



UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE ENERGIAS ALTERNATIVAS E RENOVÁVEIS  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA



Virgínia Vieira Aires  
Engenharia Elétrica / CEAR

DESENVOLVIMENTO DE TÉCNICAS DE ANÁLISE E PREVISÃO DE CONSUMO  
ELÉTRICO EM UNIDADES DE ALTA TENSÃO

João Pessoa  
Junho de 2022

Virgínia Vieira Aires

**DESENVOLVIMENTO DE TÉCNICAS DE ANÁLISE E  
PREVISÃO DE CONSUMO ELÉTRICO EM UNIDADES DE  
ALTA TENSÃO**

Trabalho de Conclusão de Curso submetido à Coordenação do Curso de Engenharia Elétrica da Universidade Federal da Paraíba como parte dos requisitos necessários para a obtenção do título de Engenheiro Eletricista.

**Orientador:** Prof. Dr. Juan Moises Mauricio Villanueva

João Pessoa  
Junho de 2022

**Catálogo na publicação**  
**Seção de Catalogação e Classificação**

A298d Aires, Virgínia Vieira.

Desenvolvimento de Técnicas de Análise e Previsão de  
Consumo Elétrico em Unidades de Alta Tensão / Virgínia  
Vieira Aires. – João Pessoa, 2022.

69 f.

Orientação: Juan Moises Mauricio Villanueva.  
TCC (Graduação) – UFPB/CEAR.

1. Consumo Elétrico. 2. Previsão de Consumo  
Elétrico. 3. Inteligência Artificial. I. Villanueva,  
Juan Moises Mauricio. II. Título.

UFPB/CT/BSCT

CDU 621.3(043.2)

Virgínia Vieira Aires

**DESENVOLVIMENTO DE TÉCNICAS DE ANÁLISE E  
PREVISÃO DE CONSUMO ELÉTRICO EM UNIDADES DE  
ALTA TENSÃO**

Trabalho de Conclusão de Curso submetido à Coordenação do Curso de Engenharia Elétrica da Universidade Federal da Paraíba como parte dos requisitos necessários para a obtenção do título de Engenheiro Eletricista.



---

Prof. Dr. Juan Moises Mauricio Villanueva  
Universidade Federal da Paraíba  
Orientador



---

Prof<sup>a</sup>. Dr<sup>a</sup>. Camila Seibel Gehrke  
Universidade Federal da Paraíba  
Avaliador



---

Prof. Dr. Felipe Vigolvino Lopes  
Universidade Federal da Paraíba  
Avaliador

João Pessoa  
Junho de 2022

# AGRADECIMENTOS

Agradeço a **Deus** por me guiar e dar forças para enfrentar as diferentes situações e os anos de curso; por me ensinar a ter paciência, fé e alegria.

Agradeço aos meus pais, **Josiane** e **José Ildo**, por me incentivarem a seguir meu caminho, por apoiarem em todas as minhas fases e me darem as condições para seguir meus sonhos.

Agradeço as minhas irmãs, **Olívia** e **Beatriz**, por me darem suporte e a animação necessária, por cuidarem de mim mesmo morando longe, por aguentarem meus desabafos, minhas encrencas e meus estresses.

Agradeço aos meus avós, **Hermínio** *em lembrança* e **Maria Paula**, **José** *em lembrança* e **Iraci**, por serem exemplos de determinação, simplicidade e coragem na minha vida.

Agradeço ao amigo que a vida me trouxe de forma inesperada, **Victor Santos**, que sempre me incentivou a nunca desistir, obrigada por todas as conversas, todos os puxões de orelhas e brincadeiras.

Agradeço ao meu cunhado **Matheus de Oliveira**, por mesmo distante me apoiar e fazer os momentos mais alegres.

Agradeço a meus amigos, **Bianca** e **Mateus**, por me darem apoio ao longo de toda minha caminhada, foram longos onze anos de amizade, obrigada por dividirem comigo minhas conquistas e sempre estarem presentes nos momentos difíceis.

Agradeço aos amigos que a UFPB me trouxe e que irei levar para a vida, **Ilaneide**, **Júlia**, **Valeska**, **Michael**, **Lucas Barbosa** e **Joseito**, esses anos dentro da universidade não teria sido os mesmos sem vocês.

Agradeço aos professores, **Camila Seibel** e **Felipe Lopes**, por aceitarem o convite de banca examinadora.

Por fim agradeço ao meu orientado **Prof. Dr. Juan Moises Mauricio Villanueva**, pela oportunidade e pelos ensinamentos passados. Obrigada por todo o incentivo, orientação e apoio prestado durante a iniciação científica e durante o TCC.



*Nem tudo está perdido como parece... sabe, coisas extraordinárias só acontecem a pessoas extraordinárias, vai ver é um sinal que você tem um destino extraordinário, algum destino maior do que você pode ter imaginado.*

**As Crônicas de Nárnia - C.S. Lewis**

# Resumo

O aumento do consumo de energia elétrica no Brasil, juntamente com a alta das tarifas e impostos sobre as contas, desencadeou a procura pela eficiência energética e a forma de gerar economia a partir dela. Na tentativa de buscar melhores opções econômicas e confiáveis de análise energética, técnicas de eficiência foram desenvolvidas ao longo dos anos, desde simples trocas de equipamentos antigos e uso consciente e responsável de energia elétrica, até modelos complexos como, por exemplo, aprendizado de máquinas e aplicação de redes neurais artificiais, para monitoramento e controle do consumo ou da demanda energética. O presente trabalho visou desenvolver um modelo de acompanhamento e previsão do consumo elétrico, a partir de dados reais disponibilizados em unidades de alta tensão, espalhados por todo o estado da Paraíba, Brasil. Os dados originais coletados permitiram verificar o desempenho das métricas utilizadas durante o ano de 2021. O presente estudo contou com ferramentas do tipo de clusterização, baseado no método K-Means, algoritmo para acompanhamento mensal de ultrapassagens, tanto em kWh como em reais, e um algoritmo supervisionado de Rede Neural Recorrente (RNN), que previram dados do consumo elétrico com taxa de erro de treinamento variando de 7,02% até 25,60%. O modelo utilizado neste trabalho demonstrou ser confiável como metodologia para previsão de consumo elétrico.

**Palavras-chave:** Consumo Elétrico, Previsão de Consumo Elétrico, Inteligência Artificial

# Abstract

The increase in electricity consumption in Brazil, herewith the rise in tariffs and taxes on bills, triggered the search for energy efficiency and a way to generate savings from it. In an attempt to seek better economic and reliable options for energy analysis, efficiency techniques have been developed over the years - from simple replacement of old equipment and conscious and responsible use of electrical energy, to complex models such as machine learning and application of artificial neural networks for monitoring and controlling energy consumption or demand. The present work aimed to develop a model for monitoring and forecasting electrical consumption, based on real data available in high voltage units, spread throughout the state of Paraíba, Brazil. The original data collected allowed to verify the performance of the metrics used during the year 2021. The present study used clustering tools, based on the K-Means method, an algorithm for monthly monitoring of overruns, both in kWh and in reais, and a supervised Recurrent Neural Network (RNN) algorithm, which predicted electrical consumption data with training error rates ranging from 7,02% até 25,60%. The model used in this work proved to be reliable as a methodology for forecasting electrical consumption.

**Keywords:** Electricity Consumption, Electricity Forecast Consumption, Artificial Intelligence



# Lista de Abreviaturas e Siglas

CAGEPA Companhia de Água e Esgotos da Paraíba

K-Means *K-Means Clustering*

kWh Quilowatt-hora

LSTM Memória de Curto Pazo Longa (*Long Short-Term Memory*)

MAE Erro Médio Absoluto (*Mean Absolute Error*)

MAPE Erro Percentual Absoluto Médio (*Mean Absolute Percentage Error*)

MSE Erro Médio Quadrático (*Mean Squared Error*)

RNA Rede Neural Artificial

RNN Rede Neural Recorrente (*Recurrent Neural Network*)

UC Unidade Consumidora

## Lista de Figuras

1	Matriz Elétrica Brasileira . . . . .	8
2	Exemplo de Série Temporal . . . . .	11
3	Comparação RNA e um Neurônio Biológico . . . . .	13
4	Marcos do desenvolvimento das RNA . . . . .	13
5	Alguns Modelos de RNA . . . . .	14
6	Rede Neural Feed-Forward . . . . .	15
7	Rede Neural Recorrente . . . . .	15
8	RNA Modelo de Hopfield . . . . .	16
9	RNA Modelo Máquina de Boltzmann . . . . .	16
10	Cluster Exclusivo . . . . .	17
11	Cluster Hierárquico . . . . .	18
12	Cluster Sobreposto . . . . .	18
13	Fluxograma looping K-Means . . . . .	19
14	Exemplo K-Means . . . . .	20
15	Bibliotecas Utilizadas . . . . .	22
16	Exemplo Sklearn . . . . .	23
17	Dados Recebidos . . . . .	24
18	Medidor Eletrônico . . . . .	24
19	Dados 96 Medições . . . . .	25
20	Dados Colunas . . . . .	25
21	Fluxo de Escolha de Dados . . . . .	27
22	Curva Diária 96 Medições - Segunda-Feira . . . . .	28
23	Curva Diária 24 Medições - Segunda-Feira . . . . .	29
24	Curva Limite UC . . . . .	30
25	Tarifação 1 . . . . .	30
26	Tarifação 2 . . . . .	31
27	Método do Cotovelo - Jan . . . . .	32
28	Modelo LSTM . . . . .	34
29	Modelo Proposto LSTM . . . . .	34
30	Amostras Treinamento e Teste . . . . .	35
31	Curvas "Training Loss"+ "Validation Loss"Modelo 1 MAE . . . . .	36
32	Curvas "Training Loss"+ "Validation Loss"Modelo 2 MSE . . . . .	37
33	Separação DataFrames Válidos . . . . .	38
34	Correção Nome das Colunas . . . . .	39
35	Separação Dia e Mês . . . . .	39
36	Dados de Entrada . . . . .	40
37	Dados Corrigidos - 96 Medições . . . . .	40

38	Dados Corrigidos - intervalo 24 horas . . . . .	41
39	Ponta e Fora Ponta - Segunda-Feira . . . . .	42
40	Ultrapassagem Diária - Segunda-Feira . . . . .	43
41	Ultrapassagem Semanal . . . . .	44
42	Dados de Entrada . . . . .	44
43	DataFrame Financeiro . . . . .	45
44	Curva Financeira . . . . .	46
45	Resultado Clusterização Mensal K=3 . . . . .	47
46	Resultado Clusterização Mensal K=4 . . . . .	48
47	Resultado Clusterização Anual K=3 . . . . .	49
48	Resultado Clusterização Anual K=4 . . . . .	50
49	Dados de Treinamento e Teste . . . . .	50
50	Summary Model 1 . . . . .	51
51	Model 1 - Métrica MAE . . . . .	52
52	Model 1 - Métrica MSE . . . . .	52
53	Summary Model 2 . . . . .	53
54	Model 2 - Métrica MAE . . . . .	54
55	Model 2 - Métrica MSE . . . . .	54
56	Summary Model 3 . . . . .	55
57	Model 3 - Métrica MAE . . . . .	55
58	Model 3 - Métrica MSE . . . . .	56
59	Comparação Modelos Métrica MAE . . . . .	57
60	Comparação Modelos Métrica MSE . . . . .	57
61	Dados Mês de Janeiro . . . . .	60
62	Dados Meses Completos . . . . .	61

## Lista de Tabelas

1	Modelo Métrica 1 . . . . .	35
2	Modelo Métrica 2 . . . . .	35

## Sumário

<b>1</b>	<b>Introdução</b>	<b>8</b>
1.1	PERTINÊNCIA E MOTIVAÇÃO DO TRABALHO . . . . .	8
1.2	Objetivos . . . . .	10
1.2.1	Objetivo Específico . . . . .	10
1.3	Organização do Trabalho . . . . .	10
<b>2</b>	<b>Fundamentação Teórica</b>	<b>11</b>
2.1	Séries Temporais . . . . .	11
2.2	Consumo Elétrico . . . . .	12
2.3	Redes Neurais Artificiais . . . . .	12
2.4	Algoritmo de Clusterização . . . . .	16
2.4.1	K-Means . . . . .	18
<b>3</b>	<b>METODOLOGIA</b>	<b>21</b>
3.1	Linguagem Python . . . . .	21
3.1.1	Bibliotecas . . . . .	21
3.2	Estudo de Caso . . . . .	23
3.2.1	Conjunto de Dados . . . . .	24
3.3	Determinação da Curva Limite . . . . .	26
3.4	Determinação Financeira . . . . .	30
3.5	Número de Clusters K-Means . . . . .	31
3.6	Métricas de Avaliação . . . . .	32
3.6.1	Erro Médio Quadrático . . . . .	32
3.6.2	Erro Médio Absoluto . . . . .	33
3.7	Modelo de Previsão de Consumo . . . . .	33
<b>4</b>	<b>Resultados</b>	<b>38</b>
4.1	Correção do Banco de Dados . . . . .	38
4.2	Curva de Consumo Diário . . . . .	41
4.3	Curva Financeira . . . . .	44
4.4	Clusterização K-Means . . . . .	46
4.4.1	Cenário 1: Único Mês . . . . .	46
4.4.2	Cenário 2: Meses Completos . . . . .	48
4.5	Modelo de Previsão LSTM . . . . .	50
4.5.1	Modelo 1 - Duas Camadas Ocultas . . . . .	51
4.5.2	Modelo 2 - Três Camadas Ocultas . . . . .	53
4.5.3	Modelo 3 - Quatro Camadas Ocultas . . . . .	54
4.5.4	Comparação Entre modelos . . . . .	56

---

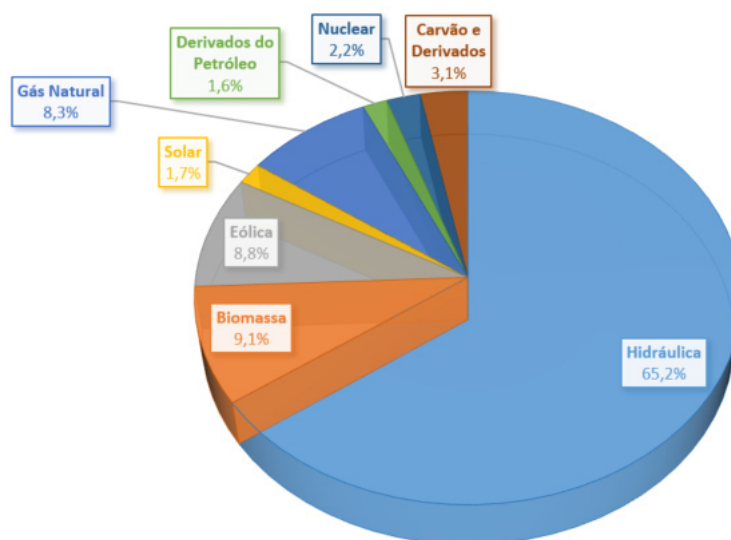
<b>5</b>	<b>CONCLUSÃO</b>	<b>59</b>
<b>6</b>	<b>ANEXO I - Dados Aplicados</b>	<b>60</b>

# 1 Introdução

## 1.1 PERTINÊNCIA E MOTIVAÇÃO DO TRABALHO

Desde sua descoberta, a eletricidade se encontra atrelada ao desenvolvimento humano. Na modernidade, toda e qualquer atividade, da simples até a mais complexa, necessita da energia elétrica. Nos últimos anos, o país tem atravessado um período de aumento acelerado do consumo elétrico. Este aumento pode ser resultado de diversos fatores, como, por exemplo, o surgimento de novas tecnologias, expansão da urbanização e crescimento populacional (TOLMASQUIM et al., 2007). Este período de alta despertou uma série de preocupações as geradoras em relação a produção e distribuição de energia elétrica, sem a ocorrência de distúrbios que viessem ocasionar danos à sociedade, como, por exemplo apagões, sobrecargas de tensão em equipamentos conectados a rede, instabilidade na frequência de rede. Comumente, os distúrbios resultam em anseios e preocupações sobre a necessidade de se gerar mais eletricidade, por meio da construção de novas usinas elétricas, como hidroelétricas, principal fonte da matriz energética brasileira (Figura 1)

Figura 1: Matriz Elétrica Brasileira



Fonte: Empresa de Pesquisa Energética - EPE GOV

Apesar de ser considerada uma fonte renovável, a energia hidrelétrica gerada é diretamente dependente da disponibilidade de recursos hídricos e de suas renovações, sendo dependente da quantidade de chuvas do período úmido e do volume estocado das águas por reservatórios.

No ano de 2021, o Brasil enfrentou um período de crise hídrica, o que procedeu em aumento nas tarifas presentes na conta de luz (SOARES; COSTA, 2022). Esta alta foi consequência da necessidade da operadora do sistema elétrico em procurarem novas fontes de energia e que, geralmente, eram mais caras, como termoeletricas, que produzem energia

através da queima de combustíveis.

Dessa forma, o aumento dos custos elétricos, mais uma vez, compeliu com a necessidade de se desenvolver e aplicar metodologias de eficiência energética, principalmente em indústrias e grandes consumidores dessa fonte.

Para propósito de planejamento e gestão de energia elétrica é importante que as empresas sejam capazes de conhecer o consumo elétrico de suas unidades, por meio do armazenamento dos valores no determinado período de tempo que se deseja estudar. Em posse desses dados, as empresas podem trabalhar com o acompanhamento e com os ajustes dos indicadores disponíveis, possibilitando a evolução e o desenvolvimento de técnicas para previsão do consumo de energia elétrica e gerando maior controle dos gastos financeiros referentes ao consumo elétrico, de acordo com a modalidade tarifária contratada para a unidade consumidora, podendo ser entre as modalidades Verde, Azul - Serviço Público para as Tensões distintas de 13,8 kV e 69 kV. O resultado é uma estimativa de consumo futuro de forma adequada, que permitirá a realização de medidas mitigadoras e que possibilita que as empresas planejem eventuais custos mais elevados ou que consigam se adaptar a demandas menores de energia elétrica em períodos de escassez, evitando, assim, pagar pelo não consumido.

Esta monografia tem por objetivo discutir a análise e a previsão do consumo elétrico em unidades atendidas por alta tensão, por meio da curva de energia entregue, em diversas regiões do estado da Paraíba. As curvas foram dadas por meio de séries temporais (MONTGOMERY et al., 2015) e as observações foram registradas ao longo do ano de 2021.

Saber explorar os dados disponibilizados de consumo é necessário para a tomada de decisão de qualquer empresa. Caso os valores de consumo elétrico estejam acima ou abaixo do consumo padrão, a empresa está sujeita a contratar, de forma errônea, demandas desnecessárias, que conduzirão, também, a maiores gastos sem necessidade, com ultrapassagens ou limites não atingidos de consumo. Assim sendo, é importante que previsões sejam estabelecidas de maneira correta para o planejamento eficiente do consumo empresarial de energia.

O presente trabalho aborda a metodologia de clusterização, agrupamento dos valores, baseado em K-Means devido ao seu modelo de agrupamento de dados semelhantes, classificação de ultrapassagem a partir da comparação entre as curvas de energia média recebida e o limite determinado para as unidades consumidoras, e para modelo de rede neural artificial, o tipo Rede Neural Recorrente LSTM no desenvolvimento de previsões de consumo elétrico, por possuir células especializadas para armazenamento de dados com longos períodos temporais. Com a rede neural proposta, foram feitos três modelos diferenciados e seus resultados foram avaliados por meio de comparações.

O estudo do tema viabilizou o uso dos modelos propostos, certificando suas vantagens e desvantagens, assim como relatando nos resultados práticos através da aplicação de dados

reais, fornecidos pela Companhia de Água e Esgotos da Paraíba, em diversas regiões desse estado.

## 1.2 Objetivos

O objetivo geral deste trabalho consiste no desenvolvimento de um algoritmo para realizar a análise e previsão de consumo elétrico, comparando diferentes modelos de redes neurais artificiais e resultados através de levantamento bibliográfico, com a finalidade de auxiliar a gestão e o operacional para eficiência energética. Além de abordar as curvas históricas das unidades.

