EVENTO ON-LINE
III CIGELUBRA
Congresso Internacional de Gestão e Engenharia Luso-Brasileiro
02 a 04 de dezembro de 2026
Inovação e Transformação Digital como Estratégia Competitiva nas Engenharias

UM ESTUDO DE CASO DE CLUSTERS POR EVENTOS CLIMÁTICOS EXTREMOS PARA APOIAR AÇÕES DE LOGÍSTICA HUMANITÁRIA NO ESTADO DO PARANÁ

Gustavo Alarcão
Departamento de Engenharia de Produção e Sistemas - DEPS/UFSC

Ricardo Villarroel Dávalos
Departamento de Engenharia de Produção e Sistemas - DEPS/UFSC

Daniel Lorenzetto
Coordenadoria Estadual de Defesa Civil do Estado do Paraná – CEDC/PR

Fabiana Santos Lima
Programa de Pós-Graduação em Educação Científica e Tecnológica – UFSC/CED/PPGECT

Resumo: O estado do Paraná enfrenta uma ampla diversidade de desastres naturais anualmente, que são responsáveis por prejuízos financeiros expressivos, ferimentos e eventuais óbitos. Entre 2010 e 2024, os desastres naturais foram responsáveis por 48 óbitos e aproximadamente 18.740 feridos e enfermos no estado, além de R$ 2,45 bilhões em prejuízo público e R$ 33,57 bilhões em prejuízo privado. Durante esse período, ocorreram 1598 desastres, com uma distribuição por grupos de 53,19% para hidrológicos, 23,40% meteorológicos e 21,96% climatológicos (BRASIL, 2023). Este artigo tem como principal objetivo desenvolver um procedimento para apoiar a formação de *clusters* baseados nas ocorrências de desastres no estado do Paraná ocorridos entre 2010 e 2024. Para isso, é utilizada uma base de dados fornecida pela Coordenadoria Estadual de Defesa Civil do Paraná (CEDC/PR). Neste estudo, foram utilizados conceitos de descoberta de conhecimento em bases de dados com o intuito de estabelecer uma análise exploratória dos dados que possa contribuir para a melhoria da logística humanitária no estado. A aplicação do algoritmo K-Means às 39 microrregiões, considerando três atributos definidos pelo agrupamento de desastres com perfis semelhantes de recebimento de materiais, resultou na seleção de sete *clusters*. Foram identificados perfis com baixa frequência de ocorrências, concentrações de eventos associados a alagamentos, chuvas intensas, enxurradas e inundações, além de uma microrregião com elevada frequência nas três categorias e outra com destaque para estiagens. A interpretação desses perfis, junta aos registros de materiais enviados, forneceu subsídios para diferenciar as prioridades de preparação e disponibilidade de suprimentos entre as microrregiões.

Palavras-chave: Logística humanitária, Desastres naturais, Análise por *clusters*, *K-Means*, Descoberta de conhecimento em bases de dados.

A Case Study of Extreme Weather Event Clusters to Support Humanitarian Logistics Actions in the State of Paraná

Abstract: The state of Paraná, Brazil, experiences diverse natural disasters each year, causing substantial financial losses, injuries, and deaths. Between 2010 and 2024, these disasters caused 48 deaths and approximately 18,740 cases of injury or illness, alongside BRL 2.45 billion in public

losses and BRL 33.57 billion in private losses. During this period, 1,598 disasters were recorded; hydrological, meteorological, and climatological events accounted for 53.19%, 23.40%, and 21.96%, respectively (BRASIL, 2023). This study aimed to develop a procedure for forming clusters based on disaster occurrences in Paraná between 2010 and 2024. Data provided by the Paraná State Civil Defense Coordination Office (CEDC/PR) were analyzed using the Knowledge Discovery in Databases framework to support exploratory analysis and improve humanitarian logistics in the state. The K-Means algorithm was applied to 39 geographical microregions using three attributes constructed by grouping disaster types with similar profiles of relief supplies received. The analysis led to the selection of seven clusters. The identified profiles included low disaster frequencies; concentrations of urban flooding, heavy rainfall, flash floods, and river flooding; one microregion with high frequencies across all three categories; and another characterized by frequent droughts. Interpreting these profiles alongside records of dispatched relief supplies provided a basis for differentiating preparedness and supply availability priorities across microregions.

Keywords: Humanitarian logistics, Natural disasters, Cluster analysis, K-Means, Knowledge discovery in databases.

1. Introdução

O cenário global atual está marcado por um aumento na frequência e intensidade de eventos climáticos extremos, impulsionados pelas mudanças climáticas e pela ocupação desordenada do solo. Nesse contexto, os desastres naturais representam uma das maiores ameaças à estabilidade socioeconômica e à segurança humana, exigindo respostas rápidas e estratégias de mitigação eficientes.

O Estado do Paraná apresenta uma vulnerabilidade significativa a eventos climáticos extremos, como secas, inundações, enxurradas, deslizamento de terra, incêndios florestais, tempestades fortes e outros. Um levantamento do TCE-PR (Tribunal de Contas do Estado do Paraná) junto às prefeituras aponta que 60% destas não possuem estrutura de Defesa Civil e que 70% não contam com sistema de alerta de riscos para moradores. Também adverte que mais da metade dos 399 municípios paranaenses não possui estrutura física para enfrentar desastres climáticos (TCE-PR, 2026).

