



Ferramenta de Machine Learning para identificação de perfis de consumidores de energia do grupo A utilizando clusterização

Tema: Comunicação e Relacionamento com o cliente

Autores: Carlos José Alves Santos (Way2); Pedro Ivo Anastácio Pellegrini (Sinapsis); Ricardo de Araújo Elias (Celesc)

Co-Autores: -

Empresa: Way2 Technology; Sinapsis Energia; Celesc Distribuição

Resumo

O artigo apresenta uma solução para comercializadoras varejistas de energia, utilizando técnicas de clusterização – técnica de aprendizado de máquina não supervisionado - para segmentar clientes e identificar padrões de consumo energético. A solução proposta permite associar cada cliente a um conjunto de produtos e serviços alinhados aos seus comportamentos de consumo e características. A análise foi aplicada aos clientes do grupo A da Celesc, na área de concessão de Santa Catarina.

É apresentado o contexto do mercado livre brasileiro, junto com os resultados de pesquisa de campo realizada com os clientes da Celesc, avaliando a aceitação de produtos relacionados a eficiência energética, geração distribuída e contratos com características personalizadas.

A análise utilizou o K-means como técnica de clusterização, uma vez que este resultou nos melhores valores de Silhouette. Os resultados possibilitaram a identificação de clusters com características específicas e clientes distintos, o que permite com que a comercializadora possa utilizar os resultados do algoritmo para estabelecer planos de priorização em prospecções de clientes, além de embasar abordagens com diferenciados níveis de risco associado.

1. Introdução

As recentes regulamentações da comercialização do mercado de energia trouxeram grandes mudanças para o Setor Elétrico Brasileiro (SEB) e, consequentemente, para as empresas de energia elétrica, que passaram a lidar com um maior volume de adesão ao Ambiente de Contratação Livre (ACL).

Somado a este ponto, com a modernização do SEB e o aumento da relevância dos dados, há uma tendência insurgente do protagonismo que comercializadoras e consumidores considerados varejistas passam a assumir. Este protagonismo se alinha com a necessidade de maior inteligência na análise dos dados e com a oportunidade da geração de novos produtos e serviços que atendam os diferentes perfis de consumo. Para tornar o consumidor protagonista dentro dos produtos que o são fornecidos, técnicas de análise de dados, envolvendo Machine Learning, são cada vez mais exploradas para personalizar o atendimento aos clientes. O atual momento do mercado de contratação de energia abre margem para novas formas de utilizar estes dados.

Neste contexto, o presente trabalho consiste em resultados e insights obtidos com as saídas parciais do P&D “Matchmaking de produtos de energia e ofertas complementares para clientes varejistas” (PD-05697-0523) no qual foram utilizadas técnicas de clusterização para identificar padrões de comportamento dos clientes da Celesc e gerar oportunidades em sua segmentação. A Figura 1 introduz o conceito do projeto, onde serão utilizadas informações dos clientes, para que, a partir da utilização do algoritmo de clusterização e de um algoritmo de matchmaking, sejam sugeridos produtos para os clientes da comercializadora.

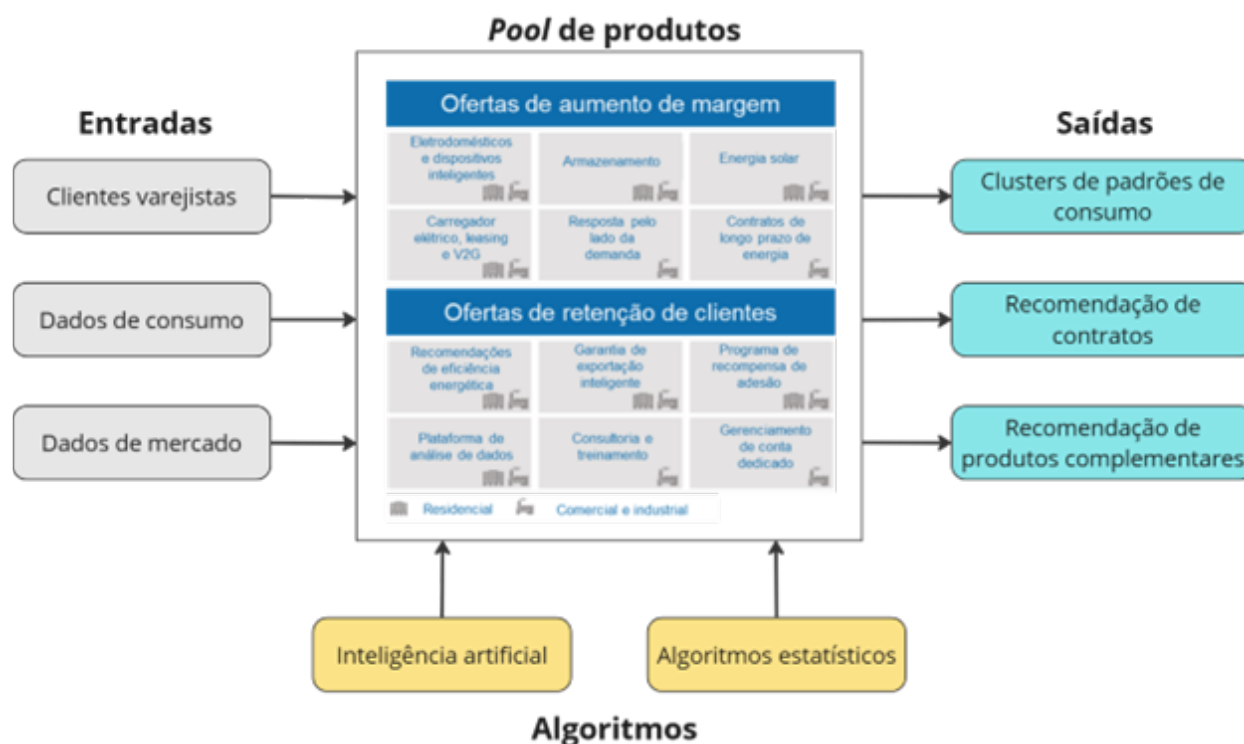


Figura 1 - Conceito do P&D “Matchmaking de produtos de energia e ofertas complementares para clientes varejistas”

Ao longo deste artigo, é apresentado o embasamento teórico para a escolha do algoritmo de clusterização, seguido dos resultados obtidos por meio da base de dados de clientes faturados no Grupo A da Celesc. Por fim, é realizado um paralelo entre os resultados e os produtos levantados, trazendo um racional sobre a correlação de cliente – cluster.