### 1.2.1 Objetivo Específico

- a) Análise da curva de consumo para identificação de ultrapassagem;
- b) Desenvolvimento de algoritmo de agrupamento baseados em técnica K-Means;
- c) Desenvolvimento de algoritmo de previsão de consumo utilizando análise de grandes volumes de dados;
- d) Comparação entre os resultados, verificando qual o modelo que melhor se adapta ao conjunto de dados e que possui menor erro.

## 1.3 Organização do Trabalho

A estrutura deste trabalho será composto por cinco capítulos. O primeiro apresentou uma introdução, abordando a motivação do tema desenvolvido, bem como a definição dos objetivos gerais e específicos.

No segundo capítulo será abordado de forma teórica os principais conceitos relacionados ao tema, como séries temporais, o consumo elétrico, conceito das redes neurais artificiais e o modelo abordado ao longo deste trabalho.

No terceiro capítulo, será abordado a metodologia empregada, partindo da descrição das técnicas aplicadas e seus funcionamentos, de forma geral e específica.

O quarto capítulo apresentará os resultados encontrados, comparando as performances de cada etapa através de gráficos, de forma individual e geral.

E, para finalizar, o capítulo cinco constará a conclusão do trabalho, comparando os resultados obtidos com os traçados nos objetivos. Além de citar possíveis aplicações para trabalhos futuros.



## 2 Fundamentação Teórica

Este capítulo irá abordar os conceitos gerais necessários para o desenvolvimento do projeto, apresentando de forma clara suas definições.

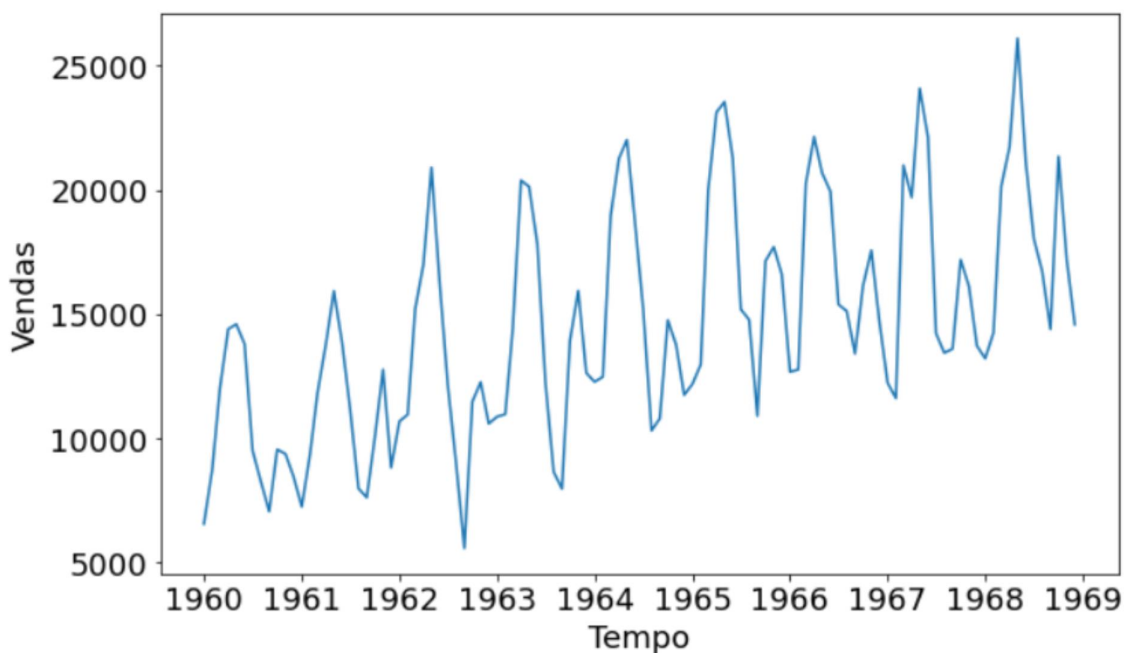
### 2.1 Séries Temporais

Uma série temporal é estabelecida como sendo um conjunto de observações feitas de forma sequencial ao longo de um período de tempo, onde estas ocorrem de forma ordenada e em intervalos igualmente espaçados (HYNDMAN; ATHANASOPOULOS, 2018). Se levarmos em consideração a ordem temporal podemos classificar elas em dois modelos (EDUCACIONAL, 2020), a primeira é dita discreta quando o intervalo de observações pertence a um conjunto discreto, em tempos fixos e equidistante; já uma série temporal contínua ocorre quando as observações ocorrem através de um intervalo de tempo, por exemplo as amostras que foram medidas em qualquer tempo dentro do intervalo  $T=[0,1]$ . Normalmente uma série temporal é representada por uma sequência ordenada de observações em um vetor, como mostrado na equação 1, já na Figura 2 tem-se um exemplo de série temporal contínua.

$$Y(t) = \{Y_1(t), Y_2(t), \dots, Y_n(t)\} \quad (1)$$

- **Y(t):** É um vetor de n-dimensões;
- **t:** É o tempo em que a amostra foi medida.

Figura 2: Exemplo de Série Temporal



Fonte:Artigo (Séries Temporais e Suas Aplicações (SPADINI, 2021))

O estudo de séries temporais tem sido bastante utilizada em aplicações envolvendo previsão, pois permite que seja analisado comportamentos passados para então prever o comportamento futuro dos problemas onde são aplicadas. De acordo com (OLIVEIRA, 2019), as séries temporais possuem alguns tipos de padrões sendo elas:

1. **Tendência:** Este padrão ocorre quando existe um aumento ou diminuição a longo prazo nos dados, relacionando a repetições de padrões em intervalos de tempos.
2. **Sazonal:** ocorre quando os dados da série é afetada por fatores sazonais, como por exemplo dados que mudam relativamente a cada estação do ano em que foram medidos.
3. **Cíclico:** Este padrão varia com uma frequência não fixa, muitas vezes pode ser confundida com o sazonal.
4. **Irregular:** Este padrão agrega os efeitos que não foram apresentados pelos outros padrões.

## 2.2 Consumo Elétrico

De forma simples e direta podemos definir o consumo elétrico como sendo a quantidade de energia elétrica empregada em um determinado período de tempo e dado pela unidade Quilowatt-hora (kWh). Portanto, quanto maior a potência e o tempo de uso do equipamento, maior será o gasto com despesas envolvendo energia elétrica.

Podemos calcular o consumo através da simples equação 2 (MATTEDE, 2020):

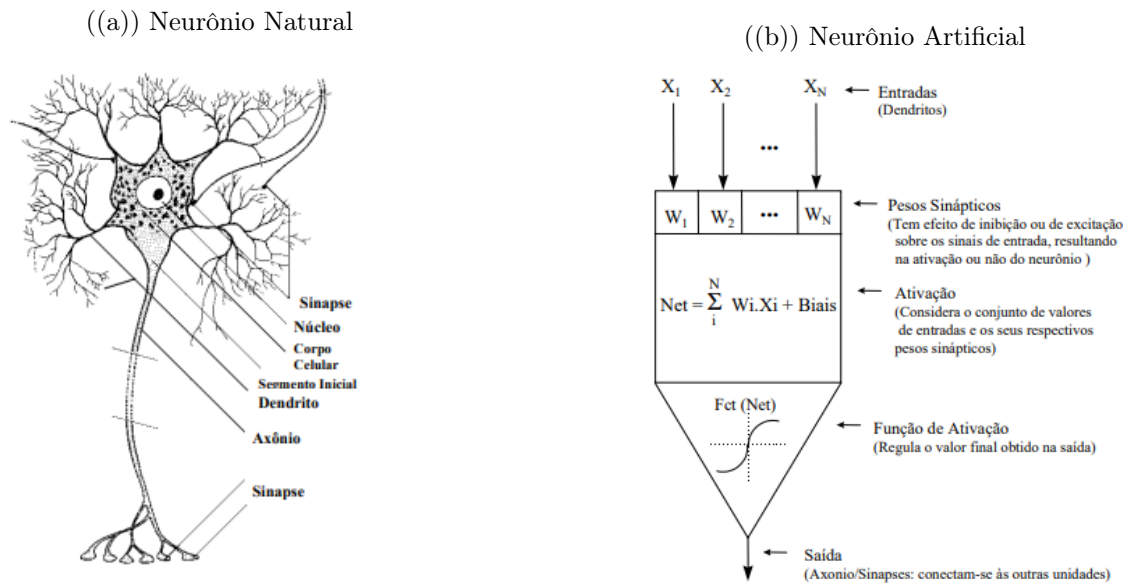
$$C(kWh) = \frac{Pot(w) \times Hr(h) \times D}{1000} \quad (2)$$

- **C:** É o consumo final em Quilowatt-hora;
- **Pot:** É a potência do equipamento;
- **Hr:** O número de horas que o equipamento se encontra em uso;
- **D:** É o número de dias em que o equipamento esta em funcionamento.

## 2.3 Redes Neurais Artificiais

Redes Neurais Artificiais (RNA) são modelos computacionais que apresentam inspirações em estruturas neurais de organismos inteligentes. São modelos que conseguem adquirir e desenvolver os conhecimentos através de experiências, podemos dizer que elas são modeladas de forma a imitar a capacidade de aprendizado humano (HYKIN, 2008). Pode-se observar na Figura 3 uma comparação entre um modelo desenvolvido com um neurônio biológico.

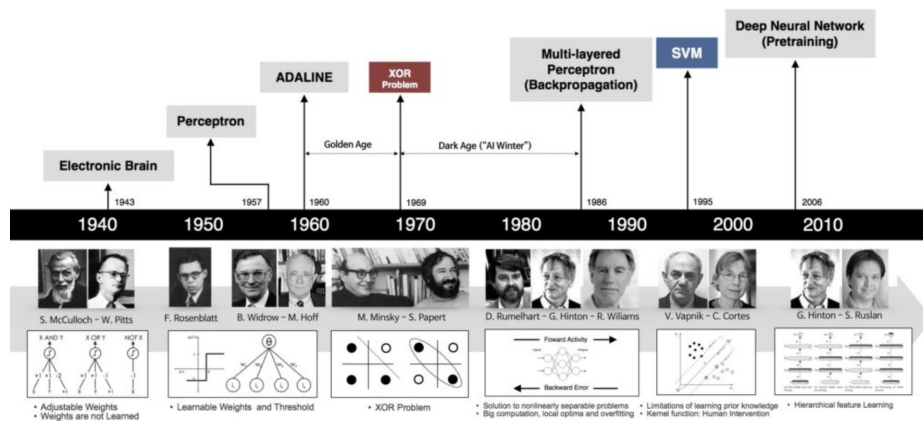
Figura 3: Comparação RNA e um Neurônio Biológico



Fonte: Sistemas Inteligentes baseados em Redes Neurais Artificiais aplicados ao Processamento de Imagens - (OSÓRIO; BITTENCOURT, 2000)

O primeiro modelo desenvolvido ocorreu no ano de 1943, quando o neurofisiologista Warren McCulloch e o matemático Walter Pitts escreveram um artigo modelando uma rede neural simples, mas que permitiu a abertura de novas pesquisas envolvendo este tema (FLECK et al., 2016). Na Figura 4 é possível observar alguns marcos na história das Redes Neurais, iniciando com o modelo de McCulloch-Pitts (ALMADA, ); o modelo Perceptron, desenvolvido em 1958 por Frank Rosenblatt (ALMADA, ); Durante o período de tempo entre os anos de 1969 a 1982 ocorreu a redução do interesse em estudos associados as RNAs, devido a limitações dos modelos Perceptron comprovados por Minsky e Papert; e então a retomada com o desenvolvimento do modelo de rede criada por John Hopfield em 1982 (FLECK et al., 2016).

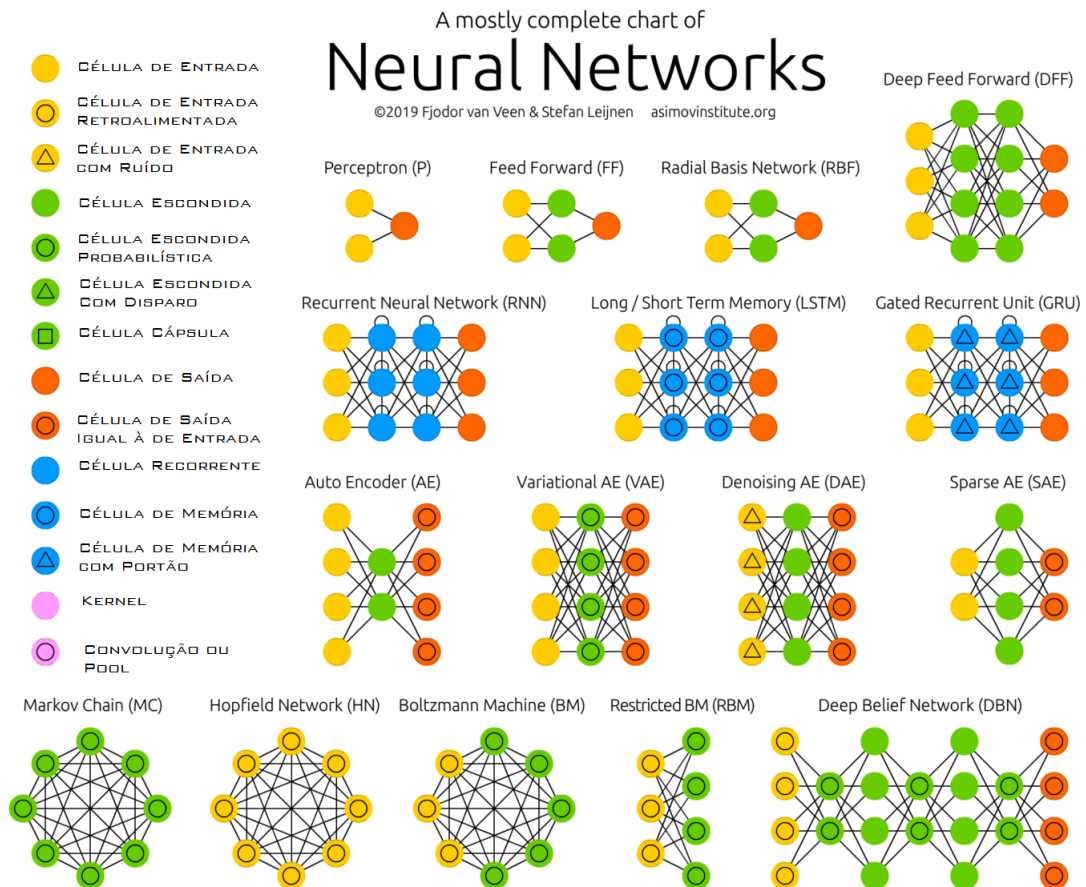
Figura 4: Marcos do desenvolvimento das RNA



Fonte: Livro (Cap. 2 - Uma Breve História das Redes Neurais Artificiais (ACADEMY, 2018d))

Na Figura 5 visualiza-se de forma resumida algumas arquiteturas de RNA, suas estruturas e os seus tipos de células, logo após serão expostas três das principais categorias de redes.

Figura 5: Alguns Modelos de RNA



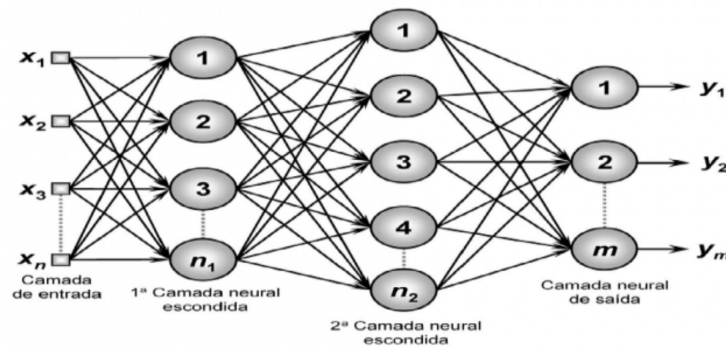
Fonte: Livro (Cap. 9 - A Arquitetura das Redes Neurais (ACADEMY, 2018a))

Pode-se classificar em três principais categorias:

1. **Redes Neurais *Feed-Forward*:** Considerado o tipo mais básico. Os dados fluem de forma linear, sendo mais utilizados em problemas que não possuem dependências temporais.

Este tipo de rede tem a característica de não relacionar os neurônios atuais com neurônios anteriores, por isso o nome *Feed-Forward*, a Figura 6 mostra um exemplo de RNA *Feed-Forward* com múltiplas camadas.

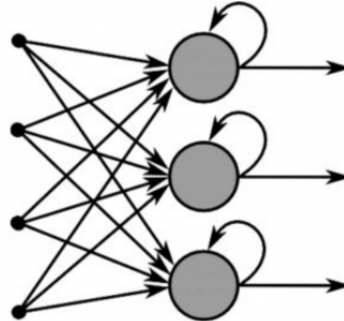
Figura 6: Rede Neural Feed-Forward



Fonte:Livro - (GOMIDE, 2012)

2. **Redes Neurais Recorrentes:** São redes que possuem pelo menos uma camada com realimentação, ou seja, eles recebem como entrada amostras passadas ao mesmo tempo que recebem a própria amostra atual, possuindo duas fontes de entrada. A Figura 7 exemplifica um modelo de recorrente e como ocorre a realimentação da amostra atual.

Figura 7: Rede Neural Recorrente



Fonte:Livro (Cap. 48 - Redes Neurais Recorrentes (ACADEMY, 2018c))

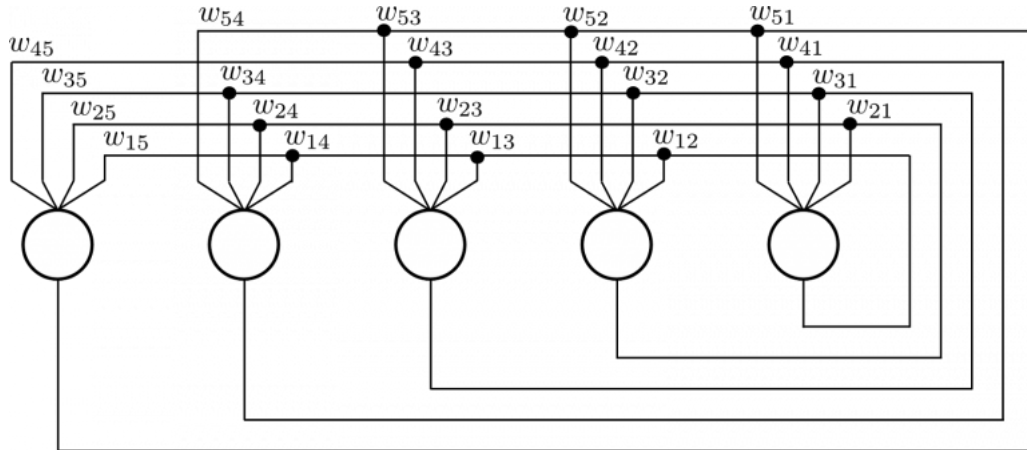
Estes tipos de rede possuem a habilidade de "guardar" memórias passadas, por isso são mais semelhantes ao cérebro humano. Além disto, a memória permite que sejam aplicados para reconhecimento de padrões e previsões de séries temporais.

No presente trabalho iremos abordar uma variação da rede recorrente, chamada *Long Short-Term Memory* que será esclarecida na seção 3.6.

3. **Redes Neurais Conectadas Simetricamente:** Assim como as redes recorrentes, possuem loops de conexões, recebendo amostras atuais e passadas, mas o que difere das recorrentes são os pesos das conexões, que se mantêm constantes em todas as direções de passagem dos dados.

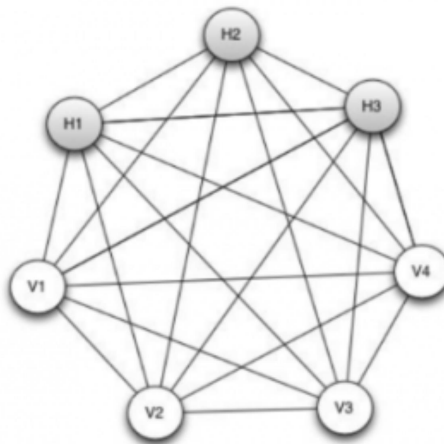
Pode-se dividir as redes simétricas em dois tipos: Redes Hopfield, quando não existe camada oculta exemplificada na Figura 8, e Máquinas de Boltzmann, quando existem camadas ocultas no modelo que será exemplificada na Figura 9.