Neste contexto, a Logística Humanitária (LH) assume um papel imprescindível na prevenção e mitigação visando reduzir a vulnerabilidade das populações e minimizar danos antes que eventos adversos ocorram. No entanto, a capacidade de resposta e prevenção ainda é limitada devido à falta de metodologias avançadas que integrem análise de dados históricos, previsões climáticas e tecnologias de ponta (SHARMA; RAVINDRANATH, 2019).

A análise de *clusters* é uma técnica que procura numa base de dados grupos que possuem características semelhantes como, por exemplo, na LH, grupos de municípios que apresentam os mesmos desastres naturais com determinadas solicitações de itens humanitários para aquisição junto aos fornecedores. Estes *clusters* podem ser definidos a partir de diferentes atributos, tais como localização geográfica, tipos de desastres naturais, tipos de itens humanitários solicitados, dentre outros.

No Brasil, cada estado possui uma Coordenadoria Estadual da Defesa Civil - CEDC que é responsável pela coordenação das ações envolvidas na gestão dos desastres e atua em dois segmentos: prevenção e resposta. Conta com vários centros de distribuição de itens humanitários estando localizados em várias microrregiões do estado.

Assim, para o Estado do Paraná, para cada ocorrência de qualquer tipo de desastre natural existem solicitações para a *CEDC/PR* de diferentes itens de primeira necessidade, sendo estes por exemplo: cestas básicas, colchões de solteiro, kits de limpeza, galões de água, etc.

Este artigo tem como principal objetivo desenvolver uma abordagem para formar *clusters* por combinação de frequências de desastres no estado do Paraná, a qual pode ser utilizada para auxiliar na coordenação estratégica, na definição de prioridades e no compartilhamento de experiências em determinadas microrregiões do Estado do Paraná.

2. Revisão Bibliográfica

2.1 Desastres Naturais

Desastres naturais são divididos em grupos, que por sua vez são divididos em subgrupos e então em subdivisões menores. Os grupos de desastres naturais são: geológico, hidrológico, meteorológico, climatológico e biológico (CLASSIFICAÇÂO, 2012). Este projeto direciona sua análise a subgrupos, pois isso permite uma interpretação mais abrangente do que através da classificação em grupos. Neste texto, cada subgrupo de desastres será denominado simplesmente “desastre".

2.2 Logística Humanitária

De acordo com Thomas e Kopczak (2005, p. 2), a logística humanitária é o processo de planejamento, implementação e controle do fluxo e da armazenagem de bens e materiais, de maneira eficiente e economicamente eficaz, além da informação relacionada, do ponto de origem ao ponto de consumo.

Em dezembro de 2004, um terremoto na Indonésia e o subsequente tsunami chamaram a atenção do mundo. Esse evento foi responsável pela morte de 230.000 pessoas. Embora mais de 40 países e 700 ONGs tenham enviado ajuda, a magnitude do desastre trouxe significativos desafios operacionais e logísticos para a assistência humanitária (LEIRAS et al., 2017 apud NEGREIROS, 2022).

Para aprimorar a assistência humanitária, os *clusters* surgiram como mecanismo de coordenação para esse processo. O conceito de *cluster* envolve organizar a ajuda humanitária de acordo com certos setores e com uma liderança predeterminada (JAHRE; JENSEN, 2010, p. 6). É importante notar a diferença entre esse conceito de *cluster* e os algoritmos de clusterização, sendo o segundo o foco deste estudo.

2.3 Descoberta de Conhecimento em Bases de Dados

Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996, p. 39-41), a descoberta de conhecimento em bases de dados, do inglês *Knowledge Discovery in Databases* (KDD), é o processo geral (multidisciplinar) de obtenção de conhecimento a partir de dados. O termo processo implica a presença de várias etapas, e um resumo destas pode ser visto na Figura 1.

Figura 1 – Processo de KDD



A aplicação do processo começa com uma escolha precisa da base de dados que servirá de objeto de estudo. Além disso, a etapa de seleção também pode consistir em separar o conjunto de dados apenas nos subconjuntos de interesse. Para que essa etapa possa ser eficiente, é importante entender o objetivo do processo de KDD a partir do ponto de vista do cliente final.

O pré-processamento consiste na preparação inicial dos dados e inclui operações que não alterem o significado analítico dos dados, como formatação, remoção de ruído, definição da abordagem para lidar com informações ausentes e agrupamento dos dados de diferentes fontes. Além do mais, esse passo também inclui o gerenciamento das estruturas de dados, como para datas e valores numéricos.

A partir disto, a transformação engloba todas as operações que deixarão os dados prontos para serem usados como entrada de algum algoritmo de mineração. Essa etapa então cuida da redução dos dados a atributos que possam ser úteis para o objetivo final. Além disso, a aplicação de mudanças estruturais, como a padronização por escore-z nesse projeto, é necessária para aplicar os modelos de mineração de maneira adequada.

É importante notar a diferença entre KDD e mineração de dados, termos que são utilizados como sinônimos. O processo de KDD envolve passos adicionais como preparação, seleção e limpeza dos dados, incorporação de conhecimento a priori e a interpretação dos resultados obtidos na mineração. A mineração consiste na definição de um algoritmo (por exemplo, de clusterização) e de seus parâmetros, na execução adequada desse algoritmo e na inspeção e validação dos resultados retornados.

Por fim, segue a etapa de interpretação, em que se avaliam os resultados obtidos na mineração e os preparam para servir de base direta para tomada de ações. O processo de KDD pode envolver significativa iteração e conter ciclos entre quaisquer duas etapas (FAYYAD et al., 1996, p. 42).