1.1 Mercado livre e produtos

Em 1995, com a publicação da Lei federal nº 9.074 (BRASIL, 1995), foram criados os consumidores livres, os quais podiam escolher o próprio fornecedor de energia elétrica. Ainda era um universo limitado e representava a concepção inicial da liberdade de escolha do fornecedor de energia ao consumidor, criando um mercado competitivo. Até o final da década de 1990, outras leis surgiram – Lei federal nº 9.427 de 1996 (BRASIL, 1996) e Lei federal nº 9.648 de 1998 (BRASIL, 1998) – e ampliaram o poder de escolha para os consumidores com demanda contratada entre 500 kW e 3.000 kW, os quais foram denominados consumidores especiais, reforçando a intenção inicial.

A partir de 2004, foram segregados dois mercados: o Ambiente de Contratação Regulado (ACR) e o Ambiente de Contratação Livre (ACL). No ACR estão os consumidores cujo portfólio de fornecedor de energia é aquele que a distribuidora comprou, a partir de leilões regulados pelo poder concedente. No ACL estão os clientes que podem escolher diretamente o fornecedor de energia na Câmara de Comercialização de Energia Elétrica e foram divididos em duas classes: consumidores especiais e consumidores livres.

A partir de 2018, iniciaram-se as regulamentações para igualar as condições entre consumidores livres e consumidores especiais, culminando na Portaria MME 465/2019 (MME, 2019), a qual determinou que todos os consumidores com potência acima de 500 kW eram consumidores livres, encerrando a divisão em duas classes no formato citado anteriormente. Essa abertura gradual do mercado foi parte de uma transição para inclusão de unidades consumidoras de energia elétrica conectadas em alta tensão e com demanda abaixo de 500 kW, o que foi regulamentado pela Portaria MME nº 50/2022 (MME, 2022) em um modelo chamado de varejista.

Com esse movimento, abre-se a possibilidade para mais de 100 mil unidades consumidoras em todo o território nacional aderirem ao mercado livre de energia elétrica, e no limite será uma possibilidade para todas as unidades consumidoras do Brasil, que são mais de 80 milhões. Essas unidades consumidoras ficarão sob responsabilidade do comercializador varejista.

A figura do comercializador varejista foi introduzida em 2011 pela Câmara de Comercialização de Energia Elétrica (CCEE), conforme descrito na Nota Técnica nº 55/2012 (ANEEL, 2012). Esse marco deu início às discussões para a evolução normativa que viabilizasse a implementação dessa figura. Um dos principais objetivos era consolidar as operações de consumidores de energia elétrica em um único agente, simplificando processos e reduzindo custos operacionais.

O modelo foi pensado, sobretudo, para atender consumidores especiais, cujo consumo de energia, embora significativo em termos absolutos, não justificava os custos operacionais envolvidos na gestão individualizada. A crescente demanda por sistemas e atendimento por parte da CCEE, impulsionada pelo aumento no número de consumidores especiais, evidenciou a necessidade de uma solução mais eficiente. Nesse contexto, o comercializador varejista foi concebido para assumir o conjunto de obrigações financeiras e operacionais desses consumidores, reduzindo o volume de interações e operações gerenciadas diretamente pela CCEE.

Apesar das vantagens propostas, a adesão inicial ao modelo foi limitada. Somente com a recente obrigatoriedade de representação varejista para mais de 100 mil unidades consumidoras do grupo de alta tensão, a partir de janeiro de 2024, o modelo começou a ganhar maior relevância. Essa nova dinâmica trouxe um papel mais central ao comercializador varejista na estruturação e funcionamento do mercado.

No contexto atual, em que o mercado de energia elétrica se orienta pelos pilares da descarbonização, descentralização e digitalização, o papel dos comercializadores varejistas evolui para atender às novas demandas competitivas. Essas empresas buscam desenvolver e oferecer soluções inovadoras que alinhem suas operações a essas três frentes, logo, como um primeiro passo, com base em *benchmarkings* nacionais e internacionais, esse trabalho propõe produtos ligados a essa temática e divididos em duas partes: retenção de clientes e aumento de margem.

Do lado de produtos que resultam em aumento de margem, foram selecionados serviço de análise e troca de equipamentos por critério de eficiência energética (i), armazenamento de energia (ii), instalação de energia solar (iii), instalação de carregador elétrico (iv), instalação de banco de capacitores (v) e contratos de energia (vi). Do lado de produtos de retenção de clientes, foram selecionados: recomendações de eficiência energética (a), oferta de certificados de energia renovável (b), programas de recompensa (c), plataforma de análise de dados (d), treinamentos e consultoria (e), gerenciamento dedicado de conta (f).

