

Rumo à Sustentabilidade Global: Análise exploratória mediante técnicas de *machine learning* não supervisionado

Miguel Ângelo Lellis Moreira (Universidade Federal Fluminense) miguellellis@id.uff.br

Maria Teresa Pereira (Instituto Superior de Engenharia do Porto) mtp@isep.ipp.pt

Marcos dos Santos (Instituto Militar de Engenharia) marcosdossantos@ime.eb.br

Carlos Francisco Simões Gomes (Universidade Federal Fluminense) cfsg1@bol.com.br

Resumo

No contexto do desenvolvimento sustentável em todo o mundo, as questões relacionadas à política públicas para construção de um mundo sustentável têm desempenhado um papel significativo no avanço não apenas de aspectos sustentáveis naturais, mas econômico, e tecnológico em alto potencial. Neste cenário, este estudo se baseia em uma base de dados consolidada a partir de duas diferentes fontes de informação, onde sua integração compreende 19 variáveis socioeconômicas e de sustentabilidade frente à 145 países. Mediante técnicas de machine learning não supervisionado, foi conduzida uma análise exploratória que se fundamentou na modelagem estatística para construção de clusters entre as nações a fim de perceber países com características similares no âmbito do desenvolvimento sustentável mundial. Tal análise visa elucidar as correlações existentes entre variáveis, identificando aquelas que exercem maior influência e são mais impactadas nesse cenário de avaliação sustentável. Mediante a implementação do algoritmo *k-means*, associando ao apoio computacional, resultou na construção de oito clusters como sendo uma configuração ideal para representar o contexto sustentabilidade no mundo. Isso permitiu identificar países com características quantitativas semelhantes em relação à sustentabilidade. Por fim, são apresentadas as conclusões derivadas dos resultados encontrados, bem como as limitações identificadas neste estudo. Além disso, são delineadas propostas para pesquisas futuras, visando explorar a fundo questões de sustentabilidade em um caráter global e superando as lacunas identificadas nesta análise.

Palavras-Chaves: Sustentabilidade Global, *Machine Learning* Não Supervisionado, *Clustering*.

1. Introdução

Nas últimas décadas, tem-se observado uma crescente conscientização e ação global em direção à sustentabilidade, impulsionada pelas mudanças climáticas, escassez de recursos e a necessidade premente de preservar o meio ambiente para as gerações futuras (Mcdougall; Wagner; Macbryde, 2022). Esse contexto tem levado muitos países a intensificarem seus esforços em direção à sustentabilidade, não apenas mediante de investimentos financeiros, mas também ao adotarem medidas para desenvolver fontes de energia renováveis, reduzir emissões de gases de efeito estufa e promover a preservação de ecossistemas vitais (Lin *et al.*, 2022).

A preservação ambiental e a busca por um desenvolvimento sustentável tornaram-se elementos cruciais para a estabilidade global (Uralovich *et al.*, 2023). Assim como antes se investia em estratégias apenas sob uma ótica econômica e social, agora há um movimento crescente em direção a práticas sustentáveis, visando à proteção da biodiversidade, a promoção da economia circular e a adaptação às mudanças climáticas (Stephenson e Damerell, 2022). A proteção dos recursos naturais, como água e terra, além da transição para uma economia de baixo carbono, tornaram-se preocupações centrais para a maioria das nações (Jayarathna, Agdas e Dawes, 2023).

Deve ser ressaltado que a sustentabilidade transcende a mera preservação dos recursos naturais; ela abrange igualmente as práticas sociais e econômicas adotadas em diferentes contextos (Ahmad *et al.*, 2023). É fundamental compreender que a busca por um futuro sustentável não se restringe apenas à conservação do meio ambiente, mas também requer a promoção de sistemas sociais justos e inclusivos, assim como modelos econômicos que equilibrem o crescimento com a responsabilidade ambiental e social (Almeida *et al.*, 2022).

O interesse em alcançar índices favoráveis sob a perspectiva sustentável também está impulsionando avanços tecnológicos, tais como inovações em energia limpa, agricultura sustentável e eficiência energética (Zhang, Saydaliev e Ma, 2022). A implementação de inteligência artificial, big data e tecnologias ambientais está sendo vista como uma forma de abordar os desafios atuais, garantindo a eficiência na gestão dos recursos naturais e impulsionando um desenvolvimento sustentável (Castro *et al.*, 2020).

Contudo, junto com esses esforços, há uma crescente preocupação com a degradação ambiental, perda de biodiversidade e os impactos das mudanças climáticas na segurança alimentar e hídrica (Nguyen *et al.*, 2023). A diplomacia ambiental e acordos internacionais

têm se mostrado fundamentais para enfrentar esses desafios, promovendo a cooperação entre países e a implementação de políticas ambientais globais (Zahoor e Mushtaq, 2023).

Nesse cenário, compreender e analisar os indicadores de sustentabilidade mundial tornou-se crucial para tomar decisões estratégicas de alto nível (Ferreira *et al.*, 2019). A identificação de tendências, a compreensão das relações entre diferentes variáveis e similaridades entre nações com abordagens sustentáveis com características semelhantes são de grande valor na compreensão e prospecção de cenários quanto a promoção de agendas sustentáveis (Saini *et al.*, 2022).

Por intermédio de uma abordagem exploratória, este estudo concentra-se na análise de indicadores de sustentabilidade e socioeconômicos mediante técnicas de *machine learning* não supervisionado. Será utilizado um conjunto de dados abrangendo 145 países, integrando métricas socioeconômicas e ambientais, a fim de compreender as correlações entre variáveis e realizar a clusterização desses países, identificando padrões e agrupamentos com base em suas abordagens e desempenhos em sustentabilidade.

O artigo se divide em quatro seções. Estando contextualizado já nesta seção, a seção 2 se dedica a exploração dos fundamentos teóricos e metodológicos que fornecem base à pesquisa a ser realizada. A seção 3 se dedica à aplicação das técnicas de *machine learning* não supervisionado para exploração das correlações entre variáveis e construção dos possíveis clusters de países com características quantitativas similares. Por fim, a seção 4 apresenta as considerações finais do estudo, juntamente com as propostas de trabalhos futuros.