A aplicação de técnicas de clusterização à logística humanitária já era explorada em 2011. Lima, Oliveira e Gonçalves (2011) propuseram a formação de *clusters* de regiões de Santa Catarina com base na frequência de diferentes tipos de desastres, situando essa abordagem na etapa de mineração de dados do processo de KDD.

2.4 Clusterização

Neste trabalho, todos os vetores são considerados vetores coluna, salvo indicação em contrário. O sobrescrito T indica a transposição.

A análise por *clusters* pode ter uma grande variedade de objetivos, todos relacionados à divisão de um conjunto de objetos em subconjuntos (*clusters*), tal que elementos de um subconjunto são mais próximos uns dos outros do que de elementos de outro subconjunto (HASTIE et al., 2009, p. 501). Para medir a proximidade entre dois pontos x_i e $x_{i'}$, uma ferramenta comum é a distância euclidiana $\left\| x_i - x_{i'} \right\|_2$, que é a utilizada nesse trabalho.

Segundo Zaki e Meira Jr. (2020, p. 334), para uma base de dados $D \in R^{n \times d}$ com *n* objetos $x_i^T \in R^d$ e d atributos $X_j \in R^n$, o objetivo da clusterização baseada em representantes é particioná-la em um *clustering* $C = \{C_0, C_1, \ldots, C_{k-1}\}$. Assim, para cada *cluster* $C_q \in C$ existe um ponto representativo, normalmente escolhido como a média (centroide) μ_q de todos os seus pontos, isto é definido pela equação a seguir.

$$\mu_q = \frac{1}{|C_q|} \sum_{x_i \in C_q} x_i. \tag{1}$$

O algoritmo *K-Means* faz parte da classe de métodos de clusterização baseada em representantes, e é um método bastante popular. De acordo com Zaki e Meira Jr. (2020, p. 334-335), dados todos os possíveis *clusterings* C em D, o objetivo deste algoritmo é achar o *clustering* C^* que minimiza a soma dos quadrados dos erros, conforme definido na equação a seguir.

$$C^* = \arg\min_{C} \sum_{q=0}^{k-1} \sum_{x_i \in C_q} \left\| x_i - \mu_q \right\|_2^2. \tag{2}$$

Existem várias maneiras de medir a validade de uma clusterização, sendo estas divididas em 3 principais categorias: externa, interna e relativa. As medidas de validação interna utilizam critérios derivados da base de dados, enquanto as de validação externa não. Já as de validação relativa focam em comparar diferentes *clusterings*, variando os parâmetros do modelo (ZAKI; MEIRA JR., 2020, p. 426).

Nesse projeto, analisou-se a compacidade e a separação dos *clusters*, através de coeficientes de silhueta (ROUSSEEUW, 1987, p. 55). Para um objeto $x_i \in C_q$, seu coeficiente de silhueta é dado pela Equação 3.

$$s(x_i) = \frac{b(x_i) - a(x_i)}{max\{a(x_i), b(x_i)\}}, \tag{3}$$

em que $a(x_i)$ é a distância média entre x_i e os outros objetos de C_q enquanto $b(x_i)$ é a distância média de x_i para os pontos de um *cluster* vizinho. Segue que $-1 \leq s(x_i) \leq 1$, para todo x_i, e um valor próximo a $+1$ indica uma boa clusterização do ponto, enquanto um coeficiente próximo de -1 indica que provavelmente houve uma má classificação do objeto.

Os coeficientes de silhueta são essencialmente uma medida de validação interna, mas podem ser utilizados para uma validação relativa. Uma maneira apropriada de escolher o número de *clusters* k é escolher o valor de k que maximiza a largura média global da silhueta, definida na Equação 4 (ROUSSEEUW, 1987, p. 59),

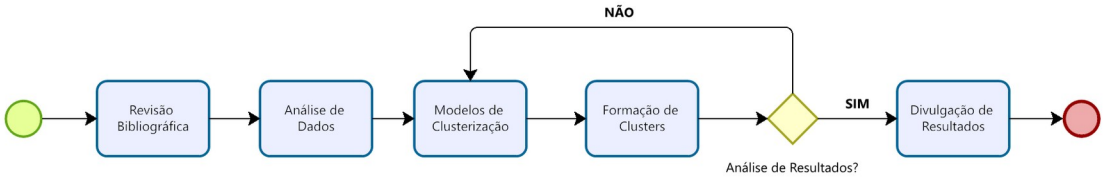
$$\bar{s}(k) = \frac{1}{n}\sum_{i=1}^{n} s(x_i). \tag{4}$$

O presente estudo utiliza a média global $\bar{s}(k)$ para identificar valores candidatos de k, e, em seguida, analisa os gráficos de silhueta (ROUSSEEUW, 1987, p. 58) para cada um desses valores de k.

3. Metodologia

O procedimento metodológico utilizado encontra-se definido como um estudo de caso para analisar dados históricos de situações de desastres extremos e identificar padrões e microrregiões do estado paranaense. Considerando o campo multidisciplinar que engloba diversas técnicas, algoritmos e o uso de metodologias quantitativas, foi possível adotar uma abordagem mais eficiente na identificação de *clusters* georreferenciados e na padronização de métodos de prevenção, contribuindo para um planejamento integrado e eficaz diante da iminência destes eventos climáticos (MIGUEL, 2007). A Figura 2 ilustra as etapas metodológicas utilizadas para este artigo.

