DOI: 10.48011/asba.v2i1.1164

Aprendizado de Máquina Aplicado a Classificação de Consumidores de Energia Fotovoltaica no Estado do Maranhão.

Thamires C. Coutinho [*] **Davi C. Nascimento** [**]
Lindomar J. de Souza [***] **Rogean M. C. Leite** [****]

[*] *Instituto Federal de Educação, Ciencia e Tecnologia do Maranhão, São Luís, MA, (e-mail: tha.cutrim@gmail.com).*
[**] *Universidade Federal do Maranhão, MA, (e-mail: davi.nascimento777@gmail.com)*
[***] *Instituto Federal de Educação, Ciencia e Tecnologia do Maranhão, São Luís, MA, (e-mail: lindomar@ifma.edu.br)*
[****] *Instituto Federal de Educação, Ciencia e Tecnologia do Maranhão, São Luís, MA, (e-mail: gean.mauro.cl@gmail.com)*

Abstract: The growing demand for electricity all over the world, combined with the scarcity and decrease in production by traditional generation sources, makes the search for new sources of energy necessary. Most of the large power generation plants in Brazil are far from large consumption centers, implying high expenses with the implementation and maintenance of transmission systems. An alternative capable of mitigating unwanted impacts, is the implantation of energy sources, preferably renewable within the consumer units, even if they are connected to the networks of the energy concessionaires. There is a tendency for a gradual increase in the search and implementation of these systems, in Brazil it is observed that in addition to the majority dependence on an energy source, there are also the high costs of their tariffs, which makes investment in alternative sources attractive. As a result, it is important to analyze characteristics to verify the profiles of consumers who have already purchased a photovoltaic generation system. This work proposes the extraction of data with web scraping tools, together with machine learning in order to analyze if there are characteristics that define the profiles of consumers in the state of Maranhão. The results obtained went through validation metrics, with an accuracy of 99.89 %.

Resumo: A crescente demanda por energia elétrica em todo mundo aliada à escassez e diminuição na produção por parte das fontes de geração tradicionais, torna a busca por novas fontes de energia necessária. A maioria das grandes usinas de geração elétrica do Brasil está distante dos grandes centros de consumo, implicando em elevados gastos com a implantação e manutenção dos sistemas de transmissão. Uma alternativa capaz de atenuar impactos não desejados, consiste na implantação de fontes de energia, preferencialmente renováveis dentro das unidades consumidoras, mesmo que estejam ligadas às redes das concessionárias de energia. Há uma tendência de aumento gradual na busca e implantação desses sistemas, no Brasil observa-se que além da dependência majoritária de uma fonte de energia, há ainda os altos custos de suas tarifas, o que torna o investimento em fontes alternativas atrativo. Com isso, torna-se importante a análise de características para verificação de perfis de consumidores que já adquiriram sistema de geração fotovoltaica. Este trabalho propõe a extração de dados com ferramentas de web scraping, juntamente a aprendizado de máquina visando analisar se existem características que definem os perfis de consumidores no estado do Maranhão. Os resultados obtidos passaram por métricas de validação, com acurácia de 99,89%.

Keywords: Solar energy; Machine learning; Distributed generation; Clustering; Renewable energy.
Palavras-chaves: Energia solar; Aprendizado de máquina; Geração distribuída; Clusterização; Energias renováveis.

1. INTRODUÇÃO

No Brasil, entre as fontes de energia utilizadas, existe a predominância de uma matriz renovável com produção de 66,6% proveniente de fonte hidrelétrica, Energética (2019). Apesar de ser considerada uma fonte de energia "limpa", as usinas hidrelétricas enfrentam questionamentos devido aos impactos socioambientais decorrentes da sua construção, com grandes áreas alagadas e populações ribeirinhas desalojadas. Além disso, por depender da água das chuvas, recurso cada vez menos abundante no Brasil, o país tem sido obrigado a aumentar a geração através de usinas termelétricas, causando um aumento no custo da energia e da emissão dos gases do efeito estufa, ocasionados pela queima dos combustíveis fósseis. Sendo assim, é importante que o país invista em outros tipos de energia renováveis, que causem menos impactos, ambientais e econômicos, para a população, Rocha et al. (2018). A cada ano, a queima de combustíveis fósseis gera mais de 5 bilhões de toneladas de carbono na atmosfera e este número aumenta cerca de 80 milhões de toneladas por ano, Hinrichs et al. (2010).