2. Material e métodos

Os modelos de *machine learning* não supervisionado oferecem um conjunto de técnicas estatísticas voltadas para análise de dados que não estão previamente rotulados, permitindo a exploração e identificação de correlações e associações entre variáveis e observações (Zaadnoordijk, Besold e Cusack, 2022). Essas abordagens são frequentemente empregadas para descobrir padrões em conjuntos de dados, revelando inter-relações que podem não ser prontamente perceptíveis em análises iniciais (Moreira *et al.*, 2023).

Como mencionado em (Wang e Biljecki, 2022), esses modelos estatísticos desempenham um papel crucial na análise preliminar de conjuntos de dados, auxiliando na compreensão das variáveis e suas correlações, especialmente no contexto da sustentabilidade (Ishikiriya, Miro e Gomes, 2015). Isso permite a observação da influência de variáveis e indicadores nos desafios ambientais, sociais e econômicos, sendo importante ressaltar que, embora as técnicas

não supervisionadas ofereçam uma exploração significativa dos dados, elas não permitem inferências estatísticas completas sobre um conjunto de dados específico, limitando-se à análise do estado atual dos dados (Ferreira *et al.*, 2019).

Adotando uma visão mais abrangente, as técnicas não supervisionadas, quando aplicadas a dados quantitativos, têm a capacidade de reduzir a complexidade entre variáveis altamente correlacionadas e de agrupar observações em diferentes classes (Kounta *et al.*, 2022). Esta última abordagem é a técnica específica adotada neste estudo. Entre os principais modelos utilizados estão a Análise de Componentes Principais (*Principal Component Analysis* - PCA), *K-means*, Análise de Correspondência Múltipla (*Multiple Correspondence Analysis* - ACM) e outras abordagens estatísticas, contribuindo significativamente para a compreensão e análise de indicadores de sustentabilidade em um contexto global (Alloghani *et al.*, 2020).

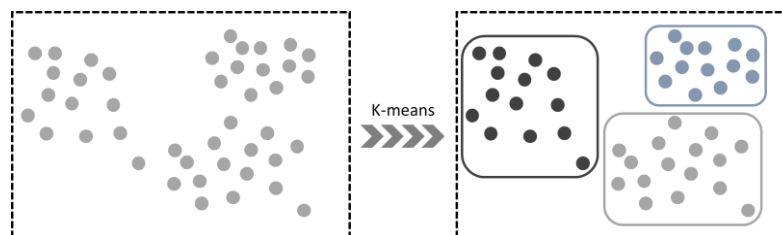
2.1. Clustering

As técnicas agrupamento, um dos principais modelos de machine learning não supervisionado, têm como objetivo dividir um conjunto de dados em grupos ou clusters baseados na similaridade entre os objetos ou pontos de dados (Moreira *et al.*, 2022). Diversas abordagens de agrupamento estão disponíveis, cada uma com suas características específicas. Neste estudo em análise, focalizaremos a técnica estatística conhecida como *k-means*, a qual pressupõe a definição prévia de n clusters (Alloghani *et al.*, 2020).

O algoritmo *k-means* inicia-se selecionando aleatoriamente K centroides, um para cada cluster. Em seguida, ele associa cada ponto de dados ao centroide mais próximo, estabelecendo assim os primeiros clusters. Posteriormente, o algoritmo recalcula os centros de cada cluster, determinando a média dos pontos pertencentes a esse cluster. Este processo de retribuição dos pontos aos centroides mais próximos e recálculo dos centroides é repetido até que não haja mais mudanças nos centroides (Ikotun *et al.*, 2023).

Neste estudo em particular, a métrica utilizada para calcular a distância entre os pontos de dados e os centroides é a distância Euclidiana, considerando a normalização dos dados quantitativos pelo Z-Score. É importante notar que o algoritmo *k-means* é iterativo e pode ser sensível à escolha inicial dos centroides (Faisal e Zamzami, 2020). Por essa razão, é comum executar o algoritmo várias vezes com diferentes seleções iniciais dos centroides e escolher a solução que melhor se adapte aos dados. A Figura 1 exemplifica uma característica gráfica do *K-means*.

Figura 1 – Exemplo da técnica K-means



Para auxiliar na determinação do número ideal de clusters, será empregada a técnica gráfica conhecida como método *elbow* (Cui, 2020), que busca identificar o ponto de inflexão na curva, representando o ponto onde o aumento no número de clusters não resulta em uma redução significativa na variância.

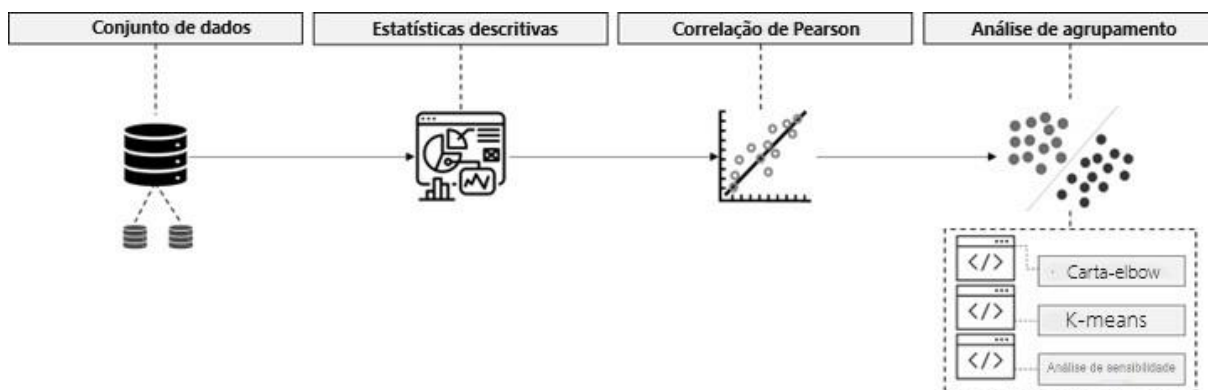
2.2. Processo metodológico

O processo metodológico adotado para análise, se integra por quatro etapas macro. A primeira fase abrangerá a integração de dois conjuntos de dados distintos. O primeiro conjunto apresentará dados quantitativos referentes às informações econômicas dos países (Casas e Cozzi, 2023). O segundo conjunto conterá informações respectivas à indicadores de sustentabilidade sob uma ótica mundial (Sachs *et al.*, 2023). Ressalta-se que dos dados coletados referenciam-se à desempenhos obtidos sob uma gama de indicadores no ano de 2023, não representando referências passadas de longa data ou prospecções futuras.

