São José dos Campos - SP
LV SBPO
LV Simpósio Brasileiro de Pesquisa Operacional
São José dos Campos, SP - 6 a 9 de novembro de 2023

UMA MODELAGEM HÍBRIDA COM K-MÉDIAS E ELECTRE-III PARA ANALISAR PAÍSES CONSIDERANDO DIFERENTES CONJUNTOS DE DADOS SOCIOECONÔMICOS

Igor Danilo Costa Matos
Programa de Pós-Graduação em Engenharia de Produção - Universidade Federal Fluminense
Rua Passo da Pátria, 156, Campus Praia Vermelha, Bloco D - sala 309, São Domingos, Niterói-RJ
matosigor@id.uff.br

Diogo Ferreira de Lima Silva, Helder Gomes Costa, Marcos Costa Roboredo
Departamento de Engenharia de Produção, Universidade Federal Fluminense
Rua Passo da Pátria, 156, Campus Praia Vermelha, Bloco D - sala 306, São Domingos, Niterói-RJ
diogofls@id.uff.br, heldergc@id.uff.br, mcroboredo@id.uff.br

RESUMO

O presente artigo utiliza uma abordagem híbrida para avaliação de países, aplicando em conjunto o algoritmo k-médias, usado para a obtenção de *clusters* nos conjuntos de dados, e o método de ordenação multicritério ELECTRE-III, utilizado para ordenar os *clusters* obtidos. A abordagem foi aplicada a dois conjuntos de dados distintos, referentes ao índice de liberdade econômica da Fundação *Heritage* e ao índice de prosperidade do Instituto *Legatum*. Os resultados incluem visualizações dos *clusters* de países, as relações de preferência definidas pelo ELECTRE-III e a ordenação obtida. Além disso, uma análise posterior é apresentada com o uso de 10.000 simulações considerando variações nos parâmetros do ELECTRE-III e o uso de distribuições de probabilidade. Os resultados mostraram-se consistentes com as expectativas e a robustez das ordenações foi confirmada pelas estatísticas obtidas com as simulações.

PALAVRAS-CHAVE. K-médias, ELECTRE-III, Avaliação de países.

Tópicos (IC – Inteligência Computacional, ADM - Apoio à Decisão Multicritério)

ABSTRACT

This article employs a hybrid approach to evaluate countries, combining the k-means algorithm for clustering in the datasets with the ELECTRE-III multicriteria sorting method to rank the obtained clusters. The approach is applied to two distinct datasets, namely the Economic Freedom Index from the Heritage Foundation and the Prosperity Index from the Legatum Institute. The results include visualizations of country clusters, preference relations defined by ELECTRE-III, and the resulting ranking. Furthermore, a subsequent analysis is presented using 10,000 simulations that consider variations in the ELECTRE-III parameters and the utilization of probability distributions. The results demonstrate consistency with expectations, and the robustness of the rankings is confirmed by the statistics obtained from the simulations.

KEYWORDS. K-means, ELECTRE-III, Country evaluation.

Paper topics (IC - Computational Intelligence, ADM - Multicriteria Decision Aiding)

DOI: 10.59254/sbpo-2023-175028
http://live-proceedings.pantheonsite.io/p/175028

1. Introdução

A análise de indicadores de países frequentemente envolve vários fatores que podem ser conflitantes. Nesse cenário, o uso de modelos matemáticos capazes de combinar múltiplas dimensões torna-se interessante para decisores e analistas. A partir do uso destas técnicas, formuladores de políticas públicas de um país podem, por exemplo, analisar a situação do país perante os países vizinhos, ou mesmo em relação a outros países com situação econômica semelhante. Nesse cenário, destacam-se as ferramentas de áreas como Aprendizado de Máquina (AM) [Faceli et al., 2021] e Apoio à Decisão Multicritério (ADM) [Greco et al., 2016].

Algoritmos de Aprendizado de Máquina podem ser classificados de acordo com a forma com que o computador aprende através dos dados a realizar determinada tarefa. Por exemplo, em Aprendizado Supervisionado, aprende-se a rotular objetos a partir de um conjunto de objetos previamente rotulados (conjunto de treinamento). Já no caso do Aprendizado Não Supervisionado, procura-se descobrir e entender padrões de um conjunto de dados, sem a presença inicial de rótulos [Rani et al., 2023; Faceli et al., 2021]. As técnicas utilizadas em Aprendizado Não Supervisionado incluem, por exemplo, agrupamento de objetos, estatística descritiva e redução de dimensionalidade [Géron, 2022].