Dentre as energias renováveis disponíveis, observa-se a inserção da geração fotovoltaica, com uma tendência de crescimento no Brasil ao longo dos anos. Esta é a energia obtida da irradiação solar por meio da conversão direta da luz em eletricidade, Campos (2017). Nesse contexto, modificações nas políticas energéticas e incentivos financeiros à Geração Distribuída (GD) foram concebidos aos consumidores brasileiros a partir de 2012, após a publicação da Resolução Normativa (REN) no 482 e da REN nº 687/2015 da Agência Nacional de Energia Elétrica (ANEEL), que viabilizou a conexão da GD aos sistemas de distribuição, Dranka et al. (2018).

Partindo destas prerrogativas, analisar como se distribuem os consumidores com geração distribuída de fonte fotovoltaica é importante para quantificar o potencial e como ele se distribui, para assim analisar a existência de perfis determinantes para aquisição do sistema. Buscando essas respostas, a extração de dados tem sua relevância para contornar as dificuldades em modelos de publicação das informações, que permite aos seus usuários um modo informal de publicá-las, que remete a não garantia da consistência desses dados, dificultando sua filtragem para subsidiar diversas áreas de conhecimentos, Pontolio (2015).

Neste trabalho foi proposto a extração de informações presentes no site da ANEEL sobre geração distribuída fotovoltaica referente ao estado do Maranhão para analisar seu perfil distribuitivo por meio de algoritmos de aprendizado de máquina.

⋆ Os autores agradecem ao Programa de Bolsas Institucionais do Instituto Federal de Educação, Ciência e Tecnologia do Maranhão - Campus Monte Castelo pelo suporte financeiro.

2. REFERENCIAL TEORICO

2.1 Energia solar

A energia solar fotovoltaica é a energia obtida através da conversão direta da luz em eletricidade, também conhecido como efeito fotovoltaico. Este efeito, relatado por Edmond Becquerel, em 1839, é o aparecimento de uma diferença de potencial nos extremos de uma estrutura de material semicondutor, produzida pela absorção da luz, Cresesb (2007b). No Brasil a geração distribuída, em 2019, chegou a um equivalente de 1 GW de potência instalada, sendo desse total, 870 MW pertencendo a fonte fotovoltaica Aneel (2020). Para o ano de 2020, o número de usinas conectadas na modalidade de geração distribuída alcançou 201.773 unidades, sendo 201,3 mil do tipo solar fotovoltaica e potência instalada de 2.3 GW, Energia (2020). Nesse sentido, a geração distribuída fotovoltaica tem ganhado destaque na matriz de energia elétrica brasileira e mundial, Dranka et al. (2018). De acordo com PCE (2015), a estimativa para o ano de 2050, é que a geração fotovoltaica representará entre 7.3% e 15.7% da capacidade instalada no Brasil, contribuindo ainda para a diversificação do mix energético do país.

O sistema pode operar em on-grid e off-grid. O que define o primeiro é sua conexão direta com a rede. Isso significa que a energia produzida é enviada para a concessionária que, por um sistema de créditos, abate parte da sua conta de luz. Vale lembrar que não é possível zerar a conta de luz utilizando o on-grid, uma vez que nela incidem tributos mínimos para a manutenção do sistema de fornecimento da concessionária. O sistema off-grid por outro lado, não possui conexão à rede, a energia é armazenada diretamente em baterias que irão alimentar o domicílio, Ferreira (2020). Sistemas on-grid são mais baratos em comparação com o off-grid e portanto, se uso mais difundido entre os consumidores de regiões urbanas. O outro sistema pode ser encontrado em localidades onde a distribuição de energia tem dificuldade em alimentar o domicilio ou não há distribuição de consecionária energética. Um exemplo de sistema on-grid pode ser visto na Figura 1.

Segundo Lau (2020), existem dois grandes grupos no que diz respeito às modalidades tarifárias no Brasil, o grupo A (consumidores atendidos em alta tensão) e o grupo B (consumidores atendidos em baixa tensão). Os primeiros são consumidores atendidos com tensão acima de 2.300 Volts, são tipicamente indústrias e grandes complexos comerciais. Pode ser subdividido em 6 subgrupos, conforme mostrado na Tabela 1.

O segundo se refere a consumidores atendidos em tensão abaixo de 2.300 V. É composto tipicamente por residências, lojas, grande parte dos edifícios comercias, e imóveis rurais. O grupo B é subdividido em mais 4 subgrupos, que estão mostrados na Tabela 2.

