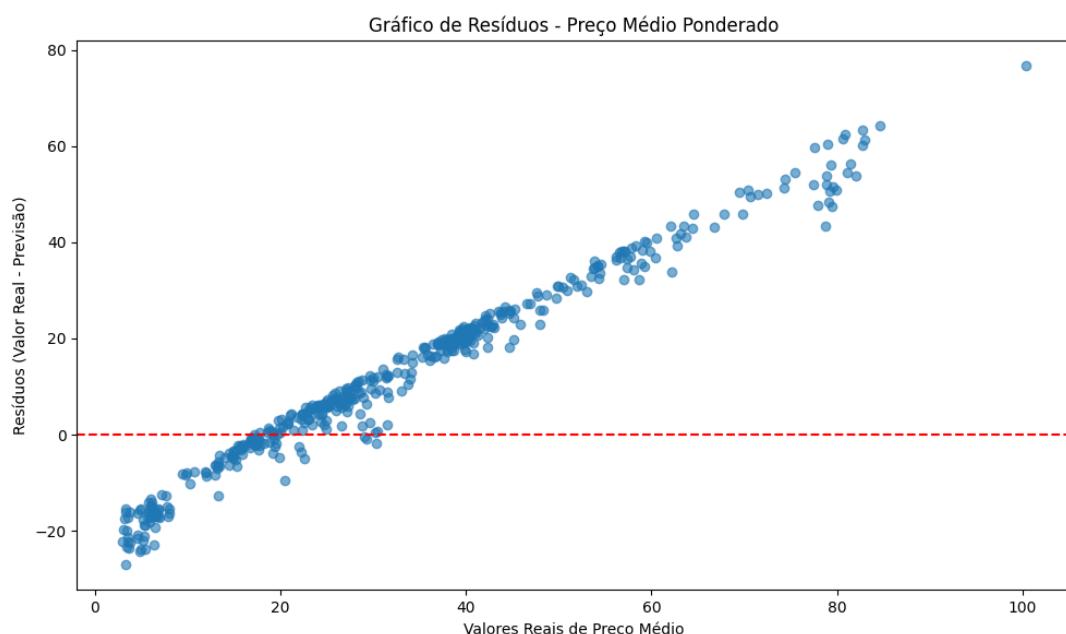
Fonte: Elaborado pelos autores.

Ao analisar o gráfico, observa-se que os pontos relacionados aos valores previstos estão muito mais concentrados, formando quase uma linha. Esse padrão sugere que o modelo

tem dificuldade em capturar a variabilidade dos dados, o que pode indicar que ele está gerando previsões centradas em torno de um valor médio ou constante.

Outra abordagem para avaliar o desempenho do modelo é por meio da análise da distribuição dos resíduos (JAMES et al., 2013), apresentada na Figura 15. Um gráfico de resíduos ideal exibe pontos dispersos de forma aleatória em torno da linha horizontal (que representa resíduos iguais a zero), sem padrões visíveis. Isso indicaria que o modelo está capturando bem a relação entre as variáveis e que os erros são aleatórios, sem tendências sistemáticas.

Figura 15 – Gráfico de Resíduos



Fonte: Elaborado pelos autores.

No entanto, como pode ser observado, os pontos apresentam um agrupamento evidente. Esse comportamento sugere que o modelo não está capturando algum aspecto relevante dos dados, possivelmente devido à ausência de variáveis explicativas adicionais ou a limitações na estrutura da rede neural.

Observou-se que a rede estava gerando previsões excessivamente concentradas. Em resposta a isso, foram testadas diferentes configurações de redes neurais, utilizando técnicas como aumento da complexidade da rede, normalização e *dropout*. No total, foram avaliados cinco tipos de redes para cada uma dessas técnicas isoladas, além de outras cinco redes combinando todos os métodos, para todas o número de iterações usados foi 200. Os parâmetros de desempenho, como R quadrado, MAE, MSE e RMSE, foram calculados conforme as métricas descritas na literatura, e os resultados são apresentados na Tabela 6.

