

# Segmentação de Clientes com Particle Swarm Optimization e Modelo RFM

Tiago Santos Mol<sup>1</sup>

<sup>1</sup>Universidade de Federal de Ouro Preto / Ciência da Computação  
Ouro Preto – MG, Brasil

tiago.mol@aluno.ufop.edu.br

**Abstract.** Customer segmentation supports personalized marketing strategies by grouping consumers with similar purchasing behavior. This work formulates customer clustering as an optimization problem and applies Particle Swarm Optimization (PSO) to discover cluster centroids based on the RFM model (Recency, Frequency, Monetary). The solution is validated using internal clustering metrics and visual analysis, resulting in coherent behavioral profiles and well-separated groups in a PCA projection.

**Resumo.** A segmentação de clientes auxilia estratégias de marketing personalizadas ao agrupar consumidores com comportamento de compra semelhante. Este trabalho formula o agrupamento como um problema de otimização e aplica Particle Swarm Optimization (PSO) para encontrar centróides de clusters com base no modelo RFM (Recency, Frequency, Monetary). A solução é validada por métricas internas e análise visual, resultando em perfis comportamentais coerentes e boa separação dos grupos em uma projeção via PCA.

## 1. Introdução

Diversos trabalhos têm explorado a utilização de algoritmos de Inteligência Coletiva para segmentação de clientes. Por exemplo, Chan et al. demonstraram a eficácia do Particle Swarm Optimization combinado ao modelo RFM para identificar clientes mais lucrativos em campanhas de marketing [1].

A segmentação de clientes é uma estratégia fundamental para compreender padrões de consumo e direcionar campanhas de marketing de forma eficiente. Entre as abordagens utilizadas, o modelo RFM (Recency, Frequency e Monetary) destaca-se por representar o comportamento de compra por meio do tempo desde a última compra, frequência de transações e valor monetário associado ao cliente, sendo amplamente utilizado na caracterização de perfis.

Técnicas de Inteligência Computacional podem automatizar a identificação de grupos, especialmente quando a definição manual de perfis é inviável. Nesse contexto, algoritmos de Inteligência Coletiva, como o Particle Swarm Optimization (PSO), realizam busca global no espaço de soluções e podem ser empregados para encontrar centróides que gerem agrupamentos com baixa dispersão intra-cluster.

Neste trabalho, aplica-se PSO para segmentar clientes com base no modelo RFM. O problema é formulado como uma tarefa de otimização, na qual cada partícula representa um conjunto de centróides e a qualidade da solução é avaliada pela soma dos erros quadráticos intra-cluster. Os resultados são analisados por métricas de validação interna e pela interpretação dos perfis obtidos.

## 2. Metodologia

### 2.1. Base de dados e pré-processamento

Foi utilizado um conjunto de dados sintético com  $n = 300$  clientes e três atributos do modelo RFM: *Recency* (dias desde a última compra), *Frequency* (número de compras no período) e *Monetary* (valor monetário associado ao cliente). As variáveis foram padronizadas com *StandardScaler* para evitar dominância numérica de atributos com maior escala.

### 2.2. PSO para agrupamento

O agrupamento foi modelado como um problema de otimização. Cada partícula representa uma solução candidata formada por  $K$  centróides em um espaço de dimensão  $D$  ( neste caso,  $D = 3$ ). Assim, a posição de uma partícula pode ser representada por uma matriz  $K \times D$ .

Dada uma partícula, cada cliente é atribuído ao centróide mais próximo (distância Euclidiana). A função objetivo (fitness) minimizada foi a soma dos erros quadráticos intra-cluster (SSE), definida pela soma das distâncias quadráticas entre cada ponto e o centróide ao qual foi atribuído. Soluções que produzem clusters vazios recebem penalização.

As partículas são atualizadas conforme o modelo *gbest* do PSO, combinando termo de inércia, componente cognitiva (melhor posição individual) e componente social (melhor posição global). Foram utilizados  $w = 0,72$ ,  $c_1 = 1,49$  e  $c_2 = 1,49$ , com 30 partículas e 80 iterações. Para reduzir efeito de aleatoriedade, o PSO foi executado 5 vezes com sementes distintas e selecionou-se a execução com menor SSE.

