

INSTITUTO FEDERAL DE EDUCAÇÃO CIÊNCIA E TECNOLOGIA DE MINAS
GERAIS – *CAMPUS* BAMBUÍ
BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO

Ana Caroline Coutinho Carvalho

**APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA PREDIÇÃO DO
CONSUMO DE ENERGIA ELÉTRICA EM AMBIENTE DOMÉSTICO**

BambuÍ - MG

2024

ANA CAROLINE COUTINHO CARVALHO

**APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA PREDIÇÃO DO
CONSUMO DE ENERGIA ELÉTRICA EM AMBIENTE DOMÉSTICO**

Trabalho de conclusão de curso apresentado ao
Curso Bacharelado em Engenharia da
Computação do Instituto Federal de Minas
Gerais - *Campus* Bambuí para obtenção do grau
de bacharel em Engenharia da Computação.
Orientador: Me. Felipe Lopes de Melo Faria
Coorientador: Me. Calebe Giaculi Júnior

Bambuí - MG
2024

Catálogo na Fonte Biblioteca IFMG - Campus Bambuí

C331a Carvalho, Ana Caroline Coutinho.

Aplicação de técnicas de mineração de dados para predição do consumo de energia elétrica em ambiente doméstico. / Ana Caroline Coutinho Carvalho. – 2024.

53 f.; il.: color.

Orientador: Me. Felipe Lopes de Melo Faria.

Trabalho de Conclusão de Curso (graduação) - Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais – Campus Bambuí, MG, Curso Bacharelado em Engenharia da Computação, 2024.

1. Mineração de dados. 2. Predição. 3. Energia elétrica doméstica. I. Faria, Felipe Lopes de Melo. II. Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais – Campus Bambuí, MG. III. Título.

CDD 519.50285



MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS GERAIS
Campus Bambuí

Diretoria de Ensino
Departamento de Engenharia e Computação
Faz. Varginha - Rodovia Bambuí/Medeiros - Km 05 - Caixa Postal 05 - CEP 38900-000 - Bambuí - MG
37 3431 4900 - www.ifmg.edu.br

ANA CAROLINE COUTINHO CARVALHO

**APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA PREDIÇÃO
DO CONSUMO DE ENERGIA ELÉTRICA EM AMBIENTE DOMÉSTICO**

Trabalho de conclusão de curso apresentado ao Curso Bacharelado em Engenharia de Computação do Instituto Federal de Minas Gerais - *Campus* Bambuí para obtenção do grau de Bacharel em Engenharia de Computação.

Aprovado em 10/04/2024

BAMBUÍ-MG
2024

Bambuí, 10 de abril de 2024.



Documento assinado eletronicamente por **Felipe Lopes de Melo Faria**, **Professor**, em 10/04/2024, às 16:45, conforme Decreto nº 10.543, de 13 de novembro de 2020.



Documento assinado eletronicamente por **Calebe Giaculi Junior**, **Professor**, em 10/04/2024, às 16:45, conforme Decreto nº 10.543, de 13 de novembro de 2020.



Documento assinado eletronicamente por **Carlos Renato Nolli**, **Professor**, em 10/04/2024, às 16:46, conforme Decreto nº 10.543, de 13 de novembro de 2020.



Documento assinado eletronicamente por **Álvaro Antonio Fonseca de Souza**, **Professor Substituto**, em 10/04/2024, às 16:46, conforme Decreto nº 10.543, de 13 de novembro de 2020.



A autenticidade do documento pode ser conferida no site <https://sei.ifmg.edu.br/consultadocs> informando o código verificador **1898630** e o código CRC **CC6B8113**.

Dedico este trabalho primeiramente a Deus e aos meus queridos pais, Cristiana e Ronei. Sou profundamente grata por todo o suporte incondicional e pelos valores que me transmitiram ao longo da minha jornada. Esta conquista é dedicada a vocês, pois são a luz que guia meus passos e me mantém motivada ao meu crescimento.

AGRADECIMENTOS

Agradeço a todos que estiveram presentes e me apoiaram durante a jornada da minha monografia. Aos meus pais, familiares e amigos, não existem palavras suficientes para descrever o quanto sou grata por todo o amor, encorajamento e suporte que me deram. Cada palavra de estímulo, abraços e momentos compartilhados contribuíram para o meu sucesso. Também desejo agradecer aos meus orientadores e colegas de estudo pelo seu apoio e orientação. Essa conquista é nossa, e dedico a vocês o meu agradecimento por fazerem parte da minha jornada.

“Nenhum país pode desabrochar de verdade se ele poda o potencial das suas mulheres e se priva das contribuições de metade da sua população”

Michelle Obama

RESUMO

A presente monografia tem como objetivo a aplicação de técnicas de Mineração de Dados, como RNA, KNN e *Random Forests*, para a predição e análise de resultados, a fim de possíveis tomadas de decisão, na área de energia elétrica no ambiente doméstico. A base de dados utilizada engloba informações como: consumo de energia elétrica em ambientes domésticos e dados meteorológicos, sendo que a residência, que é o objeto de estudo, está localizada na Bélgica. Os registros foram coletados e anotados por Candanedo, Feldheim e Deramaix (2017). Métodos matemáticos foram aplicados, como: análise de correlação para compreender as relações entre atributos/variáveis, e normalização para melhorar a visualização da base, padronizando a escala dos dados. A avaliação e validação dos resultados obtidos deram-se por meio de métricas como, MAE, MAPE e RMSE. Destaca-se que os resultados deste estudo superaram os obtidos na pesquisa base, apresentando resultados de 22,83 para MAE, 22,14% para MAPE e 54,52 para RMSE. E a pesquisa base apresentou resultados de 31,36 para MAE, 29,76% para MAPE e 70,74 para RMSE. Para trabalhos futuros, diversas oportunidades destacam-se para aprimorar e expandir os resultados obtidos. Verificar se o WEKA trata *Overfitting*. Aplicação de novas técnicas de predição e novos métodos de validação. Outra possível abordagem: análise dos tipos de eletrodomésticos utilizados e os horários de maior demanda de energia, podendo analisar em períodos e intervalos diferentes, como em semanas ou em meses, com foco na identificação de economias na energia e fazer comparações nacionais, se todas as residências economizassem. Por fim, replicar o estudo no contexto brasileiro.

Palavras-chave: Mineração de Dados. Predição. Energia Elétrica Doméstica.

ABSTRACT

This monograph aims to apply Data Mining techniques, such as ANN, KNN and Random Forests, for the prediction and analysis of results, in order to make possible decisions, in the area of electrical energy in the domestic environment. The database used includes information such as: electrical energy consumption in domestic environments and meteorological data, and the residence, which is the object of study, is located in Belgium. The records were collected and annotated by Candanedo, Feldheim and Deramaix (2017). Mathematical methods were applied, such as: correlation analysis to understand the relationships between attributes/variables, and normalization to improve the visualization of the database, standardizing the data scale. The evaluation and validation of the results obtained were carried out using metrics such as MAE, MAPE and RMSE. It is noteworthy that the results of this study surpassed those obtained in the base research, presenting results of 22.83 for MAE, 22.14% for MAPE and 54.52 for RMSE. And the base survey presented results of 31.36 for MAE, 29.76% for MAPE and 70.74 for RMSE. For future work, several opportunities stand out to improve and expand the results obtained. Check if WEKA treats Overfitting. Application of new prediction techniques and new validation methods. Another possible approach: analysis of the types of household appliances used and the times of greatest energy demand, which can be analyzed at different periods and intervals, such as weeks or months, with a focus on identifying energy savings and making national comparisons, if all households saved. Finally, replicate the study in the Brazilian context.

Keywords: *Data Mining. Prediction. Household Electricity.*

LISTA DE FIGURAS

Figura 1- Processo de descoberta de conhecimento em bancos de dados (KDD).....	26
Figura 2 - Planta baixa do primeiro andar.....	29
Figura 3 - Planta baixa do segundo andar.....	30
Figura 4 – Correlação em relação ao <i>Appliances</i>	32
Figura 5 – Correlação em relação ao primeiro cômodo da casa.....	34
Figura 6 - Correlação em relação ao <i>Appliances</i>	35
Figura 7 - Correlação em relação ao <i>Lights</i>	36
Figura 8 - Correlação de todos os atributos através da média de temperatura e umidade.....	36
Figura 9 - Correlação em relação ao <i>Appliances</i> com base normalizada.....	37
Figura 10 - Correlação em relação ao primeiro cômodo da casa com base normalizada.....	38
Figura 11 - Correlação em relação ao <i>Appliances</i> com base normalizada.....	39
Figura 12 - Correlação em relação ao <i>Lights</i> com base normalizada.....	40
Figura 13 - Correlação de todos os atributos através da média de temperatura e umidade com base normalizada.....	41
Figura 14 - Gráfico dos resultados das técnicas.....	43

LISTA DE QUADROS

Quadro 1 - Classificação MAE, MAPE e RMSE.....	21
Quadro 2 - Eletrodomésticos da casa separados por cômodo.....	31

LISTA DE TABELAS

Tabela 1 - <i>Product Backlog</i>	28
Tabela 2 – <i>Sprint 1</i>	28
Tabela 3 - Resultados das técnicas na base original.....	42
Tabela 4 - Tabela de resultados da predição do trabalho base.....	43
Tabela 5 - Resultados das técnicas contendo somente dados noturnos.....	44
Tabela 6 - Resultados das técnicas contendo somente dados diurnos.....	44
Tabela 7 – <i>Sprint 2</i>	51
Tabela 8 – <i>Sprint 3</i>	51
Tabela 9 – <i>Sprint 4</i>	51
Tabela 10 – <i>Sprint 5</i>	51

LISTA DE ABREVIATURAS E SIGLAS

CV – Coeficiente de Variação

EPE – Empresa de Pesquisa Energética

GBM – *Gradient Boosting Machines*

GWh – *Gigawatt-hora*

IEA – *International Energy Agency*

KDD – Descoberta de Conhecimento em Base de Dados

KNN - K – *Nearest Neighbors*

Kwh - *Quilowatt-hora*

LM – Regressão Linear Múltipla

MAE - Erro Médio Absoluto

MAPE - Erro Percentual Absoluto Médio

MBE – Erro Médio de Viés

MLP – Perceptron Multi-Camadas

PIB – Produto Interno Bruto

R^2 – Coeficiente de Determinação

RF - *Random Forests*

RMSE - Raiz do Erro Quadrático Médio

RNA – Redes Neurais Artificiais

SVM – *Support Vector Machine*

SVR – *Support Vector Regression*

Wh – *Watt-hora*

WEKA – *Waikato Environment for Knowledge Analysis*

SUMÁRIO

1. INTRODUÇÃO.....	14
1.1 Objetivos.....	14
<i>1.1.1 Objetivo Geral.....</i>	<i>14</i>
<i>1.1.2 Objetivos Específicos.....</i>	<i>15</i>
<i>1.1.3 Resultados Esperados.....</i>	<i>15</i>
1.2 Justificativa.....	15
2 REFERENCIAL TEÓRICO.....	17
2.1 Fundamentação teórica.....	17
<i>2.1.1 Técnicas de predição.....</i>	<i>18</i>
<i>2.1.2 WEKA.....</i>	<i>18</i>
<i>2.1.3 Correlação.....</i>	<i>19</i>
<i>2.1.4 Normalização.....</i>	<i>19</i>
<i>2.1.5 Classificação.....</i>	<i>20</i>
2.2 Estado-da-arte.....	21
3 METODOLOGIA.....	25
3.1 Classificação do trabalho.....	25
3.2 Solução.....	25
3.3 Materiais e Tecnologias.....	27
<i>3.3.1 Materiais.....</i>	<i>27</i>
<i>3.3.2 Processo de Desenvolvimento.....</i>	<i>27</i>
4 RESULTADOS E DISCUSSÕES.....	29
4.1 Caracterização da casa.....	29
4.2 Caracterização da Base.....	31
<i>4.2.1 Correlação dos atributos.....</i>	<i>32</i>
<i>4.2.2 Correlação da média dos atributos.....</i>	<i>34</i>
<i>4.2.3 Correlação dos atributos com base normalizada.....</i>	<i>37</i>
<i>4.2.3.1 Correlação média atributos com base normalizada.....</i>	<i>39</i>
4.3 Aplicação das técnicas.....	41
<i>4.3.1 Divisão da base para aplicação das técnicas.....</i>	<i>43</i>
5 CONCLUSÃO.....	46
REFERÊNCIAS BIBLIOGRÁFICAS.....	47
APÊNDICE A – RESTANTE DAS SPRINTS DA MONOGRAFIA.....	51
APÊNDICE B – CARACTERIZAÇÃO DA BASE.....	52

1. INTRODUÇÃO

De acordo com IEA – *International Energy Agency* (2020), o consumo de energia elétrica doméstica é um tema de grande importância, pois está diretamente relacionado à preservação dos recursos naturais, mitigação das mudanças climáticas, entre outros. Um outro fator de grande peso é que a economia de energia pode impactar também na economia da conta de energia. Assim, é importante que os consumidores de energia doméstica adotem medidas para reduzir o consumo de energia elétrica.