De forma a identificar a aceitação de cada produto no mercado potencial, foi realizada uma pesquisa qualitativa. Na pesquisa, foi descartado o produto de banco de capacitores, já que não é um equipamento amplamente conhecido. Chegou-se à conclusão de que o banco de capacitores deveria ser apresentado ao cliente pelo benefício financeiro de evitar a multa por excesso de reativos, logo se torna uma decisão objetiva do cliente em evitar ou não a multa. O produto (vi) gerou dúvidas nos entrevistados, apesar da contextualização do que é um contrato de energia. Sem compreender exatamente o processo, o cliente apresenta tendência de escolher um contrato com desconto garantido em relação à energia da distribuidora.

Já os produtos (ii) e (iv) são mais tangíveis, contudo, as tecnologias de armazenamento ainda são de alto custo e se aplicam em uma condição específica: alto custo de déficit energético e rede elétrica fraca, portanto, é de aplicação restrita atualmente. Dada a baixa adesão ao veículo elétrico, os carregadores elétricos são uma opção para empresas que possuem frota ou que queiram testar essa solução de mobilidade. A maior afinidade ficou por conta dos produtos (i) e (iii). O serviço de análise e implantação de eficiência energética existe há muitos anos e desperta o interesse de como economizar, especialmente se o valor a ser investido nos equipamentos possa ser pago com a economia. O produto de instalação de energia solar cativa o cliente devido a sustentabilidade e sensação de autonomia sobre o próprio consumo.

Quanto aos produtos de retenção de clientes, as ofertas (d), (e) e (f) não tiveram aceitação significativa. A pesquisa indica que os clientes que se interessam já dispõe de alguma forma de analisar esses dados, inclusive com gerenciamento de conta. Sem delinear exatamente o que seriam os treinamentos e consultorias, não há grande interesse, entretanto, é interessante ressaltar que o produto de mercado livre acaba sendo ensinado pelas comercializadoras que desejam vender a solução. A plataforma de recompensa (c) faz sentido a depender da relação de descontos que se pode ter. Do ponto de vista dos entrevistados, se houver algo relacionado a materiais de energia elétrica, os preços devem ser competitivos, além disso, algumas empresas pontuaram que evitam realizar compras pela internet, o que reduz o apreço do produto. Por fim, os certificados de energia renovável chamam atenção por melhorar a estratégia de marketing da empresa, o que se assenta no pilar de descarbonização.

Destaca-se que a pesquisa foi realizada com clientes do grupo A da região de concessão da Celesc, sendo cinco unidades consumidoras da classe industrial, 10 unidades consumidoras da classe comércio/serviço. Dessas unidades, cinco já estavam em processo de migração para o mercado livre.

Ao mesmo tempo, propõe tratamento de dados de clientes do mercado ACR a partir de um algoritmo de clusterização, de forma a identificar padrões descorrelacionados com o viés humano. Em um segundo passo, a metodologia aplica uma pesquisa qualitativa para identificar a aderência dos produtos levantados no mercado consumidor potencial. Finalmente, como último passo, se propõe a desenvolver um algoritmo de *matchmaking* que possa relacionar produtos e clientes.

Uma alternativa para que o comercializador varejista consiga proporcionar um melhor atendimento a clientes, os quais terão menor conhecimento de mercado e menor ticket médio, seria o uso de tecnologia para, em um primeiro momento, agregar consumidores de perfil semelhante e, em um segundo momento, complementar a oferta do produto energia com outros serviços ou propostas que criem valor ao usuário final.

1.2 Técnicas de clusterização

A segmentação dos clientes é utilizada amplamente pelo mercado para identificar perfis de clientes mais propensos a um determinado tipo de produto. O objetivo dessa abordagem é identificar categorizações relevantes que permitam o desenvolvimento mais adequado de estratégias, atendendo a necessidades e desejos específicos de qualquer que seja o segmento que se beneficia da utilização destas ferramentas (CHÉRON, KLEINSCHMIDT, 1985).

Paralelamente, no setor elétrico, estas técnicas podem ser utilizadas para identificar padrões de comportamento no consumo energético dos clientes, além de efetuar o cruzamento com outras informações dos clientes, como índices de qualidade do fornecimento da própria distribuidora (DIC e FIC), bem como informações históricas de pagamento do próprio cliente. Todos estes dados corroboram para que a comercializadora construa um planejamento eficaz de fornecimento de produtos personalizados para estes clientes, aumentando sua satisfação com o atendimento fornecido e possibilitando a identificação de oportunidades pelo lado da concessionária.

Para efetuar esta segmentação, foi utilizada a abordagem do Algoritmo de Clusterização. Esta é uma técnica de aprendizado de máquina utilizada para agrupar clientes de energia de acordo com seus padrões

de consumo. Ele permite entender a base de clientes e identificar grupos com características semelhantes dentro da base de dados utilizada. Um quadro ilustrativo é exibido na Figura 2, indicando principais diferenças de comportamento entre as técnicas de clusterização mais difundidas, e é possível observar que, independentemente da abordagem, um agrupamento sempre é formado. A diferença reside em como este agrupamento surge, podendo ser definido por proximidade entre outros pontos na mesma região (modelo de densidade), respeitando uma quantidade pré-estabelecida de agrupamentos (modelo de centróides) ou simplesmente realizar uma segmentação hierárquica na base (modelo hierárquico).