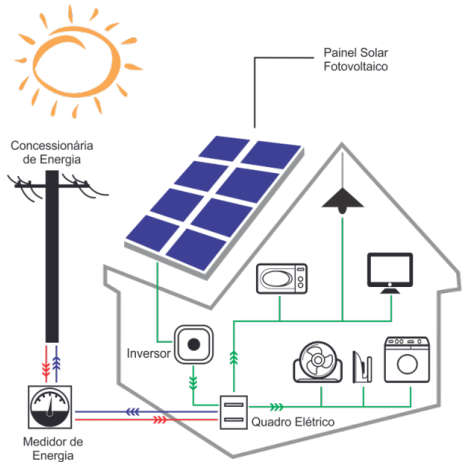

Figura 1. Sistema on-grid de geração fotovoltaica. Fonte: Solar (2020).

Tabela 1. Subgrupos de consumo enrgético para o grupo A.

Subgrupo	Tensão
A1	230 kV ou mais
A2	88 kV a 138 kV
A3	69 kV
A3a	30 kV a 44 kV
A4	2,3 kV a 25 kV
AS	sistema subterrâneo

Tabela 2. Subgrupos de consumo enrgético para o grupo B.

Subgrupo	Sistema
B1	residencial e residencial baixa renda
B2	rural e cooperativa de eletrificação rural
B3	demais classes
B4	iluminação pública

Essas subdivisões juntamente com os encargos setoriais, tributos e bandeiras tarifárias contribuem para a flutuação do valor da conta de energia repassada aos consumidores.

2.2 Irradiação Solar no Maranhão

Por estar localizado na região nordeste do Brasil, o estado do Maranhão apresenta condiçoes climaticas favoraveis à implatação de sistemas de geração fotovoltaica. A média de irradiação solar para a cidade de São Luís, capital do Maranhão, pode ser vista nas Tabela 3 e Figura 2. Nela pode-se observar os valores por mês, com abril tendo os menores e setembro os maiores de irradiação solar.

2.3 Web Scraping

Em teoria, *web scraping* é a prática de coletar dados por qualquer outro meio que não seja um programa que interaja com uma API, do inglês *Application Programming Interface*, que segundo FOLDOC (1995), a interface pela qual um programa de aplicativo acessa o sistema operacional e outros serviços. Uma API é definida no nível do código-fonte e fornece um nível de abstração entre o aplicativo e o kernel (ou outros utilitários privilegiados) para garantir a portabilidade do código. Isso geralmente é feito

Tabela 3. Irradiação solar diária média $[kWh/m^2.dia]$ para a cidade de São Luís.

Meses	São Luís 1	São Luís 2	Oceano Atlântico
Janeiro	5,16	4,96	5,24
Fevereiro	5,12	4,99	5,16
Março	4,90	4,79	4,92
Abril	4,77	4,65	4,78
Maio	4,76	4,69	4,94
Junho	4,91	4,84	5,11
Julho	5,15	5,08	5,34
Agosto	5,73	5,66	5,73
Setembro	6,16	6,00	6,00
Outubro	5,99	5,74	5,87
Novembro	5,90	5,64	5,88
Dezembro	5,72	5,48	5,67
Média	5,36	5,21	5,39

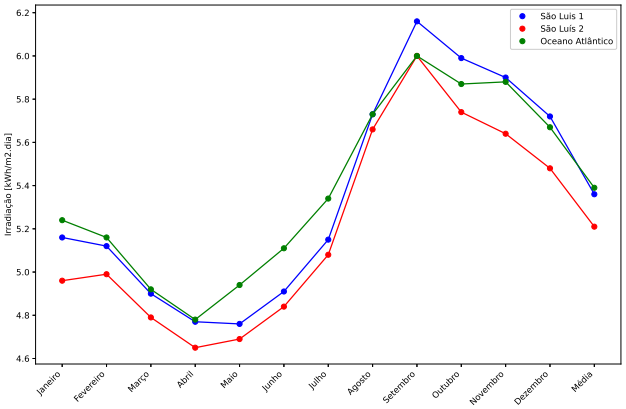

Figura 2. Irradiação solar diária média $[kWh/m^2.dia]$ para a cidade de São Luís - Maranhão. Fonte: Cresesb (2007a)

escrevendo um programa automatizado que consulta um servidor da Web, solicita dados (geralmente na forma de HTML e outros arquivos que compõem páginas da Web) e analisa esses dados para extrair as informações necessárias, Mitchell (2018). Possui utilidade quando necessitamos de um grande volume de dados extraídos recorrentemente, já que realizar todo o processo de extração de forma manual demandaria muito tempo e esforço, o que pode ser resolvido em poucos minutos pelo script automatizado, Mazini and Sato (2019). Aplicações como a citada anteriormente, fornecem e/ou guardam grandes quantidades de dados para analisar. Essa informação, quando analisada individualmente, não é útil sendo necessária uma análise conjunta de todos os dados. Neste contexto, surge a necessidade de criar ferramentas de análise de dados que permitam lidar com esse novo paradigma da informação. Um dos principais meios encontrados para lidar com esta quantidade enorme de informação foi a sua categorização em grupos ou clusters, Nunes (2016).