O presente trabalho propõe a previsão de energia elétrica, utilizando-se de técnicas de mineração de dados, através de uma base de dados, cujo nome é: *Appliances Energy Prediction Data Set* (Base de dados de previsão de energia de eletrodomésticos). Essa base de dados encontra-se disponível na UCI – *Machine Learning Repository*¹. O pré-processamento foi realizado com a finalidade de fazer as transformações necessárias e em seguida aplicou-se as seguintes técnicas de previsão de Mineração de Dados: Redes Neurais Artificiais (RNA), K – *Nearest Neighbors* (KNN) e *Random Forests* (RF). Os resultados da pesquisa através dessas técnicas de previsão foram obtidos pelo *software* WEKA. Com os resultados, averiguou-se qual técnica se destacou através de suas métricas: Erro Médio Absoluto (MAE), Erro Percentual Absoluto Médio (MAPE) e Raiz do Erro Quadrático Médio (RMSE), para possíveis tomada de decisões, no contexto de gasto de energia elétrica em residências.

1.1 Objetivos

A fim de garantir a realização efetiva desta monografia, é essencial estabelecer objetivos para definir e direcionar o seu escopo. Esses objetivos podem ser encontrados nas seções: 1.1.1, Objetivo Geral; 1.1.2, Objetivos Específicos; e na 1.1.3, Resultados Esperados.

1.1.1 Objetivo Geral

Realizar a aplicação de técnicas de Mineração de Dados em uma base de dados de previsão do consumo de energia de eletrodomésticos e avaliar os modelos gerados.

1.1.2 *Objetivos Específicos*

- Realizar o pré-processamento da base de dados;
- Aplicar técnicas de Mineração de Dados na base gerada e validar os modelos;
- Realizar uma comparação dos resultados obtidos com os resultados do trabalho realizado por Candanedo, Feldheim e Deramaix (2017).

1.1.3 *Resultados Esperados*

Espera-se que as técnicas aplicadas à mineração de dados contribuam na geração de modelos preditivos no escopo de uma base de dados do consumo de energia elétrica de eletrodomésticos, podendo-se avançar no escopo de economia do consumo de energia de eletrodomésticos. Assim, almeja-se que o trabalho produzido possa contribuir com a tomada de decisão que impacte na economia do consumo de energia. Visto que, segundo a Empresa de Pesquisa Energética (2018), a economia de energia elétrica ajuda na preservação do meio ambiente.

1.2 *Justificativa*

De acordo com relatório da IEA – *International Energy Agency* (2020) ou Agência Internacional de Energia, o estudo do consumo de energia elétrica tem grande importância a partir do momento em que se sabe quais atitudes causam impacto na eficiência energética. Segundo a Empresa de Pesquisa Energética (2018), saber o que fazer para consumir menos energia é fundamental para manter os recursos naturais e contribuir para a melhoria das mudanças climáticas. Ainda conforme a empresa, a energia consumida em grande parte do mundo, atualmente, vem de energias não renováveis. Sendo o rendimento desse tipo de energia elevado, geram empregos e possuem bons preços. Porém, são prejudiciais ao meio ambiente e à saúde, devido à emissão de gases de efeito estufa.

A Mineração de Dados, de acordo com Tan, Steinbach e Kumar (2009), auxilia a sociedade na tomada de decisão em diversas áreas como: a de Negócios, Medicina, Ciências e Engenharia. Na de Negócios, a coleta de dados auxilia clientes a criarem perfis, direcionarem as vendas, detectarem anomalias e, conseqüentemente, aumentarem suas vendas e

lucratividade. Nas áreas da Medicina, Ciências e Engenharia, permite aos cientistas estudarem milhares de genes de uma vez, uma vez que isso não seria possível antes dessa tecnologia. Com a Mineração de Dados é possível isolar genes e identificar doenças. Já, no âmbito institucional, conforme Carvalho *et al.* (2022), o estudo da predição permite saber a geração de energia elétrica de uma usina fotovoltaica.

2 REFERENCIAL TEÓRICO

O referencial teórico do presente trabalho será subdividido em duas seções. A Seção 2.1 conceitua a área de computação em relação a Mineração de Dados, cuja área utilizada neste trabalho, junto à explicação do domínio deste estudo, é a energia elétrica no âmbito doméstico. Já a Seção 2.2 trata-se do Estado-da-arte.

2.1 Fundamentação teórica

De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), a Mineração de Dados originou-se da intersecção de três áreas distintas: banco de dados, estatística e inteligência artificial. A combinação dessas três áreas deu origem à Mineração de Dados, com o objetivo de extrair informações úteis e significativas, a partir de grandes conjuntos de dados, tornando-se uma área de pesquisa ativa nos anos 90, com o desenvolvimento de novas técnicas e algoritmos para análise de dados. De acordo com Tan, Steinbach e Kumar (2009), o processo de Mineração de Dados é composto por várias etapas, incluindo a seleção e preparação dos dados, aplicação de técnicas de Mineração de Dados, interpretação dos resultados e avaliação da qualidade das descobertas, processo nomeado como KDD - *knowledge discovery in databases* (Descoberta de Conhecimento em Bases de Dados). Ainda de acordo com os autores, existem várias tarefas de Mineração de Dados, como classificação, *clustering*, associação de regras, análise de séries temporais, predição, entre outras.

Segundo Singh e Yassine (2018), a previsão do consumo de energia elétrica doméstica é um desafio importante, uma vez que o consumo pode ser influenciado por fatores como o comportamento do usuário, a condição do clima e o uso de aparelhos elétricos. Sendo assim, a Mineração de Dados pode ser usada para analisar os padrões históricos de consumo de energia elétrica em domicílios e identificar padrões que possam ser usados para prever o consumo futuro. Essas previsões podem ajudar os consumidores a ajustarem seu consumo de energia (Wh) e reduzirem o uso desnecessário de energia elétrica.

Os consumidores podem optar por escolherem eletrodomésticos e eletrônicos que possuem o selo “A”, visto que, de acordo com o site Mundo Educação, esses selos variam de “A” a “E”, sendo que “A” são os que consomem menos energia e “E”, os que consomem mais energia.

2.1.1 Técnicas de predição

A Mineração de Dados compreende diversas tarefas, sendo uma delas a predição. Dentro da tarefa de predição, há uma variedade de técnicas disponíveis. De acordo com Tan, Steinbach e Kumar (2009), RNAs são um conjunto de algoritmos de aprendizado de máquina que se inspiram no funcionamento do cérebro humano para realizar tarefas de classificação, previsão e reconhecimento de padrões. “As redes neurais artificiais (RNAs) têm sido um método alternativo e eficiente para realizar a previsão de cargas elétricas na atualidade [...]” (ABREU *et al.*, 2017, p. 02).

A segunda técnica a ser citada é o KNN. Para Tan, Steinbach e Kumar (2009), o KNN é um algoritmo de aprendizado supervisionado, baseado em instâncias que classificam objetos com base na similaridade com outros objetos previamente rotulados. Ele não requer uma fase de treinamento explícita e, em vez disso, utiliza diretamente um conjunto de dados de treinamento existente. Quando um novo objeto é apresentado ao algoritmo, ele determina os k objetos mais próximos no conjunto de dados de treinamento e atribui a classe mais comum entre esses objetos ao novo objeto. O livro também discute algumas das principais questões e considerações envolvidas na implementação do KNN, como a escolha do valor de k , a medida de distância a ser usada e o tratamento de dados ausentes ou incompletos.

A terceira técnica a ser citada é o RF. Segundo Tan, Steinbach e Kumar (2009), a técnica RF, funciona criando várias Árvores de Decisão, a partir do conjunto de dados de treinamento, cada uma utilizando um subconjunto aleatório dos recursos (variáveis independentes ou atributos) e dos exemplos (instâncias) para o treinamento. Quando um novo exemplo é apresentado ao algoritmo, ele passa pelo conjunto de árvores de decisão e determina a classe mais comum entre todas as árvores.

2.1.2 WEKA

A ferramenta utilizada é o WEKA¹ ou *Waikato Environment for Knowledge Analysis*, um *software* livre de Mineração de Dados que oferece uma ampla gama de ferramentas e algoritmos para análise de dados. Ele foi desenvolvido na Universidade de Waikato, na Nova Zelândia, sendo utilizado para fins acadêmicos.

1 <https://www.cs.waikato.ac.nz/ml/weka/>

O software WEKA incorpora uma funcionalidade denominada *Forecast*, a qual, conforme consta na documentação do *Pentaho Data Mining*, é capaz de prever valores futuros em séries temporais, fundamentando-se em padrões históricos previamente identificados. Esse módulo não apenas oferece a representação gráfica dos resultados das previsões, como também permite a avaliação dos modelos produzidos.

2.1.3 Correlação

Um ponto importante da caracterização realizada na base de dados foi a análise da correlação dos atributos em relação ao atributo *Appliances* (consumo de energia elétrica dos eletrodomésticos) para a posterior configuração dos experimentos em relação a quais atributos utilizar.

De acordo com Sousa (2019), a correlação de *Pearson* é uma medida que varia entre -1 e +1, possuindo uma interpretação simplificada, pois não é afetada pelas unidades de medida dos atributos. Quando o resultado for zero, indica a ausência de uma relação linear entre as duas variáveis. Vale ressaltar que quanto mais próximo de 1 for o valor absoluto do coeficiente, mais forte é a relação linear entre elas. O sinal do coeficiente indica o sentido da relação: um valor positivo indica que as variáveis variam na mesma direção, sendo então diretamente proporcional, enquanto um valor negativo indica que variam em sentido inverso, ou seja, inversamente proporcional. Além disso, o resultado do coeficiente representa a intensidade dessa relação linear entre as variáveis.

Para fazer a interpretação da correlação de *Pearson*, de acordo com Cohen (1988), os valores menores que 0,29 significa correlação fraca, já os valores entre 0,30 e 0,49 significa correlação moderada e os valores entre 0,50 e 1,0 significa correlação forte.

2.1.4 Normalização

De acordo com Han, Kamber e Pei (2012), a normalização consiste em dar peso igual aos atributos que estão sendo utilizados, visto que utilizar variadas unidades de medidas pode atrapalhar a análise dos dados. A normalização faz com que os dados sejam transformados, que, no caso da presente monografia, foram transformados em intervalos de [0,1]. Para isso, a Equação 1 foi utilizada. Ainda de acordo com Han, Kamber e Pei (2012), a equação adaptada

do autor, possui as variáveis x , \min e \max dispostas do valor real, o valor mínimo daquele atributo e valor máximo daquele atributo, respectivamente, de cada instância.

$$x = \frac{x - \min}{\max - \min} \quad (1)$$

2.1.5 Classificação

Neste trabalho, utilizou-se das métricas MAE, MAPE e RMSE, que segundo De Myttenaere (2017), o MAPE é uma medida de qualidade usada em modelos de regressão, amplamente utilizada devido à sua interpretação intuitiva em termos de erro relativo. Ainda segundo o autor, o MAPE é aplicado em situações em que a quantidade a ser prevista é significativamente maior que zero, sendo adequado para aplicações de previsão, quando há dados suficientes. A fórmula utilizada para o cálculo do MAPE, que pode ser vista na Equação 2, de acordo com Coquita *et al.* (2015) é dada por:

$$MAPE = \frac{100}{N} \times \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{x_i} \right| \quad (2)$$

O Erro Médio Absoluto (MAE), de acordo com Herlocker *et al.* (2004), calcula a média das diferenças absolutas entre as classificações previstas e as classificações verdadeiras. O MAE tem sido amplamente utilizado na avaliação de sistemas de recomendação, em várias situações. A fórmula utilizada para o cálculo do MAE, pode ser vista na Equação 3, de acordo com Cazella *et al.* (2009).

$$MAE = \frac{\sum_{i=1}^N |x_i - \hat{x}_i|}{N} \quad (3)$$

De acordo com Rafique *et al.* (2020), a Raiz do Erro Quadrático Médio (RMSE) mede a discrepância entre os valores medidos e previstos. Portanto, menores valores de RMSE indicam menores discrepâncias. De acordo com Santos e Da Silva (2020), a fórmula utilizada para calcular o RMSE pode ser vista na Equação 4.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2} \quad (4)$$

Ainda de acordo com Santos e Da Silva (2020), N representa o número de amostras, x_i o valor observado (ou real) e \hat{x}_i o valor predito pelo modelo, as variáveis descritas são para todas as equações (1), (2) e (3).

A categorização das métricas MAE, MAPE e RMSE, de acordo com Ağbulut, Gürel e Biçen (2021), é importante para avaliar a qualidade dos modelos de previsão. Os intervalos fornecidos proporcionam uma compreensão mais simples do desempenho do modelo, contribuindo para a interpretação da predição, que podem ser vistos no Quadro 1.

Quadro 1 - Classificação MAE, MAPE e RMSE

Métrica	Classificação
MAE	Quanto mais próximo de 0 melhor e quanto mais próximo de 100 pior
MAPE	Alta precisão de previsão: $MAPE \leq 10\%$ Boa previsão: $10\% < MAPE \leq 20\%$ Previsão razoável: $20\% < MAPE \leq 50\%$ Previsão imprecisa: $MAPE > 50\%$
RMSE	Quanto mais próximo de 0 melhor e quanto mais próximo de 100 pior

Fonte: Adaptado de Ağbulut, Gürel e Biçen, 2021.