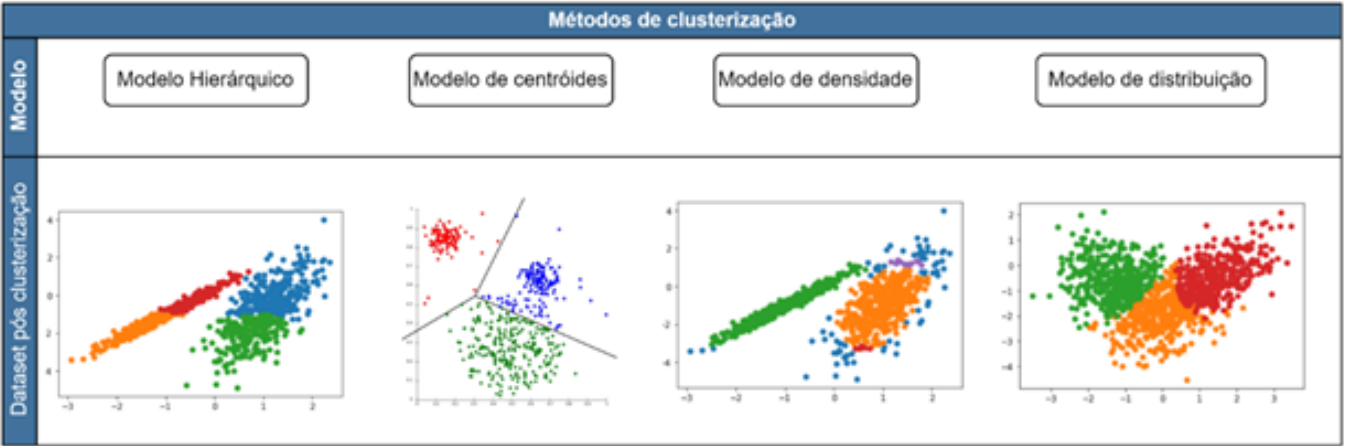


Figura 2 - Figuras ilustrativas do funcionamento de algoritmos de clusterização (FANG, 2024; TOLBA, 2020) A clusterização é realizada com base em algoritmos matemáticos que analisam dados históricos de consumo de energia dos clientes e criam grupos de clientes com perfis semelhantes. Estes grupos de clientes podem ser usados para segmentação de mercado e personalização de estratégias de gestão de energia, bem como dados de entrada para direcionar oferta de produtos adequados, incluindo produtos de energia (contratos) personalizados, comunicação segmentada e programas de eficiência energética (SAXENA, 2017).

Os agrupamentos finais são influenciados majoritariamente por dois fatores. O primeiro deles é qual técnica de clusterização será utilizada para efetuar o agrupamento. Ao longo deste artigo, três técnicas serão citadas, mas apenas os resultados da técnica que apresentou melhores resultados serão explorados no último capítulo. O segundo deles é a refinamento de parâmetros e hiper parâmetros da configuração clusterização, isto é, variáveis que podem ser modificadas para determinar quantidade de agrupamentos que serão gerados e a distância que será considerada entre os pontos para configurá-los ou não como um cluster. Em ambos os casos, foram realizados testes que permitiram identificar a abordagem com maior assertividade.

2. Desenvolvimento

Foram exploradas duas bases de dados para gerar os resultados obtidos com o método proposto no presente artigo. Ambas apresentam informações exclusivas que fornecem uma nova gama de interpretações no resultado do algoritmo de clusterização.

A primeira foi a base de dados de todos os clientes faturados no grupo A da Celesc, que fornece informações relevantes sobre consumo mensal dos clientes, além de trazer informações sobre demanda con-

tratada, energia reativa e informações de faturamento do cliente. Foram utilizados dados de 72 meses desta base, que foi requisitada dos próprios sistemas da distribuidora. Esta base foi devidamente anonimizada, em conformidade com as cláusulas estabelecidas pela Lei Geral de Proteção dos Dados (LGPD).

A **segunda base de dados** foi a Base de Dados Geográfica da Distribuidora (BDGD), que além das informações acerca do próprio cliente, como o a Classificação Nacional das Atividades Econômicas (CNAE), consumo (energia ativa e reativa) e demanda contratada, traz informações sobre tipo de curva de carga do cliente e seus indicadores DIC (Duração de Interrupção Individual por Unidade Consumidora) e FIC (Frequência de Interrupção Individual por Unidade Consumidora). Estes dados foram requisitados dos próprios dados abertos da ANEEL (ANEEL, 2024)

Como mencionado, o algoritmo de clusterização evidencia a proximidade matemática dos clientes, sem necessariamente aplicar regra de negócio. Esta abordagem possibilita a identificação de padrões ocultos no comportamento dos clientes, e principalmente, possibilita a sua segmentação e um maior entendimento sobre o potencial cliente. Por conta disso, os resultados que serão exibidos aqui têm uma característica exploratória que pode ser utilizada por comercializadoras para identificar oportunidades dentro de um grupo de clientes que estão sendo prospectados.

Preparação pré-clusterização

A principal etapa que antecede a execução da clusterização é a limpeza e tratamento dos dados utilizados. Foram identificadas variáveis redundantes como, por exemplo, data de pagamento de fatura e data de vencimento, que foram excluídas e substituídas por apenas uma coluna com a diferença entre estas duas datas. Além disso, foram utilizados processos matemáticos que identificavam correlações altas entre as colunas. Uma alta correlação entre duas colunas implica que, manter ambas na execução da clusterização pode não ser o ideal, uma vez que elas acabam sendo redundantes para o algoritmo.

Este, inclusive, é um desafio comum ao utilizar técnicas de clusterização em dados energéticos: a sua alta correlação. Isso ocorre pois os clientes sempre seguem determinados padrões de consumo em conjunto. Por exemplo, um cliente que apresenta um valor alto de consumo faturado apresentará um alto valor pago na fatura de energia, ou então um cliente que possui um valor pequeno de consumo energético no horário de ponta apresentará uma menor demanda contratada de energia. O grande valor do algoritmo é cruzar as informações categóricas dos clientes com estes dados, proporcionando segmentações que não seriam tão evidentes para especialistas.

A Figura 3 destaca os processos realizados para efetuar a clusterização.