Após a estruturação e unificação das bases de dados, serão analisadas um total de 19 variáveis respectivas à 145 nações, possibilitando assim dar-se início à avaliação mediante estatística descritiva. Este processo visa identificar os principais indicadores e suas distribuições correspondentes para cada tipo de variável analisada. Em seguida, será conduzida uma análise de correlação de Pearson para esclarecer as variáveis mais influentes no contexto em questão.

Posteriormente, será realizada a etapa de clusterização utilizando a técnica estatística *K-means*. O emprego do gráfico de *elbow* auxiliará na determinação do número ideal de clusters para a avaliação, sendo acompanhado por uma análise de sensibilidade para diferentes cenários de agrupamentos a partir de diferentes valores de *k*. A Figura 2 apresenta o processo metodológico adotado neste estudo.

Figura 2 – Processo metodológico



3. Estudo de caso

Conforme já abordado anteriormente neste estudo, o conjunto de dados a ser analisado é composto por 145 observações, representando uma nação cada, analisados sob a luz de um conjunto de 19 variáveis, todas transcritas quantitativamente. Como apoio computacional, utilizou-se da linguagem R e sua interface Estudio. Para uma compreensão geral dos dados distribuídos, a figura 3 nos expõem as relações quantitativas das variáveis. Na tabela 1, são apresentados os significados de cada variável em exploração.

Figura 3 – Resumo para dados quantitativos

Id-01		Id-02		Id-03		Id-04		Id-05		Id-06			
Min. : 0.60	Min. : 1	Min. : 0.0	Min. : 1.186e+09	Min. : 52.80	Min. : 9.740e+02	1st Qu.: 27.60	1st Qu.: 1.479e+10	1st Qu.: 66.70	1st Qu.: 4.068e+06	Median : 42.05	Median : 1.074e+07		
Median : 42.05	Median : 25	Median : 29.1	Median : 5.374e+10	Median : 74.40	Median : 4.756e+07	Mean : 41.70	Mean : 5.882e+11	Mean : 72.55	Mean : 4.756e+07	3rd Qu.: 58.10	3rd Qu.: 3.251e+07		
3rd Qu.: 58.10	3rd Qu.: 120	3rd Qu.: 42.6	3rd Qu.: 2.688e+11	3rd Qu.: 78.20	3rd Qu.: 3.251e+07	Max. : 82.60	Max. : 2.143e+13	Max. : 84.20	Max. : 1.398e+09				
Id-07		Id-08		Id-09		Id-10		Id-11		Id-12		Id-13	
Min. : 0.00	Min. : 0.090	Min. : 0.00	Min. : 19.80	Min. : 12.95	Min. : 10.60	1st Qu.: 11.80	1st Qu.: 54.39	1st Qu.: 51.83	1st Qu.: 63.61	1st Qu.: 56.01	Median : 15.45	Median : 85.22	Median : 68.29
Median : 15.45	Median : 5.110	Median : 93.77	Median : 61.06	Median : 75.66	Median : 85.22	Mean : 16.24	Mean : 60.05	Mean : 69.76	Mean : 77.11	Mean : 67.48	3rd Qu.: 20.70	3rd Qu.: 95.65	3rd Qu.: 76.25
3rd Qu.: 20.70	3rd Qu.: 8.730	3rd Qu.: 98.86	3rd Qu.: 67.49	3rd Qu.: 86.08	3rd Qu.: 95.65	Max. : 37.20	Max. : 83.40	Max. : 97.11	Max. : 99.76	Max. : 95.05			
Id-14		Id-15		Id-16		Id-17		Id-18		Id-19			
Min. : 10.90	Min. : 39.53	Min. : 10.25	Min. : 0.00	Min. : 13.82	Min. : 37.72	1st Qu.: 48.75	1st Qu.: 39.74	1st Qu.: 60.91	1st Qu.: 67.94	1st Qu.: 67.94	Median : 68.80	Median : 83.70	Median : 83.70
Median : 68.80	Median : 73.77	Median : 50.81	Median : 63.76	Median : 77.44	Median : 83.70	Mean : 62.08	Mean : 58.89	Mean : 72.47	Mean : 79.30	Mean : 79.30	3rd Qu.: 74.42	3rd Qu.: 93.59	3rd Qu.: 93.59
3rd Qu.: 74.42	3rd Qu.: 79.71	3rd Qu.: 75.52	3rd Qu.: 80.51	3rd Qu.: 86.29	3rd Qu.: 93.59	Max. : 99.55	Max. : 99.59	Max. : 99.05	Max. : 98.81	Max. : 98.81			