2.2 Estado-da-arte

Para contextualizar o trabalho desenvolvido nesta monografia, realizou-se uma revisão bibliográfica sobre o Estado-da-arte, da área de energia, apresentando as principais técnicas utilizadas e os resultados obtidos. Existem diversos esforços na área de energia elétrica para análise de geração de energia elétrica, envolvendo Mineração de Dados. Carvalho *et al.* (2022) utilizaram de técnicas de Mineração de Dados para se fazer a predição da geração de energia elétrica, em diversos *campi*. Foram realizados quatro experimentos em que o primeiro e o segundo contêm apenas dados da geração de energia elétrica. O terceiro e quarto experimentos contêm dados da geração de energia elétrica e dados meteorológicos. As técnicas utilizadas no trabalho foram: KNN, Regressão Linear Simples e RF. No primeiro experimento, feito com os *campi* Bambuí, Ouro Preto e Ribeirão das Neves, a melhor técnica

foi KNN com a base de dados de Ouro Preto com MAPE valendo 9,16%. No segundo experimento, feito com os *campi* Governador Valadares, Ouro Preto, Ribeirão das Neves e São João Evangelista, a melhor técnica foi RF com a base de dados de Ouro Preto com MAPE valendo 13,99%. No terceiro experimento, feito com os *campi* do experimento anterior só que agora junto aos dados meteorológicos, a melhor técnica foi RF com a base de dados de Bambuí com MAPE valendo 9,41%. E por fim, no quarto experimento, feito com os *campi* do segundo experimento só que agora junto aos dados meteorológicos, a melhor técnica foi RF com a base de dados de Ouro Preto com MAPE valendo 13,95%. As métricas que foram utilizadas para analisar os resultados foram: MAPE, MAE e RMSE.

Yoo e Myriam (2018) aplicaram Redes Neurais para a previsão de consumo de energia elétrica em residências. As casas analisadas estão localizadas em Seul, capital da Coreia do Sul. Para fazer a previsão, foram coletados 20 anos de dados, onde esses dados são características particulares da população, como: dados demográficos, de consumo residencial e dados meteorológicos. Foram realizados treinamentos e testes com os dados para se obter os resultados. Já as métricas utilizadas para as análises foram: Coeficiente de Variação (CV), Erro Médio de Viés (MBE), MAPE e Média das métricas. Na análise de conjunto de dados 1, o melhor resultado foi de 4,26%, obtido através da média dos valores das métricas e para encontrar esse resultado foi necessário selecionar 10% do conjunto. Na análise do conjunto de dados 2, as variáveis indicadas foram as que tiveram menor taxa de erro, através de 10% do conjunto. Na terceira análise, foi feita uma comparação de previsão dos dois primeiros conjuntos de dados. Como resultado, os dados separados por meses do ano têm desempenho superior. A quarta análise deu-se por meio das combinações de variáveis, que dependendo do mês do ano, todas as variáveis têm uma influência direta na previsão do consumo de energia.

O trabalho de Candanedo, Feldheim e Deramaix (2017) trata-se da previsão do consumo de energia elétrica em eletrodomésticos e diferentes preditores. A casa analisada por eles, foi construída em 2015, na Bélgica. Os dados foram coletados por meio de medições a cada 10 minutos, por, aproximadamente, 4,5 meses, ou seja, 137 dias. As técnicas utilizadas no trabalho foram: Regressão Linear Múltipla, *Support Vector Machine with Radial Kernel*, RF e *Gradient Boosting Machines* (GBM). A técnica que mostrou melhor resultado foi a GBM, com 97% da variância no treinamento e 57% no teste com a métrica R^2 (ou coeficiente de determinação, que de acordo com o site Minitab, é uma métrica estatística que avalia a proximidade dos dados em relação à linha de regressão ajustada). As métricas utilizadas para

se chegar à conclusão de qual se saiu melhor foram: MAPE, MAE, RMSE e R^2 .

O trabalho feito por Antonioli *et al.* (2022) trata-se da aplicação das técnicas Regressão Linear, Redes Neurais MLP e *Support Vector Machine* (SVM), no *software* WEKA, para a previsão do consumo de energia elétrica, em um frigorífico de aves. O SVM usado para regressão, torna-se SVR (*Support Vector Regression*); pelo WEKA é conhecido como SMOreg. As métricas utilizadas pelos autores foram MAE, MAPE, RMSE, com 70% dos dados para treinamento e 30% para teste. Dentre as técnicas aplicadas, a que mostrou melhor resultado foi a SMOreg, com menor valor de MAPE, valendo 3,43%.

Vasconcelos *et al.* (2017) aplicaram as técnicas de Mineração de Dados, para se fazer a previsão do consumo de energia elétrica residencial, em *Gigawatt-hora* (GWh), utilizando dados do PIB (Produto Interno Bruto), da população do Brasil e da Eletrobras. Com a aplicação das técnicas foi realizado uma comparação, sendo aplicadas SVR (*Support Vector Regression*), Redes Neurais MLP e Regressão Linear. A métrica utilizada para fazer a avaliação da primeira etapa foi o MAPE. Os resultados foram obtidos em duas etapas: a primeira etapa gerou previsões dos anos de 2004 a 2013; a segunda, gerou previsões de 2014 a 2016. Para fazer as previsões da primeira etapa utilizaram-se três grupos de dados: o de consumo, o de consumo e PIB e o de consumo, mais o de PIB e população. Após a obtenção dos resultados, foram comparados com os de outro trabalho, o de Torrini (2014), obtendo menores valores do MAPE, com a técnica SVR. Para consumo, o menor resultado foi 0,0239%; para consumo e PIB, o menor resultado foi de 0,0332%; e para consumo mais PIB e população, o menor resultado foi 0,0282%. Os resultados da segunda etapa fizeram comparações das técnicas com os dados da Empresa de Pesquisa Energética (EPE) e com o trabalho de Torrini (2014), prevendo consumo dos anos 2014, 2015 e 2016. Para o ano 2014, o melhor resultado foi da técnica Regressão Linear, possuindo junto ao consumo mais o PIB e população, resultando no valor de 132179 Gwh; a técnica SVR, junto ao consumo e o PIB, também obteve bons resultados, com o valor de 132785 GWh. No ano de 2015, as melhores técnicas foram: Regressão Linear e Rede Neural MLP, ambas utilizando somente consumo, resultando valores de, respectivamente, 131036 GWh e 132514 GWh. E no ano 2016, a técnica SVR, utilizando consumo mais PIB e população, obteve o melhor resultado, com o valor de 133316 Gwh, seguido das Redes Neurais, MLP, somente com consumo, obtendo como resultado o valor de 132457 GWh.

A presente monografia insere-se também no escopo de Mineração de Dados do uso de

energia elétrica, no âmbito doméstico, visando analisar técnicas de Mineração de Dados. A base de dados a ser utilizada é a mesma presente no Estado-da-arte, espelhando-se no trabalho de Candanedo, Feldheim e Deramaix, (2017) que obteve seus resultados comparados aos resultados obtidos nesta monografia.

3 METODOLOGIA

Este capítulo trata-se da metodologia do trabalho. Na Seção 3.1, será detalhado a Classificação do Trabalho; na Seção 3.2, a Solução; e na Seção 3.3, os Materiais e Tecnologias.

3.1 Classificação do trabalho

O trabalho caracteriza-se como natureza aplicada, de acordo com Moresi (2003), por gerar conhecimento através de práticas focadas em problemas específicos, na área de consumo de energia elétrica doméstica. Moresi (2003) afirma que a abordagem da pesquisa é quantitativa, por se tratar de dados numéricos, cujas informações são obtidas através de recursos e técnicas estatísticas, através da Mineração de Dados. A pesquisa encaixa-se como descritiva e exploratória: descritiva, por descrever como a base de dados de previsão do consumo de energia de eletrodomésticos contribui para tomada de decisão e por estabelecer correlações entre as variáveis; exploratória, por explorar os dados através de análises para descobrir informações relevantes. Quanto aos meios de investigação ou procedimentos, enquadra-se como pesquisa experimental, por se tratar de aplicação de técnicas de Mineração de Dados para análise dos modelos gerados e realizarem-se experimentos nesse escopo.

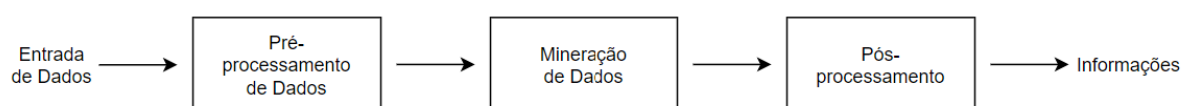
O estilo da pesquisa é presumivelmente melhor, de acordo com Wazlawick (2009), por realizarem-se testes que demonstram que a abordagem a ser utilizada pode ser melhor que as de Candanedo, Feldheim e Deramaix (2017), apresentadas no Estado-da-arte.

3.2 Solução

Na área de Mineração de Dados existe o processo de descoberta de conhecimento em banco de dados, ou simplesmente, processo KDD. De acordo com Tan, Steinbach e Kumar (2009), o KDD transforma os dados brutos em informações valiosas e relevantes. Começando pela entrada de dados até o final, ou seja, as informações obtidas. Esse processo é composto por pré-processamento, aplicação de técnicas de Mineração de Dados e o pós-processamento, conforme a Figura 1. Para a execução do KDD, neste trabalho, seguiu-se o passo a passo descrito abaixo:

- Passo 1: Seleção da base de dados, que no caso da monografia é: Base de dados de previsão de energia de eletrodomésticos;
- Passo 2: Realizar o pré-processamento da base, deixando-a apta para aplicação das técnicas no *software* WEKA;
- Passo 3: Aplicação das técnicas, que são: RNA, KNN e RF. Todas as técnicas devem ter seus parâmetros calibrados para a execução;
- Passo 4: Análise dos resultados obtidos por meio dos modelos gerados.

Figura 1- Processo de descoberta de conhecimento em bancos de dados (KDD).



Fonte: Adaptado de Tan, Steinbach e Kumar (2009).

As técnicas que foram empregadas possuem parâmetros que devem ser seguidos. Na RNA do tipo *Perceptron* Multi-Camadas serão utilizados valores padrões disponibilizados pelo WEKA. De acordo com Freire *et al.* (2014), os valores padrões são: -L, refere-se à taxa de aprendizado utilizada, cujo valor deve estar entre 0 e 1, sendo o padrão 0,3. O parâmetro -M representa a taxa de momento, cujo valor também deve estar entre 0 e 1, sendo o padrão 0,2. O parâmetro -N corresponde ao número de épocas para o treinamento da rede, com o valor padrão de 500.

Para o KNN será calibrado o valor do parâmetro de k, que se trata da quantidade de vizinhos mais próximos. Os valores a serem considerados de k são: 1, 3, 5, 7 e 10, conforme metodologia utilizada no trabalho de Faria, Pereira Jr. e Merschmann (2015).

Na técnica RF, de acordo com Breiman (2001), dois parâmetros devem ser calibrados, o *n*tree e o *m*try. O *n*tree determina o número de árvores na floresta, sendo avaliados os valores 500 e 1000; o *m*try define o número de atributos considerados na construção de cada árvore, sendo: 1/3 do número de atributos preditores, 1/6 da quantidade de atributos preditores e 2/3 da quantidade de atributos preditores.

No painel de avaliação do WEKA, existe a opção “selecionar as métricas desejadas”. Para fazer a avaliação, foi utilizada a técnica de divisão dos dados em conjuntos de treinamento e teste, podendo ser configurada manualmente no *software*, através do campo *Evaluation* e selecionar o *Evaluate on training* (PENTAHO DATA MINING, 2014). De

acordo com Dos Santos *et al.* (2019), a divisão da amostra em conjuntos de treinamento e teste é feita para avaliar se um modelo apresenta um desempenho adequado, não apenas nos dados usados para treiná-lo, como também em novas observações, que são os testes. As proporções mais comumente utilizadas são 60:40, 70:30 ou 80:20. Na presente monografia, utilizou-se o valor padrão do WEKA, 70:30, podendo ser visto no campo *Evaluate on held out training*.

Para analisar os dados obtidos nesta monografia, foram utilizadas as métricas MAE, MAPE e RMSE que, de acordo com Ağbulut, Gürel e Biçen (2021), o MAE, MAPE e RMSE quanto mais próximo de zero estiverem, melhor.

3.3 Materiais e Tecnologias

As seguintes seções têm como objetivo descrever os materiais e tecnologias utilizados na realização da monografia. Serão apresentados detalhes sobre cada material e tecnologia empregados, destacando suas características, propriedades e aplicações. A Seção 3.3.1 são os Materiais e Tecnologias e na Seção 3.3.2, o Processo de Desenvolvimento do trabalho.