2.4 K-means

O K-means é uma heurística de agrupamento não hierárquico que busca minimizar a distância dos elementos a um conjunto de k centros dado por $\chi = x_1, x_2, ..., x_k$ de forma iterativa. A distância entre um ponto pi e um conjunto de clusters, dada por $d(p_i, \chi)$, é definida como sendo a

distância do ponto ao centro mais próximo dele. A função a ser minimizada então, é dada por:

$$d(p_i, \chi) = \sum_{i=1}^{n} d(p_i, \chi)^2, \quad (1)$$

O algoritmo depende de um parâmetro (k=número de clusters) definido de forma *ad hoc* pelo usuário. Isto costuma ser um problema, tendo em vista que normalmente não se sabe quantos clusters existem *a priori*. Este algoritmo é extremamente veloz, geralmente convergindo em poucas iterações para uma configuração estável, na qual nenhum elemento está designado para um cluster cujo centro não lhe seja o mais próximo. Um exemplo da execução do algoritmo de K-Means pode ser visto na Figura 3, Linden (2009).

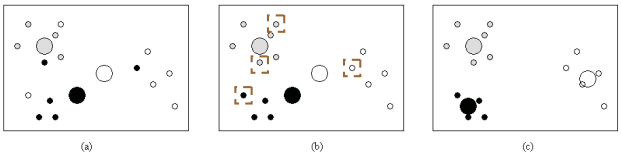

Figura 3. Exemplo de execução do algoritmo de K-Means. (a) Cada elemento foi designado para um dos três grupos aleatoriamente e os centróides (círculos maiores) de cada grupo foram calculados. (b) Os elementos foram designados agora para os grupos cujos centróides lhe estão mais próximos. (c) Os centróides foram recalculados. Os grupos já estão em sua forma final. Caso não estivessem, repetiríamos os passos (b) e (c). Fonte: Linden (2009)

A escolha do número de clusters se torna muito importante para a execussão do algoritmo. Um dos mais tradicionais é o método do *Elbow* ou método do cotovelo. Para Kodinariya and Makwana (2013), a idéia é que comece com $K = 2$ e continue aumentando a cada etapa em 1, calculando seus clusters e o custo que acompanha o treinamento. Em algum valor para K, o custo cai drasticamente e, depois disso, atinge um platô quando você aumenta ainda mais. Este é o valor K que você deseja.

2.5 Árvore de Decisão

Este método é não paramétrico e usado para classificação e regressão. Para Gama (2002), na representação da árvore de decisão cada nó de decisão contem um teste num atributo, cada ramo descendente corresponde a um possível valor deste atributo, cada folha está associada a uma classe e cada percurso na árvore (da raiz à folha) corresponde a uma regra de classificação, tal como a Figura 4.

Um dos seus critérios mais adotados é a entropia, pode ser definida como uma medida da aleatoriedade de uma variável. A entropia da variável nominal X que pode tomar i valores:

$$Entropia(X) = -\sum_{i} p_i * log_2 p_i \quad (2)$$

A entropia tem máximo ($log_2 i$) se $p_i = p_j$ para qualquer $i <> j$ e a entropia$(x) = 0$ se existe um i tal que $p_i = 1$.

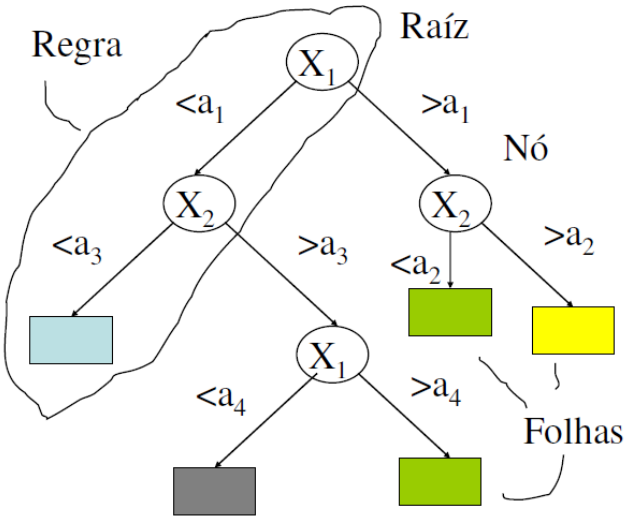

Figura 4. Exemplo de árvore de decisão com seus componentes. Fonte: Gama (2002).