Figura 2 – Procedimento metodológico



A primeira etapa representou a revisão bibliográfica na qual se procuraram conceitos importantes e aplicações de LH, formação de clusters para gerenciamento da LH e roteirização de modelos de clusterização. A etapa seguinte dedicou-se à análise e organização de dados históricos e georreferenciados para identificação de padrões e tendências em desastres causados por eventos climáticos extremos. Na terceira etapa empenhou-se em implementar em uma ferramenta computacional os modelos de clusterização georreferenciados e proceder com sua experimentação.

Na quarta etapa foi efetuada a formação de *clusters* e identificação de microrregiões semelhantes utilizando métodos quantitativos para formar *clusters* de microrregiões com características climáticas e geográficas semelhantes. Na seguinte etapa foram avaliados os resultados a partir das práticas da CEDC/PR realizadas na resposta às solicitações de itens humanitários por situação de desastre natural.

4. Estudo de Caso

Adotou-se como abordagem definir cada atributo X_j, com $j \in \{1,\ldots,d\}$, como sendo o total de ocorrências de um agrupamento de desastres. Isso será discutido com mais detalhes na seção Transformação. Já para os objetos do modelo, cada x_i^T, com $i \in \{1,\ldots,n\}$, representa uma das 39 microrregiões do estado do Paraná, de modo que $n=39$. Assim, D_{ij} representa o total de ocorrências dos desastres no agrupamento X_j, na região representada por x_i^T.

A manipulação dos dados foi feita principalmente a partir da biblioteca *Pandas*, associada à linguagem de programação *Python*. Já a base para a implementação do algoritmo *K-Means* e do cálculo dos coeficientes de silhueta foi a biblioteca *Scikit-Learn*. O código completo está disponível em https://github.com/gustavoalarcao/clustering_parana. O repositório também inclui um arquivo *.csv* associando cada microrregião a um *cluster.*

4.1 Seleção

Após reunir os dados da CEDC/PR no diretório do projeto, foi realizada a remoção de colunas não relevantes para o escopo do projeto, como o prefeito dos municípios durante a data do desastre. Além do mais, desastres biológicos foram desconsiderados, devido ao horizonte do estudo, direcionado a desastres geográficos.

4.2 Pré-Processamento

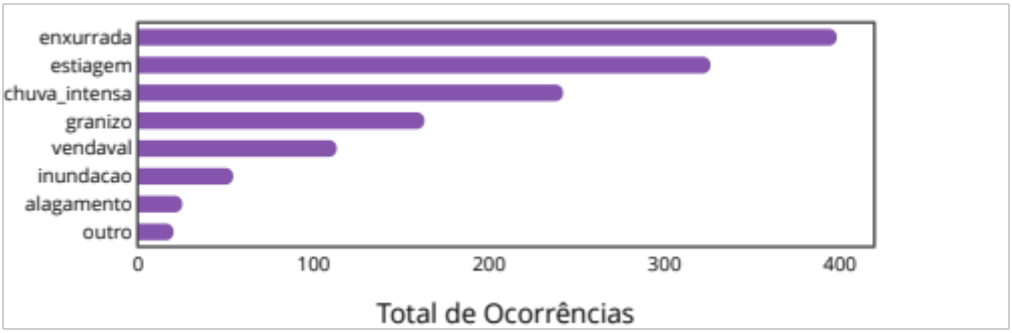
Inicialmente, foram realizadas mudanças para facilitar a manipulação dos dados. Para isso, foram feitas a remoção e substituição de caracteres especiais, caracteres maiúsculos e espaços. Além disso, parte das colunas foram renomeadas e reordenadas.

Agora, com a base de dados mais tratável, a próxima etapa foi alterar as estruturas de dados das colunas. As colunas envolvendo valores monetários precisaram ser processadas para então serem convertidas em tipos numéricos e aquelas envolvendo datas passaram por processo semelhante para então ser convertidas em tipos temporais. Em seguida, cada município foi associado a sua respectiva microrregião, com base na relação de municípios disponibilizada pelo Governo do Paraná.

4.3 Transformação

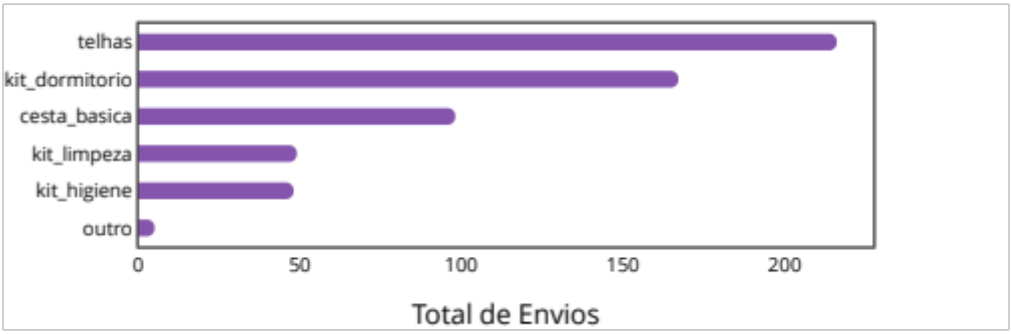
Como a base de dados apresentava categorias muito específicas de materiais enviados e de desastres, algumas com apenas uma ocorrência, foi necessário agrupá-las em categorias mais abrangentes. Os desastres foram agrupados nas categorias apresentadas na Figura 3.