Já no campo de AMD, métodos apresentam uma abordagem sistemática para considerar avaliações de alternativas em múltiplos critérios e as preferências do(s) decisor(es) em um problema de decisão [Greco et al., 2016]. Os tipos de problema abordados (chamados de problemáticas) incluem descrição, escolha, ordenação, classificação e portfólio. A utilização desse tipo de abordagem tem mostra-se útil para avaliação de países, conforme observa-se nos estudos de Azevedo e Costa [2022]; Siksnelyte-Butkiene e Streimikiene [2022]; Ozmen [2019].

O presente artigo aplica uma técnica de AM para a tarefa de agrupamento de dados (*clustering*), o k-médias (*k-means*). Procedimentos de agrupamento identificam partições no conjunto de dados, chamadas de *clusters*, formadas por objetos que compartilham características semelhantes [Faceli et al., 2021]. Além disto, este estudo aplica uma abordagem AMD para ordenar os *clusters* obtidos pelo método k-médias em termos de preferências. Essa ordenação é realizada a partir de um método multicritério, o ELECTRE-III. Assim, é possível alcançar uma informação adicional em relação ao agrupamento dos objetos. O uso do k-médias em conjunto com o ELECTRE-III foi proposto em um trabalho recente, apresentado por de Souza e Costa [2021]. Neste trabalho os autores aplicaram a abordagem híbrida para analisar a sustentabilidade de países.

Neste artigo, ampliamos o uso da abordagem para analisar países segundo os indicadores presentes no índice de liberdade econômica da *The Heritage Foundation* e o índice de desenvolvimento sustentável do *Legatum Institute*. Além disso, adicionamos uma análise de robustez à ordenação obtida com o ELECTRE-III através de uma simulações Monte-Carlo com variação de parâmetros do ELECTRE-III, obtidos nas simulações através de distribuições de probabilidade.

O restante deste trabalho está organizado como a seguir. A Seção 2 apresenta uma breve fundamentação teórica com detalhes dos métodos k-médias e ELECTRE-III. A Seção apresenta as aplicações da abordagem híbrida nos dois conjuntos de dados relativos à análise de países. Os resultados são discutidos na Seção 4. Por fim, a Seção 5 apresenta as conclusões do trabalho e perspectivas para trabalhos futuros.

2. Fundamentação Teórica

Neste momento, definiremos a notação a ser utilizada no artigo. Alguns termos típicos das áreas de Aprendizado de Máquina e Decisão Multicritério serão considerados sinônimos para efeito do presente trabalho.

DOI: 10.59254/sbpo-2023-175028
http://live-proceedings.pantheonsite.io/p/175028

Seja $A = \{a_1, ..., a_m\}$ um conjunto de m alternativas e $G = \{g_1, g_2, ..., g_n\}$ um conjunto de n critérios. Seja $\mathbf{X} \in \mathbb{R}^{m \times n}$ uma matriz de dados (decisão), onde cada linha representa uma alternativa (objeto, exemplo) e cada coluna representa um critério (atributo, dimensão). No presente trabalho, o vetor de atributos relacionado a uma alternativa a_i será dado por $\mathbf{x}_i = [x_{i,1}, x_{i,2}, ..., x_{i,n}]$, que também é a i-ésima linha da matriz $\mathbf{X}$. O valor da alternativa a_i no critério g_j é dado por $x_{i,j}$, ou seja, pela entrada (i, j) da matriz de decisão $\mathbf{X}$.

2.1. O algoritmo k-médias

O método k-médias é considerado um algoritmo particional baseado em erro quadrático [Faceli et al., 2021]. Nesse tipo de algoritmo, o erro quadrático é dado pela soma da variação *intracluster*, calculado através da soma da distância de cada objeto ao centroide de um *cluster* C_j. Esses algoritmos, portanto, são muito uteis para achar *clusters* de formato globular.

O k-médias divide o conjunto de dados em k *clusters*, sendo k um hiperparâmetro do algoritmo. Inicialmente, seleciona-se aleatoriamente k elementos para serem os centroides iniciais dos *clusters* e, a partir deles, distâncias são obtidas para cada objeto do conjunto de dados. Cada objeto, então, é associado ao *cluster* cujo centroide esteja mais próximo. Assim, cada *cluster* incluirá objetos que estejam próximos em termos de distância euclidiana. Em seguida, atualiza-se os centroides, recalculado-os a partir dos *clusters* obtidos na etapa anterior. Então, novamente, é calculada a distância de cada objeto aos novos centroides e um novo agrupamento é realizado. Esse processo é repetido até que não haja mais modificação alguma na alocação dos objetos aos *clusters*. Esses passos são organizados no Algoritmo 1.