3.3.1 Materiais

A máquina utilizada durante o trabalho pertence ao discente, e possui as especificações: processador Intel Core I7 – 1165G7 2.80 *Gigahertz*, 8 *Gigabyte* DDR4 de memória principal (RAM), placa de vídeo NVIDIA GeForce MX350 2 *Gigabyte* e sistema operacional de 64 bits Windows 11. O *software* utilizado é o WEKA² versão 3.8.6.

3.3.2 Processo de Desenvolvimento

A metodologia ágil empregada no trabalho é a SCRUM, adaptada para esta monografia. A metodologia SCRUM possui o *Product Backlog* e as *Sprints*. Segundo Pereira, Torreão e Marçal (2007), o *Product Backlog* consiste em uma lista de itens a serem realizados, com o peso de cada uma e a ordem de priorização, bem como o tempo necessário para a realização. Feito o *Product Backlog*, as *Sprints* separam os itens em iterações que, na presente monografia, são adaptadas, e o tempo marcado em meses, não em semanas, como

relatam Pereira, Torreão e Marçal (2007). O *Product Backlog* não necessita de estar totalmente pronto, podendo ser alterado, de acordo com a necessidade. O *Product Backlog* desse trabalho é apresentado na Tabela 1, com a seguinte estrutura: a coluna um é a coluna do ID que marca a indicação de cada tarefa, a coluna dois marca o nome da tarefa, a coluna três marca o tempo de execução de cada tarefa, disposto em meses, e a coluna quatro o peso de cada tarefa que consiste na importância de cada uma.

Tabela 1 - *Product Backlog*

ID	Tarefa	Tempo estimado em Meses	Peso (1 a 10)
1	Revisão bibliográfica	11	10
2	Apresentar proposta	1	10
3	Pré-processamento da base de dados	2	9
4	Reuniões periódicas	11	9
5	Aplicação das técnicas de Mineração de Dados no WEKA	2	9
6	Fazer análise dos resultados obtidos	4	10
7	Comparação dos resultados obtidos com os do Estado-da-Arte	2	10
8	Escrita da monografia	11	10
9	Defesa da monografia	1	10

Fonte: Elaborado pela autora, 2023.

O presente trabalho é composto por cinco *Sprints*. Cada *Sprint* é descrita em uma tabela contendo o ID da tarefa na primeira coluna, o nome da tarefa na segunda coluna e o tempo estimado para sua conclusão, na terceira coluna. Todas as colunas são derivadas do *Product Backlog*. O tempo necessário para cada tarefa varia de um a onze meses. A Tabela 2 apresenta um exemplo das *Sprints*, enquanto as demais *Sprints* estão disponíveis no APÊNDICE A.

Tabela 2 – *Sprint 1*

ID	Tarefas	Meses
1	Revisão bibliográfica	11
4	Reuniões periódicas	11
8	Escrita da monografia	11

Fonte: Elaborado pela autora, 2023.

4 RESULTADOS E DISCUSSÕES

Este capítulo trata-se dos resultados obtidos e discussões em torno desses resultados. Nas Seções 4.1 e 4.2, serão explanadas a caracterização da casa e da base, respectivamente. Na Seção 4.3, será realizada a aplicação das técnicas bem como as análises.

4.1 Caracterização da casa

Segundo Candanedo, Feldheim e Deramaix (2017), a casa analisada está situada na Bélgica, na cidade de Stambruges. As Figuras 2 e 3 mostram o *layout* da casa, indicando as posições de cada sensor de temperatura (Tn), umidade (RHn) e o coordenador C que está localizado no centro da casa. A temperatura e umidade foram monitoradas através de sensores.

Figura 2 - Planta baixa do primeiro andar.



Fonte: Candanedo, Feldheim e Deramaix, 2017.

Figura 3 - Planta baixa do segundo andar.



Fonte: Candanedo, Feldheim e Deramaix, 2017.

No trabalho de Candanedo, Feldheim e Deramaix (2017), foi realizada a seleção das variáveis/atributos a serem utilizadas(os) na aplicação das técnicas de mineração de dados. De acordo com os autores, é de suma importância determinar quais atributos têm maior relevância e quais não contribuem significativamente para a melhoria na previsão do consumo de energia elétrica de eletrodomésticos.

As variáveis/atributos utilizadas(os) por Candanedo, Feldheim e Deramaix (2017) foram: date (data), *appliances* (consumo de energia em eletrodomésticos), *lights* (consumo de energia em luminárias), T1 (temperatura da cozinha), RH_1 (umidade da cozinha), T2 (temperatura da sala), RH2 (umidade da sala), T3 (temperatura da lavanderia), RH_3 (umidade da lavanderia), T4 (temperatura do escritório), RH_4 (umidade do escritório), T5 (temperatura do banheiro), RH_5 (umidade do banheiro), T6 (temperatura do exterior da casa), RH_6 (umidade do exterior da casa), T7 (temperatura da sala de passar roupas), RH_7 (umidade da sala de passar roupas), T8 (temperatura quarto adolescente), RH_8 (umidade quarto adolescente), T9 (temperatura quarto dos pais), RH_9 (umidade quarto dos pais),

T_{out} (temperatura do exterior da casa), Press_{mm_hg} (pressão), RH_{out} (umidade do exterior da casa), *Windspeed* (velocidade do vento), *Visibility* (visibilidade), *Tdewpoint* (ponto de orvalho), rv1 (variável aleatória), rv2 (variável aleatória).

De acordo com Candanedo, Feldheim e Deramaix (2017), a casa é ocupada por quatro pessoas, sendo dois adolescentes e dois adultos. Um dos adultos tem um emprego que envolve trabalho regular em casa (*home office*). Essa casa possui diversos eletrodomésticos que podem ser vistos no Quadro 2.

Quadro 2 - Eletrodomésticos da casa separados por cômodo

Sala	Equipamento
Lavanderia	Geladeira pequena, freezer vertical, adega, máquina de lavar roupa, secadora, roteador de internet, <i>hub</i> de internet, armazenamento em rede anexado (NAS).
Cozinha	Geladeira, <i>cooktop</i> de indução, exaustor de cozinha, microondas, forno, lava-louças, cafeteira
Sala de jantar/estar	Repetidor Wi-Fi, TV, gabinete para disco rígido, leitor de DVD, decodificador de cabo, <i>laptop</i> , impressora, persianas elétricas
Escritório	2 computadores <i>desktop</i> , 3 monitores, roteador, <i>laptop</i> , impressora/copiadora, persianas elétricas
Sala de passar roupas	Relógio, rádio, ferro de passar roupas, persianas elétricas
Quarto dos pais	Relógio, rádio, persianas elétricas, 2 luminárias
Quarto 2	Computador de mesa, monitor, despertador, persiana elétrica
Banheiros	2 escovas de dente elétricas, secador de cabelo

Fonte: Adaptado de Candanedo, Feldheim e Deramaix, 2017.

4.2 Caracterização da Base

Os autores Candanedo, Feldheim e Deramaix (2017) têm o objetivo de analisar a base de dados por eles caracterizada, a fim de se prever o consumo de energia dos eletrodomésticos. Os autores realizaram as medições e anotações dos valores obtidos e assim criaram a base de dados. Os dados dispostos são das categorias: energia em *watt*-hora (Wh); dados meteorológicos como: temperatura em graus Celsius (°C); umidade em porcentagem (%); pressão em milímetro de mercúrio (mmHg); velocidade do vento, em metro por segundo (m/s); ponto de orvalho, em graus Celsius (°C); visibilidade da estação meteorológica, em quilômetros (Km). A base de dados possui um total de 19.735 (dezenove mil, setecentos e

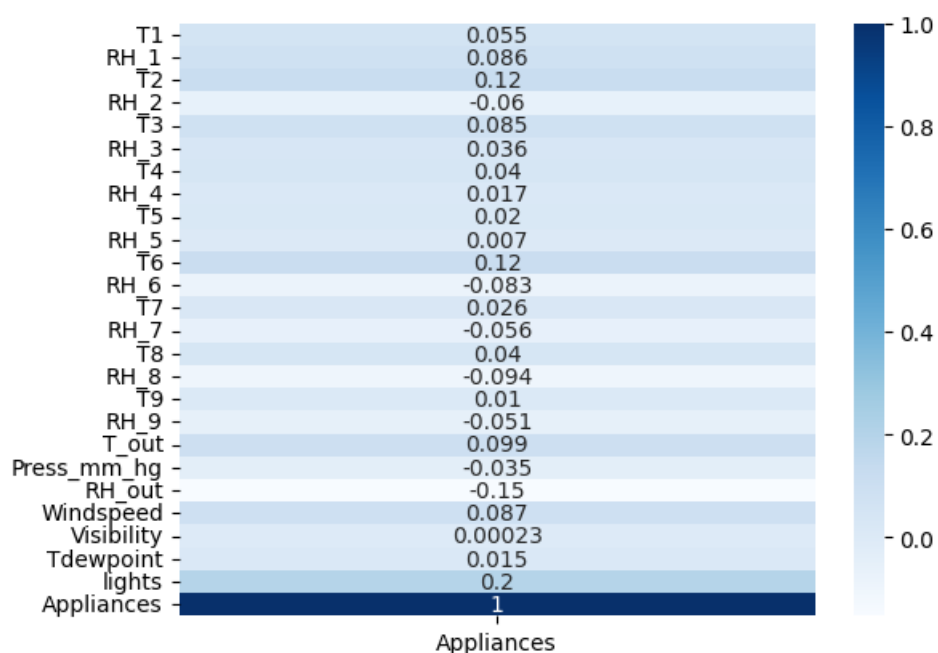
trinta e cinco) instâncias e 28 (vinte e oito) atributos. No APÊNDICE B, pode ser visto cada atributo, sua unidade de medida e seu respectivo tipo.

4.2.1 Correlação dos atributos

A Figura 4 contém os valores de correlação de todos os atributos presentes na base de dados, em relação ao *Appliances*. Os resultados mais significativos, diretamente proporcionais, de correlação, em relação do *Appliances*, foram: o *Lights*, possuindo o valor de 0,2; T2, possuindo o valor de 0,12; e T6, possuindo o valor de 0,12. Os maiores valores diretamente proporcionais dependem da quantidade de pessoas que estão presentes na casa, ou seja, quanto mais pessoas e mais luminárias acesas em cada cômodo, mais energia está sendo gasta.

Os resultados mais significativos, inversamente proporcionais, de correlação, em relação ao *Appliances*, foram: o RH_out, possuindo o valor de -0,15; o RH_8, possuindo o valor de -0,094; e RH_6, possuindo o valor de -0,083. Os maiores valores inversamente proporcionais, em relação ao *Appliances*, podem ser explicados pela quantidade de energia consumida e o valor da umidade, ou seja, quanto menos úmido, mais energia gasta, ou, quanto mais úmido, menos energia gasta.

Figura 4 – Correlação em relação ao *Appliances*

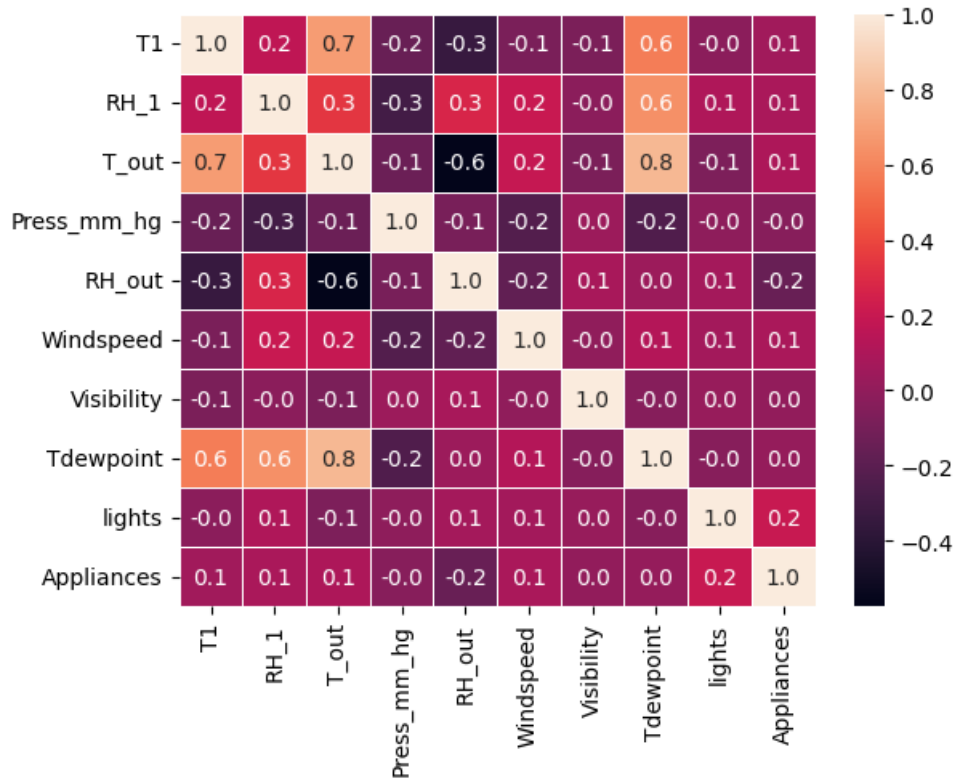


Na Figura 5, encontram-se os resultados da correlação obtidos após a exclusão de diversos atributos da base de dados, mantendo apenas os valores correspondentes ao cômodo designado como sala de estar (T1 e RH_1), junto a outros atributos relativos à base como um todo (T_out, Press-mm-hg, RH_out, *Windspeed*, *Visibility* e *Tdewpoint*). Esses valores são representados por T1 e RH_1, bem como por atributos de valores meteorológicos, como T_out, Press-mm-hg, RH_out, *Windspeed*, *Visibility* e *Tdewpoint*. Essas informações foram obtidas a partir da estação meteorológica mais próxima da casa. Além desses atributos, os dados referentes ao consumo das luminárias também foram mantidos, representados pelo atributo *Lights*.