Figura 3 – Frequência de ocorrência dos desastres



Os materiais enviados, por sua vez, foram agrupados nas categorias apresentadas na Figura 4.

Figura 4 – Frequência de envio dos materiais



Após o agrupamento apresentado nas Figuras 3 e 4, foram excluídos os registros de ocorrências e de envios classificados na categoria “outro", por apresentarem baixa frequência e não contribuírem de maneira significativa para a análise geral.

Para prosseguir com o agrupamento dos desastres agora com base no tipo de ajuda associado, foram examinadas estatísticas descritivas da distribuição das quantidades de materiais enviadas, conforme descrito na Tabela 1.

Tabela 1 – Estatísticas descritivas das quantidades de material enviadas

Material	Média	CV	Menor envio	Maior envio
Kit dormitório	102,83	0,73	10	500
Kit higiene	115,98	0,70	15	380
Kit limpeza	123,59	0,77	14	500
Cesta básica	223,66	1,15	10	2000
Telhas	2050,64	0,76	51	10000

Conforme apresentado na Tabela 1, os coeficientes de variação das quantidades de materiais por envio, entre $0,70$ e $1,15$, evidenciam uma variabilidade considerável no tamanho dos envios em cada categoria. Assim, utilizar apenas o número de envios para agrupar os desastres desconsideraria diferenças nas quantidades de materiais fornecidas.

Complementarmente, as quantidades médias por envio diferiram consideravelmente entre as categorias de materiais. Dessa forma, dividiu-se a quantidade de material de cada envio pela média da respectiva categoria, expressando os valores em escalas comparáveis, relativas ao tamanho médio dos envios. A Tabela 2 apresenta os totais normalizados de materiais enviados por tipo de desastre.

Tabela 2 – Totais normalizados de materiais enviados por tipo de desastre

Desastre	Cesta básica	Kit dormitório	Kit higiene	Kit limpeza	Telhas
Alagamento	14,56	19,07	3,04	3,80	0
Chuva intensa	7,12	17,72	4,93	10,69	0
Enxurrada	19,08	27,61	9,70	8,01	0,41
Estiagem	2,42	0	0	0	0
Granizo	29,97	58,79	12,62	11,44	162,98
Inundação	6,84	12,74	9,75	9,47	0
Vendaval	19,99	26,25	3,22	3,98	50,66

Para definir as categorias de desastres, buscou-se identificar padrões semelhantes de recebimento de materiais. Para isso, foi analisada a matriz de correlação de Pearson entre os subgrupos de desastre, considerando as quantidades normalizadas dos materiais recebidos, conforme a Tabela 3.

Tabela 3 – Matriz de correlação de Pearson entre os desastres

Desastre	Alagamento	Chuva intensa	Enxurrada	Estiagem	Granizo	Inundação	Vendaval
Alagamento	1	0,78	0,98	0,44	-0,28	0,59	-0,04
Chuva intensa	0,78	1	0,84	-0,08	-0,47	0,86	-0,34
Enxurrada	0,98	0,84	1	0,32	-0,41	0,74	-0,19
Estiagem	0,44	-0,08	0,32	1	-0,22	-0,11	-0,02
Granizo	-0,28	-0,47	-0,41	-0,22	1	-0,78	0,96
Inundação	0,59	0,86	0,74	-0,11	-0,78	1	-0,71
Vendaval	-0,04	-0,34	-0,19	-0,02	0,96	-0,71	1

As correlações entre alagamento, chuva intensa, enxurrada e inundação variaram de 0,59 a 0,98, enquanto a correlação entre granizo e vendaval foi de 0,963. Com base nesses resultados, foram definidos os seguintes agrupamentos para a construção dos atributos do modelo, ilustrados na Equação 5:

$$\begin{cases} X_1 = \{\text{alagamento}, \text{chuva intensa}, \text{enxurrada}, \text{inundação}\} \\ X_2 = \{\text{estiagem}\} \\ X_3 = \{\text{granizo}, \text{vendaval}\}. \end{cases} \quad (5)$$

Como a escala de medida dos atributos é a mesma, a transformação de D para aplicar o algoritmo K-Means foi simples. Para isso executou-se uma padronização por escore-z em cada X_j, conforme ilustrado na equação a seguir:

$$Z_j = \frac{X_j - 1_n \mu_j}{\sigma_j}, \quad (6)$$

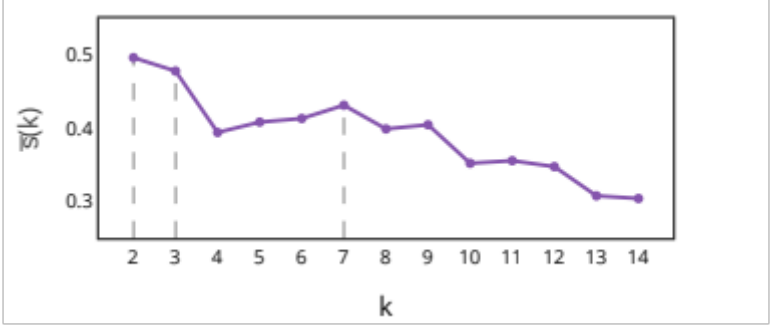
em que $\sigma_j > 0$ é o desvio-padrão do atributo, μ_j sua média e $1_n \in R^n$ é o vetor com 1 em todas as componentes. É importante esclarecer que os Z_j são utilizados apenas para ajustar o modelo de clusterização, e então as análises voltam a considerar os atributos X_j.