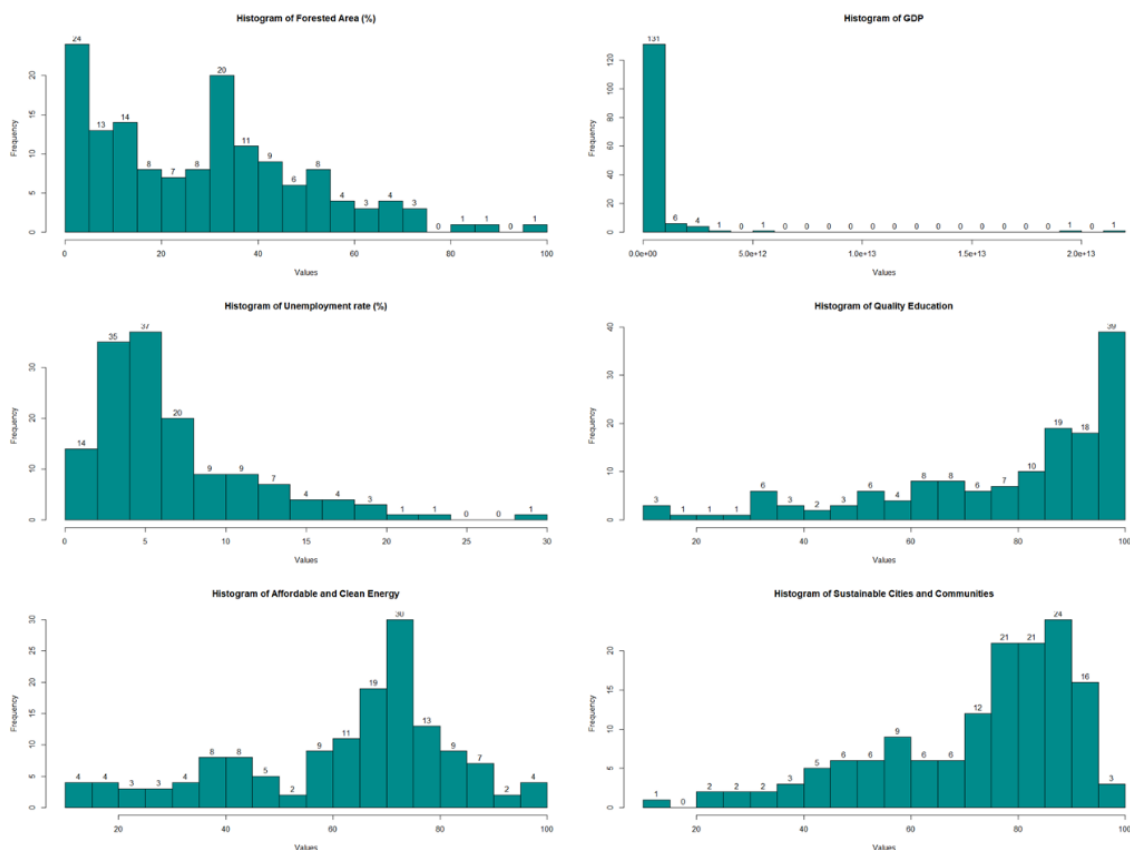
Tabela 1 - Exemplo de tabela

Variáveis	Indicador em representação
Identificação-01	Terras Agrícolas
Identificação-02	Emissões de CO2
Identificação-03	Área Florestal
Identificação-04	PIB
Identificação-05	Esperança de vida
Identificação-06	População
Identificação-07	Receitas fiscais
Identificação-08	Taxa de desemprego
Identificação-09	Sem pobreza

Identificação-10	Fome Zero
Identificação-11	Boa Saúde e Bem-Estar
Identificação-12	Educação de Qualidade
Identificação-13	Água Limpa e Saneamento
Identificação-14	Energia limpa e acessível
Identificação-15	Trabalho Decente e Crescimento Econômico
Identificação-16	Indústria, Inovação e Infraestrutura
Identificação-17	Redução das desigualdades
Identificação-18	Cidades e Comunidades Sustentáveis
Identificação -19	Consumo e Produção Responsáveis

Analisando as distribuições de algumas principais variáveis, foi identificada uma divergência entre as tendências das funções distributivas identificadas. Na figura 4, estão expostas as distribuições para as variáveis Área Florestal, PIB, Taxa de Desemprego, Educação de Qualidade, Energia Acessível e Limpa e Cidades e Comunidades Sustentáveis.

Figura 4 - Algumas distribuições quantitativas de seis variáveis.

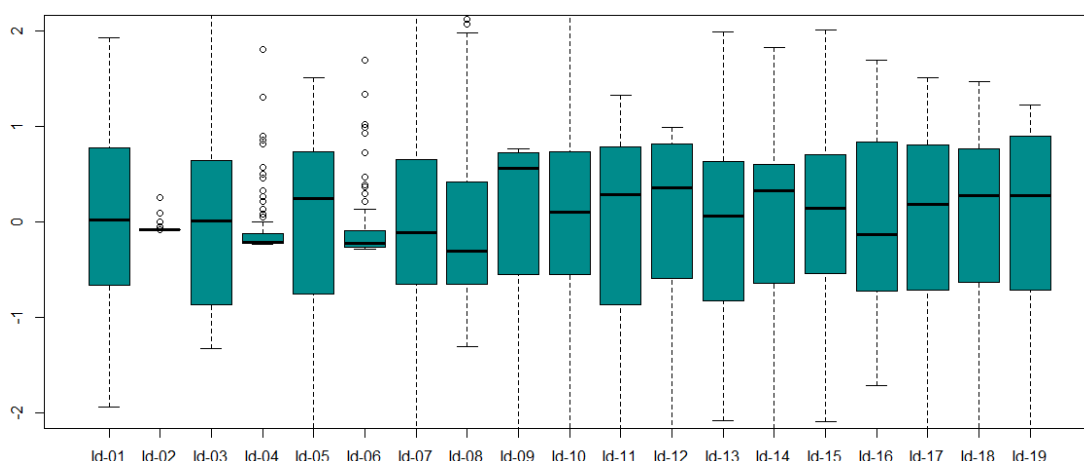


Analisando o primeiro gráfico, entende-se que menos de 70% dos países da base de dados possuem menos de 40% em áreas verdes. Uma grande discrepância é observada referente à variável PIB, onde poucos países representam um alto valor neste índice econômico. Já analisando os aspectos de desemprego, a grande maioria dos países estabelecem-se abaixo dos 10%. Quanto ao nível educacional, poucos países representaram indicadores inferiores à 60% de performance. Quanto aos dois últimos indicadores expostos, reconhece-se que as

tendências sustentáveis dos últimos anos vêm apresentando uma certa influência entre as nações, provendo uma tendência para desenvolvimento de energia limpa e construção de cidades inteligentes.