Para realizar a análise, é possível observar apenas uma parte da tabela, seja da diagonal principal para cima ou da diagonal principal para baixo, onde pode ser observada a correlação de cada atributo com todos os outros. A autora Sousa (2019) ainda diz que é importante saber que correlação não implica causalidade, ou seja, apenas porque encontramos uma relação ou associação entre os atributos, não significa que uma cause algum efeito diretamente sobre a outra.

Com os valores da correlação, cujo resultado foi diretamente proporcional, observa-se que os maiores resultados mais significativos, diretamente proporcionais, foram os valores de 0,8 para a correlação entre *Tdewpoint* e T_out e o valor de 0,7 para a correlação entre T_out e T1. Por outro lado, o resultado mais significativo, inversamente proporcional, foi -0,6 para a correlação entre RH_out e T_out. Com o maior resultado diretamente proporcional, pode-se dizer que quanto maior o ponto de orvalho, maior a temperatura externa, ou, quanto menor o ponto de orvalho, menor a temperatura. E com o maior resultado inversamente proporcional, pode-se dizer que quanto maior a umidade exterior, menor é a temperatura, por se tratar de um país com baixas temperaturas.

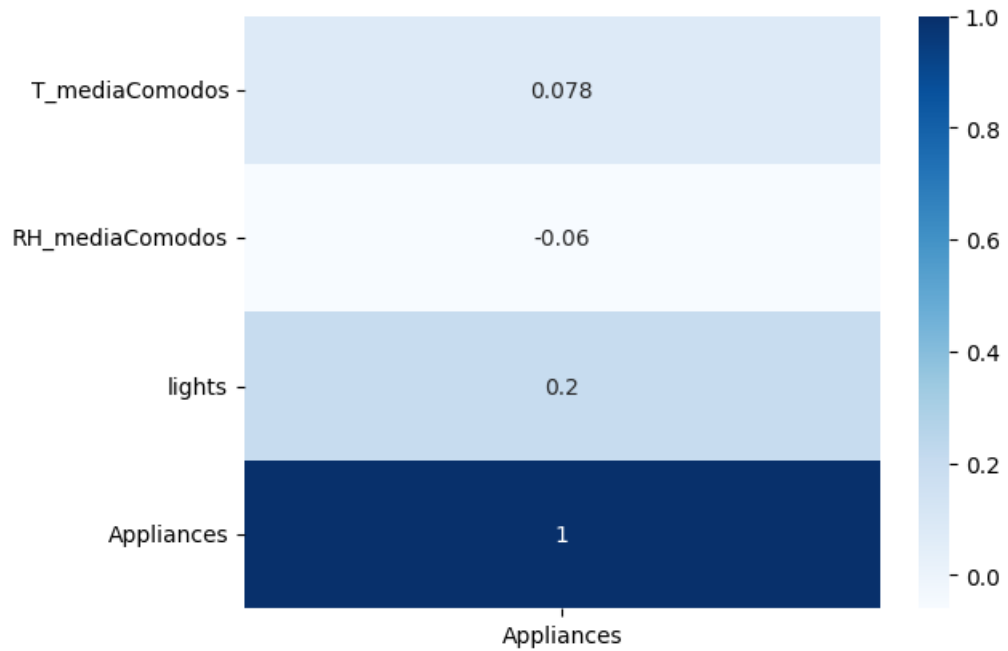
Figura 5 – Correlação em relação ao primeiro cômodo da casa



Fonte: Elaborada pela autora (2023).

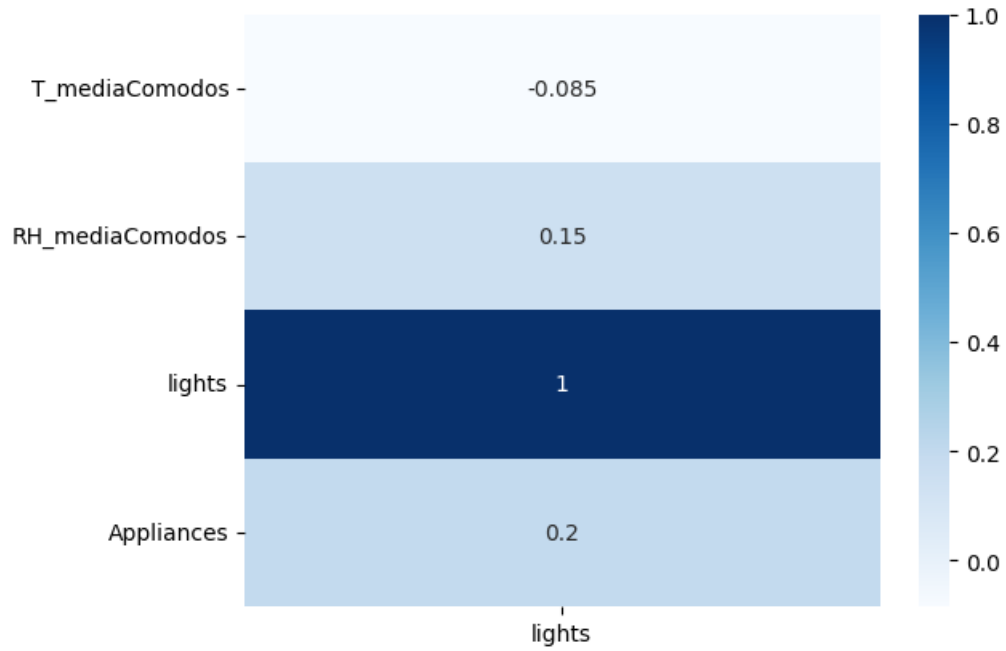
4.2.2 Correlação da média dos atributos

A Seção 4.2.1 mostra a correlação de *Pearson* dos atributos, em relação ao *Appliances* e de cada atributo com todos os outros. Na presente seção, foi realizada a média dos valores dos atributos de temperatura e umidade de cada cômodo (1 ao 9), com intuito de melhorar o resultado da correlação, em relação ao *Appliances* e em relação ao *Lights*. Na Figura 6, pode-se observar que os resultados da correlação foram: diretamente proporcionais para a correlação entre *T_mediaComodos* e *Appliances*, valendo 0,078, inversamente proporcional para a correlação entre *RH_mediaComodos* e *Appliances*, valendo -0,06 e diretamente proporcional para a correlação entre *Lights* e *Appliances*, valendo 0,2.

Figura 6 - Correlação em relação ao *Appliances*

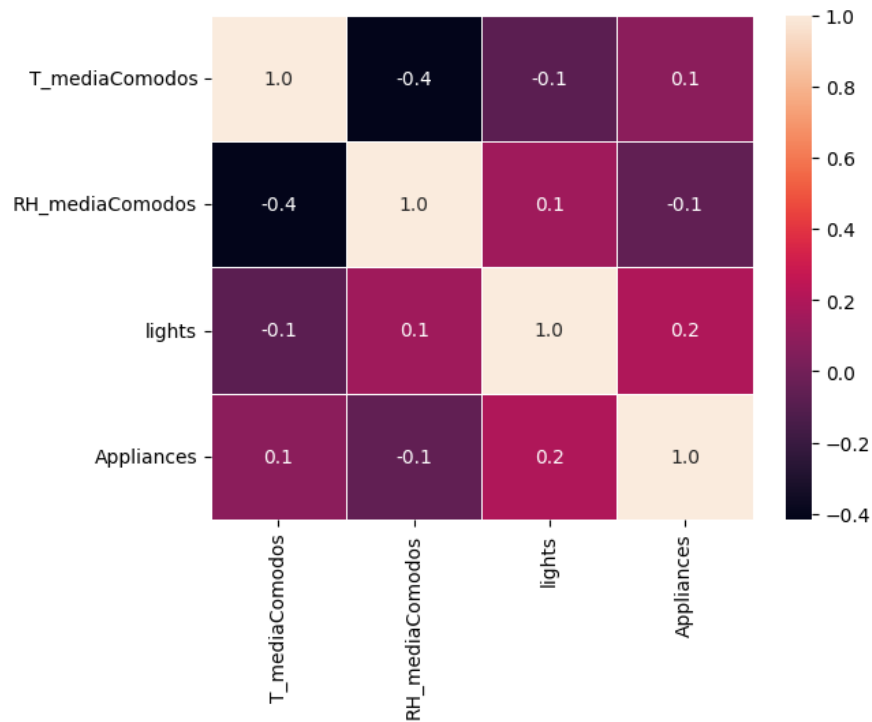
Fonte: Elaborada pela autora (2023).

Na Figura 7, pode-se observar que os resultados da correlação foram: inversamente proporcional para a correlação entre T_mediaComodos e *Lights*, valendo -0,085, diretamente proporcional para a correlação entre RH_mediaComodos e *Lights* valendo, 0,15 e diretamente proporcional para a correlação entre *Appliances* e *Lights*, valendo 0,2.

Figura 7 - Correlação em relação ao *Lights*

Fonte: Elaborada pela autora (2023).

Figura 8 - Correlação de todos os atributos através da média de temperatura e umidade



Fonte: Elaborada pela autora (2023).

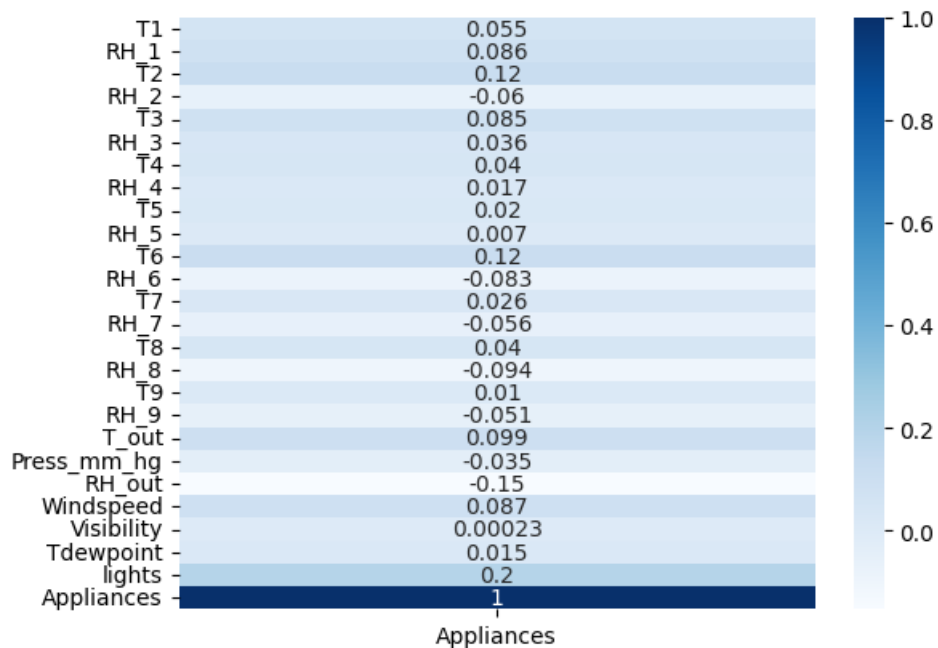
Na Figura 8, pode-se observar a correlação de todos os atributos com todos os outros, e

o resultado mais significativo da correlação positiva ou diretamente proporcional foi na relação *Lights* e *Appliances*, valendo 0,2, o resultado mais significativo da correlação negativa, ou inversamente proporcional, foi em relação ao *RH_mediaComodos* ao *T_mediaComodos*, valendo -0,6.

4.2.3 Correlação dos atributos com base normalizada

A Figura 9, contém os valores de correlação entre todos os atributos da base de dados e o consumo de energia dos eletrodomésticos (*Appliances*), com os dados normalizados. Os dados que foram utilizados foram os mesmos presentes na Seção 4.2.1. Os maiores resultados significativos diretamente proporcional de correlação com o *Appliances* foram: o *Lights*, possuindo o valor de 0,2; *T2*, possuindo o valor de 0,12; e *T6*, possuindo o valor de 0,12. Os maiores resultados significativos inversamente proporcional de correlação com o *Appliances* foram: o *RH_out*, possuindo o valor de -0,15; o *RH_8*, possuindo o valor de -0,094; e *RH_6*, possuindo o valor de -0,083.