Figura 8: RNA Modelo de Hopfield



Fonte: Artigo (Redes de Hopfield (COELHO, 2017))

Figura 9: RNA Modelo Máquina de Boltzmann



Fonte: Artigo (Redes Neurais – Redes Clássicas (REIS, 2016))

Este trabalho terá foco no estudo e desenvolvimento de modelos de redes neurais recorrentes (RNN) visando a previsão de medições de consumo para as unidades consumidoras.

## 2.4 Algoritmo de Clusterização

Nesta seção será explicado o método de agrupamento empregado para o desenvolvimento do trabalho. Primeiramente será exposto de forma simples o que são clusters, seus tipos e o método de cluster utilizado.

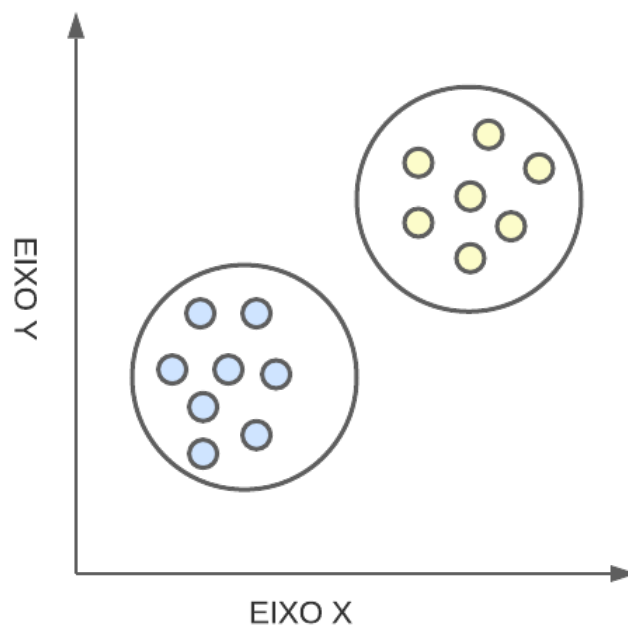
De acordo com (XU; TIAN, 2015), um cluster pode ser definido como uma metodologia de aglomeração de um grupo de dados em subgrupos (clusters), de forma que dados semelhantes fiquem no mesmo subgrupo.

Assim, pode-se dizer que a técnica de clusterização procura separar os dados de forma mais semelhante possível. Cada cluster possui um centro chamado de centróide e eles são encontrados baseados em correlação dos dados. Já o número de clusters pode ser encontrado de forma matemática e será melhor explicado na seção 3.5.

Referente ao tipo de cluster (LANDAU et al., 2011), existem diferentes tipos. Entre os principais se encontram:

- **Exclusivo:** Um cluster exclusivo refere-se a grupos únicos, como o próprio nome diz, ou seja, cada dado irá pertencer de forma única ao grupo específico com os dados semelhantes. Pode-se observar na Figura 10 um exemplo de agrupamento exclusivo.

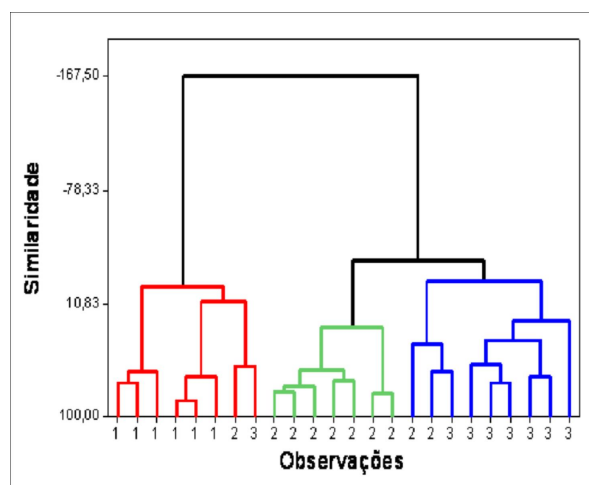
Figura 10: Cluster Exclusivo



Fonte: Autoria Própria

- **Hierárquico:** O agrupamento ocorre de forma categorizada. Neste caso os grupos podem conter outros subgrupos, com novas classificações, seguindo sempre uma hierarquia das aglomerações, como visto na Figura 11.

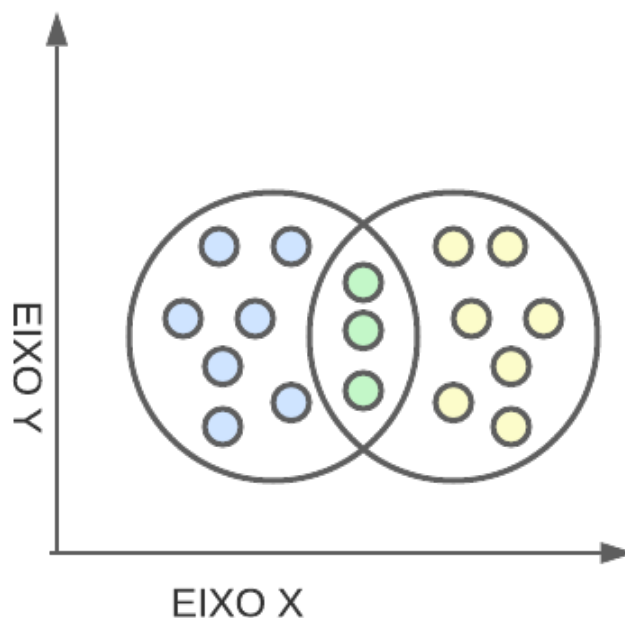
Figura 11: Cluster Hierárquico



Fonte:Artigo (Agrupamento hierárquico (LUCENA, 2019))

- **Sobreposto:** O cluster é dito sobreposto quando os dados possuem características que os permitem pertencer a mais de um grupo de classificação, como ilustrado na Figura 12.

Figura 12: Cluster Sobreposto



Fonte:Autoria Própria

#### 2.4.1 K-Means

Para o agrupamento dos valores disponibilizados, iremos utilizar a metodologia *K-Means*, ele será de importância para visualização dos dados de consumo com medições de

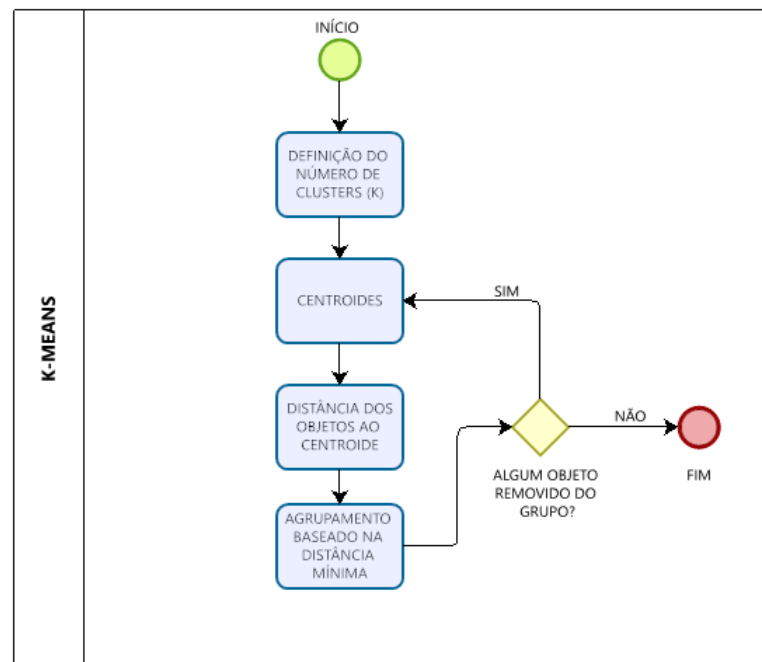


valores elevados de valores de consumo normais ou baixos. Outros tipos de distribuição, como o *boxplot*, foram descartados devido a possibilidade de interpretarem os dados mais elevados ou mais baixos como *outliers*, ou seja, como medições errôneas

K-Means é um método de agrupamento desenvolvido por James MacQueen em 1967 (MACQUEEN, 1967), é conhecido por ser um modelo simples de aprendizado não supervisionado, ou seja, não possui dados rotulados e que são compreendidos de forma própria pelo algoritmo, mas que possui uma alta precisão.

Seu funcionamento se dá através de algumas etapas (MIGUEL, 2021), que são ilustradas na Figura 13. A primeira etapa é a de seleção do número de clusters,  $k$  clusters. A segunda etapa envolve a associação de cada grupo ao centróide mais próximo, esta primeira associação se dá de forma aleatória.

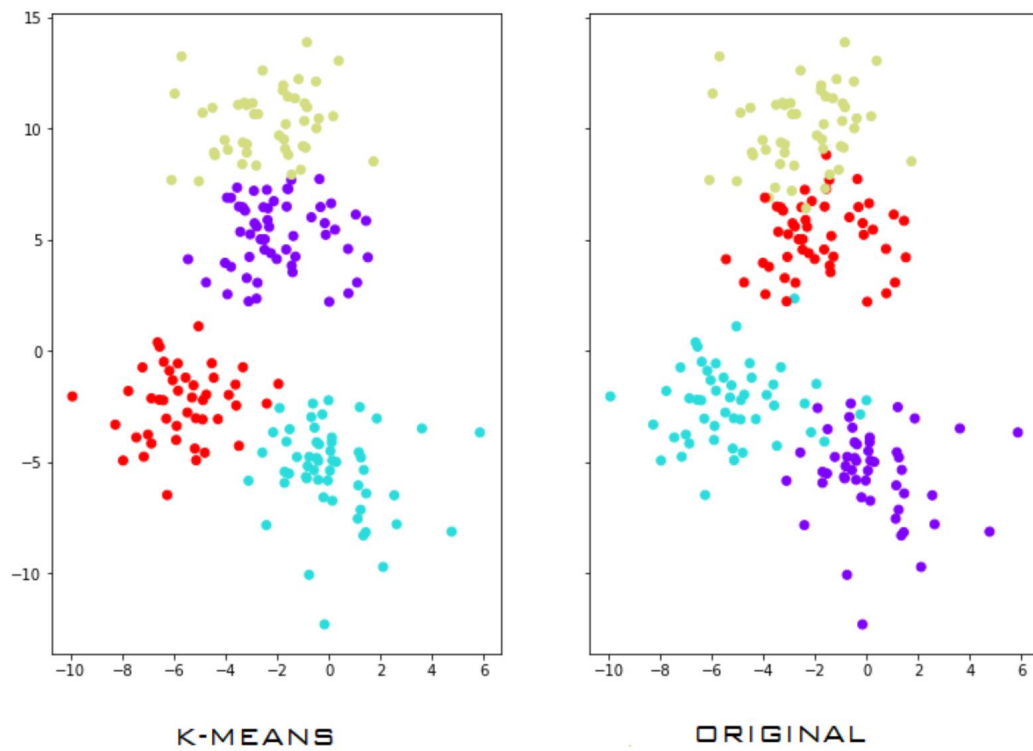
Figura 13: Fluxograma looping K-Means



Fonte: Autoria Própria

Então como próxima etapa, ocorre o cálculo de cada ponto para o centróide de menor distância Euclidiana. Em seguida os centróides são reposicionados de acordo com a média das posições dos pontos do grupo, o reposicionamento ocorre de forma repetitiva até a encontrar a localização ideal, um exemplo de agrupamento é visualizado na Figura 14.

Figura 14: Exemplo K-Means



Fonte: Autoria Própria

## 3 METODOLOGIA

Neste capítulo será realizado a apresentação da metodologia utilizada no desenvolvimento do trabalho, realizando de forma explicada os recursos necessários para alcançar o objetivo do projeto. Ele pode ser implementado de forma simples mas que possui uma alta confiabilidade

### 3.1 Linguagem Python

Para realização deste projeto foi escolhida a linguagem de programação PYTHON, por ser uma linguagem de alto nível e de fácil entendimento. Esta linguagem foi criada em 1991 pelo holandês Guido van Rossum (LUTZ, 2001), com o intuito de ajudar os programadores, por ter uma linguagem simples que permite a escrita de códigos mais legíveis e de fácil interpretação.

PYTHON possui suporte tanto para algoritmos complexos voltados a técnicas de orientação a objetos, quanto a algoritmos simples que permitem fazer diversas atividades através de poucas linhas de comando e se expandiu por ser muito expressiva, ou seja, de forma fácil é possível transcrever o raciocínio em algoritmos (SUMMERFIELD, 2010).

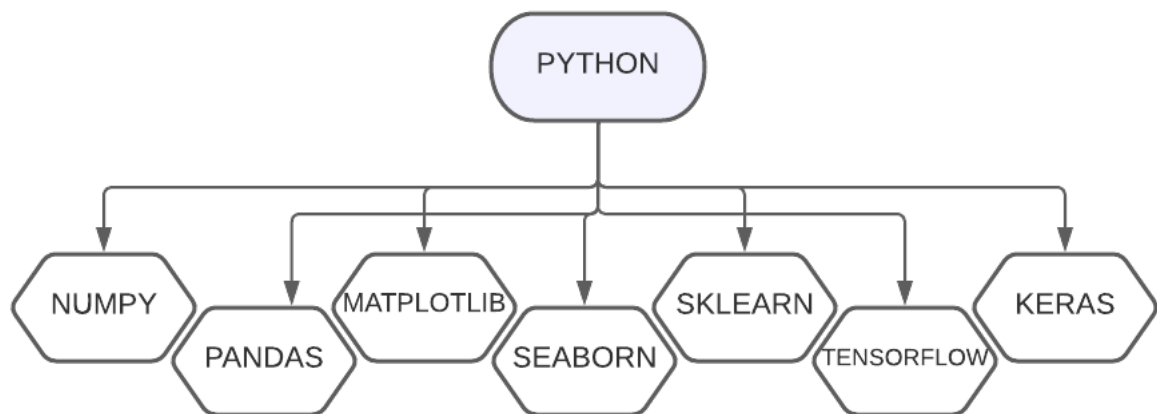
Por necessitar de bibliotecas com maior processamento, o trabalho foi desenvolvido utilizando a ferramenta Google Colaboratory, ferramenta semelhante ao Jupyter Notebook. Além de permitir acesso a diferentes bibliotecas sem a necessidade de downloads, o Collaboratory permite armazenamento em nuvem, compartilhamento e modificações de forma online e grátis.

#### 3.1.1 Bibliotecas

Para melhor desenvolvimento do trabalho, algumas bibliotecas foram utilizadas, e que são importantes serem citadas para melhor entendimento.

A linguagem apresenta uma extensa gama de bibliotecas, sendo esta composta por diversos módulos que permitem utilização das mais diferentes funcionalidades dentro dos algoritmos, facilitando o trabalho dos programadores e desenvolvedores (FOUNDATION, 2019). Neste projeto foram utilizadas as seguintes bibliotecas representas na Figura 15 e explanadas logo a seguir.

Figura 15: Bibliotecas Utilizadas



Fonte:Autoria Própria

- **Numpy:** A biblioteca desenvolvida para computação científica e cálculos numéricos, com ele tem-se a possibilidade de trabalhar com objetos multidimensionais envolvendo array(MCKINNEY, 2012).
- **Pandas:** Biblioteca desenvolvida para manipular e analisar de forma fácil grandes volumes de dados. De forma interna possui metodologias capazes de filtrar, agrupar e organizar os dados (MCKINNEY, 2012).
- **Matplotlib:** Pacote Python desenvolvido para visualização e exploração de dados a partir de diferentes tipos gráficos.
- **Seaborn:** Desenvolvido a partir do Matplotlib, o seaborn é a biblioteca para criação de gráficos estatísticos. Suas funções permitem que seja trabalhado junto aos dataframes da biblioteca Pandas e aos arrays da biblioteca Numpy (WASKOM, 2020).
- **Sklearn:** Scikit-learn, ou apenas sklearn, é uma biblioteca desenvolvida para trabalhar especificamente com machine learning. Possui módulos que variam de acordo com o momento do desenvolvimento do código e foi projetada para operar junto a bibliotecas numéricas como o Numpy.

Para este trabalho será usado o módulo de pré-processamento para que os dados sejam normalizados no intervalo entre zero e um, para que o modelo proposto de LSTM não use dados negativos ou discrepantes. A Figura 16 exemplifica um módulo utilizado para o pré-processamento.

Figura 16: Exemplo Sklearn

```

>>> X_train = np.array([[ 1., -1.,  2.],
...                      [ 2.,  0.,  0.],
...                      [ 0.,  1., -1.]])
>>> min_max_scaler = preprocessing.MinMaxScaler()
>>> X_train_minmax = min_max_scaler.fit_transform(X_train)
>>> X_train_minmax
array([[0.5       , 0.       , 1.        ],
       [1.        , 0.5      , 0.33333333],
       [0.        , 1.        , 0.        ]])

```

Fonte: Internet (Scikit Learn - *Preprocessing Data*)

- **TensorFlow:** É um Framework, open-source, para facilitar a criação de códigos de machine learning em diferentes linguagens. Desenvolvida pela Google em 2015, logo se tornou uma das fundamentais ferramentas para trabalho e treinamento de deep learning.
- **Keras:** É uma API – Application Programming Interface, de alto nível desenvolvido junto ao Tensorflow. Seu projeto permitiu que a criação de redes ocorresse de forma simplificada e abrangendo algoritmos de otimização, de normalização e de camadas de ativação.

De acordo com (GULLI; PAL, 2017) suas aplicações suportam desde redes neurais padrões até rede convolucionais e recorrentes, além de permitir que sejam produzidos modelos em smartphones ou na web.

### 3.2 Estudo de Caso

Neste trabalho foram empregados os dados telemedidos em unidades de alta tensão dispostos em todo o estado da Paraíba e disponibilizados pela concessionária a pedido da Companhia de Água e Esgotos da Paraíba (CAGEPA).

Os dados são referentes às medições do ano de 2021, dos meses de janeiro, fevereiro, maio e junho, porém, o trabalho desenvolvido poderá ser aproveitado para futuros bancos de dados com os meses completos de um ano. A Figura 17 ilustra os dados recebidos, em que cada coluna representará um dado disponibilizado. A primeira coluna visualiza os horários as quais as medições foram realizadas. A segunda, terceira, quarta e quinta coluna apresentam os dados de energia ativa e reativa medidas na UC. Suas colunas são denominadas a partir da junção do número do medidor eletrônico da unidade com o sufixo do tipo de dado apresentado. O que cada sufixo significa será abordados mais a frente na seção 3.2.1. Estas quatro colunas se repetem para o medidor de cada unidade consumidora da CAGEPA.

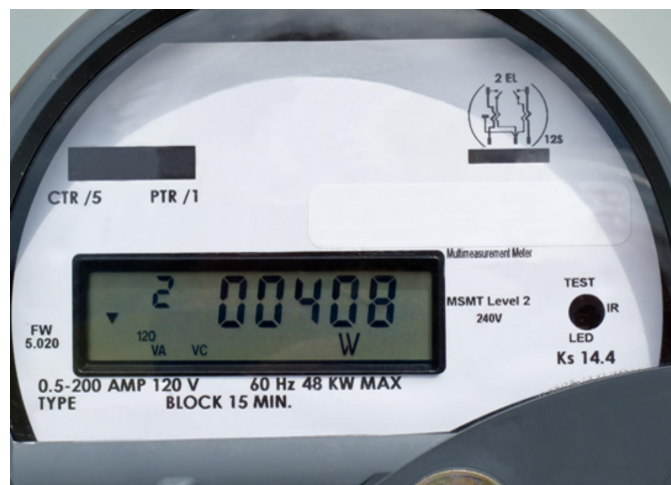
Figura 17: Dados Recebidos

DATA/Hora	00021107-EAE	00021107-EAR	00021107-ERE	00021107-ERR	00021131-EAE	00021131-EAR	00021131-ERE	00021131-ERR
01/01/2021 00:15	7,698		4,866					
01/01/2021 00:30	1,578		1,008					
01/01/2021 00:45	7,194		4,332					
01/01/2021 01:00	3,066		1,974					
01/01/2021 01:15	4,308		2,7					
01/01/2021 01:30	5,7		3,45					
01/01/2021 01:45	4,428		2,814					
01/01/2021 02:00	4,05		2,538					
01/01/2021 02:15	6,192		3,702					
01/01/2021 02:30	3,234		1,986					
01/01/2021 02:45	8,046		5,004					
01/01/2021 03:00	2,61		1,614					
01/01/2021 03:15	5,268		3,096					
01/01/2021 03:30	4,704		2,982					
01/01/2021 03:45	3,024		1,896					
01/01/2021 04:00	6,966		4,176					
01/01/2021 04:15	3,744		2,388					

Fonte: Autoria Própria

Eles foram obtidos por meio de telemedições, através de medidores eletrônicos associados um equipamento eletrônico chamado de *gateway*. Os medidores utilizados nestas unidades podem ser exemplificados como o da Figura 18.