4.4 Mineração

Para reduzir a soma dos quadrados dos erros e aumentar a chance de convergência para o *clustering* ótimo C^*, definido em (2), o algoritmo foi executado 30 vezes para cada semente, com 10 sementes diferentes. Além do mais, utilizou-se o método de inicialização *k-means*++, que reduz o tempo de convergência (SCIKIT-LEARN DEVELOPERS, 2026).

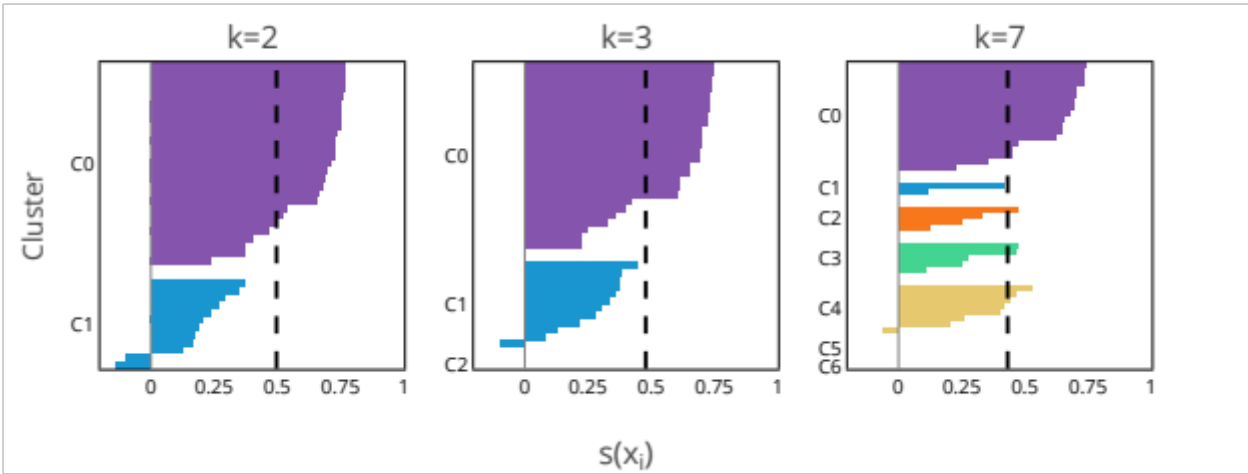
Como discutido no início da seção, cada objeto x_i^T apresentava uma contagem de ocorrências associada a cada X_j. Após a padronização avaliou-se a largura média global da silhueta de cada *clustering* C pela equação (4), conforme ilustrado na figura a seguir:

Figura 5 – Largura média global da silhueta para cada k



Os resultados exibidos na figura anterior motivaram a seleção dos valores 2, 3 e 7 como candidatos ao valor de k. Para uma análise mais aprofundada, visualizam-se as silhuetas dos *clusters* para esses valores de k, ilustradas na Figura 6.

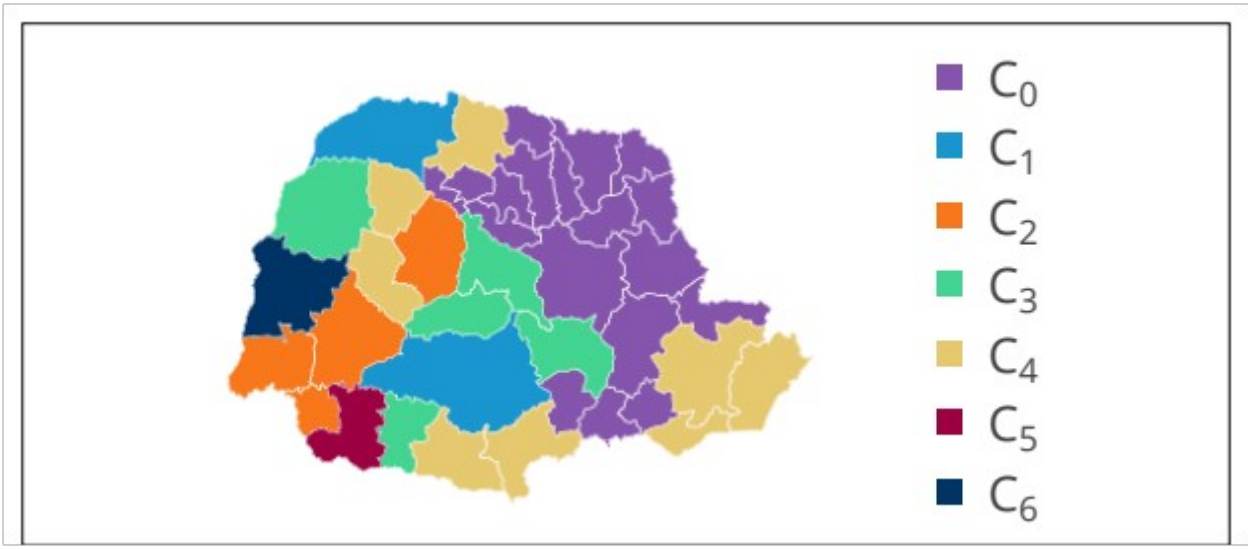
Figura 6 – Silhueta dos *clusters* para cada k de interesse



Os *clusters* sem barras visíveis correspondem a grupos com uma única microrregião, para os quais a silhueta é convencionada como zero. Para $k=7$, os valores negativos de silhueta apresentaram menor magnitude em relação às clusterizações com $k=2$ e $k=3$, sugerindo menor inadequação na atribuição dos objetos correspondentes. Além disso, essa solução permite uma segmentação mais detalhada dos perfis de desastre, com uma redução relativa de, no máximo, 13% na largura média global da silhueta $\bar{s}(k)$ em relação às demais soluções avaliadas. Portanto, este estudo define $k=7$ para ajustar o modelo *K-Means*.