Figura 9 - Correlação em relação ao *Appliances* com base normalizada



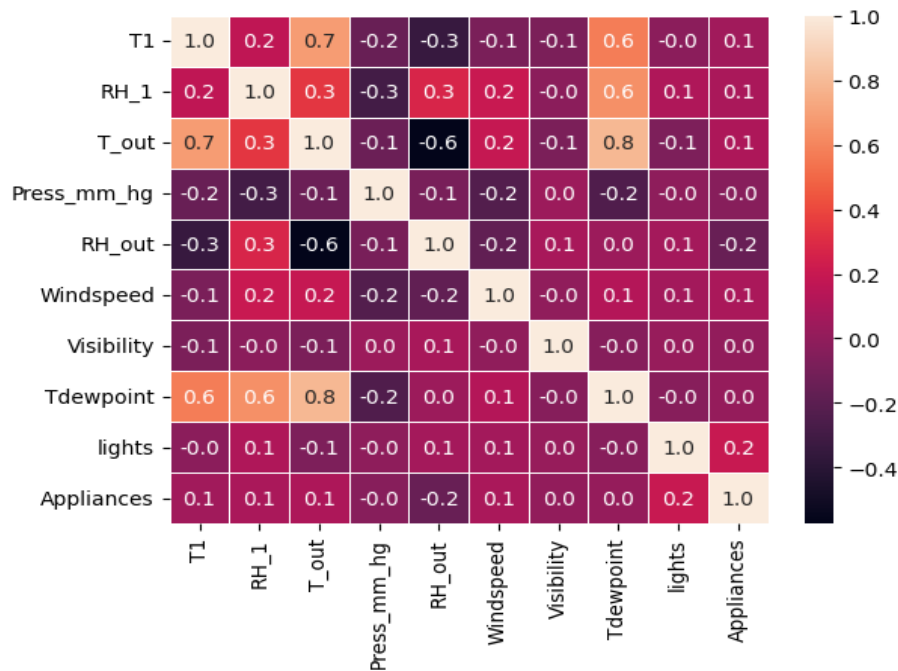
Fonte: Elaborada pela autora (2023).

Na Figura 10, encontram-se os resultados da correlação obtidos após a exclusão de

diversos atributos da base de dados. Foram mantidos apenas os atributos T1 e RH_1, T_out, Press-mm-hg, RH_out, Windspeed, Visibility e Tdewpoint, Lights e Appliances. Como mencionado anteriormente, ao realizar a análise, é possível observar apenas uma parte da tabela de correlações. Isso significa que podemos examinar a correlação de cada atributo com todos os outros, seja olhando da diagonal principal para cima, ou da diagonal principal para baixo.

Com os valores da correlação, observa-se que os resultados mais significativos, diretamente proporcionais, foram os valores de 0,8 para a correlação entre Tdewpoint e T_out e o valor de 0,7 para a correlação entre T_out e T1. Por outro lado, o resultado mais significativo, inversamente proporcional, para correlação, foi -0,6 para a correlação entre RH_out e T_out.

Figura 10 - Correlação em relação ao primeiro cômodo da casa com base normalizada



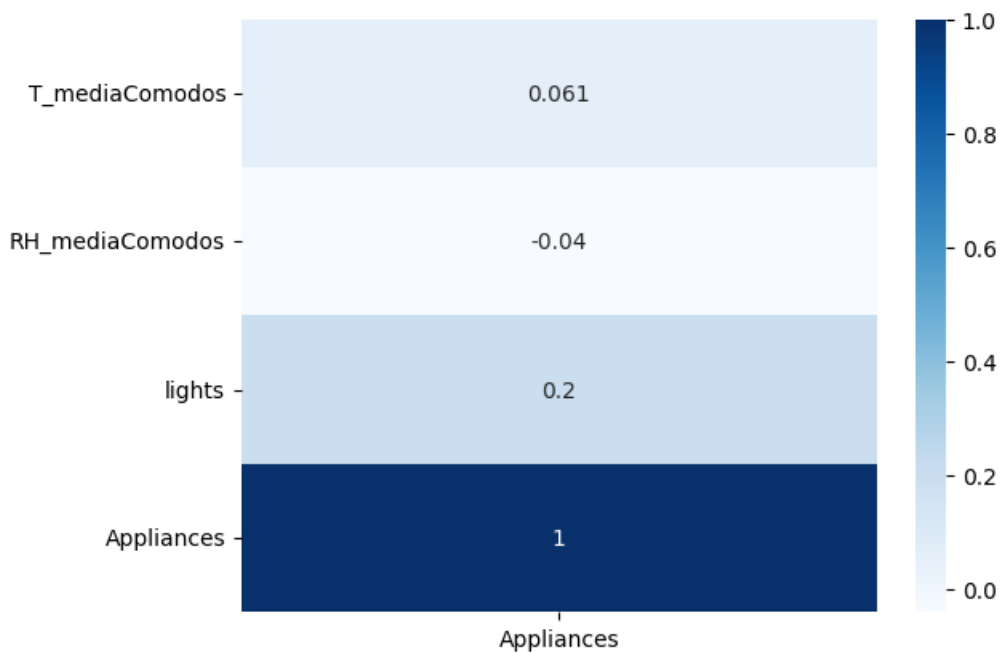
Fonte: Elaborada pela autora (2023).

Ao analisar os resultados e realizar uma comparação entre as correlações calculadas, com os dados normalizados e não normalizados, constata-se que os valores obtidos foram os mesmos em ambas as situações. Dessa forma, é possível concluir que as discrepâncias nas escalas dos valores da base de dados não exerceram influência nas duas análises.

4.2.3.1 Correlação média atributos com base normalizada

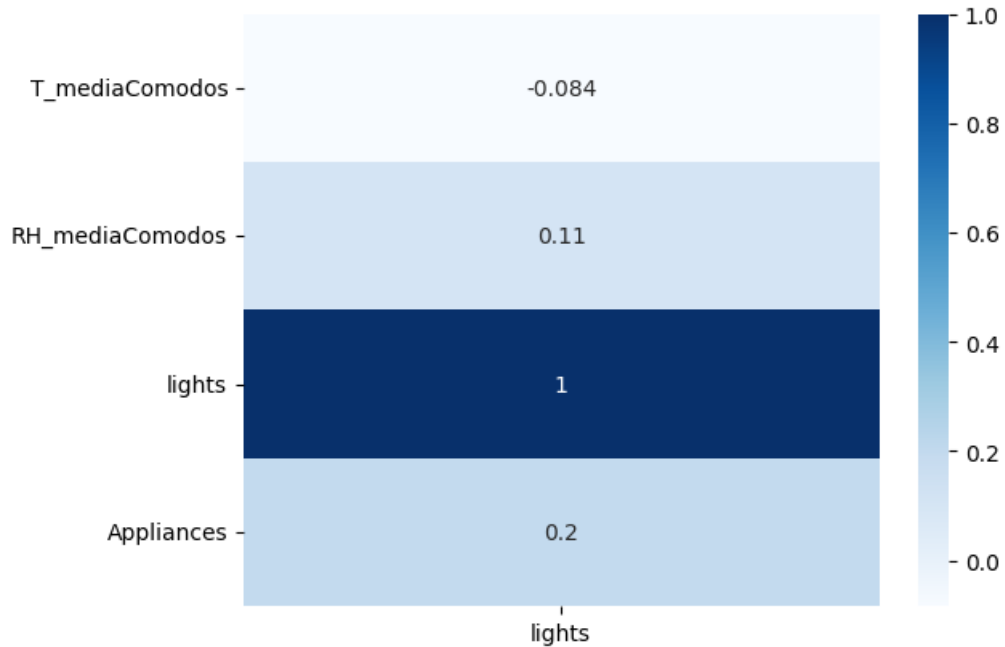
Na presente seção, conforme realizado na Seção 4.1.2, efetuou-se a média dos valores dos atributos de temperatura e umidade, de cada cômodo (1 ao 9), e os atributos *Appliances* e *Lights*. Na Figura 11, que possui correlação dos atributos em relação ao *Appliances*, pode-se observar que os resultados da correlação foram: diretamente proporcionais, para a correlação entre *T_mediaComodos* e *Appliances* valendo 0,061, inversamente proporcional para a correlação entre *RH_mediaComodos* e *Appliances* valendo -0,04 e diretamente proporcional para a correlação entre *Lights* e *Appliances* valendo 0,2.

Figura 11 - Correlação em relação ao *Appliances* com base normalizada



Fonte: Elaborada pela autora (2023).

Figura 12 - Correlação em relação ao *Lights* com base normalizada

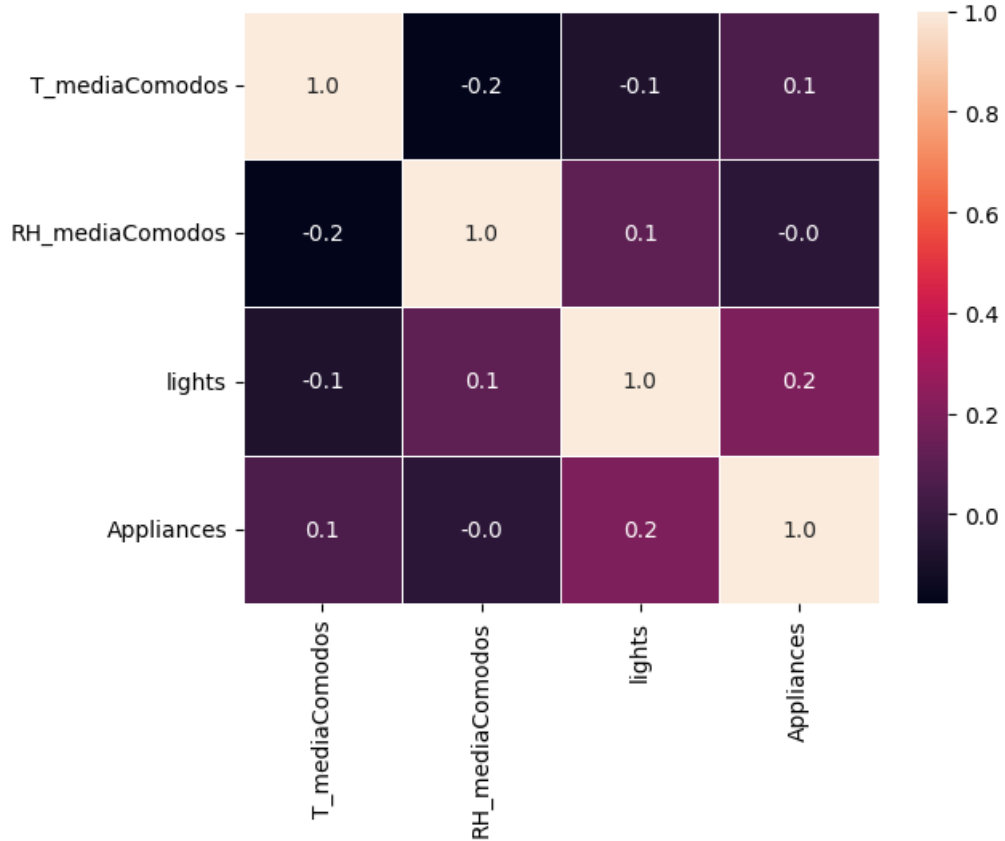


Fonte: Elaborada pela autora (2023).

Na Figura 12, pode-se observar que os resultados da correlação dos atributos em relação ao *Lights* foram: inversamente proporcionais, para a correlação entre T_mediaComodos e *Lights*, valendo -0,084; diretamente proporcionais, para a correlação entre RH_mediaComodos e *Lights*, valendo 0,11; e diretamente proporcionais, para a correlação entre *Appliances* e *Lights*, valendo 0,2.

Na Figura 13, pode-se observar a correlação de todos os atributos com todos os outros, sendo o maior resultado da correlação positiva, ou diretamente proporcional, foi em relação ao *Lights* e *Appliances*, valendo 0,2; o maior resultado da correlação inversamente proporcional foi em relação ao RH_mediaComodos ao T_mediaComodos, valendo -0,2.

Figura 13 - Correlação de todos os atributos através da média de temperatura e umidade com base normalizada



Fonte: Elaborada pela autora (2023).

Após analisar os resultados e realizar uma comparação entre as correlações calculadas, com os dados normalizados e não normalizados, constata-se que os valores obtidos foram diferentes em algumas situações, visto que foram três tipos de análises. Portanto, os dados que passaram pelo processo de normalização apresentaram resultados inferiores em certas situações, quando comparados aos dados que não foram normalizados.

4.3 Aplicação das técnicas

A presente seção apresenta os resultados obtidos, para avaliar os modelos preditivos, gerados por técnicas de regressão, sendo elas: RNA, KNN e RF. A partir das correlações da base original e da base normalizada, os resultados da base original foram mais significativos. Os dados discutidos são apenas das técnicas aplicadas nessa base. Os melhores resultados das técnicas empregadas estão sublinhado, para melhor visualização. A Tabela 3 apresenta os

resultados de todas as técnicas aplicadas. A estrutura da tabela consiste nas colunas com as técnicas e seus respectivos parâmetros e linhas com métricas de validação.