Algorithm 1 k-médias, baseado em Faceli et al. [2021]

Entradas: Conjunto de dados $\mathbf{X}$, número de *clusters* k.
Saída: Partição de $\mathbf{X}$ em k grupos.

```
Inicialize k centroides, {x̄_1, ..., x̄_k}, a partir de vetores formados por k linhas de X, escolhidas aleatoriamente.
Enquanto o critério de parada não for atingido faça
    Para cada objeto x_i ∈ X faça
        Para cada Centroide x̄_j faça
            Calcule a distância d(x_i, x̄^(j))
        fim
        Aloque x_i para o clusters C_j de centroide mais próximo.
    fim
    Atualize os centroides, calculando-os a partir dos clusters formados pelas alocações.
fim
```

2.2. ELECTRE-III

Proposto por Roy [1978], o método ELECTRE-III considera a ideia de pseudo-critério, utilizando limiares de indiferença (q_j) e preferência (p_j) para lidar com as preferências imputadas pelo(s) decisor(es). O Algoritmo 2 ilustra os principais passos do método. A fase de construção da relação de sobreclassificação compreende os Passos 1,2 e 3, onde são calculados os índices de concordância e discordância e é avaliada a credibilidade da relação de sobreclassificação entre pares de alternativas. Os passos seguintes correspondem à exploração dessas relações, a partir da qual obtém-se duas pré-ordens completas das alternativas. Para isso, duas regras são utilizadas para a ordenação das alternativas: descendente e ascendente.

DOI: 10.59254/sbpo-2023-175028
http://live-proceedings.pantheonsite.io/p/175028

Algorithm 2 Rotina do ELECTRE-III, adaptado de Rogers (2000)

Passo 1: Definição do índice de concordância ($C(a,b)$):

$$C(a,b) = \frac{\sum_{j=1}^{n} w_j c_j(a,b)}{\sum_{j=1}^{n} w_j}$$

$$c_j(a,b) = \begin{cases} 1, se\ g_j(a) + q_j(g_j(a)) \geq g_j(b) \\ 0, se\ g_j(a) + p_j(g_j(a)) < g_j(b) \\ \frac{g_j(a) - g_j(b) + p_j(g_j(a))}{p_j(g_j(a)) - q_j(g_j(a))}\ nos\ demais\ casos \end{cases}$$

Passo 2: Definição do índice de discordância ($D(a,b)$)

$$D_j(a,b) = \begin{cases} 0, se\ g_j(b) \geq g_j(a) + p_j(g_j(a)) \\ 1, se\ g_j(b) > g_j(a) + v_j(g_j(a)) \\ \frac{g_j(b) - g_j(a) - p_j(g_j(a))}{v_j(g_j(a)) - p_j(g_j(a))}\ nos\ demais\ casos \end{cases}$$

onde,

v_j: limiar de veto para o critério j

Passo 3: Definição do grau de credibilidade da sobreclassificação de b por a ($S(a,b)$)

$$S(a,b) = \begin{cases} C(a,b), se\ D(a,b) \leq C(a,b), \forall j \\ C(a,b) \prod_{j \in J(a,b)} \frac{(1 - D(a,b))}{1 - C(a,b)}\ nos\ demais\ casos \end{cases}$$

Passo 4: Cálculo da pontuação de qualificação (λ_0) e do nível de corte (λ_1)

$$\lambda_0 = \max_{a,b \in A} \{S(a,b)\}$$

$$\lambda_1 = \lambda_0 - s(\lambda_0)$$

onde $s(\lambda_0)$ é o limite de discriminação.

Passo 5: Obtenção das ordenações descendente e ascendente através de uma processo iterativo com atualização do valor de λ_1 e aplicação dos procedimentos:

$$\overline{D}_1 = \{a \in A / q_A^{\lambda_1} = \overline{q}_A = \max_{x \in A} q_A^{\lambda_1}(x)\}$$

$$\underline{D}_1 = \{a \in A / q_A^{\lambda_1} = \underline{q}_A = \min_{x \in A} q_A^{\lambda_1}(x)\}$$

onde $q_A^{\lambda_1}(a)$ é a qualificação da alternativa a em relação às demais; calculada como o número de alternativas sobreclassificadas por a sob um nível de corte λ_1 menos o número de alternativas que sobreclassificam a com esse mesmo nível de corte.

O processo detalhado da construção dos *rankings* pode ser encontrado em [Rogers et al., 2000].

DOI: 10.59254/sbpo-2023-175028
http://live-proceedings.pantheonsite.io/p/175028

3. Aplicações

Este estudo apresenta duas aplicações da abordagem híbrida proposta, que inclui o uso dos métodos k-médias e ELECTRE-III. Ambas as aplicações referem-se a análise de um grande conjunto de países considerando um conjunto de indicadores. O primeiro conjunto de indicadores refere-se aos utilizados pelo Índice *Heritage* de Liberdade Econômica 2023. Na segunda aplicação, os indicadores consistem naqueles utilizados pelo Índice *Legatum* de Prosperidade. Nos dois casos, os conjuntos incluem 12 critérios para avaliar países com pontuações entre 0 e 100. Os critérios utilizados são ilustrados na Tabela 1.