Figura 3 - Fluxo técnico implementado da clusterização

A Figura 4 mostra a matriz de correlação para a base de dados 1, que é preenchida com valores de 1 a -1. Valores positivos indicam uma maior correlação entre os dados. Valores negativos indicam menor correlação entre os dados. É possível observar que a maior parte dos valores do banco de dados possui valores positivos altamente correlacionados. Esta identificação possibilita selecionar quais dados faz mais sentido considerar na entrada do algoritmo, evitando inserção de dados redundantes que fariam a eficiência do algoritmo diminuir.

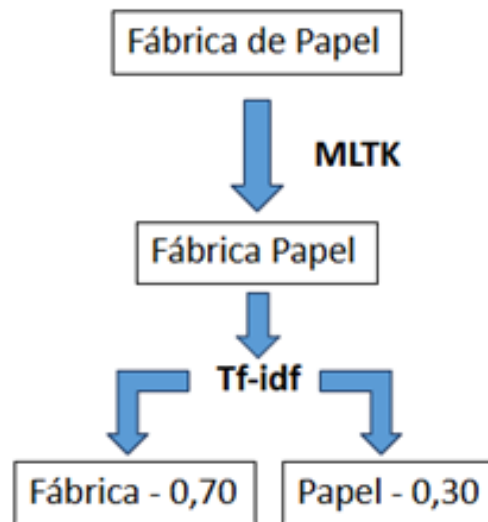


Figura 5 - Funcionamento da tokenização aplicada às colunas não numéricas das bases de dados

Por fim, é necessário identificar a quantidade ideal de clusters que o algoritmo vai gerar. Para chegar a esta decisão, é necessário a junção de dois fatores. O primeiro deles, é escolher um valor que esteja associado à regra de negócio que se busca analisar. No contexto da pesquisa, a utilização dos resultados se dará para efetuar associações ou buscar *insights* para produtos de comercializadoras varejistas. O outro fator que é levado em consideração é a análise de métricas matemáticas que evidenciam qual o melhor cluster formado. Na Figura 6 é apresentado o gráfico gerado com o método de Elbow, onde a quantidade ideal de clusters se encontra na convergência da curva, quando a inclinação é menor, assim como a distância entre os pontos.

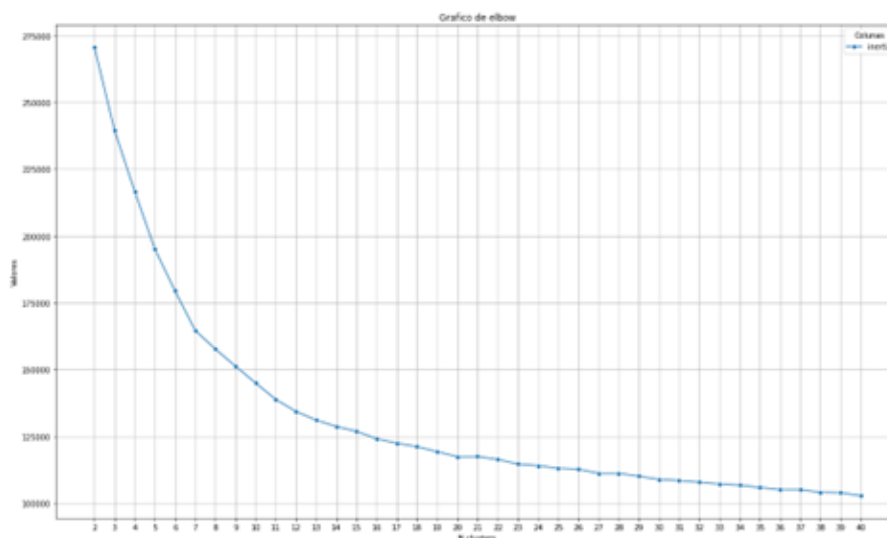


Figura 6 - Gráfico de Elbow para clusterização proposta

Pode-se notar pela Figura 3 que o algoritmo selecionado foi o k-means. Esta decisão foi tomada justamente após serem analisadas métricas de qualidade da clusterização para todos os algoritmos apresentados na seção anterior. Foram utilizadas as métricas de Calinski-Krabsz e o Silhouette Score (SHAHAPURE, NICHOLAS, 2020), que calculam a qualidade do agrupamento formado e o quanto cada agrupamento é distinto entre si. Um ponto relevante e que merece atenção, é que por conta do grande volume de dados das bases, não foi possível utilizar o modelo hierárquico.

Resultados da execução do algoritmo: case de interpretação dos resultados

Após a execução do algoritmo, foram criados sete agrupamentos de clientes do grupo A da CELESC. As imagens a seguir exploram análises efetuadas para cada um dos clusters, evidenciando características de cada um deles. Todos os resultados exibidos na Figura 7 são os valores médios obtidos para cada um dos clusters, e fornecem informações relevantes para caracterizar e destacar cada um dos clusters. Vale ressaltar que analisar valores médios de variáveis em cada um dos clusters é somente um dos vieses que podem ser utilizados para analisar os resultados e tentar extrair informações, conforme será discutido posteriormente.

Na Figura 7 (a) é possível checar que o algoritmo alocou clientes que tem tendência de atrasar o pagamento das faturas de energia no cluster 5, mesmo que nos outros agrupamentos a tendência seja um pagamento adiantado da fatura; Na Figura 7 (b) observa-se uma tendência de agrupamento de clientes com alta demanda faturada estarem alocados nos clusters 2 e 4, enquanto os clusters restantes apresentaram uma demanda faturada média no mesmo nível. A Figura 7 (c) e a Figura 7 (d) mostra a média dos indicadores DIC e FIC dos clientes de cada cluster, mostrando uma tendência de maiores valores no cluster 5, em ambos os casos.