Tabela 3 - Resultados das técnicas na base original

Métricas	RNA	KNN			RF	
		k = 1	k = 5	k = 10	<i>ntree</i> = 500	<i>ntree</i> = 1000
MAE	28,50	35,63	28,61	27,42	25,78	<u>25,73</u>
MAPE	23,99%	33,14%	26,39%	25,03%	23,77%	<u>23,73%</u>
RMSE	64,27	82,97	63,28	61,54	58,10	<u>58,02</u>

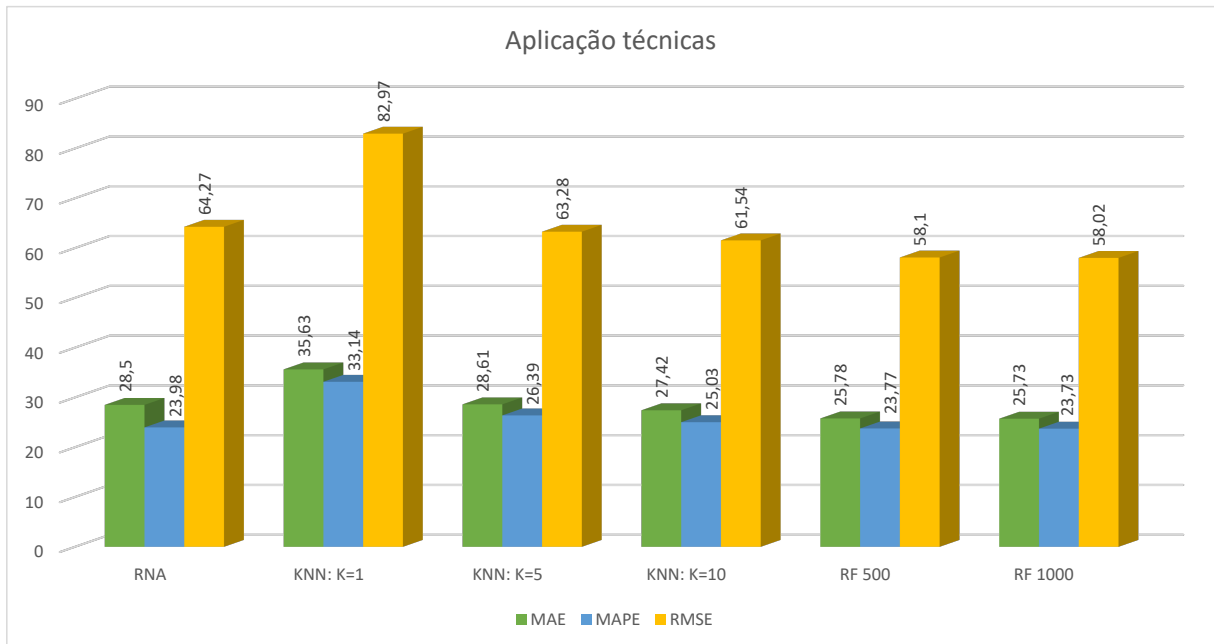
Fonte: Elaborada pela autora (2023).

Após analisar os resultados obtidos, percebe-se que os valores de MAE e RMSE diminuem ou aumentam de acordo com o valor de MAPE. O resultado mais significativo para a monografia está presente na técnica RF que, de forma simples, significa várias Árvores de Decisão. Quando aplicado a técnica KNN, outros testes foram realizados, como alterar o valor de k. Pôde-se perceber que os valores de MAPE diminuam de acordo com o aumento do valor de k. Devido à necessidade de maior poder computacional, o maior valor de k, testado, foi 30.

Na Figura 14, uma análise mais detalhada dos resultados pode ser vista através do gráfico de barras. Analisando o MAE: o algoritmo RF com *ntree*, valendo 1.000, obteve um MAE de 25,73, seguido pelo algoritmo KNN, com k valendo 10, com um MAE de 27,42, e a RNA, com um MAE de 28,50. Analisando o MAPE: o algoritmo RF, com *ntree* valendo 1000, obteve um MAPE de 23,73%, seguido pelo algoritmo RNA, com um MAPE de 23,98%, e o algoritmo KNN, com k valendo 10, com um MAPE de 25,03%. E por fim, analisando o RMSE: o algoritmo RF, com *ntree* valendo 1.000, obteve um RMSE de 58,02, seguido pelo algoritmo KNN, com k valendo 10, com um RMSE de 61,54, e a RNA com um RMSE de 64,27.

Mesmo com resultados semelhantes para *ntree* valendo 500 e 1.000, a técnica RF obteve um resultado mais significativo que o KNN e RNA. O valor mais significativo da predição foi obtido com a técnica RF e *ntree* valendo 1000, com os seguintes resultados: MAE = 25,73, MAPE = 23,73% e RMSE = 58,02.

Figura 14 - Gráfico dos resultados das técnicas



Fonte: Elaborada pela autora (2023).

Os resultados obtidos do trabalho base de Candanedo, Feldheim e Deramaix (2017) podem ser vistos na Tabela 4. Comparando os dois trabalhos (Tabelas 3 e 4), a presente monografia obteve resultados superiores, ao empregar a técnica RF, com um número de *n*tree igual a 1.000, com MAPE valendo 23,73% e para o trabalho base a melhor técnica empregada foi SMV Radial, com MAPE valendo 29,76%.

Tabela 4 - Tabela de resultados da predição do trabalho base

Teste				
Métricas	LM	SMV Radial	GBM	RF
MAE	51,97	<u>31,36</u>	35,22	31,85
MAPE	59,93%	<u>29,76%</u>	38,29%	31,39%
RMSE	93,18	<u>70,74</u>	66,65	68,48

Fonte: Candanedo, Feldheim e Deramaix, 2017.

4.3.1 Divisão da base para aplicação das técnicas

Na busca por possíveis melhores resultados através da aplicação das técnicas, a base de dados foi dividida em duas partes: noite e dia. De acordo com o site DadosMundiais.com, o

anoitecer próximo à cidade de Mons, na Bélgica, ocorre por volta das 16h47 e o amanhecer, por volta das 08h16. Ainda de acordo com o site, o inverno na Bélgica acontece entre dezembro e março, os dias nessa estação possuem menos horas e as noites são mais longas. Portanto, Mons e cidades vizinhas costumam ter no inverno 8h30 de luz solar. Assim, a base de dados foi dividida em duas partes: a primeira com dados referentes à noite e a segunda, com dados referentes ao dia.

Após a divisão, as técnicas foram aplicadas novamente, da mesma forma da base original. Os resultados obtidos estão apresentados nas Tabelas 5 e 6.

Tabela 5 - Resultados das técnicas contendo somente dados noturnos

Métricas	RNA	KNN			RF	
		k = 1	k = 5	k = 10	<i>ntree</i> = 500	<i>ntree</i> = 1000
MAE	32,57	30,56	24,12	<u>22,83</u>	22,18	22,14
MAPE	41,02%	30,96%	23,95%	<u>22,14%</u>	22,43%	22,39%
RMSE	54,73	72,50	56,34	<u>54,52</u>	50,86	50,78

Fonte: Elaborada pela autora (2023).

Tabela 6 - Resultados das técnicas contendo somente dados diurnos

Métricas	RNA	KNN			RF	
		k = 1	k = 5	k = 10	<i>ntree</i> = 500	<i>ntree</i> = 1000
MAE	<u>37,96</u>	47,71	38,98	37,63	34,80	34,83
MAPE	<u>29,52%</u>	41,19%	33,99%	32,90%	30,50%	30,54%
RMSE	<u>80,20</u>	98,88	78,23	76,43	72,64	72,68

Fonte: Elaborada pela autora (2023).

Analisando os resultados, pode-se perceber que o resultado mais significativo foi com a base contendo apenas dados noturnos, com a técnica KNN, com k igual a 10, possuindo as métricas MAE, valendo 22,83, MAPE valendo 22,14% e RMSE valendo 54,52.

De acordo com o site Countryeconomy.com, a média do valor da eletricidade na Bélgica, em 2015 e 2016, anos de coleta dos dados da base de dados, era de €0,255 por *quilowatt*-hora (Kwh), já, a média, nos anos 2022 e 2023, é de €0,442 por Kwh. Um dos países com a eletricidade mais cara da Europa, ocupando o segundo lugar do continente.

Analisando os resultados obtidos e como eles se encaixam como bons, razoáveis ou

ruins, de acordo com os autores Ağbulut, Gürel e Biçen (2021), os resultados de MAE, MAPE e RMSE são classificados de acordo com a Quadro 1 mostrado anteriormente. Os resultados obtidos, na presente monografia, encaixam-se como previsão razoável, ficando entre $20\% < \underline{22,14\%} \leq 50\%$. Mesmo com os baixos resultados de correlação tanto diretamente proporcionais e inversamente proporcionais, obteve-se resultados razoáveis, de acordo com os autores citados acima.

Por fim, comparando-se ao trabalho de Candanedo, Feldheim e Deramaix (2017), os resultados obtidos aqui, sobressaíram-se ao da pesquisa base, que resultou em 31,36 para MAE, 29,76% para MAPE e 70,74 para RMSE, e os resultados da presente monografia foram: 22,83 para MAE, 22,14% para MAPE e 54,52 para RMSE. Dessa forma, espera-se que os resultados possam ser úteis, na tomada de decisão, em relação à economia de energia elétrica doméstica.

5 CONCLUSÃO

Esta pesquisa concentra-se na aplicação das técnicas RNA, KNN e Random Forests, em uma base de dados que abrange informações, como o consumo de energia elétrica em ambientes domésticos e dados meteorológicos, cujo objeto de estudo é uma residência localizada na Bélgica. O objetivo geral da monografia foi aplicar técnicas de Mineração de Dados, em uma base de previsão do consumo de energia elétrica residencial, elaborada por Candanedo, Feldheim e Deramaix (2017).

Nos objetivos específicos, aplicou-se o pré-processamento dos dados através de técnicas matemáticas como: correlação e normalização. A correlação visa tentar entender o funcionamento da base de dados. A mais significativa, diretamente proporcional, ocorreu entre *Tdewpoint* (ponto de orvalho) e *T_out* (temperatura externa), com o valor de 0,8. O mais significativo, inversamente proporcional, ocorreu entre *RH_out* (umidade externa) e *T_out* (temperatura externa), com valor de -0,6. A normalização tem como objetivo, aprimorar a visualização dos dados, garantindo que todas as informações estejam na mesma escala. Além desse objetivo específico, a aplicação das técnicas de Mineração de Dados apresentou o melhor resultado, por meio da técnica KNN com $k = 10$, com valor $MAE = 22,83$, $MAPE = 22,14\%$ e $RMSE = 54,52$. E por fim, foi realizada a comparação entre os autores da pesquisa base, Candanedo, Feldheim e Deramaix (2017), uma vez que o resultado da pesquisa dos referidos autores, foi o que mais se aproximou do resultado da presente monografia, cujo resultado mostrou-se menor (sublinhado): 22,14% < 26,76%.

Para trabalhos futuros, diversas oportunidades destacam-se para aprimorar e ampliar os resultados obtidos, como: Verificar se o WEKA faz o tratamento de *Overfitting*; incorporação de outras ferramentas na construção dos modelos preditivos, como as linguagens de programação *Python* e/ou *R*; a aplicação de novas técnicas de predição e novos métodos de validação; análise dos tipos de eletrodomésticos utilizados e os horários de maior consumo de energia, podendo analisar em períodos e intervalos diferentes, como em semanas ou em meses, englobando também quantos porcentos de energia podem ser economizados, diminuir os gastos de energia residencial e fazer uma comparação nacional, através dessa porcentagem. Portanto, replicar este estudo no contexto brasileiro seria uma extensão valiosa para entender as particularidades e oportunidades de otimização energética nesse contexto específico.

REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, Thays et al. Previsão de cargas elétricas utilizando uma rede neural ARTMAP fuzzy com treinamento continuado. **In: Anais XIII Brazilian Congress on Computational Intelligence**, 2017. Disponível em: <http://cbic2017.org/papers/cbic-paper-88.pdf>. Acesso em: 15 abr. 2023.

AĞBULUT, Ümit; GÜREL, Ali Etem; BIÇEN, Yunus. Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. **Renewable and Sustainable Energy Reviews**, v. 135, p. 110114, 2021. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1364032120304056>. Acesso em: 29 nov. 2023.

ANTONIOLLI, Eduarda Araujo et al. PREVISÃO DO CONSUMO DE ENERGIA ELÉTRICA EM UM FRIGORÍFICO: UM ESTUDO DE CASO USANDO REGRESSÃO LINEAR, REDES NEURASIS E MÁQUINAS DE VETOR DE SUPORTE. **Tekhne e Logos**, v. 13, n. 2, p. 102-115, 2022. Disponível em: <http://revista.fatecbt.edu.br/index.php/tl/article/view/840>. Acesso em: 25 mai. 2023.

Bélgica - Preços de eletricidade doméstica. **Countryeconomy.com**. Disponível em: <https://pt.countryeconomy.com/energia-e-meio-ambiente/eletricidade-preco-domestica/belgica?year=2016>. Acesso em: 28 nov. 2023.

BREIMAN, Leo. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001. Disponível em: <https://link.springer.com/article/10.1023/A:1010933404324>. Acesso em: 24 mai. 2023.