Assim, foi possível particionar o estado do Paraná nos seguintes *clusters* ilustrados na figura a seguir:

Figura 7 – Mapa das microrregiões do Paraná



4.5 Interpretação

Para visualizar a necessidade de materiais associada a cada atributo, elaborou-se a Tabela 4.

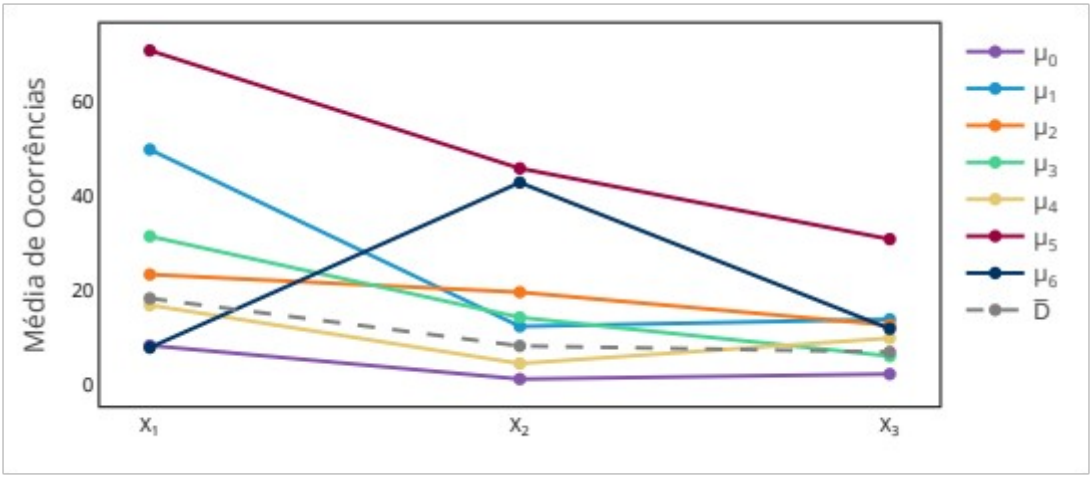
Tabela 4 – Totais normalizados de materiais enviados por atributo

Atributo	Cesta básica	Kit dormitório	Kit higiene	Kit limpeza	Telhas
X_1	47,60	77,14	27,41	31,97	0,41
X_2	2,42	0	0	0	0
X_3	49,96	85,04	15,85	15,43	213,63

Segue que os desastres das categorias X_1 e X_3 apresentam elevados volumes de recebimento de cestas básicas e kits dormitório. Em X_3, destaca-se também o envio de telhas. Os kits de higiene e de limpeza aparecem em quantidades relativas menores que as cestas básicas e os kits dormitório em ambas as categorias. Já em X_2, o fornecimento de materiais é irrisório em relação aos demais.

A Figura 8 exibe os centroides (1) de cada *cluster*, além da média geral $\overline{D}$ da base de dados (média de cada X_j).

Figura 8 – Centroide de cada *cluster*



A análise de centroides evidencia que o *cluster* 0 possui poucas ocorrências de desastres, com $\mu_0 < \overline{D}$ e que o *cluster* 4 possui frequência média ao longo dos X_j, isto é, $\mu_4 \approx \overline{D}$. Já para o *cluster* 3, observam-se valores próximos à média, exceto por $\mu_{3,1} > \overline{D}_1$. Enquanto isso, o *cluster* 2 apresenta $\mu_2 > \overline{D}$. É importante notar que as desigualdades entre vetores acima são aplicadas componente a componente. Por fim, para os *clusters* 1, 5 e 6 foram observados valores extremos: $\mu_{1,1} \gg \overline{D}_1$, $\mu_{6,2} \gg \overline{D}_2$ e μ_5 é bem maior que a média $\overline{D}$ em todos os atributos.

5. Considerações finais

Os perfis identificados fornecem subsídios para orientar a preparação de materiais nas microrregiões de cada *cluster*. Para os *clusters* 1 e 3, a maior frequência de desastres de X_1 sugere atenção à disponibilidade de cestas básicas e kits dormitório, complementados por kits de higiene e limpeza. Já para os *clusters* 2 e 5, as frequências acima da média nos três atributos sugerem uma preparação que contemple todos os materiais analisados, incluindo telhas para os eventos de X_3. Essa preparação merece atenção especial no *cluster* 5, que apresentou as maiores frequências nas três categorias.

Para o *cluster* 6, as ocorrências de X_3 também sugerem manter disponíveis telhas, cestas básicas e kits dormitório, além de kits de higiene e limpeza conforme a necessidade. Entretanto, sua elevada frequência de estiagens exige uma avaliação complementar, pois os poucos envios registrados em X_2 não permitem caracterizar as necessidades de

materiais associadas a esses eventos. Por sua vez, os *clusters* 0 e 4 não apresentaram destaque semelhante em uma categoria específica, sugerindo manter disponibilidade do conjunto de materiais e direcionar os envios conforme as características de cada ocorrência.