2.6 Métricas

Métrica de agrupamento Para verificar se a escolha dos agrupamentos foi acertiva, usa-se o método da silhueta (*silhouette*), proposto por Rousseeuw (1989). Este método se baseia em um valor adimenssional, entre -1 e 1, quanto mais próximo este valor for de 1, significa que o objeto foi bem classificado no grupo. Caso contrário, quanto mais distante de 1 e próximo de -1, significa que o objeto foi mal classificado.

Para Maciel et al. (2015), a silhueta é um gráfico do cluster C composto por um valor de silhueta $s(i), i = 1, ..., n$, que reflete a qualidade da alocação dos objetos no grupos. Cada objeto (indivíduo) do cluster é representado por i. E para cada objeto i o valor $s(i)$ é calculado:

$$s(i) = \frac{b(i) - a(i)}{max(a(i), b(i)} \quad (3)$$

Onde $a(i)$ é a dissimilaridade média do objeto i em relação a todos os objetos do mesmo grupo C, e $b(i)$ é a dissimilaridade média entre o objeto i em relação a todos os objetos do grupo vizinho mais próximo a ele.

Métricas de classificação Posteriormente, foram obtidas as porcentagens de acurácia.

$$Acuracia = \frac{VP + VN}{VP + VN + FN + FP} \times 100, \quad (4)$$

Onde VP corresponde ao número de verdadeiros positivos, VN os verdadeiros negativos, FP para falsos positivos e FN para falsos negativos. Como técnica de validação cruzada, foi utilizado o método k-fold, com $k = 10$.

3. DATA BASE

Os dados analisados neste trabalho estão disponibilizados no site da ANEEL (2020) para consulta pública, em

formato de tabela referindo-se ao Estado do Maranhão, contendo informações sobre potência instalada, município, subgrupo energetico, classe, modalidade, fonte, quandidades de unidades consumidoras que recebem crédito energético e data da conexão. Possui um total de 2794 consumidores com registro de geração distribuida de fonte solar e potência total de 34.821,50 kW, foram desconsideradas unidades que utilizam outras fontes renováveis.

4. METODOLOGIA

4.1 Aquisição de dados

Primeiramente, fez-se a aquisição dos dados utilizando um web scraping para extração, na linguagem Python. Como base também utilizou-se a biblioteca *Beautiful Soup*, ela realiza a leitura e a extração de dados de textos HTML, permitindo a busca por strings, tags, ids, classes e qualquer outro atributo que possa servir de identificação para um elemento, Richardson (2007). Para um melhor processamento, foi realizada uma verificação de possíveis dados faltosos e exclusão dos mesmos.

4.2 Definição de grupos de consumidores

Afim de definir quais grupos são contidos no data set, foi aplicado o k-means, método não supervisionadao de agrupamento, juntamente com *Elbow Method* (método do cotovelo). O k-means busca agrupar os dados tentando separa-los em n grupos de amostras com igualdade de variação. Existe a necessidade que um numero de clusters seja especificado. Caso não se saiba quantos cluters significativos os dados possuem, então é usado o *Elbow Method*. Este irá testar o k-means para várias quantidades diferentes de cluters e informar qual deles é seu número ótimo.

4.3 Agrupamento de grupos por variáveis mais significativas

Após a clusterização dos grupos, tornou-se necesssária a distinção dos mesmos através das caracteristicas mais significativas. Para isso, o classificador árvore de decisão foi utilizado com a entropia como critério adotado.

Realizada a classificação foram estimadas as métricas de classificação apartir do método de validação cruzada k-fold.

5. RESULTADOS

Após o tratamento dos dados e processo de relacionar as variáveis de interesse, foi realizada uma análise exploratória e quantitativa no conjunto de dados. Uma delas sendo a relação entre o quantitativo de consumidores de determinados subgrupos em relação as suas respectivas classes. Tal relação é ilustrada pela Figura 5.

Claramente a quantidade de um subgrupo se destaca em relação aos demais. A fim de mensurar de forma adequada tal diferença, na Figura 6 ilustra as mesmas informações da figura anterior, porém com a quantidade de consumidores em escala logarítmica.

Foi feita ainda a análise da quantidade de consumidores e seus respectivos subgrupos, sem relação com demais variáveis. Esse quantitativo é ilustrado pela Figura 8.