Tabela 1: Tabela de Critérios

g_j	Heritage (Liberdade Econômica)	Legatum (Prosperidade)
g_1	Direitos de propriedade	Segurança e Proteção
g_2	Eficácia judicial	Liberdade individual
g_3	Integridade governamental	Governança
g_4	Carga tributária	Capital social
g_5	Gastos governamentais	Ambiente de investimento
g_6	Saúde fiscal	Condições empresariais
g_7	Liberdade de negócios	Infraestrutura e acesso ao mercado
g_8	Liberdade trabalhista	Qualidade econômica
g_9	Liberdade monetária	Condições de moradia
g_{10}	Liberdade comercial	Saúde
g_{11}	Liberdade de investimento	Educação
g_{12}	Liberdade financeira	Ambiente Natural

Neste trabalho, as aplicações foram desenvolvidas usando a linguagem de programação pyhton. A biblioteca scikit-learn [Pedregosa et al., 2011] foi utilizada para a aplicação de uma normalização nos dados e para o agrupamento dos países com o método k-médias. Além disso, a biblioteca pyDecision [Pereira, 2023] foi utilizada para a aplicação do método ELECTRE-III.

3.1. Índice de Liberdade Econômica (Heritage)

O Índice de Liberdade Econômica é publicado anualmente pela *The Heritage Foundation* e tem o objetivo de avaliar o nível de liberdade na economia de diversos países. No presente trabalho, utilizou-se dados referentes ao índice em 2023, onde um total de 176 países foi analisado. A Tabela 2 ilustra um recorte da matriz de dados, incluindo os desempenhos obtidos por um subconjunto de países nos 12 indicadores.

Apesar de os critérios serem avaliados no mesmo domínio, notas entre 0 e 100, optou-se por usar uma normalização antes da aplicação do algoritmo k-médias. Para isso, a normalização MinMax foi escolhida e aplicada, conforme ilustra a Equação 1. Cada entrada da matriz de dados é transformada em um valor entre 0 e 1, conforme seja a razão entre sua distância em relação ao mínimo dentro de um critério e o intervalo observado nos dados para esse critério.

$$X_{i,j}^{norm} = \frac{X_{i,j} - \min_i X_{i,j}}{\max_i X_{i,j} - \min_i X_{i,j}} \tag{1}$$

Neste trabalho, definiu-se como hiperparâmetro para o algoritmo k-médias o $k = 5$. Assim, cinco *clusters* foram obtidos. Os seus centroides podem ser visualizados na Tabela 3. Apesar de

DOI: 10.59254/sbpo-2023-175028
http://live-proceedings.pantheonsite.io/p/175028

	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}	g_{12}
Albânia	55.9	49.2	35.5	89.1	71.0	58.2	70.7	49.7	81.3	82.8	70	70
Argélia	27.8	29.5	28.4	71.9	50.7	12.1	53.6	51.4	75.1	57.5	30	30
Angola	41.1	24.8	22.9	86.5	86.9	85.1	41.6	53.9	58.9	64.8	30	40
Argentina	34.6	56.8	42.8	66.9	53.9	33.6	55.5	55.1	36.5	61.2	55	60
Armênia	50.3	31.5	51.6	87.5	76.3	71.4	64.9	58.7	70.5	73.6	75	70
...	...	...	...	...	...	...	...	...	...	...	...	...
Vanuatu	60.3	72.2	45.5	98.0	45.5	95.4	49.8	40.1	73.8	59.7	65	40
Venezuela	0.0	3.4	6.5	74.8	95.0	12.0	30.8	34.3	0.0	42.3	0	10
Vietnã	48.7	35.1	36.4	78.6	86.5	88.7	73.4	54.1	70.5	79.4	40	50
Zâmbia	36.0	44.4	29.9	74.8	69.6	0.0	43.9	42.1	62.2	65.2	55	50
Zimbábue	20.0	15.8	20.0	65.8	92.6	89.9	39.8	35.4	0.0	54.0	25	10

Tabela 2: Matriz de dados do Economic Freedom Index

visualmente haver uma indicação de que um centroide possa ser preferível a outro, essa ordenação entre os grupos formados pelo k-médias não é parte do algoritmo. Na realidade, apenas o número de *clusters* é indicado no início do método, não existindo nenhuma indicação de que deva existir uma relação de preferências entre objetos (países) em diferentes grupos. Essa característica difere um método de agrupamento (aprendizado supervisionado) de um método de classificação multi-critério. Não obstante, pode-se visualizar os centroides dos *clusters* como possíveis alternativas de um problema de ordenação multicritério. Assim, seria possível alcançar um *ranking* dos centroides dos *clusters* obtidos pelo k-médias. Para isso, utilizou-se o método ELECTRE-III.