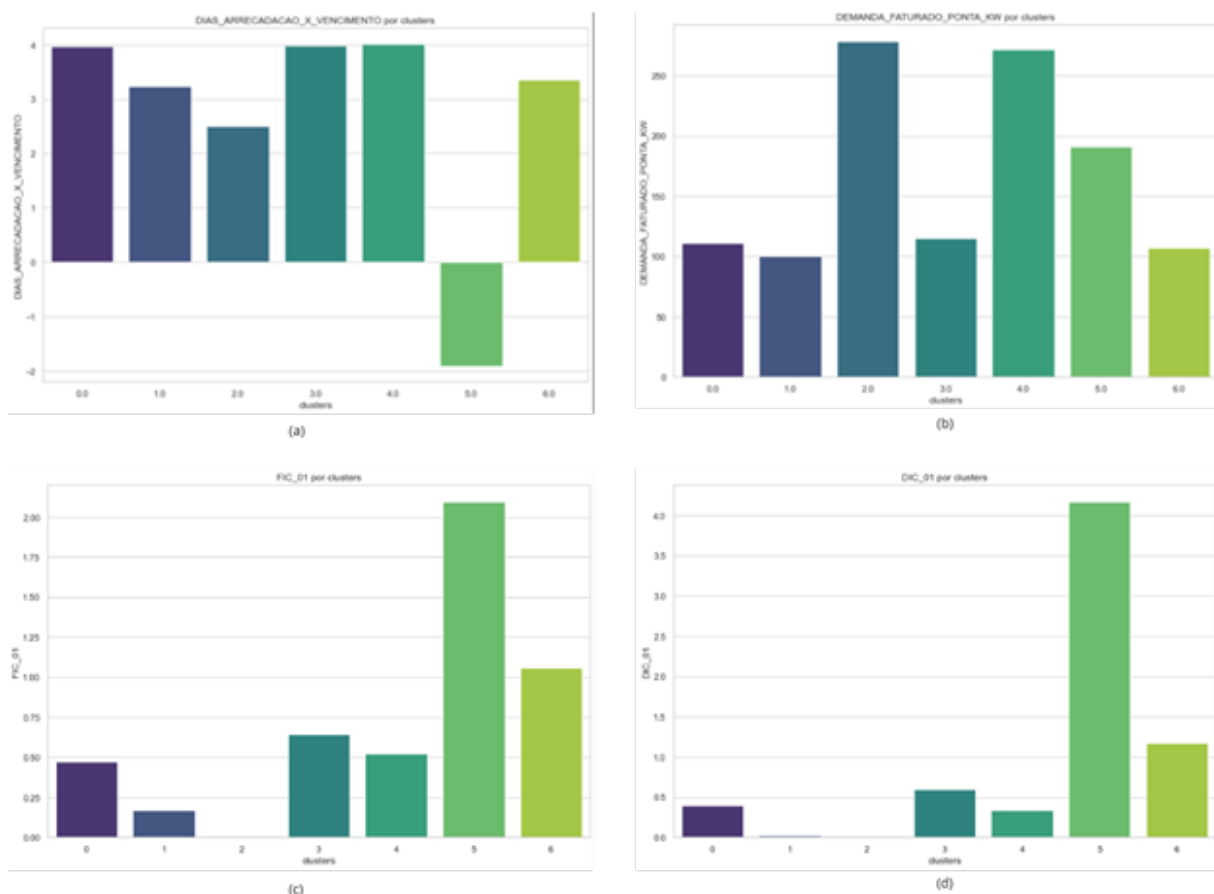


Figura 7 - Resultados gráficos selecionados da clusterização das bases de dados

Na Figura 8, é feita uma análise utilizando o fator de carga das unidades consumidoras, e para cada um dos clusters é obtido valor médio e mediana, além de trazer os valores dos quartis. Os clusters 2 e 4 se destacam com os maiores fatores de carga médios, enquanto o cluster 6 apresenta os menores valores dentre todos os agrupamentos. Do ponto de vista de um produto básico de migração para o ACL, quanto maior este valor, relativamente menor é o benefício financeiro em termos de acesso à energia incentivada frente ao valor total da fatura. Por outro lado, quanto maior este fator, mais organizada e otimizada é a UC, o que pode denotar um perfil de consumo mais estruturado.

clusters	count	mean	std	min	25%	50%	75%	max
0	156858	0,288105	0,303785	0	0,167691	0,25299	0,38334	66,8183
1	166503	0,317212	0,175736	0	0,192461	0,28456	0,417272	10,38795
2	79753	0,371577	0,355821	0	0,223821	0,351352	0,509052	82,35873
3	83821	0,298971	0,224423	0	0,179364	0,265282	0,3874	41,81601
4	66599	0,385876	4,392553	0	0,225608	0,358275	0,504894	1131,277
5	13129	0,322485	0,179372	0	0,190717	0,29573	0,436001	2,481658
6	76414	0,28349	0,188607	0	0,171164	0,247206	0,363353	22,44624

Figura 8 - Descrição da coluna "Fator de Carga" por cluster

Na Figura 9 é possível observar a mesma classe de informações para a participação da ponta na carga total das unidades consumidoras. Os maiores valores estão associados ao cluster 1 e 6, enquanto o cluster 4 apresentou os menores valores registrados. Para o produto migração, quanto maior a participação da ponta no total da carga do cliente, maior é a atratividade de acesso à energia incentivada, e, consequentemente, mais viável se torna a migração. Em contrapartida, um alto valor desta variável pode estar associado a uma maior desorganização da UC.