Figura 18: Medidor Eletrônico



Fonte: Artigo (Medidor de Energia - Eletricistas Brasil)

### 3.2.1 Conjunto de Dados

Para o andamento das análises, foram utilizados os dados recebidos de energia ativa e reativa entregues às cargas das unidades de alta tensão. As medições chegam com frequências de 15 em 15 minutos, ou seja, uma amostra é medida a cada 15 minutos, totalizando um valor de 96 medições diárias, como ilustrado na Figura 19.

Figura 19: Dados 96 Medições

	DATA/Hora	00021107- EAE
0	01/01/2021 00:15	7,698
1	01/01/2021 00:30	1,578
2	01/01/2021 00:45	7,194
3	01/01/2021 01:00	3,066
4	01/01/2021 01:15	4,308
...	...	...
91	01/01/2021 23:00	6,624
92	01/01/2021 23:15	6,966
93	01/01/2021 23:30	4,398
94	01/01/2021 23:45	7,08
95	02/01/2021 00:00	8,298

Fonte: Autoria Própria

O banco de dados é composto por colunas Energia Ativa Entregue (EAE), Energia Ativa Recebida (EAR), Energia Reativa Entregue (ERE) e Energia Reativa Recebida (ERR), sendo que apenas as colunas EAE e ERE possuem valores válidos e que serão aplicados para este trabalho, colunas estas que pode ser visualizadas na Figura 20. A coluna DATA/Hora representando a sequência temporal dos dados e será de importancia para o desenvolvimento do modelo de previsão proposto.

Figura 20: Dados Colunas

	DATA/Hora	00021107- EAE	00021107- EAR	00021107- ERE	00021107- ERR
0	01/01/2021 00:15	7,698		4,866	

Fonte: Autoria Própria

Como os dados foram disponibilizados pela CAGEPA, significa que cada unidade consumidora terá as quatro colunas já definidas anteriormente, sendo estas colunas relacio-

nadas pelo número do medidor junto com a terminação do que está sendo medido.

### 3.3 Determinação da Curva Limite

Para análise de ultrapassagem do consumo, será abordado uma técnica simples e eficiente de verificação ao se criar curvas de perfis diárias, com base na média e no desvio padrão.

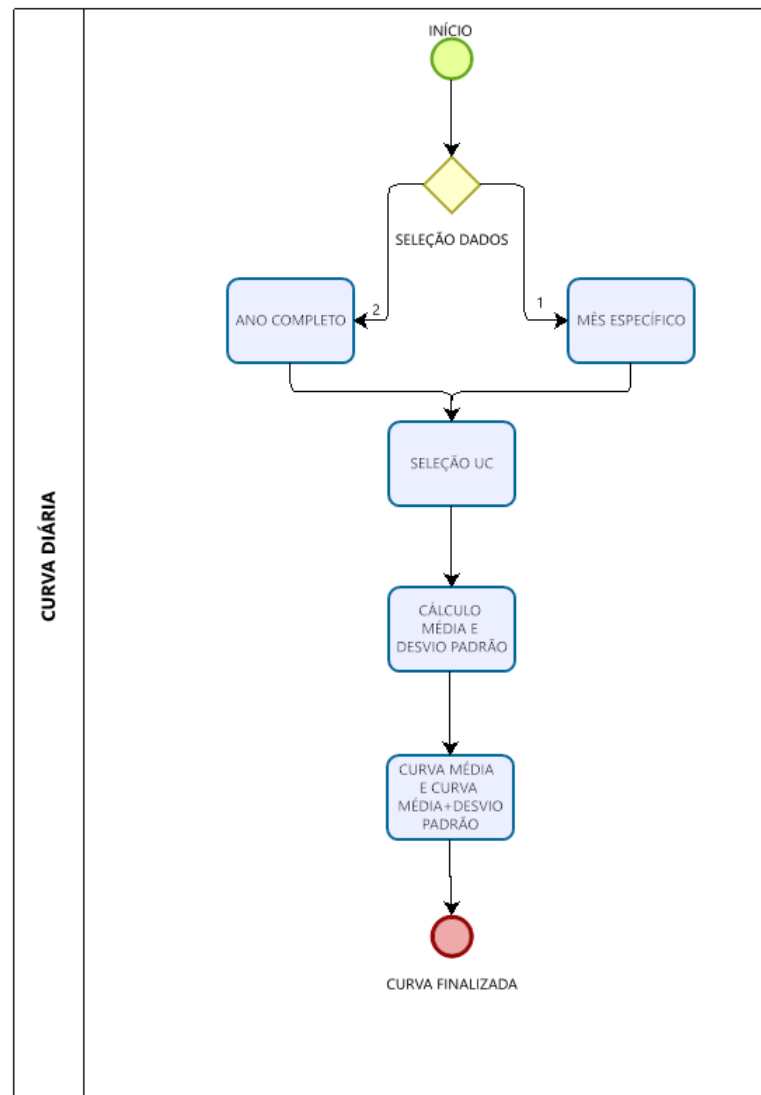
Os dados primeiramente passam por um processo de seleção por meio de entradas solicitadas ao usuário. Após a etapa interativa, as curvas serão criadas de acordo com os dados determinados, variando entre os dados dos meses completos, entrada 2 no programa, ou apenas um mês em específico a escolha, com entrada 1 no código desenvolvido.

Os valores utilizados para os cálculos das médias, desvios padrões e limites poderão ser encontrados no Anexo I.

Os gráficos criados representam as curvas diárias ou típicas, ou seja, apresentam as médias da potências consumidas para cada dia da semana, que será variado de acordo com a seleção da UC, se os dados utilizados foram de um mês isolado ou de todos os dados que compõem o conjunto de dados geral. O fluxograma da Figura 21 simplifica o caminho que de escolha dos dados.



Figura 21: Fluxo de Escolha de Dados



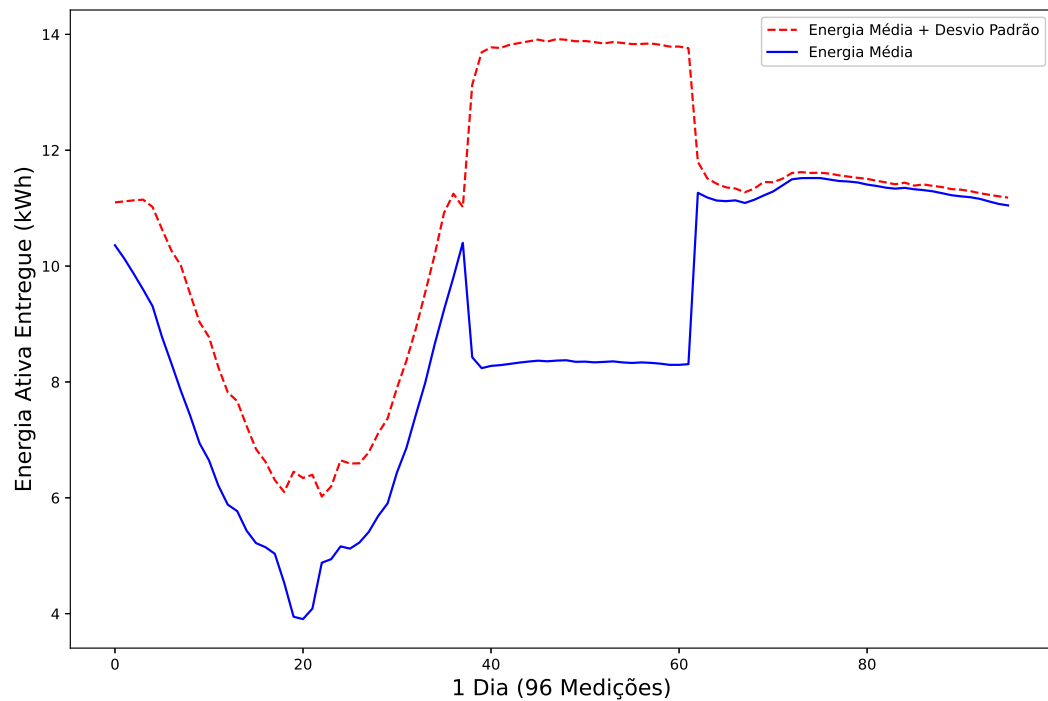
Fonte: Autoria Própria

Com os dados organizados, o programa passa para a próxima etapa, a seleção da UC a ser observada de forma mais específica, ocorrendo de forma interativa.

Então é calculada a média e o desvio padrão da energia ativa entregue, para as medições originais de 15 em 15 minutos e para os dados reorganizados para medições de 1 em 1 hora. Por fim, ocorre a plotagem das curvas.

Na Figura 22 encontram-se ilustradas a energia média da unidade em um dia da semana, contínua, e a energia média somado ao desvio padrão, curva tracejada. Neste caso, está sendo verificada a curva média para 96 horários de quatro segundas-feiras, para o mês de janeiro. A curva limite possui um comportamento diferente da curva média, justamente por conta do desvio padrão que para este intervalo é um valor alto.

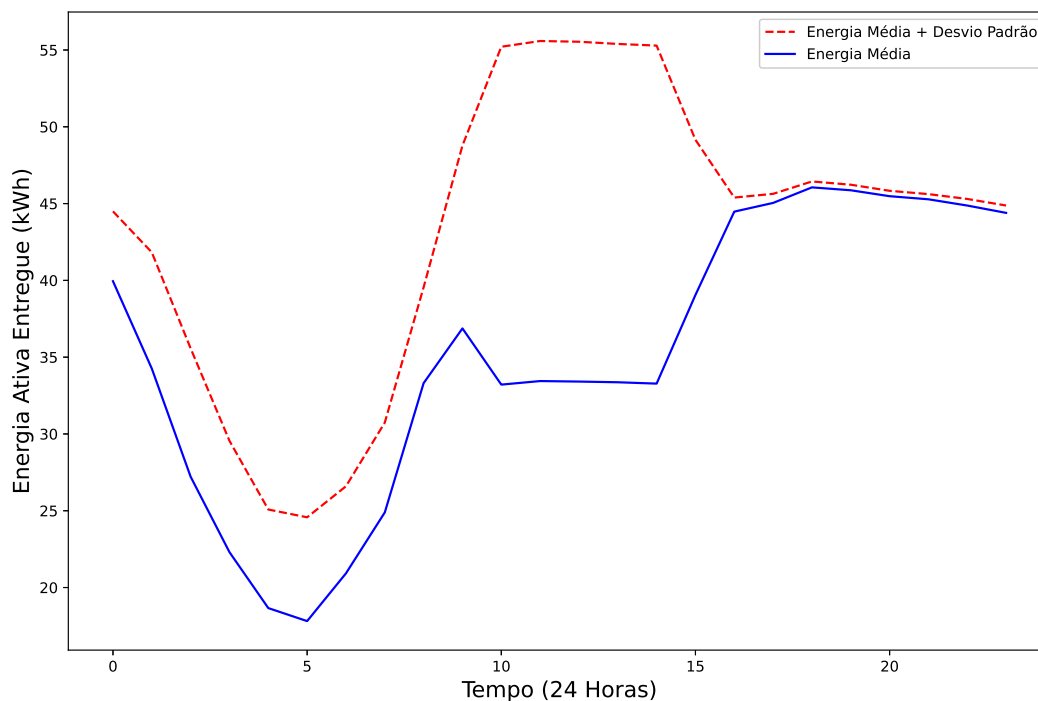
Figura 22: Curva Diária 96 Medições - Segunda-Feira



Fonte: Autoria Própria

Os mesmos dados da Figura 22 são organizados agora para um o intervalo de tempo de 1 em 1 hora. Então a média e o desvio padrão são calculados novamente. A Figura 23 apresenta a nova curva diária de energia entregue, curva contínua, junto com a nova curva média somada ao desvio padrão, curva tracejada.

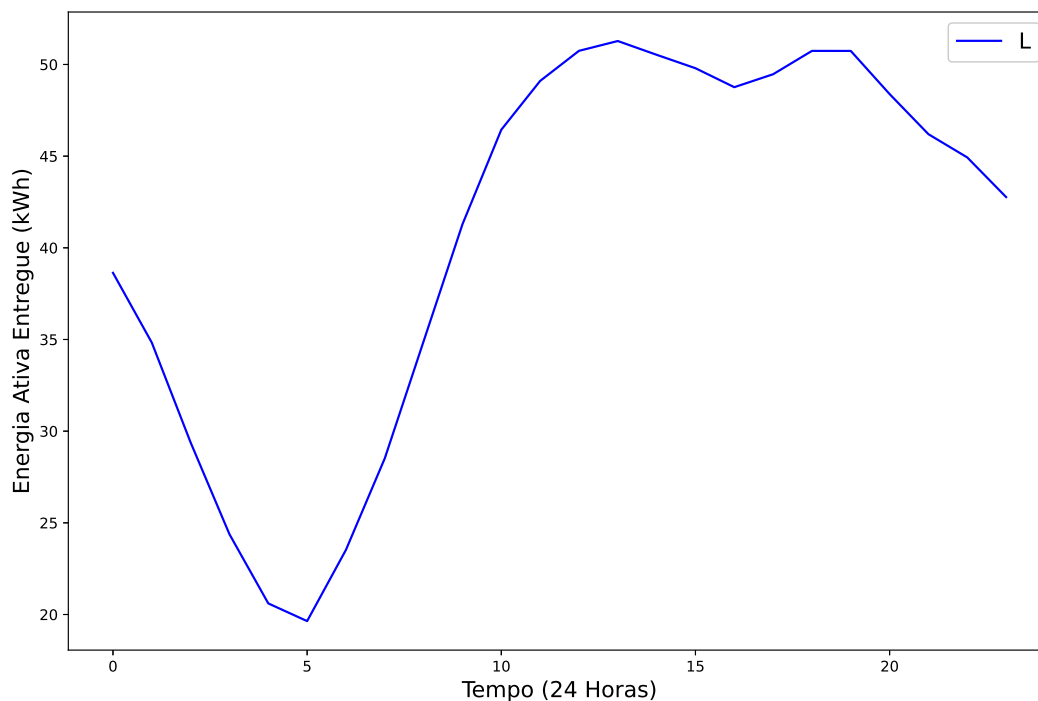
Figura 23: Curva Diária 24 Medições - Segunda-Feira



Fonte: Autoria Própria

Por fim, encontra-se a curva definida como limite da unidade consumidora que será usada para comparação com as médias da semana. Para esta curva utiliza-se a média do conjunto de dados escolhida, no caso foram utilizados todos as medições encontradas para o mês analisado de janeiro, o resultado é observado na curva da Figura 24.

Figura 24: Curva Limite UC



Fonte: Autoria Própria

### 3.4 Determinação Financeira

Além de verificar a ultrapassagem do limite diário, também será abordado uma visão financeira da curva de consumo, levando em consideração a média dos dias da semana dentro de um mês. Para isto foram utilizados os valores vigentes para as modalidades tarifárias Verde, Azul AH3 - Serviço Público Tensão de 69 kV e Azul AH4 - Serviço Público Tensão de 13,8 kV. A concessionária disponibiliza duas tabelas de tarifação dependendo do local onde a UC se encontra no estado da Paraíba. As tabelas apresentadas nas Figuras 25 e 26 a seguir mostram os valores em Reais por quilowatt-hora (kWh) utilizados para os dados são medidos.

Figura 25: Tarifação 1

TARIFA	CONSUMO PONTA (R\$/kWh)	CONSUMO FORA PONTA (R\$/kWh)	ULTRAPASSAGEM DEMANDA PONTA (R\$/kWh)	ULTRAPASSAGEM DEMANDA FORA PONTA (R\$/kWh)
AZUL AH3	0.41319	0.26768	42.96	22.34
AZUL AH4	0.43257	0.28706	104.3	52.44
VERDE	1.62414	0.28706	52.44	52.44

Fonte: Autoria Própria

Figura 26: Tarifação 2

TARIFA	CONSUMO PONTA (R\$/kWh)	CONSUMO FORA PONTA (R\$/kWh)	ULTRAPASSAGEM DEMANDA PONTA (R\$/kWh)	ULTRAPASSAGEM DEMANDA FORA PONTA (R\$/kWh)
AZUL AH4	0.38563	0.24691	74.02	36.96
VERDE	1.23178	0.24691	36.96	36.96

Fonte: Autoria Própria

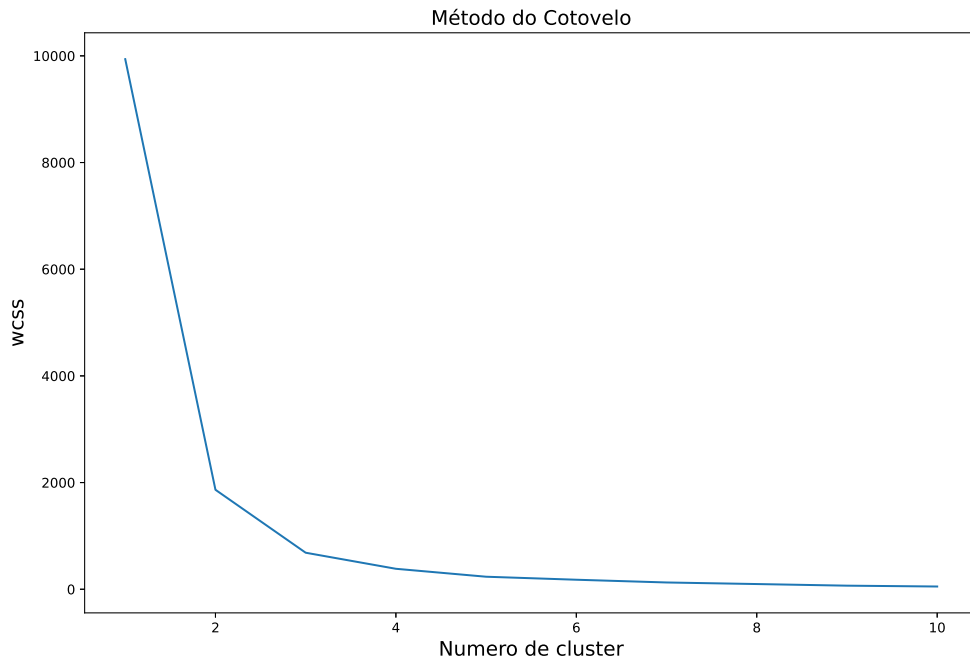
### 3.5 Número de Clusters K-Means

Para identificação de valores superiores ao limite médio, ou seja, ultrapassados foi utilizado um modelo de agrupamento baseado na técnica K-Means, como explicado na seção 2.4.1. Os grupos serão repartidos a partir dos valores de energia ativa e dos horários de medição.

Para construção do modelo proposto, é necessário a determinação do parâmetro de quantidade de clusters, K. Conforme a literatura existem diferentes metodologias para seleção do número de clusters "Ideal" (KODINARIYA; MAKWANA, 2013). Entre eles pode-se citar os métodos da regra de ouro, da *Cross-validation*, silhueta, *elbow*, sendo este último o método optado para este trabalho (KODINARIYA; MAKWANA, 2013).

O método escolhido para determinação do número de clusters ideal foi o do *elbow*, ou de forma traduzida, o método do cotovelo. Este método observa que o aumento do número de cluster reduz a soma dos quadrados intra-clusters. A partir do gráfico gerado, o ponto onde ocorre a curva mais suave e de equilíbrio é considerado o número de clusters ideal, (TEMPORAL, 2019). A curva mostrada na Figura 27 ilustra o resultado da metodologia aplicada ao nosso banco de dados.

Figura 27: Método do Cotovelo - Jan



Fonte: Autoria Própria

Sendo o eixo y referente ao valor da soma dos quadrados intra-cluster (*Within-clusters sum-of-squares* - WCSS), e o eixo x referente ao número de clusters ideiais.

Os valores escolhidos para motivos de comparação neste projeto foram os de K=3 e K=4, a partir do ponto de suavização da curva da Figura 27.