	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_10	g_11	g_12
cluster 1	0,53	0,49	0,44	0,81	0,70	0,16	0,61	0,68	0,85	0,71	0,60	0,54
cluster 2	0,87	0,86	0,74	0,68	0,40	0,61	0,82	0,77	0,91	0,83	0,79	0,73
cluster 3	0,29	0,19	0,21	0,81	0,86	0,75	0,42	0,58	0,79	0,66	0,43	0,37
cluster 4	0,51	0,43	0,40	0,85	0,71	0,75	0,64	0,67	0,85	0,75	0,62	0,56
cluster 5	0,18	0,06	0,15	0,56	0,58	0,26	0,26	0,29	0,04	0,43	0,08	0,11

Tabela 3: Matriz de centroides: Economic Freedom Index

Para a aplicação do método ELECTRE-III, foram estabelecidos os limiares de indiferença (q), limiares de preferência (p), limiares de veto (v) e pesos (w) dos critérios, conforme ilustra a Tabela 4.

Parâmetros	Valores
q	0,1
p	0,2
v	0,7
w	0,0833

Tabela 4: Parâmetros do modelo

Inicialmente, foi pesos iguais foram considerados para os critérios do modelo, seguindo o que é realizado pela fundação *Heritage*. A Tabela 5) ilustra as realações obtidas pelo ELECTRE-III

DOI: 10.59254/sbpo-2023-175028
http://live-proceedings.pantheonsite.io/p/175028

para as cinco alternativas, que se referem aos centroides organizados na Tabela 3. O símbolo $P+$ indica que a alternativa da linha sobreclassifica a da coluna, sendo o contrário indicado por $P-$. Além disso, I indica a presença de indiferença entre as alternativas enquanto R indica uma relação de incomparabilidade.

	cluster 1	**cluster 2**	**cluster 3**	**cluster 4**	**cluster 5**	**Posição**
cluster 1	-	P-	P+	P-	P+	3º
cluster 2	P+	-	P+	P+	P+	1º
cluster 3	P-	P-	-	P-	P+	4º
cluster 4	P+	P-	P+	-	P+	2º
cluster 5	P-	P-	P-	P-	-	5º

Tabela 5: Relações Obtidas para a aplicação nos dados do Índice de Liberdade Econômica

Analisando a tabela 5, percebe-se que uma ordenação completa foi obtida. Não há a presença de incomparabilidade ou mesmo indiferença entre as alternativas considerando as avaliações obtidas pelos centroides no conjunto de critérios. É interessante mencionar que o *cluster* 5 (a_5), que foi sobreclassificado por todos os outros, é formado essencialmente por economias altamente restritivas, incluindo Cuba, Venezuela, e Coreia do Norte, Sudão e Zimbabwe. Esses países receberam notas bastante baixas segundo as análises realizadas pela Fundação Heritage no conjunto de critérios da Tabela 1. Já o cluster 2, que consiste na alternativa melhor ordenada, é formado, em sua maioria, por países europeus. Os demais *clusters* são formados, em sua maioria, formado por países em desenvolvimento e subdesenvolvidos. A Figura 1 apresenta uma ilustração dos *clusters*. representando os diferentes grupos de países.

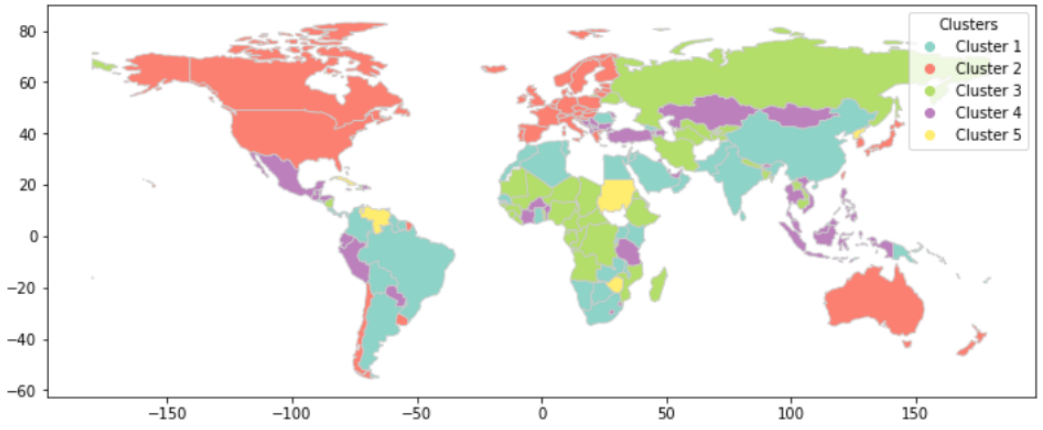


Figura 1: Agrupamento de Países - Índice de Liberdade Econômica

3.2. Índice de Prosperidade (Legatum)