Tabela 6 – Parâmetros das redes

<b>Modelo</b>	<b>MAE</b>	<b>MSE</b>	<b>RMSE</b>	<b>R^2</b>
Simples	9.80	185.65	13.62	0.14
Maior	8.57	161.93	12.72	0.25
Batch	8.21	144.89	12.03	0.33
Dropout	9.94	188.39	13.72	0.14
Combinadas	7.88	133.88	11.57	0.39

Fonte: Elaborado pelos autores

Da Tabela 6, pode-se observar que, embora os métodos aplicados resultem na redução dos erros MAE e MSE (e, consequentemente, no RMSE), os valores ainda são elevados, considerando a magnitude dos preços médios. O valor de R quadrado também está bem distante de 1, o que indica que, mesmo com a aplicação de técnicas que mitigam o *overfitting*, os resultados ainda apresentam uma concentração de valores. Isso sugere que os dados de entrada podem não ter uma relação suficientemente forte para prever com precisão os preços de saída.

## 6 CONCLUSÕES

O presente trabalho teve como objetivo analisar empresas distribuidoras do setor elétrico que estão presentes no mercado de capitais. Para isso, foram empregadas técnicas de aprendizado de máquina, com ênfase em redes neurais artificiais e métodos de clusterização. Os dados de entrada consistiram em indicadores de qualidade de energia regulamentados e disponibilizados pela ANEEL, enquanto os critérios de avaliação incluíram indicadores financeiros, com foco nos preços médios para a análise de redes neurais, e trajetórias estratégicas das empresas, conforme publicadas em seus sites de Relações com Investidores (RI), regulamentados pela CVM.

A revisão bibliográfica abordou temas fundamentais relacionados à distribuição de energia e ao sistema elétrico nacional, com ênfase nas transformações do setor impulsionadas pela Lei Geral de Concessões. Além disso, exploraram-se estudos sobre clusterização, com destaque para métodos particionais, e redes neurais, enfatizando suas aplicações no setor energético e financeiro.

Os resultados da clusterização revelaram que as trajetórias estratégicas das empresas podem ser compreendidas de maneira mais profunda por meio da análise dos indicadores de qualidade. Observou-se que empresas em fase de melhoria de desempenho, como o Grupo Equatorial, apresentam potencial de valorização por meio de estratégias de "turnaround". Por outro lado, empresas com indicadores altamente estáveis, como a CEMIG, demonstram uma consolidação gerencial robusta, com menor envolvimento em processos de fusão e aquisição. Empresas intermediárias, como a ENERGISA, tendem a buscar equilíbrio entre consolidação e gestão administrativa eficiente, seja financeira ou produtiva.

Os resultados obtidos com a rede neural indicam que o modelo tem dificuldades significativas em identificar um padrão consistente que possa prever com precisão e acurácia a valorização das empresas, com relação tanto à precisão quanto à acurácia das previsões. Embora diversas arquiteturas tenham sido testadas, ficou evidente que utilizar apenas os indicadores DEC e FEC como variáveis de entrada não é suficiente para explicar adequadamente a variabilidade dos preços no mercado. Isso sugere que outras variáveis ou indicadores podem ser necessários para uma modelagem mais robusta e precisa.

Além disso, um fator importante a ser considerado é a natureza dinâmica e em constante evolução do sistema elétrico. A mudança contínua no setor torna difícil fazer previsões precisas sobre o futuro da bolsa, pois as condições de mercado e os parâmetros regulatórios estão em constante alteração. Dessa forma, a complexidade do sistema e a necessidade de incorporar variáveis mais abrangentes podem ser elementos cruciais para aprimorar a modelagem preditiva e, possivelmente, aumentar a acurácia das previsões no futuro.

É importante ressaltar, novamente, que os métodos desenvolvidos para aquisição e tratamento de dados neste trabalho podem ser facilmente adaptados para estudos futuros, seja para aprofundar análises ou reproduzir os resultados obtidos. Com a mudança na

disponibilização dos dados de DRP e DRC de periodicidade trimestral para mensal, novas análises poderão ser realizadas conforme esses dados se tornarem disponíveis, permitindo trabalhar com uma granularidade maior e possibilitando observar os acontecimentos de maneira mais detalhada.