### 3.6 Métricas de Avaliação

Para a primeira análise comparativa entre os diferentes modelos desenvolvidos neste trabalho, levamos em consideração duas métricas de avaliação.

#### 3.6.1 Erro Médio Quadrático

Como primeira métrica temos o erro médio quadrático (MSE). Este erro é calculado a partir da média dos quadrados dos erros, isto é, a média da diferença quadrática entre o valor atual e o previsto, como podemos observar na equação 3.

$$MSE = \frac{1}{N} \times \sum_{i=1}^N (V_a - V_p)^2 \quad (3)$$

- **N** é o número de amostras;
- **Va** é o valor atual;

- **V<sub>p</sub>** é o valor previsto.

Esta métrica é muito empregada para analisar o grau de exatidão do modelo.

### 3.6.2 Erro Médio Absoluto

A segunda métrica utilizada foi a do erro médio absoluto (MAE). Podemos calcular este erro a partir da média dos seus erros absolutos, ou seja, é calculada a média do somatório do módulo de cada erro encontrado. Isto permite que o valor encontrado não seja tão afetado por pontos extremos. Na equação 4 podemos observar o cálculo desta métrica.

$$MAE = \frac{1}{N} \times \sum_{i=1}^N |V_a - V_p| \quad (4)$$

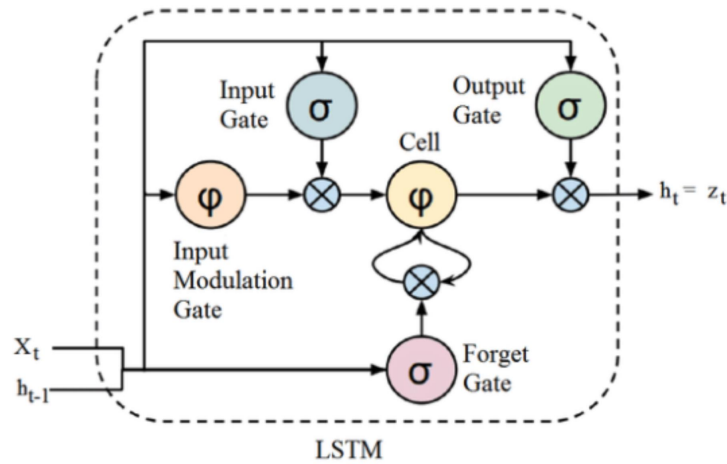
- **N** é o número de amostras;
- **V<sub>a</sub>** é o valor atual;
- **V<sub>p</sub>** é o valor previsto.

## 3.7 Modelo de Previsão de Consumo

O modelo de rede neural recorrente escolhido para realizar a previsão dos dados foi o de Long Short Term Memory (LSTM). Este modelo consegue resolver os problemas comuns das redes neurais recorrentes (RNN) comuns, problema esse que é a perda de função de aprendizado conforme ocorre o aumento do gap entre as informações (BROWNLEE, 2017).

No contexto deste trabalho, a previsão permitirá aos gestores intervirem e se planejarem para possíveis anomalias nos consumos das unidades analisadas.

Figura 28: Modelo LSTM

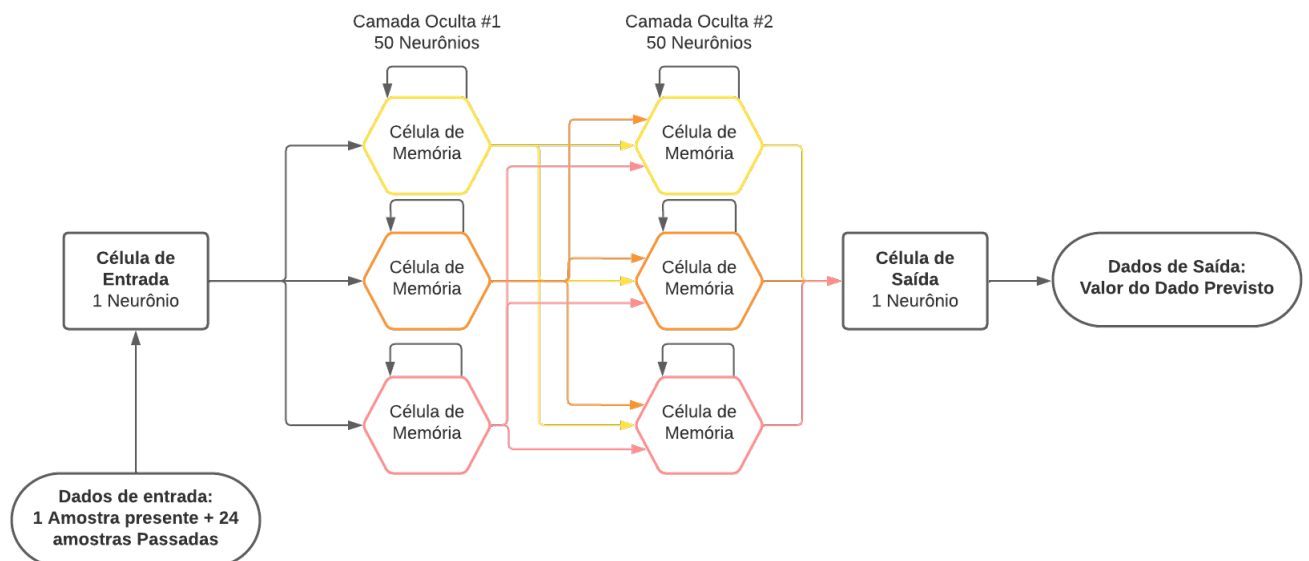


Fonte: Livro (Cap. 51 - Arquitetura de Redes Neurais Long Short Term Memory (LSTM) (ACADEMY, 2018b))

Como ilustrado na Figura 28 acima, o modelo LSTM possui três gates para manipular os dados de entrada. O primeiro gate, **Forget Gate**, responsável por decidir se a informação do período anterior será útil ou descartável; **Input Gate** é gate encarregado de tentar aprender novas informações dos dados de entrada; e por último o **Output Gate**, onde os dados são atualizados do período atual para o futuro (SALEM, 2022).

Para este trabalho foram feitos três modelos diferentes de LSTM, para então comparação dos resultados de saída. Cada modelo possui o mesmo número de dados de entrada. Serão analisadas uma amostra de entrada, com 24 amostras de histórico para chegar na previsão das amostras futuras, o que será variado apenas na quantidade de camadas ocultas. O modelo básico proposto é ilustrado na Figura 30.

Figura 29: Modelo Proposto LSTM



Fonte: Autoria Própria



As Tabelas 1 e 2 apresentam os parâmetros externos escolhidos para os modelos desenvolvidos de rede neural recorrente. O número de épocas foi escolhido a partir de simulações prévias dos modelos, ou seja, uma quantidade que apresentou melhor resultados de curva de validação de treinamento.

Tabela 1: Modelo Métrica 1

LSTM - Parâmetros de Treinamento	
Parâmetros	Valores
Épocas	50
Cost Function	MSE
Otimizador do gradient	ADAM
Função de Ativação	ReLU

Tabela 2: Modelo Métrica 2

LSTM - Parâmetros de Treinamento	
Parâmetros	Valores
Épocas	50
Cost Function	MAE
Otimizador do gradient	ADAM
Função de Ativação	ReLU

Neste método iremos separar os dados em dois grupos, o primeiro grupo será voltado para o treinamento da LSTM, será usado para este grupo os dados dos meses de janeiro, fevereiro e maio. Já o segundo grupo será voltado para o teste, e terá os dados do mês de junho. Para o cálculo dos erros foram utilizado as métricas MSE e MAE, como explanados na seção anterior 3.6. Com a comparação das curvas referentes a cada métrica, podemos avaliar que a curva MAE obteve valores mais aproximados a curva original dos dados.

A Figura 31 apresenta os dados formatados para as 24 amostras passadas e devidamente separados nos dois grupos, de treinamento e teste.

Figura 30: Amostras Treinamento e Teste

```
[ ] x_train2.shape, y_train2.shape
    ((2136, 24, 1), (2136, 1))

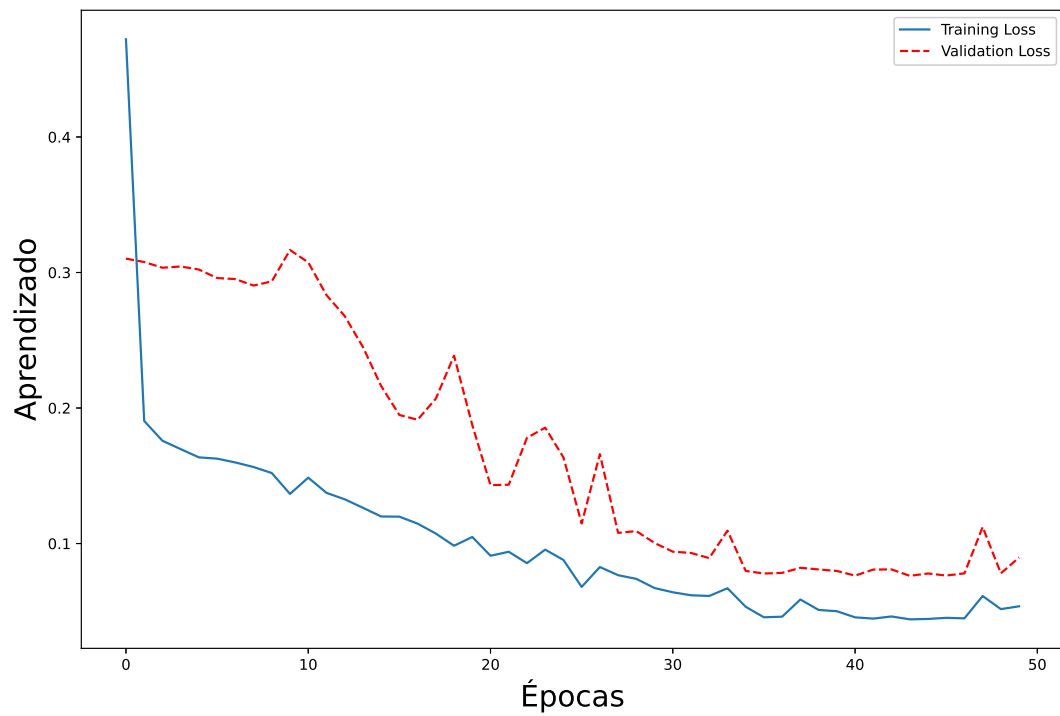
[ ] x_test2, y_test2 = data_preparation(data_test2)

[ ] x_test2.shape
    (696, 24, 1)
```

Fonte: Autoria Própria

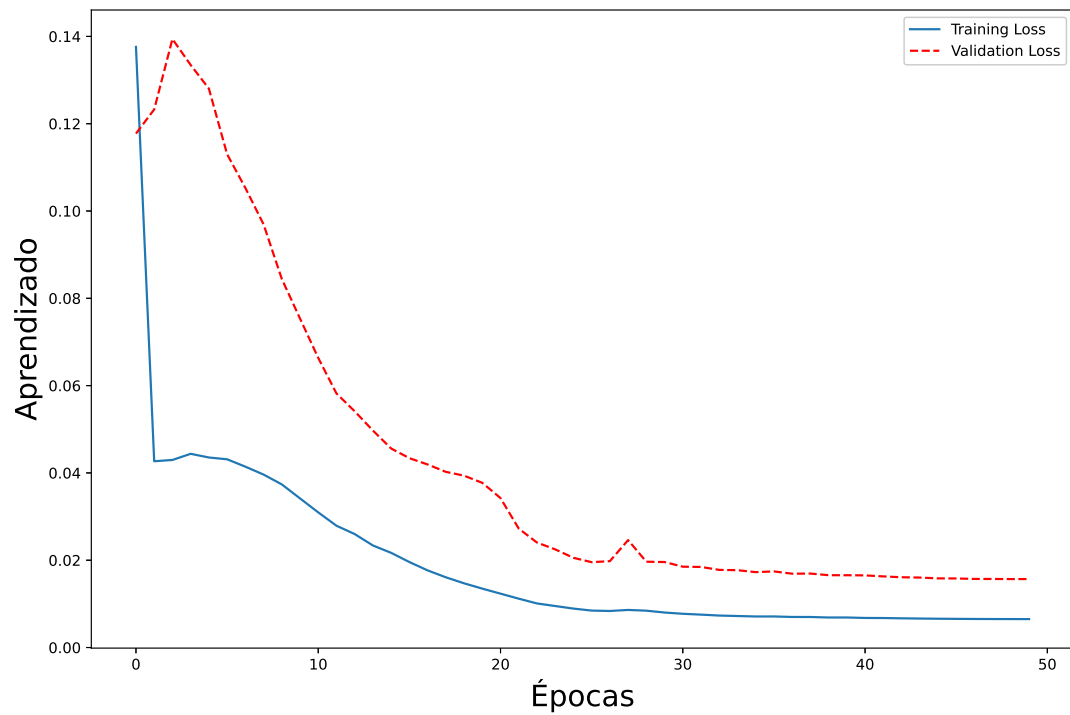
Nas Figuras 32 e 33, observamos o erro de treinamento entre o conjunto de entrada e o conjunto de saída do nosso modelo. A Figura 32 possui as curvas para a métrica MAE, e a Figura 33 as curvas quando a métrica MSE é aplicada, isto aplicado ao primeiro modelo desenvolvido.

Figura 31: Curvas "Training Loss"+ "Validation Loss"Modelo 1 MAE



Fonte: Autoria Própria

Figura 32: Curvas "Training Loss"+ "Validation Loss"Modelo 2 MSE



Fonte: Autoria Própria

A curva contínua refere-se a perda do treino, "*Training Loss*", e indica o quanto o modelo se adaptou aos valores de entrada. Já a curva tracejada refere-se à validação dos dados, "*Validation Loss*" e indica como o modelo irá se adaptar ao valores de entrada.

Para todos os modelos foi aplicada a mesma função de ativação, ReLU, que possui vantagem por não ativar todos os neurônios ao mesmo tempo (RIZZO; CANATO, 2020).

## 4 Resultados

Neste capítulo será exposto e discutido os resultados que foram obtidos neste trabalho. Inicialmente será realizado a verificação da metodologia de correção do banco de dados original.

Em seguida, será analisado os resultados encontrados nas curvas diárias e da curva financeira encontrada. Por fim, será feito a comparação do resultado obtido pelo modelo de rede neural artificial proposto para previsão com os dados originais.

### 4.1 Correção do Banco de Dados

Os dados são inicialmente recebidos no formato xlsx, que encontravam-se sem formatação ou separação bem definidas. Para facilitação do trabalho foi então feito a conversão do arquivo para csv, de forma que a leitura e tratamento dos dados foram simplificados. Após a conversão ocorre a leitura dos dados, e então entra-se devidamente na etapa de tratamento.

A primeira etapa de tratamento é a separação dos dados válidos dos nulos, vale salientar que algumas colunas podem conter estes valores zerados, mas não poderão ser considerados outliers por poderem se tratar de unidades não existentes para o período de tempo, porém, quando ocorrer a mudança de frequência dos dados de medições de 15 em 15 minutos para 1 em 1 hora, estes valores serão corrigidos pela média dos valores normais.

Agora com apenas valores válidos, ocorre a separação em dois DataFrames, o primeiro referente aos dados de energia ativa e o segundo a energia reativa, representado na Figura 33. A partir desta divisão as análises ocorrerão de formas separadas, apesar de que a rotina criada poderá ser utilizada de forma igual para os dois DataFrames. Para este trabalho será dada prioridade ao consumo de energia ativa, então todas as simulações e previsões feitas serão referentes as colunas EAE.

Figura 33: Separação DataFrames Válidos

```
#EAE ENERGIA ATIVA ENTREGUE
df_EAE= df3.loc[:, df3.columns.str.endswith('EAE')]

#EAR ENERGIA ATIVA RECEBIDA - APENAS DADOS NULOS
df_EAR= df3.loc[:, df3.columns.str.endswith('EAR')]

#ERE ENERGIA REATIVA ENTREGUE
df_ERE= df3.loc[:, df3.columns.str.endswith('ERE')]

#ERR ENERGIA REATIVA RECEBIDA - APENAS DADOS NULOS
df_ERR= df3.loc[:, df3.columns.str.endswith('ERR')]
```

Fonte: Autoria Própria

Com os dados organizados, pode-se então realizar a mudança dos nomes das colunas. A nomenclatura original vem como referência o número dos medidores. Para melhor espe-

cificação eles serão relacionados e substituídos pelos números das unidades consumidoras, através do código apresentado na Figura 34.

Figura 34: Correção Nome das Colunas

```
dict_from_list = dict(zip(x_key,y_value))
df_EAE= df_EAE.rename(columns=dict_from_list)
df_EAE['DATA']=df3['DATA']
df_EAE['HORA']=df3['HORA']
df_EAE['D/H']=df['DATA/Hora']
df_ERE= df_ERE.rename(columns=dict_from_list)
df_ERE['DATA']=df3['DATA']
df_ERE['HORA']=df3['HORA']
df_ERE['D/H']=df['DATA/Hora']
```

Fonte: Autoria Própria

A última parte de correção dos dados, se encontra na separação das datas por dia da semana e por mês, como visto na Figura 35. Isto será fundamental para as próximas etapas.

Figura 35: Separação Dia e Mês

```
#Dicionário p/ separação dos dias
DIAS = {0:'Segunda-feira',1:'Terça-feira',2:'Quarta-feira',3:'Quinta-feira',4:'Sexta-feira',5:'Sábado',6:'Domingo'}
MES = {1:'Jan',2:'Fev',3:'Mar',4:'Abr',5:'Mai',6:'Jun',7:'Jul',8:'Ago',9:'Set',10:'Out',11:'Nov',12:'Dez'}

df_EAE['DATA']=pd.to_datetime(df_EAE['DATA'])
df_EAE['DIA']=df_EAE['DATA'].dt.weekday.apply(lambda x: DIAS[x])
df_EAE['MES']=df_EAE['DATA'].dt.month.apply(lambda x: MES[x])
df_EAE.to_excel('/content/drive/MyDrive/Colab Notebooks/CONSUMO/EAE.xlsx')

df_ERE['DATA']=pd.to_datetime(df_ERE['DATA'])
df_ERE['DIA']=df_ERE['DATA'].dt.weekday.apply(lambda x: DIAS[x])
df_ERE['MES']=df_ERE['DATA'].dt.month.apply(lambda x: MES[x])
df_ERE.to_excel('/content/drive/MyDrive/Colab Notebooks/CONSUMO/ERE.xlsx')
```

Fonte: Autoria Própria

Por fim, pode-se comparar os resultados das correções a partir das Figuras a seguir.

No caso apresentado na Figura 36 temos os dados de entrada sem nenhum tipo de tratamento, nas colunas referentes à energia ativa recebida e energia reativa recebida. Pode observar que a existência de dados faltantes, resultando em valor total nulo para a UC.

Figura 36: Dados de Entrada

	DATA/Hora	00025010-EAE	00025010-EAR	00025010-ERE	00025010-ERR
0	01/01/2021 00:15	11,25		1,788	
1	01/01/2021 00:30	11,22		1,77	
2	01/01/2021 00:45	11,226		1,764	
3	01/01/2021 01:00	11,214		1,758	
4	01/01/2021 01:15	11,19		1,728	
...	...	...	...	...	...
11516	30/06/2021 23:15	10,404		1,8	
11517	30/06/2021 23:30	10,368		1,812	
11518	30/06/2021 23:45	9,966		1,722	
11519	01/07/2021 00:00	9,534		1,632	
11520	TOTAL	117097,2	0	25024,89	0

11521 rows x 5 columns

Fonte: Autoria Própria

No caso da Figura 37 tem-se os dados ao final das seguintes etapas de correções: retirada de valores nulos; correlação medidor - UC; separação da data em mês e dia da semana, para o intervalo de medição de 15 em 15 minutos. A Figura 37 irá representar apenas os dados de uma UC que foram utilizados para realizar todas as simulações deste trabalho.