A segunda aplicação segue os mesmos passos realizados para o caso do índice de Liberdade Econômica, porém, dessa vez os dados são analisados de acordo com seus desempenhos dados pelo Instituto Legatum. Os critérios são organizados na terceira coluna da Tabela 1. Os dados referem-se aos desempenhos alcançados no índice de 2023. Os centroides dos cinco *clusters* gerados a partir do algoritmo k-médias são organizados na Tabela 6.

DOI: 10.59254/sbpo-2023-175028
http://live-proceedings.pantheonsite.io/p/175028

	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}	g_{12}
cluster 1	0,67	0,65	0,50	0,56	0,51	0,53	0,56	0,49	0,74	0,72	0,64	0,53
cluster 2	0,43	0,38	0,28	0,37	0,18	0,35	0,14	0,21	0,26	0,36	0,19	0,41
cluster 3	0,66	0,23	0,42	0,52	0,57	0,58	0,63	0,61	0,79	0,79	0,68	0,36
cluster 4	0,88	0,84	0,83	0,71	0,88	0,83	0,83	0,82	0,96	0,88	0,89	0,75
cluster 5	0,51	0,34	0,33	0,44	0,33	0,42	0,36	0,33	0,56	0,60	0,42	0,38

Tabela 6: Matriz de centroides: Índice Legatum de Prosperidade

O método ELECTRE-III foi aplicado com os parâmetros ilustrados na Tabela 4 e os centroides obtidos para o novo conjunto de dados. A Tabela 7 apresenta o *ranking* parcial obtido pelo procedimento. Nesta aplicação, percebe-se que dois grupos de países foram considerados indiferentes pelo método ELECTRE-III, e dividiram a segunda posição da ordenação. O *cluster* 4, melhor posicionado, é composto majoritariamente por países desenvolvidos e em sua maioria do hemisfério norte, com algumas exceções como Uruguai, Chile e Costa Rica. Essas nações apresentam excelentes desempenhos em todos os pilares. Na última posição, o *cluster* 2 é constituído principalmente por países do Oriente Médio e África que enfrentam ainda muitos problemas referentes a segurança, acesso a saúde e educação.

	cluster 1	**cluster 2**	**cluster 3**	**cluster 4**	**cluster 5**	Posição
cluster 1	-	P+	I	P-	P+	2º
cluster 2	P-	-	P-	P-	P-	5º
cluster 3	I	P+	-	P-	P+	2º
cluster 4	P+	P+	P+	-	P+	1º
cluster 5	P-	P+	P-	P-	-	4º

Tabela 7: Relações de Preferência Obtidas e Ordenação: Índice *Legatum* de Prosperidade

A Figura 2 ilustra os *clusters* obtidos para a base de dados do Instituto Legatum. É interessante notar a semelhança nos *clusters* da primeira posição e da última posição nas aplicações dos dois estudos de caso.