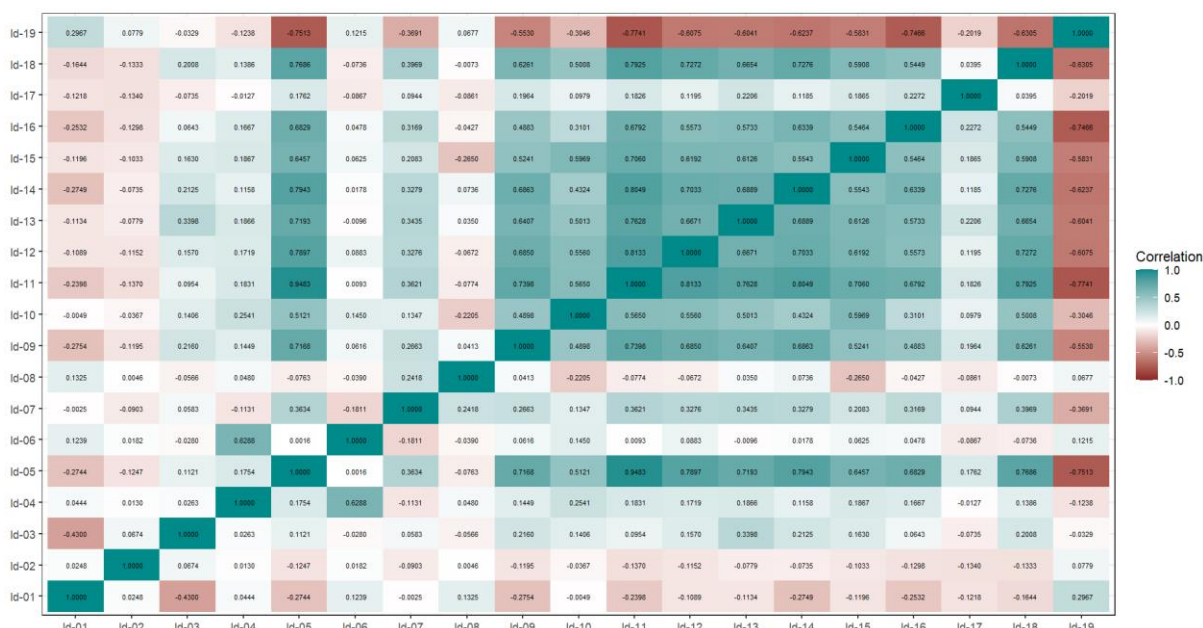
Realizando uma normalização dos dados, mediante a técnica z-score, obtêm-se uma distribuição normalizada em gráfico *boxplot*, conforme exposto na figura 5. Entende-se que as variáveis socioeconômicas (Id-01 – Id-08), apresentam em alguns casos uma baixa variabilidades e em outras parte uma alta variabilidades das observações. Já as variáveis sob aspectos sustentáveis (Id-09 – Id-19), apresentam menor concentração entre os quartis e maior variabilidade em todas as variáveis.

Figura 5 - A variabilidade dos dados diz respeito às dezenove variáveis por meio do gráfico *boxplot*



Finalizada a parte descritiva dos dados, o foco deste estudo se destina na compreensão das principais correlações entre as variáveis. Neste cenário, mediante a realização de uma análise de correlação de Pearson entre as dezenove variáveis desta pesquisa, se observa algumas tendências e em sua grande maioria, a ausência de correlação entre variáveis socioeconômicas e de sustentabilidade. A figura 6 apresenta as correlações identificadas.

Figura 6 - Correlação de Pearson entre dezenove variáveis.

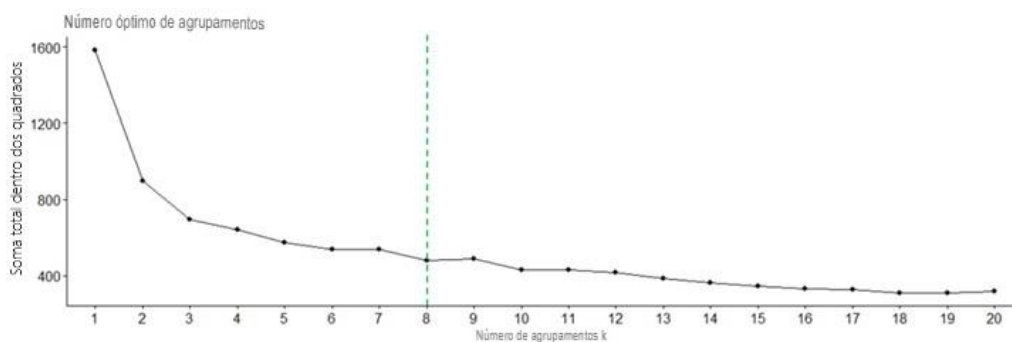


Analisando a correlação de *pearson* construída, entende-se que, com base no conjunto de dados, não há uma grande influência de variáveis socioeconômicas quanto ao investimento e inclinação para aspectos sustentabilidade. Em uma forma mais direta, não necessariamente países desenvolvidos estão propensos a se tornarem sustentáveis mais rápido que países menos desenvolvidos. Observando a variável *Life Expectation* (Id-05), foi a única que apresentou alta correlação com as variáveis de sustentabilidade, podendo assim ser um caso específico de investigação.