clusters	count	mean	std	min	25%	50%	75%
0	156082	5,93%	0,04691	0	1,74%	5,70%	8,99%
1	166413	6,89%	0,046384	0	3,15%	7,08%	9,78%
2	79607	5,90%	0,049736	0	1,23%	5,59%	9,45%
3	83767	5,90%	0,047215	0	1,61%	5,49%	9,18%
4	66297	5,29%	0,047517	0	0,96%	4,71%	8,76%
5	13102	5,63%	0,045358	0	1,45%	5,06%	8,87%
6	76378	6,24%	0,047477	0	2,10%	5,92%	9,33%

Figura 9 - Ponta/carga (%) por cluster

Na Figura 10 e na Figura 11 são apresentados valores monetários associados aos clusters, sendo que na primeira foi calculado o valor pago pela energia de cada um dos clientes cativos analisados, enquanto na segunda foram apenas refletidos os valores de fatura paga pelos clientes. Um ponto relevante de se notar aqui é que, mesmo pagando um valor alto pela energia, não há distinção significativa no valor pago entre os clusters. Além disso, é possível notar na Figura 11 que clusters com maiores consumidores de energia, consequentemente, pagaram valores maiores nas suas faturas. Porém, vale ressaltar que a fatura de energia pode ser uma variável que não entrega tanto valor ao ser analisada quanto energia ativa, por exemplo, uma vez que na fatura incidem várias outras variáveis que impactam no seu valor final e podem atrapalhar eventuais análises.

clusters	count	mean	std	min	25%	50%	75%
0	157684	R\$ 9.848,02	2015667	0	R\$ 676,27	R\$ 770,95	R\$ 928,60
1	167219	R\$ 6.214,59	1033836	0	R\$ 674,01	R\$ 759,49	R\$ 896,85
2	80599	R\$ 7.565,92	629062,7	0	R\$ 608,27	R\$ 700,47	R\$ 833,59
3	84219	R\$ 3.695,03	154594	0	R\$ 668,07	R\$ 755,57	R\$ 889,40
4	67225	R\$ 15.787,83	1812184	0	R\$ 594,83	R\$ 675,72	R\$ 802,01
5	13302	R\$ 9.862,14	471467,9	0	R\$ 640,10	R\$ 732,42	R\$ 889,07
6	76852	R\$ 3.781,08	248400,2	0	R\$ 684,22	R\$ 775,80	R\$ 925,73

Figura 10 - Valor pago pela energia (R\$/MWh) por cluster

clusters	count	mean	std	min	25%	50%	75%
0	157716	12171,72	14786,52	0	R\$ 4.542,41	R\$ 8.642,93	R\$ 16.092,60
1	167247	11719,58	10559,88	0	R\$ 4.578,28	R\$ 8.655,11	R\$ 15.615,12
2	80628	45429,26	90657,72	0	R\$ 8.058,35	R\$ 23.124,82	R\$ 46.247,12
3	84230	13998,92	19643,02	0	R\$ 5.307,36	R\$ 9.738,06	R\$ 17.856,54
4	67299	43026,57	119807,3	0	R\$ 10.551,35	R\$ 32.361,98	R\$ 59.171,40
5	13313	29617,18	89253,17	0	R\$ 5.721,67	R\$ 13.659,67	R\$ 26.106,81
6	76857	11655,62	10493,18	0	R\$ 4.623,81	R\$ 8.513,61	R\$ 15.549,75

Figura 11 - Valor pago pela fatura por cluster

Interpretação dos resultados obtidos e convergência em plano de negócio para comercializadoras

De maneira geral, os resultados forneceram alguns indicativos sobre o comportamento dos clientes de maneira geral e como estes podem ser associados a um eventual plano de organização da comercializadora. O plano pode ser feito levando em consideração o comportamento de cada um dos clusters, possibilitando um planejamento estratégico para cada um dos clientes que se encontra naquele determinado agrupamento.

Os clusters 2 e 4 apresentaram os clientes com maior volume faturado de energia, e, consequentemente, representaram os clientes com um maior fator de carga o que evidencia consumidores organizados e com perfil estruturado de demanda. Esses clientes são particularmente um ponto de atenção para comercializadora, devido aos bons padrões de consumo e menor propensão à inadimplência.

No cluster 1, a elevada taxa de Ponta/Carga destaca-se como um ponto de atenção. Estes consumidores apresentam perfis de consumo que podem ser interpretados como menos otimizados, o que os torna potenciais beneficiários de programas de eficiência energética e, até mesmo, de serviços de consultoria dedicada para redução de custos com energia. Dentro do planejamento da comercializadora, buscar fidelizar estes clientes por meio da oferta de soluções que melhorem o gerenciamento energético e otimizem seu consumo pode trazer resultados.

No cluster 6, o menor fator de carga e a alta participação da ponta indicam consumidores com perfis menos organizados e maior volatilidade no consumo. Para a comercializadora, estes clientes representam um desafio maior, mas também uma oportunidade de atuar como facilitadora de uma gestão energética mais eficiente. Estratégias para este grupo podem incluir a oferta de consultorias de eficiência energética

e condições flexíveis de pagamento para atrair e fidelizar clientes que, de outra forma, poderiam ser considerados de maior risco.

Os clientes do cluster 5 apresentaram características bem distintas dos demais agrupamentos. São os clientes que apresentam maior inadimplência, os caracterizando como UCs de alto risco, reduzindo consideravelmente a sua prioridade em uma eventual prospecção de novos clientes. Porém, os altos índices de DIC/FIC podem traduzir que a menor qualidade do fornecimento de energia elétrica para estes clientes pode explicar o comportamento de atraso de pagamento de fatura, além de caracterizar clientes que poderiam migrar com maior facilidade, caso haja insatisfação com os serviços prestados pela distribuidora.