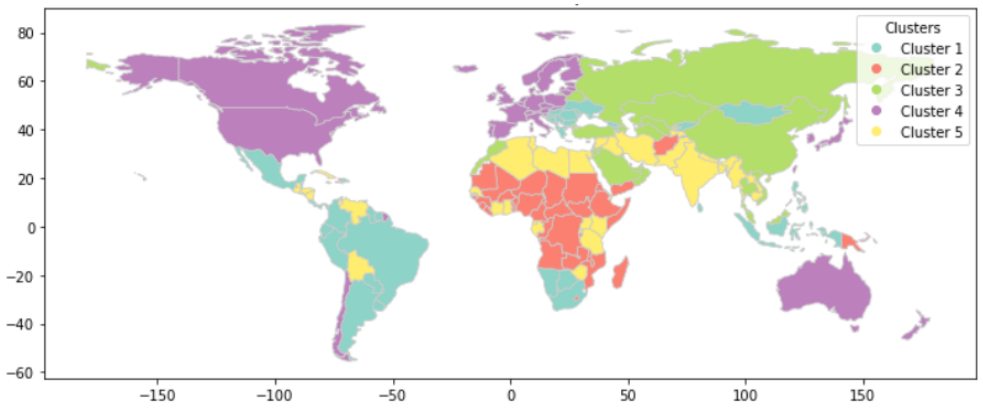


Figura 2: Agrupamento de Países - Índice de Prosperidade

DOI: 10.59254/sbpo-2023-175028
http://live-proceedings.pantheonsite.io/p/175028

4. Análise de Resultados

Como análise posterior à obtenção dos resultados, foi realizado um estudo da robustez das ordenações. Para isso, 10.000 simulações do ELECTRE-III foram realizadas variando os parâmetros de entrada do modelo (pesos, limiares de indiferença e de preferência). Os pesos seguiram a distribuição *Dirichlet*, sendo esta definida como uma distribuição multivariada sobre os pesos (w_i) que garante que $\sum_{i=1}^{n} w_i = 1$ e que $w_i \geq 0, \forall i$. Os limiares de indiferença seguiram uma distribuição Uniforme $U(0,07; 0,13)$ enquanto os limiares de indiferença foram obtidos através de uma distribuição uniforme, com valores entre 0,17 e 0,23 (ou seja, $U(0,17; 0,23)$). Perceba que a definição das distribuições impossibilita casos inconsistentes onde $q_j \geq p_j$ para algum critério g_j.

Para cada simulação, obteve-se a relação binária entre cada par de alternativas (*clusters*). As Tabelas 8 e 9 apresentam a frequência com que cada resultado é econtrado nas simulações. Nessas tabelas, é possível confirmar a robustez das soluções encontradas. Por exemplo, o *cluster 2*, o melhor ranqueado na aplicação envolvendo o índice de Liberdade Econômica, apresenta a relação binária P+ quando comparado aos demais na maioria das simulações ilustradas na Tabela 8.

Tabela 8: Resultados das simulações nos dados do Índice de Liberdade Econômica

	cluster 1	**cluster 2**	**cluster 3**	**cluster 4**	**cluster 5**
cluster 1	-	P+: 16.26% P-: 82.14% I: 0.63% R: 0.97%	P+: 80.76% P-: 10.06% I: 7.79% R: 1.39%	P-: 50.65% I: 49.35%	P+: 100%
cluster 2	P+: 82.14% P-: 16.26% I: 0.63% R: 0.97%	-	P+: 88.50% P-: 3.28% I: 0.53% R: 7.69%	P+: 64.18% P-: 27.33% I: 4.51% R: 3.98%	P+: 100%
cluster 3	P+: 10.06% P-: 80.76% I: 7.79% R: 1.39%	P+: 3.28% P-: 88.50% I: 0.53% R: 7.69%	-	P+: 0.02% P-: 96.73% I: 3.23% R: 0.02%	P+: 100%
cluster 4	P+: 50.65% I: 49.35%	P+: 27.33% P-: 64.18% I: 4.51% R: 3.98%	P+: 96.73% P-: 0.02% I: 3.23% R: 0.02%	-	P+: 100%
cluster 5	P-: 100%	P-: 100%	P-: 100%	P-: 100%	-

Por outro lado, o *cluster* 2 na aplicação envolvendo os dados de Prosperidade recebeu países com desempenhos mais baixos, ficando na quinta colocação na ordenação. Ao analisar os resultados da Tabela 8, percebe-se que a relação binária P- é obtida na vasta maioria dos casos em comparações envolvendo essa alternativa. Neste caso, a alternativa mais bem ranqueada foi o *cluster* 4, o que confirmou-se na Tabela 8. Perceba que esse *cluster*, formado por países mais prósperos (veja a Figura 3), sobreclassificou os demais nas 10000 simulações. Nessa análise podemos destacar ainda que a indiferença encontrada na Tabela 7 entre os *clusters* 1 e 3 foi confirmada na maioria das simulações geradas. Em 66,07% dos casos, houve indiferença entre as alternativas. Considerando as outras simulações, o *cluster* 1 obteve vantagem em 33,92% dos casos enquanto foi sobreclassificado pelo *cluster* 3 apenas em 0,01% dos cenários testados com a as distribuições definidas nesta análise posterior.

DOI: 10.59254/sbpo-2023-175028
http://live-proceedings.pantheonsite.io/p/175028

Tabela 9: Resultados das simulações nos dados do Índice Legatum de Prosperidade

	cluster 1	**cluster 2**	**cluster 3**	**cluster 4**	**cluster 5**
cluster 1	-	P+: 100%	P+: 33.92% P-: 0.01% I: 66.07%	P-: 100%	P+: 100%
cluster 2	P-: 100%	-	P-: 100%	P-: 99.97% I: 0.03%	P-: 100%
cluster 3	P+: 0.01% P-: 33.92% I: 66.07%	P+: 99.97% I: 0.03%	-	P-: 100%	P+: 99.91% P-: 0.06% I: 0.03%
cluster 4	P+: 100%	P+: 100%	P+: 100%	-	P+: 100%
cluster 5	P-: 100%	P+: 97.85% P-: 0.24% I: 1.91%	P+: 0.06% P-: 99.91% I: 0.03%	P-: 100%	-

5. Conclusões

Este artigo apresenta a aplicação do algoritmo k-médias em conjunto com o método de Apoio à Decisão Multicritério ELECTRE-III em duas aplicações envolvendo análise de países. Inicialmente, o k-médias é aplicado para agrupar países de acordo com seus desempenhos nos indicadores usados em cada estudo de caso. Então, os centroides dos *clusters* são ordenados com o uso do ELECTRE-III. Assim, uma informação adicional é obtida sobre os grupos formados pelo procedimento de Aprendizado Não Supervisionado.