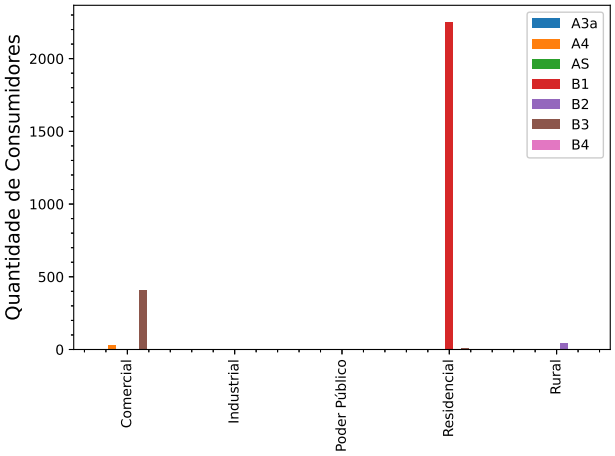

Figura 5. Quantidade de consumidores por classes em função de subgrupos. Fonte: Autores.

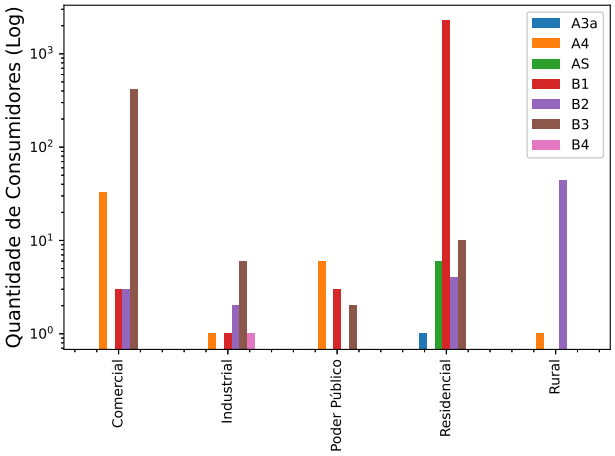

Figura 6. Quantidade de consumidores por classes em função de subgrupos, em escala logarítmica. Fonte: Autores.

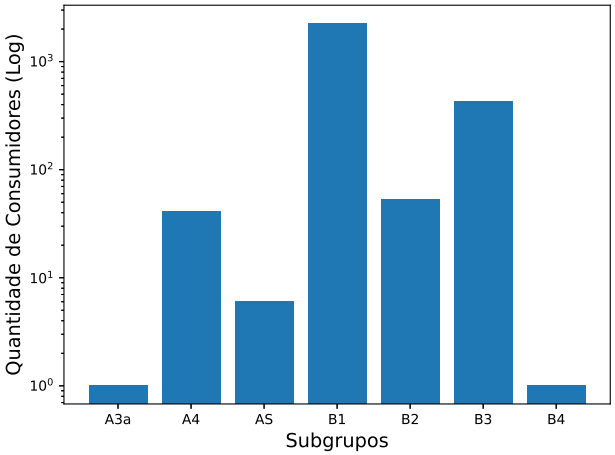

Figura 7. Quantidade de consumidores por subgrupo, em escala logarítmica. Fonte: Autores.

Novamente, o subgrupo B1 se destaca. De forma que a Figura 7 ilustra a quantidade de consumidores usando a escala logarítmica.

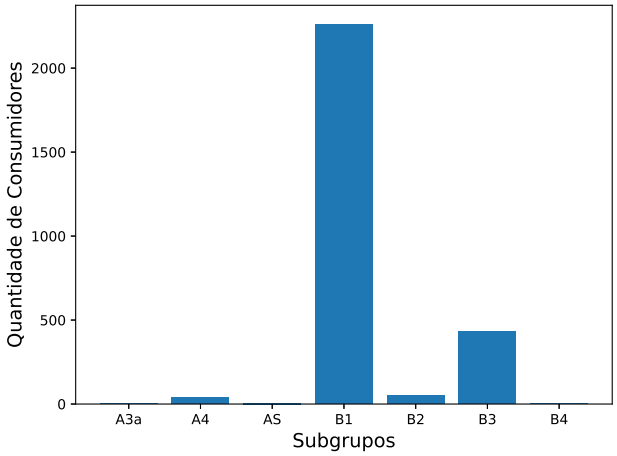

Figura 8. Quantidade de consumidores por subgrupo. Fonte: Autores.

Foi então aplicado o método do Elbow, que estimou 5 *clusters* como quantidade ideal. A partir disso, executou-se o k-means com $k = 5$. Os rótulos obtidos foram incorporados aos dados tabulados.