Figura 37: Dados Corrigidos - 96 Medições

	DATA	HORA	D/H	DIA	MES	9980091
0	2021-01-01	00:15:00	01/01/2021 00:15	Sexta-feira	Jan	11.250
1	2021-01-01	00:30:00	01/01/2021 00:30	Sexta-feira	Jan	11.220
2	2021-01-01	00:45:00	01/01/2021 00:45	Sexta-feira	Jan	11.226
3	2021-01-01	01:00:00	01/01/2021 01:00	Sexta-feira	Jan	11.214
4	2021-01-01	01:15:00	01/01/2021 01:15	Sexta-feira	Jan	11.190
...	...	...	...	...	...	...
11515	2021-06-30	23:00:00	30/06/2021 23:00	Quarta-feira	Jun	10.440
11516	2021-06-30	23:15:00	30/06/2021 23:15	Quarta-feira	Jun	10.404
11517	2021-06-30	23:30:00	30/06/2021 23:30	Quarta-feira	Jun	10.368
11518	2021-06-30	23:45:00	30/06/2021 23:45	Quarta-feira	Jun	9.966
11519	2021-07-01	00:00:00	01/07/2021 00:00	Quinta-feira	Jul	9.534

11520 rows x 6 columns

Fonte: Autoria Própria

Após todas as correções os dados finalizados são rearrumados para o segundo intervalo, 1 em 1 hora, caso exposto na Figura 38.

Figura 38: Dados Corrigidos - intervalo 24 horas

	DATA	HORA	D/H	DIA	MES	9980091
0	2021-01-01	00:15:00	01/01/2021 00:15	Sexta-feira	Jan	11.250
1	2021-01-01	00:30:00	01/01/2021 00:30	Sexta-feira	Jan	11.220
2	2021-01-01	00:45:00	01/01/2021 00:45	Sexta-feira	Jan	11.226
3	2021-01-01	01:00:00	01/01/2021 01:00	Sexta-feira	Jan	11.214
4	2021-01-01	01:15:00	01/01/2021 01:15	Sexta-feira	Jan	11.190
...	...	...	...	...	...	...
11515	2021-06-30	23:00:00	30/06/2021 23:00	Quarta-feira	Jun	10.440
11516	2021-06-30	23:15:00	30/06/2021 23:15	Quarta-feira	Jun	10.404
11517	2021-06-30	23:30:00	30/06/2021 23:30	Quarta-feira	Jun	10.368
11518	2021-06-30	23:45:00	30/06/2021 23:45	Quarta-feira	Jun	9.966
11519	2021-07-01	00:00:00	01/07/2021 00:00	Quinta-feira	Jul	9.534

11520 rows x 6 columns

Fonte: Autoria Própria

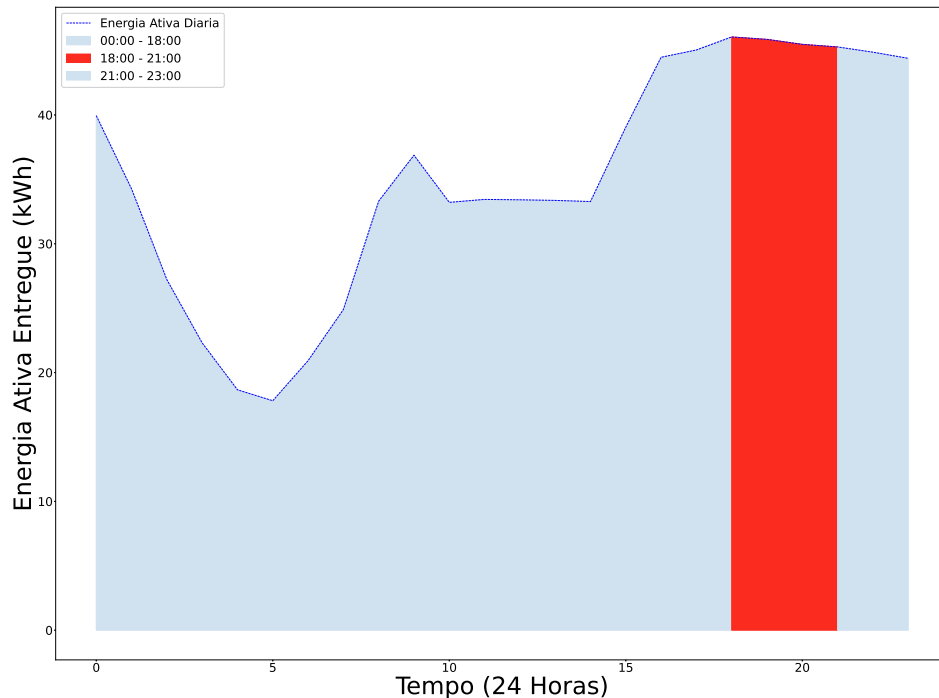
## 4.2 Curva de Consumo Diário

Para determinação da curva diária, inicialmente é feito uma seleção dos dados a serem analisados, como mostrado pelo fluxograma da seção 3.3. Para ambos os casos, em seguida é selecionada o número da unidade consumidora (UC) a ser examinada mais pontualmente. Com a UC escolhida, então é calculada a média por horário e gerado alguns gráficos.

O primeiro gráfico é para análise da energia entregue durante o período de ponta e fora ponta, para tomadas de decisões. Para a Figura 39, é apresentado a curva típica de uma segunda-feira, o limite da curva é a energia média para todas as segundas-ferias do mês analisada para o intervalo de 24 horas.

Observa-se que para o intervalo de ponta, horário entre 18 horas a 21 horas ocorreu um pico no consumo de energia ativa, isto pode indicar que sucedeu a entrada de cargas mais pesadas no sistema elétrico da UC.

Figura 39: Ponta e Fora Ponta - Segunda-Feira



Fonte: Autoria Própria

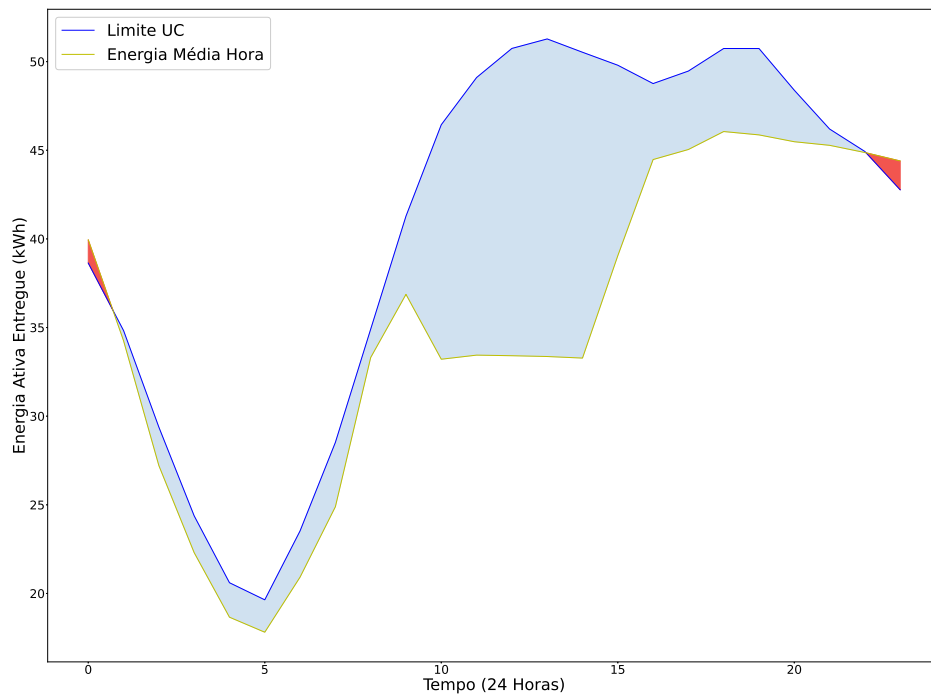
Para cada dia da semana foi encontrada sua própria curva típica, a exemplo da Figura 39.

O segundo gráfico, realiza a comparação entre o limite mensal, encontrado através da média de todos os dias do mês analisado, com a curva típica diária para a UC em questão. O modelo da Figura 40, apresenta a curva tracejada referente ao limite mensal e a curva contínua para a energia média medida no local.

Neste caso, se a média encontrada durante aquele dia for maior que o limite mensal, então a área da curva será preenchida pela cor vermelha, caso contrário a área será preenchida com a cor azul mostrando que não ocorreu ultrapassagem naquele horário. Na Figura 40, observa-se que ocorreram ultrapassagens durante o período de 23 horas e meia-noite.



Figura 40: Ultrapassagem Diária - Segunda-Feira



Fonte: Autoria Própria

De forma mais geral, o DataFrame da Figura 41 foi criado para aprestar os valores e horários onde ocorreram ultrapassagem durante o dia e para melhor visualização. As saídas que possuem NaN, são dados de valores normais e que não possuem ultrapassagem, já os valores que são visíveis estão relacionados aos dados considerados ultrapassagem.

Figura 41: Ultrapassagem Semanal

	Seg	Ter	Qua	Qui	Sex	Sab	Dom
0	39.9465	43.1985	43.4355	43.4580	41.8464	41.8596	41.9280
1	NaN	36.7455	37.8630	39.5130	41.1216	39.2052	39.7008
2	NaN	NaN	30.2370	34.1445	36.2796	33.9072	36.6432
3	NaN	NaN	24.6195	28.1280	31.7112	29.1264	32.5764
4	NaN	NaN	21.6075	22.5240	27.8148	25.4328	29.5908
5	NaN	NaN	NaN	21.8835	24.6708	23.4852	25.8420
6	NaN	27.1890	NaN	24.8070	27.2052	24.9372	24.1548
7	NaN	32.5035	NaN	29.0265	34.9956	30.3588	NaN
8	NaN	39.6525	35.8260	37.0425	41.1780	37.7700	NaN
9	NaN	44.0385	43.5570	43.8135	NaN	NaN	NaN
10	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16	NaN	NaN	NaN	NaN	NaN	NaN	NaN
17	NaN	NaN	NaN	NaN	NaN	NaN	NaN
18	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19	NaN	NaN	NaN	NaN	NaN	NaN	NaN
20	NaN	NaN	NaN	NaN	NaN	NaN	NaN
21	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22	NaN	NaN	NaN	44.9430	NaN	NaN	NaN
23	44.3955	44.3850	44.3820	44.5935	43.5480	42.9888	NaN

Fonte: Autoria Própria

### 4.3 Curva Financeira

A curva financeira é feita de acordo com a tabela de tarifas já mostrada na seção 3.4. Para uma curva mais fiel, são solicitados alguns dados ao usuário, sendo estes a concessionária à qual a unidade pertence, a constante de medidor, que pode ser encontrada no faturamento de energia elétrica, e o tipo de tarifa. Por fim é retornada a soma total, como visto na Figura 42.

Figura 42: Dados de Entrada

```
Qual a concessionária: EPB, EBO
EPB
Qual a constante do medidor:
1
Qual a tarifa: 1- AZUL AH4, 2- AZUL AH3, 3- VERDE
VERDE
```

Fonte: Autoria Própria

Com a seleção do tipo de tarifa, é então realizado o cálculo financeiro dos valores de consumo medidos. Eles são feitos através da separação das medições em relação ao horário, se são fora ponta ou em ponta, além de realizar a seleção dos valores excedentes com base no limite definido para UC. Os valores de cada horário serão multiplicados pela constante do medidor, inserido manualmente, e pela tarifa referente a modalidade

contratada.

Após encontrados, os valores foram divididos em quatro colunas como apresentado na Figura 43, a primeira relativo ao valor do horário ponta, o segundo ao valor do horário fora ponta, o terceiro ao valor de ultrapassagem em horário ponta e por último ultrapassagem em horário fora ponta.

Figura 43: DataFrame Financeiro



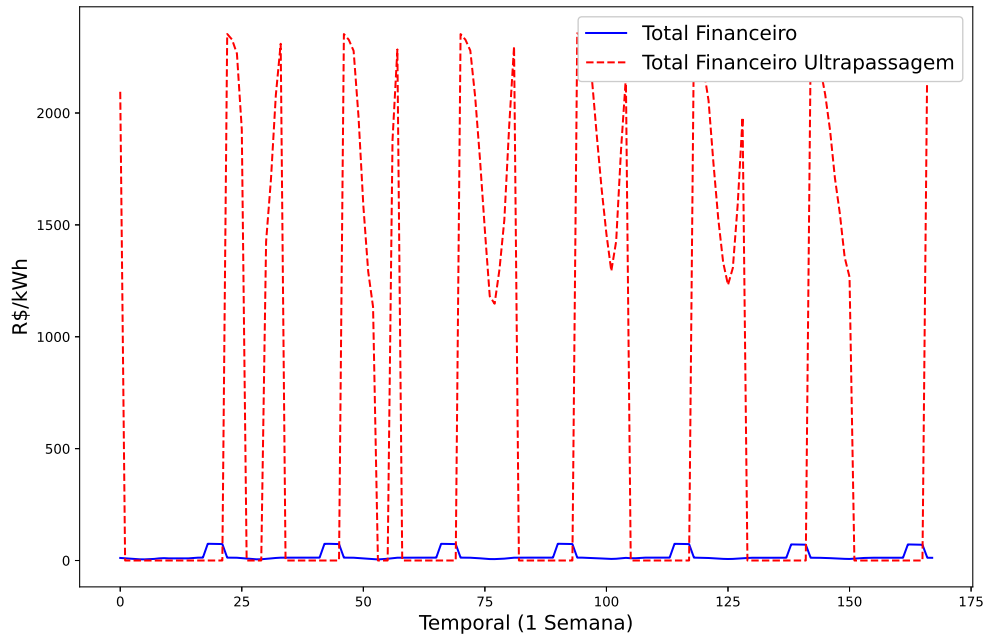
Financeiro FP	Financeiro P	Financeiro UP	Financeiro UFP
11.467042	0.000000	0.0	2094.79446
9.833814	0.000000	0.0	0.00000
7.812194	0.000000	0.0	0.00000
6.402443	0.000000	0.0	0.00000
5.356970	0.000000	0.0	0.00000
5.113687	0.000000	0.0	0.00000
6.006300	0.000000	0.0	0.00000
7.145210	0.000000	0.0	0.00000
9.561682	0.000000	0.0	0.00000
10.584763	0.000000	0.0	0.00000
9.534124	0.000000	0.0	0.00000
9.600435	0.000000	0.0	0.00000
9.590962	0.000000	0.0	0.00000
9.578044	0.000000	0.0	0.00000
9.552639	0.000000	0.0	0.00000
11.209980	0.000000	0.0	0.00000
12.767424	0.000000	0.0	0.00000
12.930187	0.000000	0.0	0.00000
0.000000	74.801392	0.0	0.00000
0.000000	74.494429	0.0	0.00000
0.000000	73.868323	0.0	0.00000
0.000000	73.539435	0.0	0.00000
12.881100	0.000000	0.0	2353.11390
12.744172	0.000000	0.0	2328.10002

Fonte: Autoria Própria

Por fim, encontra-se a curva financeira para o consumo e para os valores de ultrapassagem de acordo com a Figura 44. A curva contínua em azul representa o somatório dos valores de consumo ponta e fora ponta normais, e a curva vermelha tracejada refere-se aos valores ultrapassados em ponta e fora ponta.

Como ocorreram mais ultrapassagens em horários parecidos ao longo do mês, a curva tracejada possui picos altos devido ao valor da taxa de ultrapassagem e periódicos. A curva em contínua do valor total, possui uma curva mais suave para os horário fora ponta, e em horários de ponta alguns picos.

Figura 44: Curva Financeira



Fonte: Autoria Própria

## 4.4 Clusterização K-Means

Nesta seção iremos mostrar os resultados do agrupamento da clusterização com base no modelo K-Means, considerando dois cenários. Cada cenário foi definido com quantidades diferentes de meses analisados, sendo o primeiro apenas de um mês e o segundo considerando os todos os dados disponíveis.

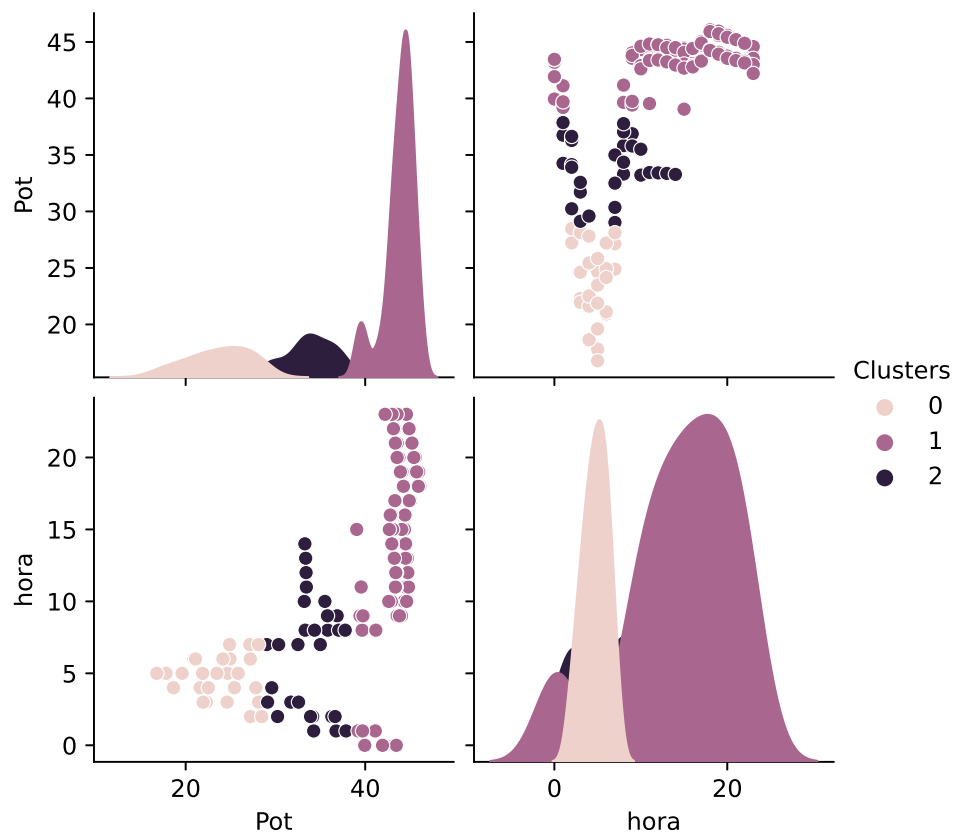
### 4.4.1 Cenário 1: Único Mês

Para o primeiro cenário, foi necessário a definição do K ideal para ser utilizado. Como dito na seção 3.5, o K ideal seria 3 ou 4, então foi feita uma análise para cada um dos K.

O agrupamento levou em conta os níveis de potência, se o dado foi de uma medição muito mais alta que o normal ou se foi de medições mais baixas que o normal.

Na Figura 45 pode-se observar o processo de clusterização para K=3. Podemos verificar que a separação dos valores de potência em relação ao horário ocorreu de forma correta, porém, valores excedentes foram agrupados com valores normais em maior frequência. De forma geral esta classificação irá permitir verificar os tipos de níveis das medições armazenadas.

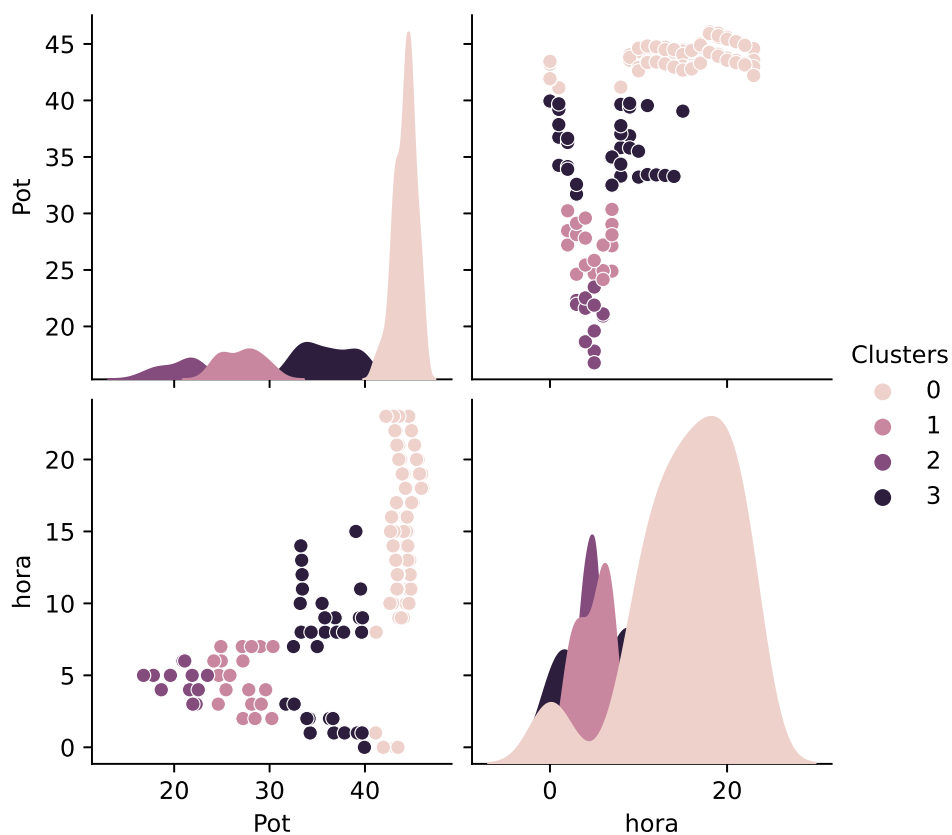
Figura 45: Resultado Clusterização Mensal K=3



Fonte: Autoria Própria

Na Figura 46, nota-se que para  $K=4$  ocorreu uma melhor separação, ou seja, os valores de potência excedentes foram mais bem definidos em relação aos clusters, entre dois dos quatro agrupamentos sendo estes os agrupamentos 0 e 2, da mesma forma que as potências abaixo do limite forma separados nos outros dois restantes, nos agrupamentos 1 e 3.