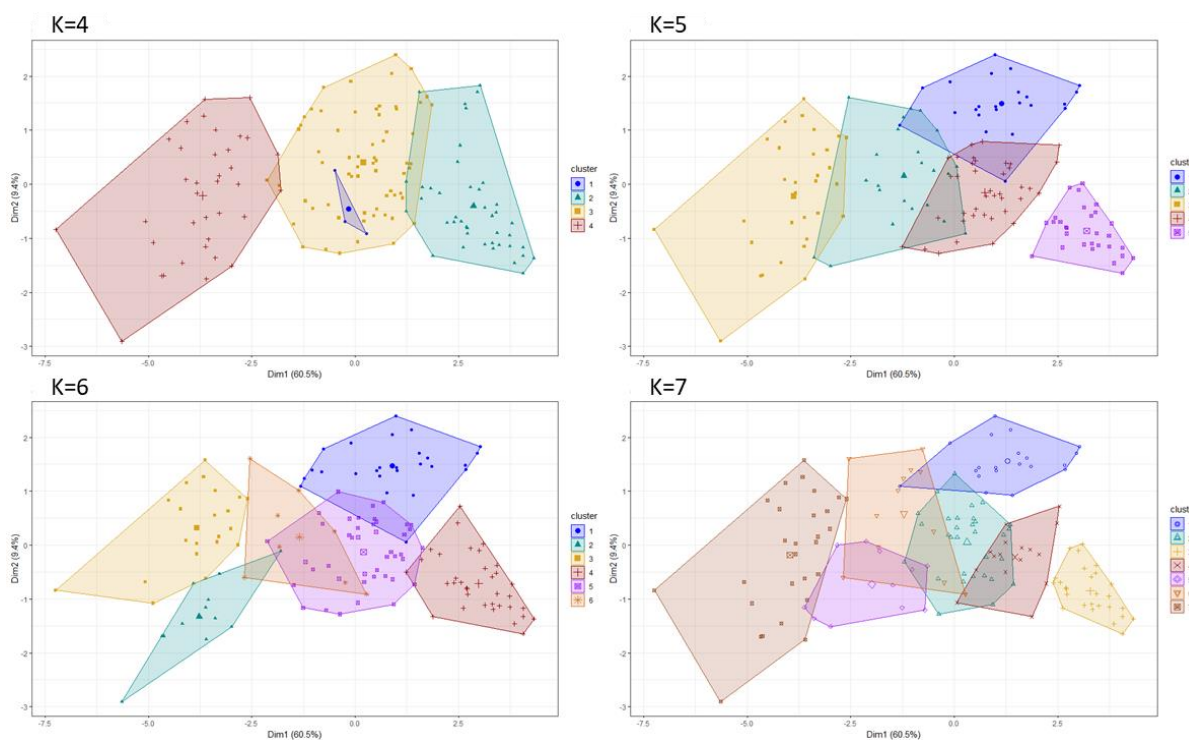
Desta forma, tendo as concentrações de correlações apenas entre as variáveis de sustentabilidade, este estudo removerá as variáveis socioeconômicas, para que não seja apresentada uma influência divergente da realidade necessária, realizando assim a construção dos clusters apenas mediante as variáveis de sustentabilidade.

Tendo em consideração a implementação da técnica de clusterização K-means, utilizou-se como apoio a metodologia de *Elbow*, possibilitando a compreensão de número favorável de clusters de forma a expor a divergência característica entre variáveis. Mediante a técnica gráfica, compreende-se que $k=8$ é o quantitativo favorável de clusters, conforme exposto na figura 7.

Figura 7 - Correlação de Pearson entre dezenove variáveis

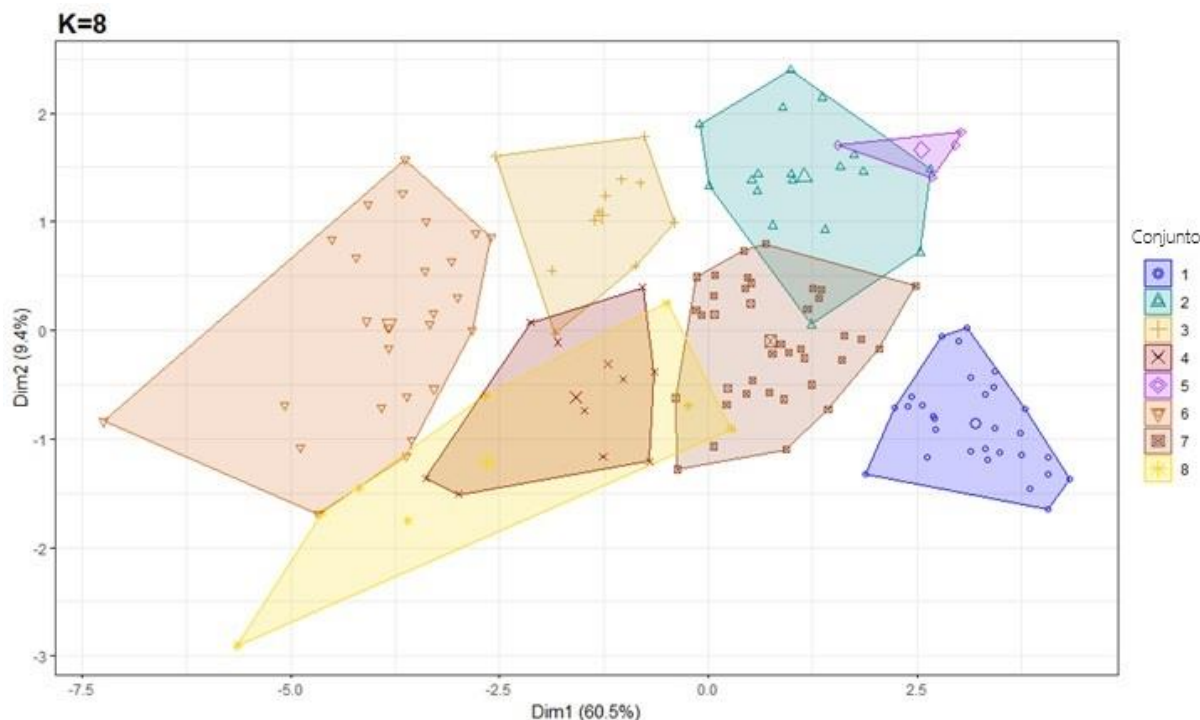


Mesmo sendo indicado a construção de oito clusters, antes de avaliarmos $k=8$, iremos observar o desenvolvimento dos agrupamentos para $k=4$ até $k=7$, conforme exposto na Figura 8.