O coeficiente de silhueta, baseado em distância euclediana,foi utilizado como métrica de validação do k-means. Sabendo que, o pior resultado para tal métrica é próximo de -1 e o melhor dos resultados é aproximado de 1, para estes dados foi obtido um coeficiente igual a $0,87$.

Em seguida, os dados foram classificados empregando a árvore de decisão. Para tal, a métrica de validação cruzada adotada foi o k-fold, tendo $k = 10$. Sendo assim, foi obtida a acurácia média de 99,89%.

A partir de toda metodologia feita, foi possível identificar as variáveis mais significativas entre os dados. Que por sua vez contribuem para separação dos grupos de consumidores. A árvore de decisão obtida é ilustrada pela Figura 9.

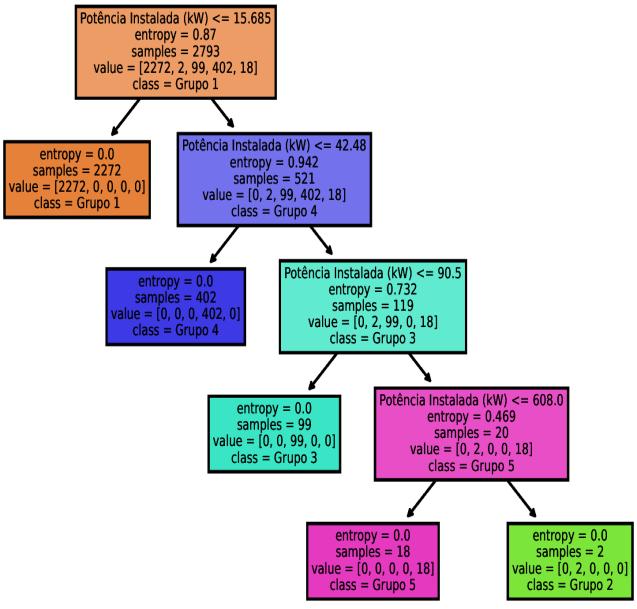

Figura 9. Árvore de decisão resultante. Fonte: Autores.

6. DISCUSSÃO

A partir da análise quantitativa foi possível verificar que no estado do Maranhão não há consumidores dos subgrupos A1, A2, e A3. Ou seja, não há um consumo de altas tensões no estado proveniente de energia fotovoltáica.

Um outro traço observado foi que a grande maioria dos consumidores maranhenses de energia fotovoltaíca são predominantemente do subgrupo B1, cerca de 80.95% do total de consumidores. Logo, a maioria do consumo dessa fonte de energia é para fins de uso residencial.

Tendo aplicado o k-Means e o validando com o coeficiente de silhueta (igual 0.87), percebe-se que a determinação dos grupo de consumidores foi feita satisfatoriamente.

Diante dos resultados obtidos pela árvore decisão, pode ser visto que a característica que divide de melhor forma os grupos é a potência instalada. Apesar desta ser muitas vezes específica para cada consumidor, pode-se relacionar com os subgrupos de consumo energético em vigência. Logo, a potência instalada corresponde a faixa de tensão requerida para a classificação que se encontra determinado consumidor em um subgrupo.

Baseado nos resultados obtidos, tanto em clusterização quanto em classificação, essa metodologia pode ser estendida. De forma que seja possivel identificar características e padrões de grupos de consumidores de demais estados brasileiros, ou mesmo mensurar um panorama de perfis mais específicos de consumidores de energia fotovoltáica a nível de Brasil.

7. CONCLUSÃO

Neste trabalho, buscou-se implementar técnicas de extração e algoritmos de aprendizado de máquina para analisar dados referentes a consumidores que já possuem geração de energia de fonte solar. Com a aplicação da metodologia proposta, foi possível observar características importantes sobre a disposição de consumidores com geração distribuída de fonte fotovoltaica. Foi possivel visualizar um panorama mais específico sobre os consumidores maranhenses e que potncia instalada é a principal caracterítica observada. Os resultados obtidos da implementação foram considerados aceitaveis dentro das métricas de validação usadas. Por se tratar de um estudo em andamento, para trabalhos futuros pensa-se em ampliar a janela de análise, buscando comparar os perfis encontrados com os de outros estados. Consequentemente contribuindo para os estudos na área de energia renovável.

AGRADECIMENTOS

Os autores agradecem ao Programa de Bolsas Institucionais do Instituto Federal de Educação, Ciência e Tecnologia do Maranhão - Campus Monte Castelo pelo apoio financeiro.

REFERÊNCIAS

Aneel (2020). Agência nacional de energia elétrica - retrospectiva 2019. http://www.aneel.gov.br/documents/656877/15495819/Retrospectiva+ANEEL+

-+2019/73fd2b23-c540-8548-f7bd-554702f74133?version=1.0. Online; Acesso em: 07-Setembro-2020.