A primeira aplicação foi realizada com a base de dados da fundação *Heritage*, que usa um conjunto de indicadores voltados à análise de países em relação ao grau de liberdade econômica. Por sua vez, a segunda aplicação envolve indicadores de prosperidade disponibilizados pelo instituto *Legatum*. Ambas as aplicações envolvem conjuntos de dados com 12 atributos (critérios) e avaliam mais de 100 países.

Os resultados alcançados foram consistentes com as expectativas. Ilustrações apresentadas mostram que os melhores e piores *clusters* são formados por conjuntos semelhantes de países nas duas aplicações. Além disso, uma análise de robustez foi realizada posteriormente variando parâmetros do ELECTRE-III de acordo com distribuições de probabilidades, o que gerou resultados para 10.000 simulações. Foi possível confirmar através dessa análise a consistência das ordenações obtidas e das relações binárias resultados do método ELECTRE-III, conforme apresentado nas Tabelas 8 e 9. Trabalhos futuros podem aplicar a abordagem em diferentes conjuntos de dados, assim como elaborar comparações com o uso de diferentes procedimentos, seja na etapa de agrupamento dos dados por meio de diferentes algoritmos não supervisionados ou na etapa de ordenação dos *clusters* com o uso de outros métodos de Apoio à Decisão Multicritério. Ademais, uma análise a respeito da influência da definição do número de *clusters* em alterações nas relações de sobreclassificação obtidas pode ser contemplada em trabalhos futuros.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, FAPERJ (E-26/200.974/2022) e CNPq (314953/2021-3).

DOI: 10.59254/sbpo-2023-175028
http://live-proceedings.pantheonsite.io/p/175028

Referências

Azevedo, G. D. C. e Costa, H. G. (2022). Applying non-compensatory multicriteria methods to build better life index countries ranking. In *2022 IEEE LATIN AMERICAN CONFERENCE ON COMPUTATIONAL INTELLIGENCE (LA-CCI)*, p. 53–58. ISBN 978-1-6654-8858-7. IEEE Latin American Conference on Computational Intelligence (LA-CCI), Montevideo, URUGUAY, NOV 23-25, 2022.

de Souza, R. L. S. e Costa, H. G. (2021). K-Means and Multicriteria Decision Aid Applied to Sustainability Evaluation. p. 1198–1208. URL https://link.springer.com/10.1007/978-3-030-71187-0_111.

Faceli, K., Lorena, A. C., Gama, J., de Almeida, T. A., e de Leon Ferreira de Carvalho, A. C. P. (2021). *Inteligência Artificial: uma abordagem de aprendizado de máquina*. LTC, 2 edition.

Greco, S., Ehrgott, M., e Figueira, J. R., editors (2016). *Multiple Criteria Decision Analysis*, volume 233 of *International Series in Operations Research Management Science*. Springer New York, New York, NY, 2 edition. ISBN 978-1-4939-3093-7. URL http://link.springer.com/10.1007/978-1-4939-3094-4.

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 3 edition.

Ozmen, M. (2019). Logistics competitiveness of oecd countries using an improved todim method. *SADHANA-ACADEMY PROCEEDINGS IN ENGINEERING SCIENCES*, 44(5). ISSN 0256-2499.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., e Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pereira, V. (2023). pyDecision. URL https://github.com/Valdecy/pyDecision/.

Rani, V., Nabi, S. T., Kumar, M., Mittal, A., e Kumar, K. (2023). Self-supervised learning: A succinct review. *ARCHIVES OF COMPUTATIONAL METHODS IN ENGINEERING*. ISSN 1134-3060.

Rogers, M., Bruen, M., e Maystre, L.-Y. (2000). *ELECTRE and Decision Support*. Springer US. ISBN 978-1-4419-5108-3.

Roy, B. (1978). Electre iii: Algorithme de classement basé sur une représentation floue des préférences en présence de critères multiples. *Cahiers du CERO*, 20:3–24.

Siksnelyte-Butkiene, I. e Streimikiene, D. (2022). Sustainable development of road transport in the eu: Multi-criteria analysis of countries' achievements. *ENERGIES*, 15(21).

DOI: 10.59254/sbpo-2023-175028
http://live-proceedings.pantheonsite.io/p/175028