Além disso, destaca-se a disparidade temporal entre os conjuntos de dados utilizados: enquanto os indicadores DEC e FEC, possuem registros desde o início dos anos 2000, os dados de DRP e DRC começaram a ser coletados apenas a partir de 2010. Essa diferença de abrangência temporal reforça a necessidade de integrar e adaptar os métodos desenvolvidos para explorar possíveis tendências de longo prazo e captar nuances específicas do comportamento recente das empresas no mercado.

## REFERÊNCIAS

- ABADI, M. et al. *TensorFlow: A System for Large-Scale Machine Learning*. [S.I.]: OSDI, 2016.
- ABREU, Y. V. d. *A reestruturação do setor elétrico brasileiro: questões e perspectivas*. Dissertação (Dissertação (Mestrado em Energia)) — Universidade de São Paulo, São Paulo, 1999. Páginas 47-51.
- ANEEL. *Manual de interpretação do arquivo TXT*. 2024. Disponível em: <https://dadosabertos.aneel.gov.br/dataset/indicadores-coletivos-de-continuidade-dec-e-fec4>.
- (ANEEL), A. N. de E. E. *Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST: Módulo 8 – Qualidade da Energia Elétrica*. Brasília: PRODIST, 2024. Disponível em: <https://www.aneel.gov.br/modulo8prodist>. Acesso em: 25 nov. 2024.
- ARISTOTELES. *Partes dos Animais*. Portugal: Imprensa Nacional-Casa da Moeda, 350 a.C. Tiragem de 800 exemplares. Concepção gráfica: UED/INCM. ISBN 978-972-27-1888-2.
- B3. *Manual de interpretação do arquivo TXT*. 2024. Acesso em: 04 nov. 2024. Disponível em: [https://www.b3.com.br/pt\\_br/market-data-e-indices/servicos-de-dados/market-data/historico/mercado-a-vista/cotacoes-historicas/](https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/historico/mercado-a-vista/cotacoes-historicas/).
- B3. *Consultas de classificação setorial: setor utilidade pública, subsetor energia elétrica*. s.d. [https://www.b3.com.br/pt\\_br/produtos-e-servicos/negociacao/renda-variavel/acoes/consultas/classificacao-setorial/](https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/acoes/consultas/classificacao-setorial/). Acesso em: 28 nov. 2024.
- BERTRAND, J. W. M.; FRANSOO, J. C. *Modelling and Simulation: Researching Operations Management*. [S.I.]: Routledge, 2010.
- BODIE, Z.; KANE, A.; MARCUS, A. J. *Investimentos [recurso eletrônico]*. 10. ed. Porto Alegre: AMGH, 2015.
- BRASIL. *Lei nº 8.987, de 13 de fevereiro de 1995: Dispõe sobre o regime de concessão e permissão da prestação de serviços públicos previsto no art. 175 da Constituição Federal, e dá outras providências*. 1995. Promulgada em Brasília. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/l8987cons.htm](https://www.planalto.gov.br/ccivil_03/leis/l8987cons.htm).
- CHOLLET, F. *Deep Learning with Python*. [S.I.]: Manning Publications, 2017.
- DROSOS, D. et al. Impact of power quality on the stock market: An empirical study in the electricity sector. *Journal of Electrical Engineering*, v. 71, n. 1, p. 12–20, 2020.
- Equatorial Energia S.A. *apresentacao\_institucional\_4t12*. 2013. Apresentação institucional para investidores. Disponível em: <https://api.mziq.com/mzfilemanager/d/62b21cba-838c-49a4-aaef-e0fb2350c169/77c4b50a-3fba-474e-9b39-cd3dc431bc3?origin=1>.
- Equatorial Energia S.A. *apresentacao\_institucional\_1t22*. 2022. Apresentação institucional para investidores. Disponível em: <https://api.mziq.com/mzfilemanager/v2/d/62b21cba-838c-49a4-aaef-e0fb2350c169/830e2c47-ca80-2c65-7ab2-cdef2e337af3?origin=1>.