Figura 8 – Clusters para $K=4$, $K=5$, $K=6$, e $K=7$


Analisando os agrupamentos gerados pelo modelo computacional, é observado que já em $k=4$, é construído um cluster com apenas 3 países. Já em $k=5$, é observada uma maior distribuição entre as nações, entretanto, expondo dois países no cluster 5, estando significativamente afastado do centroide, onde esses mesmos dois países são separados em clusters diferentes em $k=6$ e unificados novamente em $k=7$, onde se tem uma maior separação entre as nações. Para $k=8$, o resultado é exposto na figura 9.

Figura 9 – Clusters para K=8



Quando analisado os clusters para $k=8$, é identificado a singularidade em alguns países, como no cluster 8, possuindo apenas 8 países em seu agrupamento. O Conjunto 1, em azul à direita, é representado por sua maioria por países da Europa, muitos membros da OECD. Já o cluster 6 é representado por países africanos. O ponto de relevância é que características de sustentabilidade podem ser aproximadas por países vizinhos, provendo características de desempenhos similares entre si para fins de indicadores sustentáveis. Para um melhor detalhamento dos países presentes em cada cluster construído, a tabela 2 expõe as respectivas informações.

Tabela 2. Descrição dos países pertencentes a cada cluster para $k=8$

Países	Cluster	Países	Cluster	Países	Cluster
Austrália	1	Botsuana	3	Albânia	7
Áustria	1	Essuatíni	3	Argélia	7
Canadá	1	Gana	3	Bahrein	7
Croácia	1	Guatemala	3	Barbados	7
Chipre	1	Honduras	3	Belize	7
Dinamarca	1	Índia	3	Butão	7
Estônia	1	Quênia	3	Bósnia e Herzegovina	7
Finlândia	1	Namíbia	3	Bulgária	7
França	1	Nicarágua	3	Cuba	7
Alemanha	1	Filipinas	3	República Dominicana	7
Grécia	1	África do Sul	3	El Salvador	7
Hungria	1	Bangladesh	4	Fiji	7
Israel	1	Camboja	4	Geórgia	7
Itália	1	Gabão	4	Guiana	7

Japão	1	Iraque	4	Indonésia	7
Letônia	1	Mauritânia	4	Jordânia	7
Lituânia	1	Mongólia	4	Cazaquistão	7
Luxemburgo	1	Myanmar	4	Kuaite	7
Malta	1	Nepal	4	Líbano	7
Países Baixos	1	Paquistão	4	Maurícia	7
Nova Zelândia	1	Senegal	4	Moldávia	7
Noruega	1	Zâmbia	4	Montenegro	7
Polônia	1	Bélgica	5	Marrocos	7
Portugal	1	Islândia	5	Macedônia do Norte	7
Cingapura	1	Eslovênia	5	Omã	7
Espanha	1	Emirados Árabes Unidos	5	Catar	7
Suécia	1	Angola	6	Romênia	7
Suíça	1	Benim	6	Sérvia e Montenegro	7
Reino Unido	1	Burquina Faso	6	Sri Lanka	7
Estados Unidos	1	Burundi	6	Suriname	7
Argentina	2	Camarões	6	Tajiquistão	7
Armênia	2	Comores	6	Tailândia	7
Belarus	2	Djibuti	6	Trinidad e Tobago	7
Bolívia	2	Etiópia	6	Tunísia	7
Brasil	2	Guiné	6	Turquemenistão	7
Chile	2	Haiti	6		
China	2	Lesoto	6	Uzbequistão	7
Colômbia	2	Libéria	6	Vietname	7
Costa Rica	2	Madagáscar	6	Afeganistão	8
Equador	2	Malawi	6	Azerbaijão	8
Jamaica	2	Mali	6	República Centro-Africana	8
Malásia	2	Moçambique	6	Chade	8
México	2	Níger	6	Maldivas	8
Panamá	2	Nigéria	6	Papua-Nova Guiné	8
Paraguai	2	Ruanda	6	Arábia Saudita	8
Peru	2	Serra Leoa	6	Somália	8
Ucrânia	2	Sudão do Sul	6		
Uruguai	2	Sudão	6		
		Tanzânia	6		
		Togo	6		
		Uganda	6		
		Zimbábue	6		

Ressalta-se que o estudo em questão não buscar por obter análises preditivas, mas sim, compreender correlações entre variáveis de sustentabilidade frente a indicadores socioeconômicos, tornando possível o agrupamento mediante características quantitativas similares entre os países de cada cluster construído, não possibilitando neste primeiro momento alguma inferência estatística.

Mediante a exploração realizada, países com aspectos geográficos próximos e com similaridades em desenvolvimento econômico e tecnológico foram colocados em um mesmo cluster, validando assim percepções de associações entre países parceiros ou de características políticas e sociais similares, evidenciando uma percepção mediante modelagem estatística.

4. Conclusão

O objetivo primordial deste estudo consistiu em realizar uma análise exploratória dos dados por meio da aplicação de técnicas de *machine learning* não supervisionado, visando identificar correlações e associações entre as diversas variáveis analisadas no contexto de desenvolvimento de indicadores de sustentabilidade. Ao examinar uma base de dados composta por 19 variáveis e 145 nações, observou-se uma correlação limitada entre os indicadores de sustentabilidade e os índices econômicos e sociais. Essa constatação foi crucial para compreender que um país com bons índices de sustentabilidade, não necessariamente é rico ou com alto nível de desenvolvimento tecnológico, da mesma forma que fica a percepção de que países com alto desenvolvimento econômico, não possui, obrigatoriamente, bons índices de avanço no aspecto sustentável.

Posteriormente, explorou-se a formação de clusters utilizando a técnica estatística K-means, apoiada pelo gráfico de *elbow* para determinar o número ideal de clusters para o conjunto de observações. Os clusters obtidos refletem percepções relacionadas ao contexto sustentável, agrupando nações com proximidades geográficas ou características políticas e econômicas similares. Nesse contexto, pode-se concluir que este estudo foi conduzido de maneira favorável, revelando associações presentes no cenário do desenvolvimento e aprimoramento de aspectos sustentáveis no mundo.