ANEEL (2020). Unidades consumidoras com geração distribuída da unidade da federação: Ma. http://www2.aneel.gov.br/scg/gd/gd_estadual_detalhe.asp?uf=MA&pagina=1. Online; Acesso em: 30-Abril-2020.

Campos, E.T.d. (2017). Estudo de viabilidade da implantação da microgeração de energia elétrica por sistema de painéis fotovoltaicos em residência de taubaté/sp.

Cresesb (2007a). Base de dados de radiação solar incidente (irradiação solar). http://www.cresesb.cepel.br/index.php?section=sundata. Online; Acesso em: 15-Junho-2020.

Cresesb (2007b). Energia solar fotovoltaica. http://www.cresesb.cepel.br/index.php?section=com_content&cid=tutorial_solar. Online; Acesso em: 10-Junho-2020.

Dranka, G.G., de Lima, J.D., da Costa Bonotto, R., and Machado, R.H.S. (2018). Economic and risk analysis of small-scale pv systems in brazil. *IEEE Latin America Transactions*, 16(10), 2530–2538.

Energética, E.E.D.P. (2019). Balanço energético nacional: Relatório síntese, ano base 2018. *Rio de Janeiro: Ministério de Minas e Energia*.

Energia, C. (2020). Geração distribuída atinge 201,7 mil conexões no brasil. http://GeraÃğÃčodistribuÃ■daatinge201,7milconexÃµesnoBrasil. Online; Acesso em: 07-Setembro-2020.

Ferreira, L. (2020). Você sabe a diferença entre on-grid e off-grid? https://enetec.unb.br/blog/voce-sabe-a-diferenca-entre-on-grid-e-off-grid. Online; Acesso em: 19-Maio-2020.

FOLDOC (1995). Application program interface. http://foldoc.org/Application+Program+Interface. Online; Acesso em: 07-Setembro-2020.

Gama, J. (2002). Árvores de decisão. *Palestra ministrada no Núcleo da Ciência de Computação da Universidade do Porto, Porto*.

Hinrichs, R., KLEINBACH, M., and REIS, L. (2010). Energia e meio ambiente. 4ª. *LF Flávio Maron Vichi, Trad.) Cengage Learning, São Paulo*.

Kodinariya, T.M. and Makwana, P.R. (2013). Review on determining number of cluster in k-means clustering. *International Journal*, 1(6), 90–95.

Lau, L. (2020). Entenda as modalidades de tarifa de energia elétrica. https://sharenergy.com.br/entenda-as-modalidades\-de-tarifa-de-energia-eletrica/. Online; Acesso em: 10-Junho-2020.

Linden, R. (2009). Técnicas de agrupamento. *Revista de Sistemas de Informação da FSMA*, 4(4), 18–36.

Maciel, A., Vinhas, L., and Câmara, G. (2015). Algoritmos de clustering para separação de culturas agrícolas e tipos de uso e cobertura da terra utilizando dados de sensoriamento remoto. *Simpósio Brasileiro de Sensoriamento Remoto (SBSR)*, 17, 4620–4627.

Mazini, D.N. and Sato, R.C. (2019). Extração de dados financeiros com um web scraper: um estudo sobre a rentabilidade dos dividendos.

Mitchell, R. (2018). *Web scraping with Python: Collecting more data from the modern web*. "O'Reilly Media, Inc.".

Nunes, D.H.F. (2016). *Um breve estudo sobre o algoritmo K-means*. Master's thesis.

PCE (2015). Cenários para a matriz elétrica 2050: Aporte ao debate energético nacional e ao planejamento participativo de longo prazo – pce brasil 2050. http://www.nextrans.com.br/wp-content/uploads/2015/03/Energia-Cenarios-2050.pdf. Online; Acesso em: 10-Fevereiro-2020.

Pontolio, L.S. (2015). Plataforma de extração e recuperação de dados na web no contexto de big data.

Richardson, L. (2007). Beautiful soup documentation. *April*.

Rocha, B.C., Dias, D.H.N., Borba, B.S.M.C., and Ferreira, P.F.V. (2018). Economic evaluation of photovoltaic microgenaration in the southeast region of brazil. In *2018 Simposio Brasileiro de Sistemas Eletricos (SBSE)*, 1–6. IEEE.

Rousseeuw, J. (1989). A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational Application Math*.

Solar, F. (2020). Focus solar: Energia e serviços especializados. http://focussolar.eco.br/. Online; Acesso em: 15-Junho-2020.