FERREIRA, C. K. L. Privatização do setor elétrico no brasil. In: PINHEIRO, A. C.; FUKASAKU, K. (Ed.). *A privatização no Brasil: o caso dos serviços de utilidade pública*. Rio de Janeiro: Banco Nacional de Desenvolvimento Econômico e Social, 2000. p. 179–220.

GOMES, A. C. S. et al. O setor elétrico. Dba, 2002.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.I.]: MIT Press, 2016.

HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*. 3rd. ed. United States: Morgan Kaufmann, 2011. ISBN 978-0-12-381479-1.

HARTIGAN, J. A. *Clustering Algorithms*. United States of America: Wiley, 1975. (Wiley Series in Probability and Mathematical Statistics). ISBN 0-471-35645-X.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. [S.I.]: Prentice Hall, 1998.

IPEA, I. de P. E. A. *Infra-estrutura: perspectivas de reorganização; Setor Elétrico*. Brasília: IPEA, 1997. 27 p. 141 p.

JAMES, G. et al. *An Introduction to Statistical Learning: With Applications in R*. 1. ed. New York: Springer, 2013. (Springer Texts in Statistics). Available online at <https://www.statlearning.com>. ISBN 9781461471370.

LINDEN, R. Técnicas de agrupamento. *Revista de Sistemas de Informação da FSMA*, v. 4, p. 18–36, jul./dez. 2009.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: LECAM, L.; NEYMAN, J. (Ed.). *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, Califórnia: University of California Press, 1967. v. 1, p. 281–297.

Neoenergia Distribuicao Brasilia S.A. *Comunicado ao Mercado - Apresentação de Resultados 4T19 e 2019*. 2020. Apresentação institucional para investidores. Disponível em: <https://api.mziq.com/mzfilemanager/v2/d/62b21cba-838c-49a4-aaef-e0fb2350c169/830e2c47-ca80-2c65-7ab2-cdef2e337af3?origin=1>

RAUBER, T.; RUEDEL, F.; PADILHA, A. G. *Redes neurais artificiais*. 2024. Acesso em: 17 nov. 2024. Disponível em: [https://www.researchgate.net/profile/Thomas-Rauber-2/publication/228686464\\_Redes\\_neurais\\_artificiais/links/02e7e521381602f2bd000000/Redes-neurais-artificiais.pdf](https://www.researchgate.net/profile/Thomas-Rauber-2/publication/228686464_Redes_neurais_artificiais/links/02e7e521381602f2bd000000/Redes-neurais-artificiais.pdf).

RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 4. ed. Hoboken, NJ: Pearson, 2020.

S.A., G. E. *Comunicado ao Mercado - Apresentação de Resultados 4T18*. 2019. Apresentação institucional para investidores. Disponível em: [https://mz-filemanager.s3.amazonaws.com/60f49a2d-bd8c-4fd9-95ab-bdf833097a83/apresentacaoescentral-de-downloads/a3ce17be2fc0c5d177be2a83dff54b7484e0ca9f85a8481ceeb54204b0cc984/apresentacao\\_dos\\_resultados\\_do\\_4\\_trimestre\\_de\\_2018.pdf](https://mz-filemanager.s3.amazonaws.com/60f49a2d-bd8c-4fd9-95ab-bdf833097a83/apresentacaoescentral-de-downloads/a3ce17be2fc0c5d177be2a83dff54b7484e0ca9f85a8481ceeb54204b0cc984/apresentacao_dos_resultados_do_4_trimestre_de_2018.pdf).

TAUD, H.; MAS, J. Multilayer perceptron (mlp). In: OLMEDO, M. C. et al. (Ed.). *Geomatic Approaches for Modeling Land Change Scenarios*. [S.I.]: Springer, 2018, (Lecture Notes in Geoinformation and Cartography). p. 533–545.