É importante ressaltar que esta análise se baseou em dados quantitativos restritos ao ano de 2023, não se considerando o desenvolvimento ao longo do tempo das décadas passadas, assim como não busca fins preditivos para análise. As informações utilizadas refletem características numéricas mediante dados públicos, onde as técnicas de avaliação apresentadas não incorporam aspectos de subjetividade dos autores do estudo. Para pesquisas futuras, pretende-se empregar outras técnicas não supervisionadas, como análise fatorial, a fim de reduzir a dimensionalidade das variáveis e potencialmente criar um ranking mediante análise multicritério, possibilitando a ponderação e construção de pesos aos indicadores de sustentabilidade em avaliação.

REFERÊNCIAS

- AHMAD, M. *et al.* Policy framework considering resource curse, renewable energy transition, and institutional issues: Fostering sustainable development and sustainable natural resource consumption practices. *Resources Policy*, v. 86, p. 104173, 2023.
- ALLOGHANI, M. *et al.* A systematic review on supervised and unsupervised machine learning algorithms for data science. *Supervised and unsupervised learning for data science*, p. 3–21, 2020.
- CASAS, T.; COZZI, G. Elite Quality Report 2023: Country Scores and Global Rankings. *Zurich: Seismo*, 2023.
- CASTRO, C. *et al.* Logistics reorganization and management of the ambulatory pharmacy of a local health unit in Portugal. *Evaluation and Program Planning*, v. 80, p. 101801, jun. 2020. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0149718919303076>>.
- CUI, M. Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, v. 1, n. 1, p. 5–8, 2020.
- DE ALMEIDA, I. D. P. *et al.* Assisting in the choice to fill a vacancy to compose the PROANTAR team: Applying VFT and the CRITIC-GRA-3N methodology. *Procedia Computer Science*, v. 214, p. 478–486, 2022. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1877050922019123>>.
- FAISAL, M.; ZAMZAMI, E. M. Comparative analysis of inter-centroid K-Means performance using euclidean distance, canberra distance and manhattan distance. 2020, [S.l.]: IOP Publishing, 2020. p. 12112.
- FERREIRA, C. *et al.* iLeanDMAIC—A methodology for implementing the lean tools. *Procedia Manufacturing*, v. 41, p. 1095–1102, 2019.
- FERREIRA, S. *et al.* KPI development and obsolescence management in industrial maintenance. *Procedia Manufacturing*, v. 38, p. 1427–1435, 2019. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2351978920301463>>.
- IKOTUN, A. M. *et al.* K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, v. 622, p. 178–210, 2023.
- ISHIKIRIYAMA, C. S.; MIRO, D.; GOMES, C. F. S. Text Mining Business Intelligence: a small sample of what words can say. *Procedia Computer Science*, v. 55, p. 261–267, 2015.
- JAYARATHNA, C. P.; AGDAS, D.; DAWES, L. Exploring sustainable logistics practices toward a circular economy: A value creation perspective. *Business Strategy and the environment*, v. 32, n. 1, p. 704–720, 2023.
- KOUNTA, C. A. K. A. *et al.* Review of AI-based methods for chatter detection in machining based on bibliometric analysis. *The International Journal of Advanced Manufacturing Technology*, v. 122, n. 5–6, p. 2161–2186, 2022.
- LIN, C.-Y. *et al.* Development of renewable energy resources by green finance, volatility and risk: empirical evidence from China. *Renewable Energy*, v. 201, p. 821–831, 2022.
- MCDUGALL, N.; WAGNER, B.; MACBRYDE, J. Leveraging competitiveness from sustainable operations: frameworks to understand the dynamic capabilities needed to realise NRBV supply chain strategies. *Supply Chain Management*, v. 27, n. 1, p. 12–29, 2022.
- MOREIRA, M. Â. L. *et al.* Defense Perception in the Geopolitical Scope: An exploratory study through unsupervised machine learning. *Procedia Computer Science*, v. 221, p. 689–696, 2023.
- MOREIRA, M. Â. L. *et al.* Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems. *Procedia Computer Science*, v. 214, p. 117–124, 2022. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1877050922018634>>.
- NGUYEN, T. T. *et al.* Security risks from climate change and environmental degradation: implications for sustainable land use transformation in the Global South. *Current Opinion in Environmental Sustainability*, v. 63, p. 101322, 2023.
- SACHS, J. D. *et al.* Sustainable development report 2023: Implementing the SDG Stimulus. 2023.
- SAINI, N. *et al.* Non-financial disclosures and sustainable development: A scientometric analysis. *Journal of Cleaner Production*, v. 381, p. 135173, 2022.
- STEPHENSON, P. J.; DAMERELL, A. Bioeconomy and Circular Economy Approaches Need to Enhance the Focus on Biodiversity to Achieve Sustainability. *Sustainability*, v. 14, n. 17, p. 10643, 2022.



URALOVICH, K. S. *et al.* A primary factor in sustainable development and environmental sustainability is environmental education. *Caspian Journal of Environmental Sciences*, v. 21, n. 4, p. 965–975, 2023.

WANG, J.; BILJECKI, F. Unsupervised machine learning in urban studies: A systematic review of applications. *Cities*, v. 129, p. 103925, 2022.

ZAADNOORDIJK, L.; BESOLD, T. R.; CUSACK, R. Lessons from infant learning for unsupervised machine learning. *Nature Machine Intelligence*, v. 4, n. 6, p. 510–520, 2022.

ZAHOOR, I.; MUSHTAQ, A. Water pollution from agricultural activities: A critical global review. *Int. J. Chem. Biochem. Sci*, v. 23, p. 164–176, 2023.

ZHANG, L.; SAYDALIEV, H. B.; MA, X. Does green finance investment and technological innovation improve renewable energy efficiency and sustainable development goals. *Renewable Energy*, v. 193, p. 991–1000, 2022.