Figura 46: Resultado Clusterização Mensal K=4



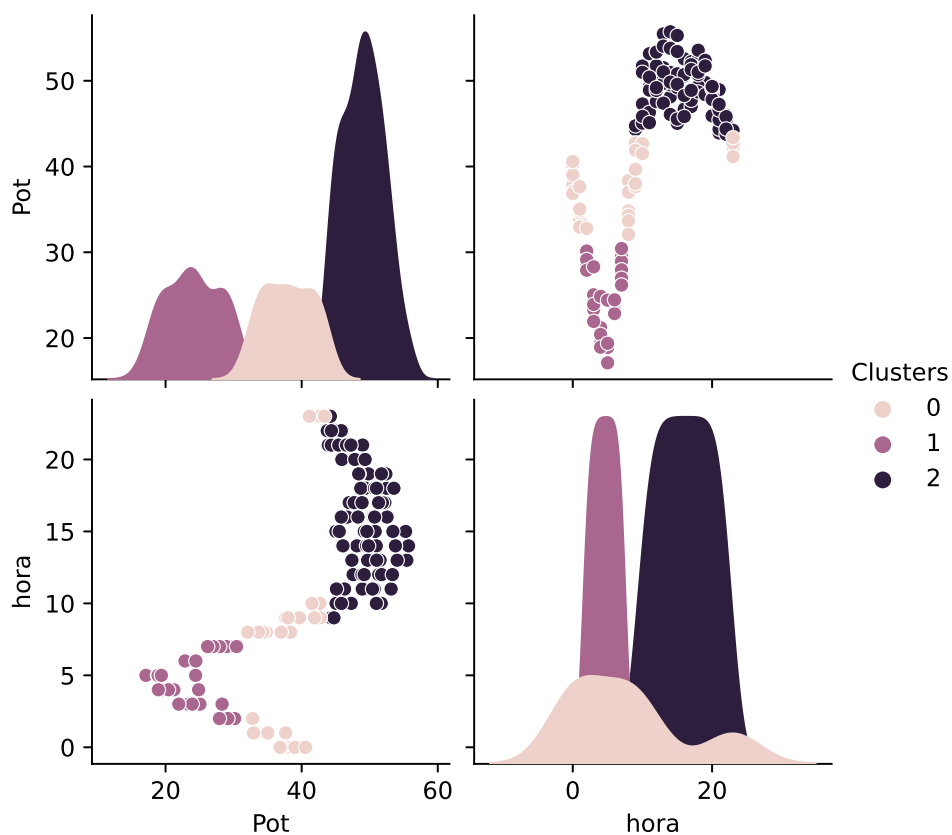
Fonte: Autoria Própria

#### 4.4.2 Cenário 2: Meses Completos

Neste cenário foram considerados todos os meses armazenados dentro do banco de dados do trabalho. A partir da técnica do cotovelo podemos determinar que o número de *clusters* permanece igual ao primeiro caso, com  $K=3$  ou  $K=4$ .

Para  $K=3$  podemos observar uma semelhança ao mesmo número de clusters do cenário 1, os dados de potência ultrapassada e normais não foram bem definidos entre as classes. Na Figura 47 pode-se observar isto.

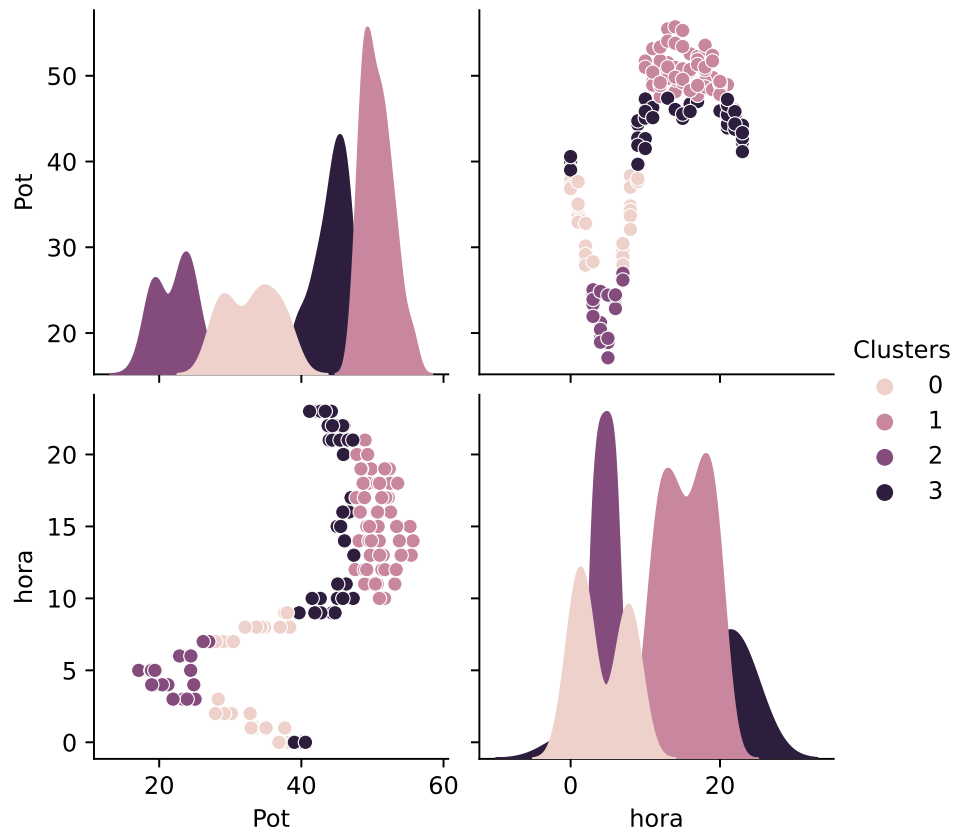
Figura 47: Resultado Clusterização Anual K=3



Fonte: Autoria Própria

Na Figura 48, temos  $K=4$ , pode-se observar uma maior uniformidade, ou seja, o agrupamento ocorreu de uma forma correta, o que se compara ao  $K=4$  do cenário 1.

Figura 48: Resultado Clusterização Anual K=4



Fonte: Autoria Própria

## 4.5 Modelo de Previsão LSTM

Como dito no capítulo anterior, foram feitos três modelos LSTM e variando as quantidades de camadas ocultas.

Para o treinamento foram usados os dados de três meses e para o teste apenas os dados do mês de junho, como pode ser visto na Figura 49. Os valores aplicados passaram por um processo de normalização, para que não ocorresse a aparição de dados negativos ou discrepantes, ou seja, eles foram padronizados em um intervalo de zero a um.

Figura 49: Dados de Treinamento e Teste

```
data_train2 = M_RESAMPLE.loc['2021-01':'2021-05',uc].to_numpy()
data_test2 = M_RESAMPLE.loc['2021-06',uc].to_numpy()

data_train2
array([33.696, 44.784, 44.61 , ..., 65.718, 48.312, 40.032])

data_test2
array([38.112, 33.882, 30.354, 26.226, 22.14 , 21.426, 24.24 , 30.252,
       49.638, 66.816, 66.774, 66.564, 66.528, 66.258, 65.544, 64.824,
       64.602, 65.778, 66.798, 66.222, 59.436, 40.962, 40.302, 35.646,
       30.534, 53.076, 64.302, 64.212, 64.14 , 64.146, 48.84 , 40.836,
```

Fonte: Autoria Própria



Logo após, os valores são reorganizados em matrizes tridimensionais para que os modelos propostos possam realizar o treinamento de modo correto. Neste caso modelamos os dados de treinamento de forma que a matriz de saída possua um histórico de 24 amostras, para previsão das amostras seguintes.

#### 4.5.1 Modelo 1 - Duas Camadas Ocultas

O primeiro modelo desenvolvido possui uma camada visível, com apenas uma entrada, duas camadas ocultas, cada camada oculta possuindo cinquenta neurônios e com a primeira camada tendo realimentação dos dados, para então gerar na camada de saída a previsão do valor de potência futura. O modelo pode ser expresso de acordo com o sumário da Figura 50.

Figura 50: Summary Model 1

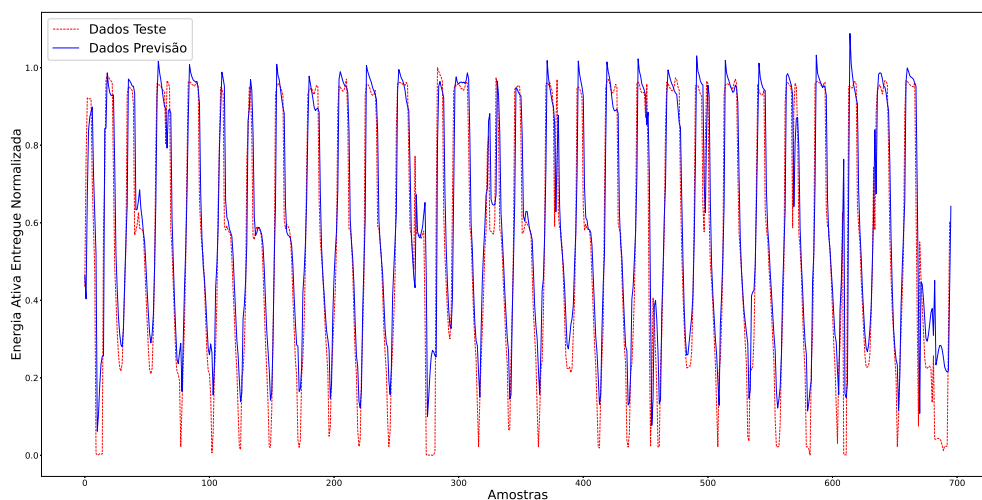
Model: "sequential_1"		
Layer (type)	Output Shape	Param #
=====	=====	=====
lstm_3 (LSTM)	(None, 24, 50)	10400
lstm_4 (LSTM)	(None, 50)	20200
dense_1 (Dense)	(None, 1)	51
=====	=====	=====
Total params: 30,651		
Trainable params: 30,651		
Non-trainable params: 0		

Fonte: Autoria Própria

Na Figura 51, observa-se os valores previstos em comparação aos valores originais, curva tracejada, com a métrica MAE e com seus dados ainda normalizados.

A curva contínua, com os dados previstos, acompanha de forma aproximada os dados de teste, curva tracejada, com variação de alguns picos, acima ou abaixo e um erro percentual médio (MAPE) de 10,99%.

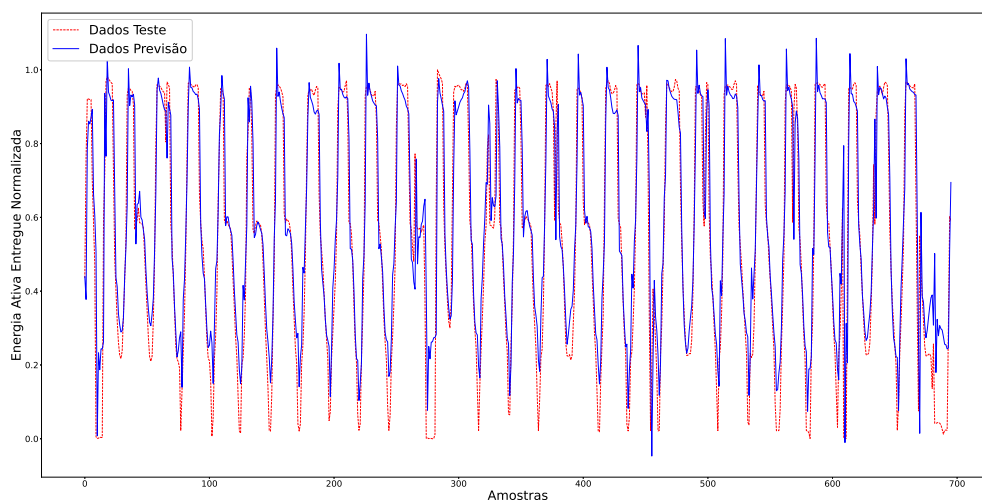
Figura 51: Model 1 - Métrica MAE



Fonte: Autoria Própria

A Figura 52 segue o mesmo padrão da Figura 51, mudando apenas a métrica utilizada, que neste caso é a MSE. Neste caso, a curva de previsão apresentou um erro percentual de 8,65%. Apesar de ter um erro menor que a métrica MAE para este modelo, a curva de previsão possui uma maior quantidade de picos em relação aos dados originais de teste.

Figura 52: Model 1 - Métrica MSE



Fonte: Autoria Própria

#### 4.5.2 Modelo 2 - Três Camadas Ocultas

O segundo modelo desenvolvido possui uma camada visível, com apenas uma entrada, três camadas ocultas, cada camada oculta possuindo cinquenta neurônios, agora com realimentação nos neurônios da primeira e segunda camadas, para então chegar na camada de saída a previsão do valor de potência. O modelo está expresso de acordo com de acordo com o sumário da Figura 53.

Figura 53: Summary Model 2

```
Model: "sequential_2"
```

Layer (type)	Output Shape	Param #
lstm_5 (LSTM)	(None, 24, 50)	10400
lstm_6 (LSTM)	(None, 24, 50)	20200
lstm_7 (LSTM)	(None, 50)	20200
dense_2 (Dense)	(None, 1)	51

```

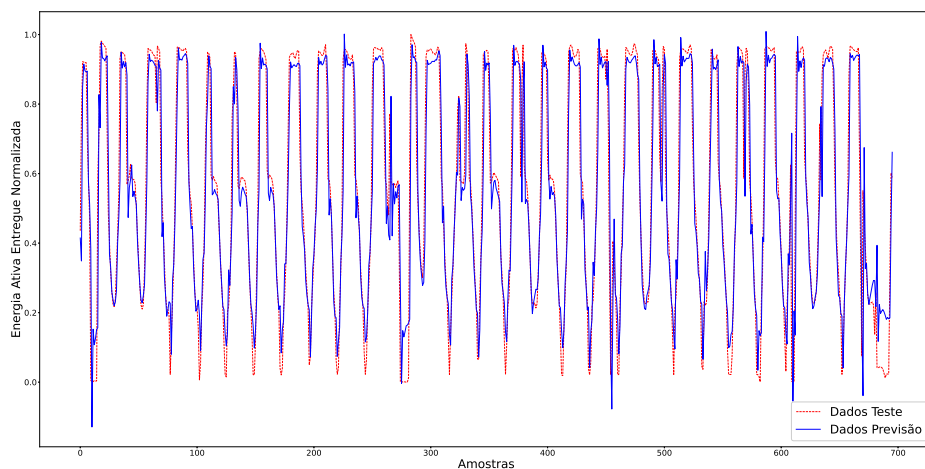
=====
Total params: 50,851
Trainable params: 50,851
Non-trainable params: 0

```

Fonte: Autoria Própria

Na Figura 54, encontra-se os valores normalizados da curva de previsão (contínua e em azul), em comparação a curva original (tracejada e em vermelho), com a métrica MAE. Para esta simulação o MAPE, erro percentual médio, encontrado chegou a um valor de 7,02%. Neste caso a previsão ocorreu de fomar mais fiel aos dados originais.

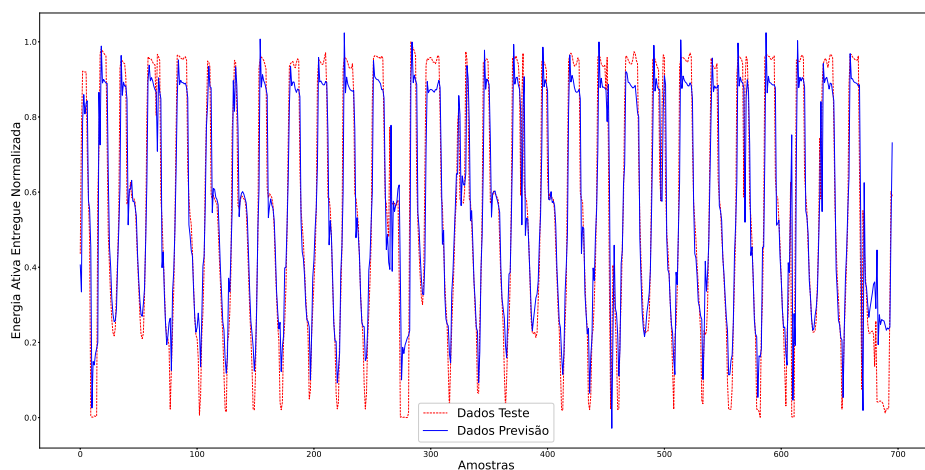
Figura 54: Model 2 - Métrica MAE



Fonte: Autoria Própria

A Figura 55 segue o mesmo padrão da Figura 54, mudando apenas a métrica utilizada, que neste caso foi MSE. A curva de previsão neste caso apresentou uma compatibilidade parecida com a simulação anterior da métrica MAE e com os dados originais, apresentando um erro médio percentual de 8,24%.

Figura 55: Model 2 - Métrica MSE



Fonte: Autoria Própria

#### 4.5.3 Modelo 3 - Quatro Camadas Ocultas

O terceiro modelo desenvolvido segue em mesmo princípio os outros dois modelos anteriores, possuindo uma camada visível de uma entrada, agora quatro camadas ocultas,

cada uma possuindo cinquenta neurônios, e agora com realimentação nos neurônios nas três primeiras camadas, para então chegar na camada de saída a previsão do valor de potência, de acordo com de acordo com o sumário da Figura 56.

Figura 56: Summary Model 3

Model: "sequential\_3"

Layer (type)	Output Shape	Param #
lstm_8 (LSTM)	(None, 24, 50)	10400
lstm_9 (LSTM)	(None, 24, 50)	20200
lstm_10 (LSTM)	(None, 24, 50)	20200
lstm_11 (LSTM)	(None, 50)	20200
dense_3 (Dense)	(None, 1)	51

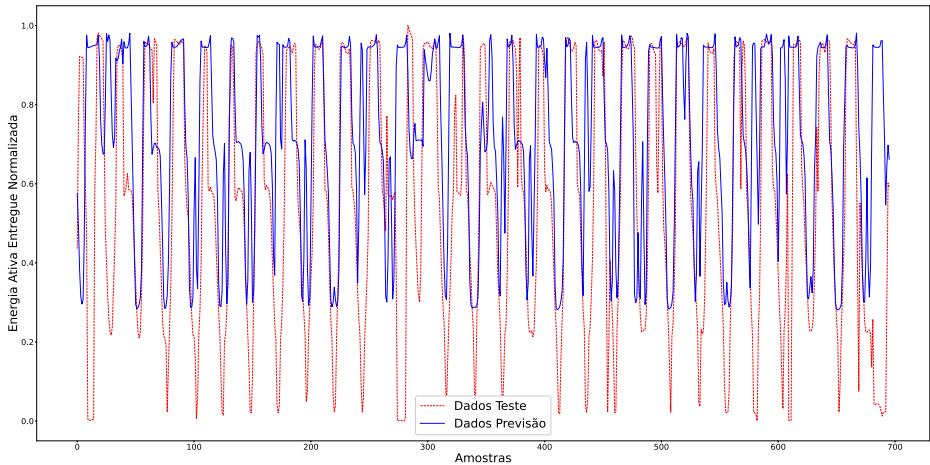
=====

Total params: 71,051  
Trainable params: 71,051  
Non-trainable params: 0

Fonte: Autoria Própria

As curvas apresentadas na Figura 57 pertencem a comparação entre a previsão (curva contínua) com os dados teste (curva traceja), para a métrica MAE. Observa-se que a previsão não acompanha os valores de teste, tendo falhas mais graves e um erro percentual de 25,60%.