Os clusters 0 e 3 apresentam características que são medianas em relação à maioria das variáveis analisadas, o que corrobora para que este seja entendido como um cluster composto por clientes com comportamento padrão, sem grandes desvios de perfil em termos de consumo, fator de carga ou pagamento. Para a comercializadora, este grupo representa estabilidade no portfólio e pode ser priorizado para ofertas de produtos padrão, como contratos com características brandas que reduzam custos gerenciais e busquem mitigar risco.

3. Conclusão

O resultados parciais obtidos evidenciam que o algoritmo de clusterização, mesmo sem se utilizar de conhecimentos prévio de especialistas de comercialização de energia, agrupou os clientes com características marcantes de comportamento de consumo energético de maneira que pode-se observar tendências de consumo destas Unidades Consumidoras e foi possível categorizá-las em grupos prioritários para o foco de planejamento de uma comercializadora varejista, permitindo um planejamento estratégico mais preciso. Os grupos trazem informações de risco associado ao padrão comportamental de consumo do cliente, que puderam ser observados a partir do cálculo do seu fator de carga e participação da ponta no total consumido do cliente. Os resultados permitem categorização das UCs em grupos prioritários, possibilitando a comercializadora alinhar ofertas de produtos, reduzir riscos financeiros e direcionar investimentos em programas de eficiência energética ou, até mesmo, fidelização de clientes.

Próximos passos envolvem a aplicação prática da metodologia e mensuração de resultados em empresas do setor elétrico, explorando a aplicabilidade do algoritmo para diferentes bases de clientes e analisando os impactos estratégicos em operações comerciais.

4. Referências bibliográficas

BRASIL. Lei nº 9.074, de 7 de julho de 1995. Estabelece regras para outorga e prorrogações das concessões e permissões de serviços públicos, entre outras providências. **Lei nº 9.074, de 7 de julho de 1995.** Brasília, DF, 8 jul 1995. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9074cons.htm. Acesso em: 05 nov. 2024.

BRASIL. Lei nº 9.427, de 26 de dezembro de 1996. Institui a Agência Nacional de Energia Elétrica - ANEEL, disciplina o regime das concessões de serviços públicos de energia elétrica e dá outras providências. **Lei nº 9.427, de 26 de dezembro de 1996.** Brasília, DF, 27 dez 1996. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/L9427cons.htm. Acesso em: 05 nov. 2024.

BRASIL. Lei nº 9.648, de 27 de maio de 1998. Altera dispositivos das Leis no 3.890-A, de 25 de abril de 1961, no 8.666, de 21 de junho de 1993, no 8.987, de 13 de fevereiro de 1995, no 9.074, de 7 de julho de

1995, no 9.427, de 26 de dezembro de 1996, e autoriza o Poder Executivo a promover a reestruturação da Centrais Elétricas Brasileiras - ELETROBRÁS e de suas subsidiárias e dá outras providências. **Lei nº 9.648, de 27 de maio de 1998.** Brasília, DF, 28 maio 1998. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/L9648cons.htm. Acesso em: 04 nov. 2024.

MME. **Portaria nº 465, de 12 de dezembro de 2019.** Brasília, DF, 16 dez. 2019. Disponível em: <https://www2.aneel.gov.br/cedoc/prt2019465mme.pdf>. Acesso em: 04 nov. 2024.

MME. **Portaria Normativa nº 50/GM/MME, de 27 de setembro de 2022.** Brasília, DF, 28 set. 2022. Disponível em: <https://www2.aneel.gov.br/cedoc/prt2019465mme.pdf>. Acesso em: 04 nov. 2024.

ANEEL. **Nota Técnica nº 55/2012-SEM/ANEEL.** Proposta de resolução normativa que estabeleça os requisitos e procedimentos atinentes à comercialização varejista de energia elétrica no Sistema Interligado Nacional – SIN. Disponível em: <https://antigo.aneel.gov.br/web/guest/consulta-processual> sob número de processo 48500.005476/2011-93. Acesso em: 20/11/2024.

CHÉRON, Emmanuel; KLEINSCHMIDT, E. J. **A review of industrial market segmentation research and a proposal for an integrated segmentation framework.** International Journal of Research in Marketing, v. 2, p. 101–115, 1985. DOI: 10.1016/0167-8116(85)90027-8.

FANG, Jonathan Shinray. **Let's know more about Distance Measures: A small comparison between Euclidean Distance, Cosine Similarity, Manhattan Distance, and Mahalanobis Distance.** Medium, 10 jun. 2024. Disponível em: <https://medium.com/@j.shinray/lets-know-more-about-distance-measures-c6002474c356>. Acesso em: 10 dez. 2024.

TOLBA, Amr; AL-MAKHADMEH, Zafer. **An improved density-based single sliding clustering algorithm for large datasets in the cultural information system.** Personal and Ubiquitous Computing, v. 24, 2020. DOI: 10.1007/s00779-019-01258-5.

SAXENA, Amit et al. **A review of clustering techniques and developments.** Neurocomputing, v. 267, p. 664–681, 2017. DOI: 10.1016/j.neucom.2017.06.053. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0925231217311815>. Acesso em: 10 dez. 2024.

ANEEL. **Base de Dados Geográfica da Distribuidora - BDGD.** Disponível em: <https://dadosabertos.aneel.gov.br/dataset/base-de-dados-geografica-da-distribuidora-bdgd>. Acesso em: 10 dez. 2024.

SHAHAPURE, K. R.; NICHOLAS, C. **Cluster Quality Analysis Using Silhouette Score.** In: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA). Sydney, NSW, Australia, 2020. p. 747–748. DOI: 10.1109/DSAA49011.2020.00096.