CANDANEDO, Luis M.; FELDHEIM, Véronique; DERAMAIX, Dominique. Data driven prediction models of energy use of appliances in a low-energy house. **Energy and buildings**, v. 140, p. 81-97, 2017. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0378778816308970>. Acesso em: 5 abr. 2023.

CARVALHO, Ana Caroline Coutinho et al. APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA GERAÇÃO DE MODELOS PREDITIVOS DE GERAÇÃO DE ENERGIA ELÉTRICA: UM ESTUDO DE CASO NOS CAMPI IFMG. **Seminário de Iniciação Científica (SIC), 2022**. Disponível em: https://sistemas.bambui.ifmg.edu.br/open_conference/index.php/sic/sic2022/paper/viewFile/628/223. Acesso em: 15 abr. 2023.

CAZELLA, Sílvio César et al. Recomendação de objetos de aprendizagem empregando filtragem colaborativa e competências. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2009. Disponível em: http://www.niee.ufrgs.br/eventos/SBIE/2009/conteudo/artigos/completos/61791_1.pdf. Acesso em: 24 mai. 2023.

COHEN, J. **Statistical power analysis for the behavioral sciences**, 1 ed., 1988.

CÔRTEZ, Sérgio da Costa; PORCARO, Rosa Maria; LIFSCHITZ, Sérgio. **Mineração de dados-funcionalidades, técnicas e abordagens**. PUC, 2002. Disponível em: <https://n9.cl/adpj3>. Acesso em: 21 jun. 23.

COQUITA, Kássio et al. Sistema de Previsão do Tempo de Chegada dos Ônibus Baseado em Dados Históricos Utilizando Modelos de Regressão. In: **Anais do XI Simpósio Brasileiro de Sistemas de Informação**. SBC, 2015. p. 119-126. Disponível em: <https://sol.sbc.org.br/index.php/sbsi/article/view/5808>. Acesso em: 22 mai. 2023.

DE MYTTENAERE, Arnaud De; GOLDEN, Boris; LE GRAND, Bénédicte; ROSSI, Fabrice. Mean absolute percentage error for regression models. **Neurocomputing**, v. 192, p. 38-48, 2017. Disponível em: <https://arxiv.org/abs/1605.02541>. Acesso em 22 mai. 2023.

DOS SANTOS, Hellen Geremias; NASCIMENTO, Carla Ferreira do; IZBICKI, Rafael; DUARTE, Yeda Aparecida de Oliveira; CHIAVEGATTO FILHO, Alexandre Dias Porto. Machine learning para análises preditivas em saúde: exemplo de aplicação para prever óbito em idosos de São Paulo, Brasil. **Cadernos de Saúde Pública**, [S.L.], v. 35, n. 7, 2019. UNIFESP (SciELO). Disponível em: <https://www.scielo.br/j/csp/a/jyhKL6G4dZhcbchMD6bcS8s/?lang=pt>. Acesso em: 17 mai. 2023.

EDITOR, M. B. **Análise de regressão: Como interpretar o R-quadrado e avaliar a qualidade de ajuste?** Disponível em: <https://blog.minitab.com/pt/analise-de-regressao-como-interpretar-o-r-quadrado-e-avaliar-a-qualidade-de-ajuste>. Acesso em: 15 abr. 2024.

Eficiência energética. Eficiência energética e economia de energia. Disponível em: <https://mundoeducacao.uol.com.br/fisica/eficiencia-energetica.htm>. Acesso em: 14 abr. 2024.

Energy Efficiency. Disponível em: https://iea.blob.core.windows.net/assets/59268647-0b70-4e7b-9f78-269e5ee93f26/Energy_Efficiency_2020.pdf. Acesso em: 20 abr. 2023.

Empresa de Pesquisa Energética - EPE. **ABCDEnergia**, 2018. Disponível em: <http://www.epe.gov.br/pt/ABCDEnergia>. Acesso em: 15 abr. 2023.

FARIA, Felipe Lopes de Melo; PEREIRA JR, Álvaro Rodrigues; MERSCHMANN, Luiz Henrique Campos. **Prediction of Artists' Rankings by Regression**. 2015. Disponível em: <https://aisel.aisnet.org/sbis2015/87/>. Acesso em: 22 mai. 2023.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37-37, 1996. Disponível em: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1230>. Acesso em: 15 abr. 2023.

FREIRE, Jean Arouche et al. Avaliação de Desempenho de Classificadores de Ciclos Hidrológicos em Reservatórios de Água na Região Amazônica. **Anais do V Workshop de Computação Aplicada a Gestão do Meio Ambiente e Recursos Naturais**. SBC, 2014. p. 25-34. Disponível em: <https://sol.sbc.org.br/index.php/wcama/article/view/10908>. Acesso em: 22 mai. 2023.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. Data mining concepts and techniques third edition. **University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University**, 2012. Disponível em: l1nq.com/frLAI. Acesso em: 15 abr. 2023.

HERLOCKER, Jonathan L. et al. Evaluating collaborative filtering recommender systems. **ACM Transactions on Information Systems (TOIS)**, v. 22, n. 1, p. 5-53, 2004. Disponível em: <https://dl.acm.org/doi/abs/10.1145/963770.963772>. Acesso em: 24 mai. 2023.

MAIA, Anderson et al. Avaliação de Técnicas de Mineração de Dados para Predição de Desligamentos em Sistemas Elétricos de Potência. **Anais do XI Simpósio Brasileiro de Sistemas de Informação**, 2015. p. 155-162. Disponível em: <https://sol.sbc.org.br/index.php/sbsi/article/view/5812>. Acesso em: 22 mai. 2023.

MORESI, Eduardo. Metodologia da pesquisa. **Brasília: Universidade Católica de Brasília**, v. 108, n. 24, p. 5, 2003. Disponível em: <http://www.inf.ufes.br/~pdcosta/ensino/2010-2-metodologia-de-pesquisa/MetodologiaPesquisa-Moresi2003.pdf>. Acesso em: 15 abr. 2023.

MUKAKA, Mavuto M. A guide to appropriate use of correlation coefficient in medical research. **Malawi medical journal**, v. 24, n. 3, p. 69-71, 2012. Disponível em: <https://www.ajol.info/index.php/mmj/article/view/81576>. Acesso em: 22 jun. 23.

O nascer e o pôr-do-sol na Bélgica. **DadosMundiais.com**. Disponível em: <https://www.dadosmundiais.com/europa/belgica/por-do-sol.php>. Acesso em: 28 nov. 2023.

PEREIRA, Paulo; TORREÃO, Paula; MARÇAL, Ana Sofia. Entendendo Scrum para gerenciar projetos de forma ágil. **Mundo PM**, v. 1, p. 3-11, 2007. Disponível em: encr.pw/Nf9R9. Acesso em: 17 abr. 2023.

RAFIQUE, Muhammad; TAREEN, Aleem Dad Khan; MIR, Adil Aslim; NADEEM, Malik Sajjad Ahmed; ASIM, Khawaja M.; KEARFOTT, Kimberlee Jane. Delegated Regressor, A Robust Approach for Automated Anomaly Detection in the Soil Radon Time Series Data. **Scientific Reports**, [S.L.], v. 10, n. 1, 20 fev. 2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41598-020-59881-9>. Acesso em: 22 mai. 2023.

SANTOS, Daiana da Silva; DA SILVA, Robson Mariano. Aplicação de redes neurais mlp na predição da evapotranspiração de referência. **Brazilian Journal of Development**, v. 6, n. 5, p. 25293-25305, 2020. Disponível em: <https://ojs.brazilianjournals.com.br/ojs/index.php/BRJD/article/view/9674>. Acesso em: 24 mai. 2023.

SINGH, Shailendra; YASSINE, Abdulsalam. Big data mining of energy time series for behavioral analytics and energy consumption forecasting. **Energies**, v. 11, n. 2, p. 452, 2018. Disponível em: <https://www.mdpi.com/1996-1073/11/2/452>. Acesso em: 15 abr. 2023.

SOUSA, Áurea. Coeficiente de correlação de Pearson e coeficiente de correlação de Spearman. O que medem e em que situações devem ser utilizados?. **Correio dos Açores: Matemática**, 21 de Março de 2019, p. 19. Disponível em: <http://hdl.handle.net/10400.3/5365>. Acesso em: 20 jun. 23.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao datamining: mineração de dados**. Ciência Moderna, 2009.

Time Series Analysis and Forecasting with Weka. **Pentaho Data Mining**. Disponível em: <https://abre.ai/jtAk>. Acesso em: 17 abr. 2023.

TORRINI, Fabiano Castro. **Modelos de lógica fuzzy para a previsão de longo prazo de consumo de energia**. 2014. Tese de Doutorado. Tese (Doutorado)—PUC-Rio. Disponível em: https://www2.dbd.puc-rio.br/pergamum/tesesabertas/1213314_2014_completo.pdf. Acesso em: 25 mai. 2023.

UCI Machine Learning Repository. Disponível em: <https://archive.ics.uci.edu/ml/index.php>. Acesso em: 16 abr. 2023.

VASCONCELOS, Leonardo Pio et al. Uma comparação de técnicas de regressão para a previsão de consumo de energia residencial no cenário nacional. **XLIX Simpósio Brasileiro de Pesquisa Operacional**, 2017. Disponível em: <http://www.din.uem.br/sbpo/sbpo2017/pdf/169301.pdf>. Acesso em: 25 mai. 2023.

WAZLAWICK, Raul S. **Metodologia de Pesquisa para Ciência da Computação**. 6. ed. Rio de Janeiro: Elsevier, 2009. WOLTERS, Erik Ch. Non-motor extranigral signs and symptoms in Parkinson 's disease. *Parkinsonism & related disorders*, v. 15, p. S6-S12, 2009. Disponível em: 11nq.com/1T378. Acesso em: 15 abr. 2023.

WEKA. **The University of Waikato**. Software. Versão 3.8.6, 2022. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/>. Acesso em: 15 abr. 2023.

YOO, Sang G.; MYRIAM, Hernández-Á. Predicting residential electricity consumption using neural networks: A case study. **Journal of Physics: Conference Series**. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/1072/1/012005>. Acesso em: 21 abr. 2023.

APÊNDICE A – Restante das Sprints da monografia

Tabela 7 – *Sprint 2*

ID	Tarefas	Meses
2	Apresentação da proposta	1

Fonte: Elaborado pela autora, 2023.

Tabela 8 – *Sprint 3*

ID	Tarefas	Meses
3	Pré-processamento da base	2
5	Aplicação das técnicas	2

Fonte: Elaborado pela autora, 2023.

Tabela 9 – *Sprint 4*

ID	Tarefas	Meses
6	Análise dos resultados	4
7	obtidos Comparação com o Estado- da-arte	2

Fonte: Elaborado pela autora, 2023.

Tabela 10 – *Sprint 5*

ID	Tarefas	Meses
9	Defesa da monografia	1

Fonte: Elaborado pela autora, 2023.

APÊNDICE B – Caracterização da Base

ID	Atributo	Unidade de medida	Classificação do atributo
1	Data	Minuto	Numérico
2	Consumo de energia em eletrodomésticos	Watt-hora (Wh)	Numérico
3	Consumo de energia em luminárias	Watt-hora (Wh)	Numérico
4	Temperatura da cozinha	Celsius (°C)	Numérico
5	Umidade da cozinha	Porcentagem (%)	Numérico
6	Temperatura da sala	Celsius (°C)	Numérico
7	Umidade da sala	Porcentagem (%)	Numérico
8	Temperatura da lavanderia	Celsius (°C)	Numérico
9	Umidade da lavanderia	Porcentagem (%)	Numérico
10	Temperatura do escritório	Celsius (°C)	Numérico
11	Umidade do escritório	Porcentagem (%)	Numérico
12	Temperatura do banheiro	Celsius (°C)	Numérico
13	Umidade do banheiro	Porcentagem (%)	Numérico
14	Temperatura do exterior da casa	Celsius (°C)	Numérico
15	Umidade do exterior da casa	Porcentagem (%)	Numérico
16	Temperatura da sala de passar roupas	Celsius (°C)	Numérico
17	Umidade da sala de passar roupas	Porcentagem (%)	Numérico
18	Temperatura do quarto do adolescente	Celsius (°C)	Numérico
19	Umidade do quarto do adolescente	Porcentagem (%)	Numérico

Fonte: Elaborado pela autora, 2023.

Continuação caracterização da Base

ID	Atributo	Unidade de medida	Classificação do atributo
20	Temperatura do quarto dos pais	Celsius (°C)	Numérico
21	Umidade do quarto dos pais	Porcentagem (%)	Numérico
22	Temperatura externa pela estação meteorológica	Celsius (°C)	Numérico
23	Pressão	Milímetro de mercúrio (mmHg)	Numérico
24	Umidade externa pela estação meteorológica	Porcentagem (%)	Numérico
25	Velocidade do vento	Metro por segundo (m/s)	Numérico
26	Visibilidade	Quilômetros (km)	Numérico
27	Ponto de Orvalho	Celsius (°C)	Numérico
28	Umidade do quarto dos pais	Porcentagem (%)	Numérico

Fonte: Elaborado pela autora, 2023.