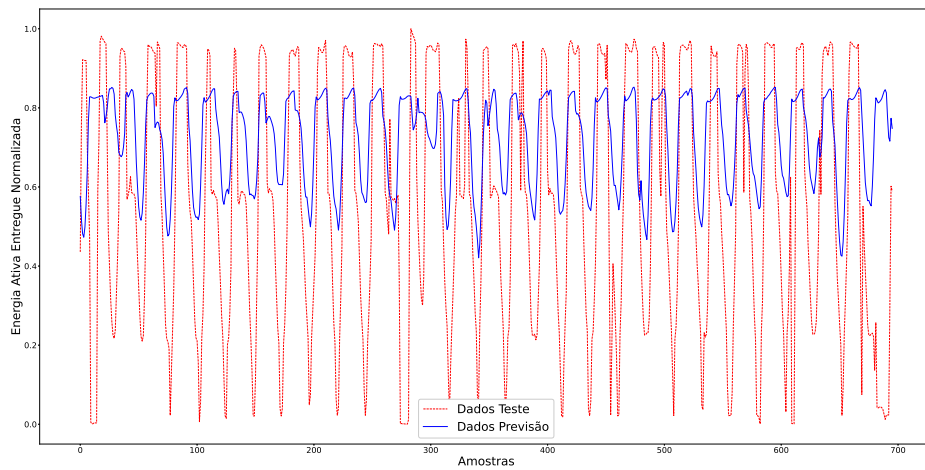
Figura 57: Model 3 - Métrica MAE



Fonte: Autoria Própria

A Figura 58 segue o mesmo padrão da Figura 57, mudando apenas a métrica utilizada, que neste caso é a MSE. Esta previsão apesar de ter uma compatibilidade maior com os dados que a simulação para a métrica MAE, apresenta ainda uma elevada porcentagem de erro médio, 14,60%. Logo, ocorre uma discrepância entre os valores originais e os previstos.

Figura 58: Model 3 - Métrica MSE



Fonte: Autoria Própria

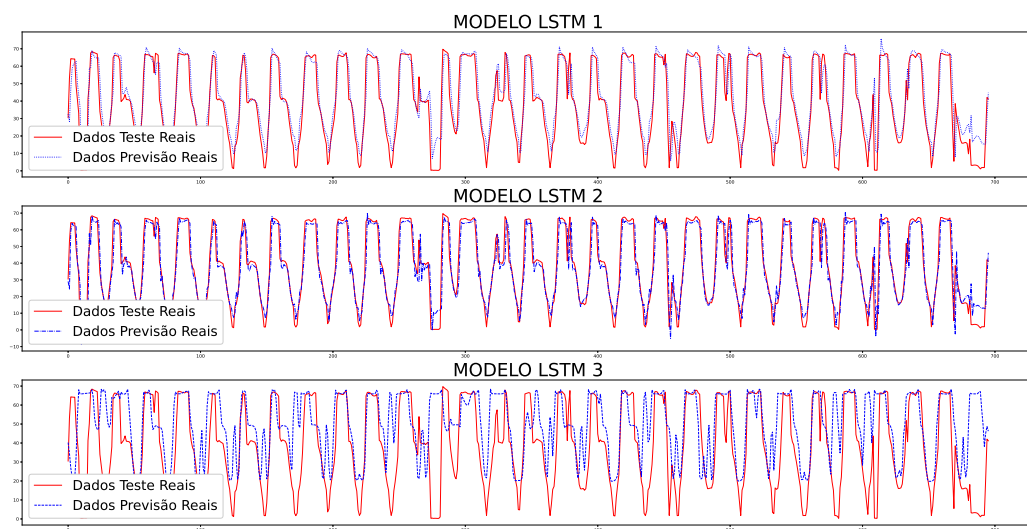
#### 4.5.4 Comparação Entre modelos

Por fim, com a conclusão das simulações dos modelos propostos, podemos então realizar as comparações.

Levando em consideração as métricas, a MSE apresentou curvas com características constantes. Porém, a MAE apresentou um modelo com a menor porcentagem de erro entre a previsão e a original, além de possui maior adaptabilidade para entrada de novos dados.

Na Figura 59, temos a comparação lado a lado entre os três modelos para a primeira métrica. É possível observar que a curva relativa ao modelo 2 obteve valores mais aproximados aos originais, seguido pela curva do modelo 1 e por fim a curva do modelo 3, que apresentou o maior valor de erro entre todas as simulações realizadas neste trabalho.

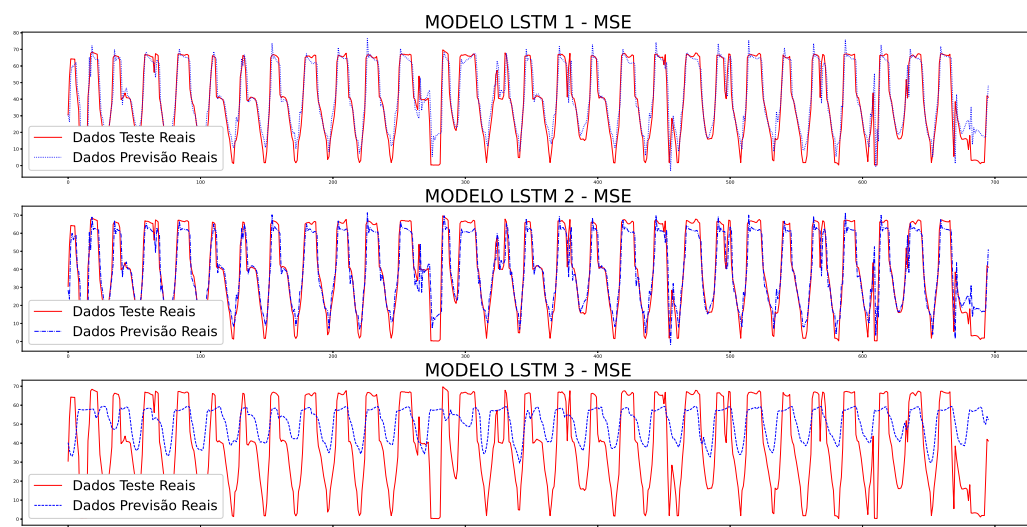
Figura 59: Comparação Modelos Métrica MAE



Fonte: Autoria Própria

Para o caso da segunda métrica, a Figura 60 apresenta a comparação entre os três modelos. Neste caso a curva do modelo 2, com três camadas ocultas, exibiu um melhor resultado de previsão. Observando as curvas, percebe-se que o modelo 1 e 2 se aproximam muito do resultado desejado, já a curva o modelo 3 apresenta o segundo pior resultado entre todas as simulações.

Figura 60: Comparação Modelos Métrica MSE



Fonte: Autoria Própria

Assim, todos os modelos desenvolvidos mostraram-se capazes de se adequar aos dados originais e então realizarem as tarefas propostas, de previsão do consumo elétrico para meses futuros. Evidenciando os dois segundos modelos como os mais confiáveis no uso das métricas MAE e MSE, respectivamente.



## 5 CONCLUSÃO

Este trabalho permitiu apresentar diferentes métodos para previsão de consumo, além de mostrar uma técnica simples e fácil de analisar a ultrapassagem do consumo. O objetivo dele ficou concentrado na elaboração de um modelo de previsão fácil de se entender e de se manipular, para utilização do pessoal que trabalha na empresa CAGEPA.

Foram construídos modelos de agrupamentos não supervisionados baseados em K-Means, e modelos de previsão supervisionados baseados em Redes Neurais Recorrentes do tipo LSTM. Os modelos foram postos a prova em um conjunto de dados selecionado e tratados, para funcionamento correto.

Para validação dos modelos de *deep learning LSTM* foram usados as métricas MAE e MSE.

A escolha do modelo de previsão se deu de forma mais complexa, visto que ocorre a necessidade de que este possuísse uma dependência temporal entre os dados. Como o modelo simples de RNN possui algumas problemas com a perda de função de aprendizado, foi então escolhido o modelo LSTM.

Os modelos de LSTM foram comparados entre si e com base na curva de valores originais, sendo o segundo modelo, com três camadas ocultas, se mostrando superior aos outros dois modelos que possuem respectivamente duas e quatro camadas ocultas, estes resultados se deram a partir do pré processamento realizado, junto como a escolha dos parâmetros internos destes.

Para um trabalho futuro sugere-se:

- Análise da curva financeira com base na variação do tempo;
- Otimização dos parâmetros;
- Desenvolvimento de novos modelos;

Por fim, podemos afirmar que o trabalho desenvolvido contribuiu de forma direta para a evolução do conhecimento assimilado durante o curso de engenharia elétrica do discente, principalmente na área de sistemas inteligentes, programação e gestão energética.

## 6 ANEXO I - Dados Aplicados

Figura 61: Dados Mês de Janeiro

```

1 - Selecionar mes de analise ou 2 - analise meses completos
1
Digite o mes a ser analisado:
Jan
      21107  9981353  9980091  9980619  1366670  9980543  1798224  9998194  \
0      7.698    0.000    11.250    0.0357    11.98      6.87    0.007    0.450
1      1.578    0.000    11.220    0.0039    11.94      7.85    0.006    0.448
2      7.194    0.000    11.226    0.0039    11.60     11.43    0.006    0.448
3      3.066    0.000    11.214    0.0036    11.62     11.42    0.007    0.449
4      4.308    0.000    11.190    0.0039    11.64     11.50    0.007    0.450
...      ...      ...      ...      ...      ...      ...      ...
2970    7.548    0.119    9.204    0.0048    11.75      8.54    0.007    0.440
2971    5.886    0.120    9.168    0.0048    11.78      8.56    0.007    0.438
2972    9.102    0.119    9.180    0.0045    11.76      8.55    0.007    0.438
2973    3.786    0.119    9.138    0.0048    11.76      8.58    0.008    0.438
2974    8.952    0.071    9.060    0.0048    11.76      8.56    0.007    0.427

      9980259  9980529  ...  1840142  9980574  183806  1039460  9980245  \
0      0.438    5.538  ...    0.398    0.036  1.60000  0.004  0.1050
1      0.438    6.984  ...    0.397    0.036  1.61250  0.004  0.1028
2      0.434   12.414  ...    0.398    0.042  1.60000  0.004  0.1026
3      0.435   12.426  ...    0.398    0.036  1.60000  0.000  0.1026
4      0.437   12.444  ...    0.397    0.036  1.71250  0.004  0.1026
...      ...      ...  ...      ...      ...      ...      ...
2970    0.322    8.640  ...    0.390    0.036  1.57500  6.308  0.1098
2971    0.324    5.472  ...    0.389    0.036  1.81250  6.324  0.1096
2972    0.322    5.472  ...    0.390    0.036  1.86875  6.328  0.1098
2973    0.323    9.996  ...    0.389    0.036  1.83125  6.324  0.1100
2974    0.323   12.396  ...    0.390    0.030  1.80625  6.328  0.1100

```

Fonte: Autoria Própria

Figura 62: Dados Meses Completos

1 - Selecionar mes de analise ou 2 - analise meses completos								
2								
	21107	9981353	9980091	9980619	1366670	9980543	1798224	9998194 \
0	7.698	0.000	11.250	0.0357	11.98	6.87	0.007	0.450
1	1.578	0.000	11.220	0.0039	11.94	7.85	0.006	0.448
2	7.194	0.000	11.226	0.0039	11.60	11.43	0.006	0.448
3	3.066	0.000	11.214	0.0036	11.62	11.42	0.007	0.449
4	4.308	0.000	11.190	0.0039	11.64	11.50	0.007	0.450
...	...	...	...	...	...	...	...	...
11515	0.000	0.471	10.440	0.1932	11.55	8.57	0.010	0.458
11516	0.000	0.469	10.404	0.1932	11.55	8.55	0.010	0.457
11517	0.000	0.470	10.368	0.1881	11.56	8.57	0.009	0.459
11518	0.000	0.472	9.966	0.0048	11.60	8.59	0.010	0.460
11519	0.000	0.470	9.534	0.0039	11.60	8.56	0.009	0.458
	9980259	9980529	...	1840142	9980574	183806	1039460	9980245 \
0	0.438	5.538	...	0.398	0.036	1.60000	0.004	0.1050
1	0.438	6.984	...	0.397	0.036	1.61250	0.004	0.1028
2	0.434	12.414	...	0.398	0.042	1.60000	0.004	0.1026
3	0.435	12.426	...	0.398	0.036	1.60000	0.000	0.1026
4	0.437	12.444	...	0.397	0.036	1.71250	0.004	0.1026
...	...	...	...	...	...	...	...	...
11515	0.311	12.540	...	0.393	0.030	1.72500	6.828	0.1928
11516	0.311	12.510	...	0.392	0.030	1.69375	6.820	0.1928
11517	0.311	12.522	...	0.393	0.030	1.66250	6.828	0.1928
11518	0.313	9.678	...	0.392	0.030	1.64375	6.824	0.1930
11519	0.311	5.370	...	0.392	0.030	1.65625	6.828	0.1936

Fonte: Autoria Própria

## Referências

- ACADEMY, D. S. *A Arquitetura das Redes Neurais*. 2018. <<https://www.deeplearningbook.com.br/a-arquitetura-das-redes-neurais/>>. Acesso em: 31 de Maio de 2022. page.1414
- ACADEMY, D. S. *Arquitetura de Redes Neurais Long Short Term Memory (LSTM)*. 2018. <<https://www.deeplearningbook.com.br/arquitetura-de-redes-neurais-long-short-term-memory/>>. Acesso em: 31 de Maio de 2022. page.3434
- ACADEMY, D. S. *Redes Neurais Recorrentes*. 2018. <<https://www.deeplearningbook.com.br/redes-neurais-recorrentes/>>. Acesso em: 31 de Maio de 2022. page.1515
- ACADEMY, D. S. *Uma Breve História das Redes Neurais Artificiais*. 2018. <<https://www.deeplearningbook.com.br/uma-breve-historia-das-redes-neurais-artificiais/>>. Acesso em: 31 de Maio de 2022. page.1313
- ALMADA, M. Modelos de neurônios em redes neurais artificiais. page.1313
- BROWNLEE, J. *Long short-term memory networks with python: develop sequence prediction models with deep learning*. [S.l.]: Machine Learning Mastery, 2017. page.3333
- COELHO, M. *Redes De Hopfield*. 2017. <<http://www2.decom.ufop.br/imobilis/redes-de-hopfield/>>. Acesso em: 01 de Junho de 2022. page.1616
- EDUCACIONAL, P. *Análise de Séries Temporais*. 2020. <<https://proeducacional.com/ead/curso-cga-modulo-i/capitulos/capitulo-4/aulas/analise-de-series-temporais/>>. Acesso em: 01 de Junho de 2022. page.1111
- FLECK, L. et al. Redes neurais artificiais: Princípios básicos. *Revista Eletrônica Científica Inovação e Tecnologia*, v. 1, n. 13, p. 47–57, 2016. page.1313
- FOUNDATION, P. S. *A Biblioteca Padrão do Python*. 2019. <<https://docs.python.org/pt-br/3/library/>>. Acesso em: 2 de Junho de 2022. page.2121
- GOMIDE, F. A. *Redes neurais artificiais para engenharia e ciências aplicadas: curso prático*. [S.l.]: SciELO Brasil, 2012. page.1515
- GULLI, A.; PAL, S. *Deep learning with Keras*. [S.l.]: Packt Publishing Ltd, 2017. page.2323
- HYKIN, S. *Neural Networks and Learning Machines*. [S.l.]: Pearson Education, Inc., 2008. page.1212
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. [S.l.]: OTexts, 2018. page.1111
- KODINARIYA, T. M.; MAKWANA, P. R. Review on determining number of cluster in k-means clustering. *International Journal*, v. 1, n. 6, p. 90–95, 2013. page.3131
- LANDAU, S. et al. *Cluster analysis*. [S.l.]: John Wiley & Sons, 2011. page.1717

- LUCENA, W. *Agrupamento hierárquico*. 2019. <<https://medium.com/@will.lucena/agrupamento-hier%C3%A1rquico-329e30a9f32d>>. Acesso em: 01 de Junho de 2022. page.1818
- LUTZ, M. *Programming python*. [S.l.]: "O'Reilly Media, Inc.", 2001. page.2121
- MACQUEEN, J. Classification and analysis of multivariate observations. In: *5th Berkeley Symp. Math. Statist. Probability*. [S.l.: s.n.], 1967. p. 281–297. page.1919
- MATTEDE, H. *Consumo de energia – Como calcular em Kwh!* 2020. <<https://www.mundodaeletrica.com.br/consumo-de-energia-como-calcular-em-kwh/>>. Acesso em: 29 de Maio de 2022. page.1212
- MCKINNEY, W. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. [S.l.]: "O'Reilly Media, Inc.", 2012. page.2222
- MIGUEL, T. *K-Means Clustering (Agrupamento k-means)*. 2021. <<https://aprenderdatascience.com/k-means-clustering-agrupamento-k-means/#como-funciona-o-k-means>>. Acesso em: 31 de Maio de 2022. page.1919
- MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. *Introduction to time series analysis and forecasting*. [S.l.]: John Wiley & Sons, 2015. page.99
- OLIVEIRA, B. *CARACTERÍSTICAS DAS SÉRIES TEMPORAIS*. 2019. <<https://operdata.com.br/blog/caracteristicas-das-series-temporais/>>. Acesso em: 31 de Maio de 2022. page.1212
- OSÓRIO, F. S.; BITTENCOURT, J. R. Sistemas inteligentes baseados em redes neurais artificiais aplicados ao processamento de imagens. In: *I WORKSHOP DE INTELIGÊNCIA ARTIFICIAL UNISC–Universidade de Santa Cruz do Sul Departamento de Informática-Junho*. [S.l.: s.n.], 2000. page.1313
- REIS, B. *Redes Neurais – Redes Clássicas*. 2016. <<http://www2.decom.ufop.br/imobilis/redes-neurais-redes-classicas/>>. Acesso em: 01 de Junho de 2022. page.1616
- RIZZO, I. V.; CANATO, R. L. C. Inteligência artificial: funções de ativação. *Prospectus (ISSN: 2674-8576)*, v. 2, n. 2, 2020. page.3737
- SALEM, F. M. *Recurrent Neural Networks: From Simple to Gated Architectures*. [S.l.]: Springer Nature, 2022. page.3434
- SOARES, M.; COSTA, H. K. de M. O segmento de distribuição de energia elétrica no brasil: uma avaliação das crises hídricas enfrentadas em 2001 e 2021. *Conjecturas*, v. 22, n. 2, p. 307–321, 2022. page.88
- SPADINI, A. S. *Séries temporais e suas aplicações*. 2021. <<https://www.alura.com.br/artigos/series-temporais-e-suas-aplicacoes>>. Acesso em: 2 de Junho de 2022. page.1111
- SUMMERFIELD, M. *Programming in Python 3: a complete introduction to the Python language*. [S.l.]: Addison-Wesley Professional, 2010. page.2121
- TEMPORAL, J. *Como definir o número de clusters para o seu KMeans*. 2019. <<https://medium.com/pizzadedados/kmeans-e-metodo-do-cotovelo-94ded9fdf3a9>>. Acesso em: 2 de Junho de 2022. page.3131

TOLMASQUIM, M. T.; GUERREIRO, A.; GORINI, R. Matriz energética brasileira: uma prospectiva. *Novos estudos CEBRAP*, SciELO Brasil, n. 79, p. 47–69, 2007. page.88

WASKOM, M. *Introdução ao Seaborn*. 2020. <<https://seaborn.pydata.org/introduction.html>>. Acesso em: 2 de Junho de 2022. page.2222

XU, D.; TIAN, Y. A comprehensive survey of clustering algorithms. *Annals of Data Science*, Springer, v. 2, n. 2, p. 165–193, 2015. page.1717