Essas indicações se baseiam no histórico de ocorrências e de materiais enviados, não permitindo determinar quantidades de estoque. Assim, a definição dos materiais e das quantidades a enviar deve considerar as necessidades da população afetada, uma vez que o fornecimento registrado não representa necessariamente toda a demanda existente.

Para trabalhos futuros, pretende-se aplicar medidas de validação externa, a partir da construção de uma ou mais tabelas de contingência. E, para complementar a análise, recomenda-se a implementação de outras técnicas de clusterização, como métodos hierárquicos. Além do mais, uma alternativa para basear a logística humanitária nas novas divisões geográficas do estado é utilizar as 29 regiões geográficas imediatas como objetos do modelo.

Referências

BRASIL. Ministério da Integração e do Desenvolvimento Regional. **Atlas Digital de Desastres no Brasil**. Brasília, DF: MIDR, 2023. Disponível em: https://atlasdigital.mdr.gov.br/paginas/graficos.xhtml. Acesso em: 13 set. 2026.

CLASSIFICAÇÃO e codificação brasileira de desastres (COBRADE). *[S. l.: s. n.]*, 2012. Disponível em: https://defesacivil.rs.gov.br/upload/arquivos/202105/04095316-cobrade-classificacao-e-codificacao-brasileira-de-desastres.pdf. Acesso em: 29 ago. 2026.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, Menlo Park, v. 17, n. 3, p. 37–54, Fall 1996. DOI: https://doi.org/10.1609/aimag.v17i3.1230. Disponível em: https://doi.org/10.1609/aimag.v17i3.1230. Acesso em: 23 set. 2026.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer, 2009. DOI: https://doi.org/10.1007/978-0-387-84858-7. Disponível em: https://doi.org/10.1007/978-0-387-84858-7. Acesso em: 23 set. 2026.

JAHRE, M.; JENSEN, L.-M. Coordination in humanitarian logistics through clusters. **International Journal of Physical Distribution & Logistics Management**, Bradford, v. 40, n. 8/9, p. 657–674, 2010. DOI: https://doi.org/10.1108/09600031011079319. Disponível em: https://doi.org/10.1108/09600031011079319. Acesso em: 23 set. 2026.

LIMA, F. S.; OLIVEIRA, D.; GONÇALVES, M. B. A formação de clusters na logística humanitária utilizando mineração de dados. *In*: CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES, 25., 2011, Belo Horizonte. **Anais [...]**. Rio de Janeiro: ANPET, 2011. p. 1638–1649. Disponível em: https://www.researchgate.net/publication/265796836_A_FORMACAO_DE_CLUSTERS_NA_LOGISTICA_HUMANITARIA_UTILIZANDO_MINERACAO_DE_DADOS. Acesso em: 23 set. 2026.

MIGUEL, P. A. C. Estudo de caso na engenharia de produção: estruturação e recomendações para sua condução. **Produção**, São Paulo, v. 17, n. 1, p. 216-229, 2007.

NEGREIROS, F. What is Humanitarian Logistics, and why does it matter? *In*: INSTITUTO BRASIL LOGÍSTICA (org.). **Humanitarian Logistics Guide**. [Brasília, DF]: IBL-Social, [2022]. p. 8–16. Disponível em: https://ibl.org.br/wp-content/uploads/2021/08/HumanitarianLogisticsGuideIBL2022.pdf. Acesso em: 23 set. 2026.

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, Amsterdam, v. 20, p. 53–65, Nov. 1987. DOI: https://doi.org/10.1016/0377-0427(87)90125-7. Disponível em: https://www.sciencedirect.com/science/article/pii/0377042787901257. Acesso em: 23 set. 2026.

SCIKIT-LEARN DEVELOPERS. KMeans. *In*: SCIKIT-LEARN DEVELOPERS. **Scikit-learn documentation**. *[S. l.]*: scikit-learn developers, c2026. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html. Acesso em: 13 set. 2026.

SHARMA, J.; RAVINDRANATH, N. H. Applying IPCC 2014 framework for hazard-specific vulnerability assessment under climate change. ***Environmental Research Communications***, v. 1, n. 5, 051004, 2019.

TCE-PR, Paraná: Mais da metade dos municípios não está preparada para desastres climáticos**.** Curitiba: TCE-PR, 21 jul. 2026. Disponível em: https://www.tce.pr.gov.br/noticias/pr-mais-da-metade-dos-municipios-nao-esta-preparado-para-desastres-climaticos.html. Acesso em: 8 set. 2026.

THOMAS, A. S.; KOPCZAK, L. R. **From Logistics to Supply Chain Management**: The Path Forward in the Humanitarian Sector. [San Francisco]: Fritz Institute, 2005. Disponível em: https://alnap.org/help-library/resources/from-logistics-to-supply-chain-management-the-path-forward-for-the-humanitarian-sector. Acesso em: 23 set. 2026.

ZAKI, M. J.; MEIRA JR., W. **Data Mining and Machine Learning**: Fundamental Concepts and Algorithms. 2nd ed. Cambridge: Cambridge University Press, 2020. DOI: https://doi.org/10.1017/9781108564175. Disponível em: https://dataminingbook.info/book_html/. Acesso em: 23 set. 2026.