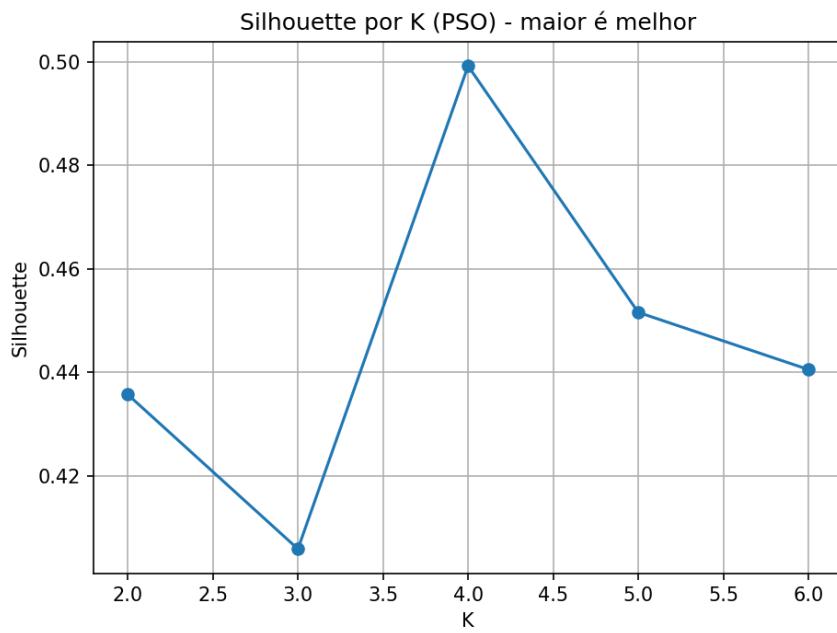
### 2.3. Seleção do número de clusters

O número de clusters foi selecionado testando-se  $K \in \{2, 3, 4, 5, 6\}$  e avaliando-se as métricas de validação interna *Silhouette* (maior é melhor) e *Davies–Bouldin* (menor é melhor). O valor escolhido foi  $K = 4$ , pois apresentou a maior *Silhouette* e o menor *DBI* entre as opções testadas.

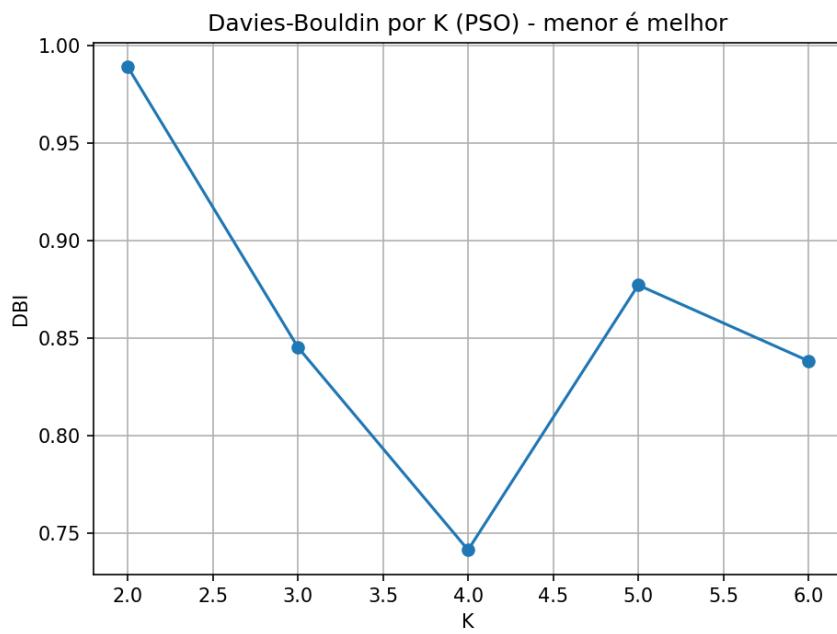
## 3. Resultados e Discussão

### 3.1. Avaliação para diferentes valores de $K$

As Figuras 1 e 2 apresentam a variação das métricas por  $K$ . Observa-se que  $K = 4$  apresenta o melhor compromisso entre separação e coesão.



**Figura 1.** Índice Silhouette para diferentes valores de  $K$  (PSO).



**Figura 2.** Índice Davies–Bouldin para diferentes valores de  $K$  (PSO).

### 3.2. Perfis de clientes obtidos

Com  $K = 4$ , foram obtidos quatro perfis interpretáveis. A Tabela 1 apresenta o tamanho de cada cluster e as médias dos atributos RFM.

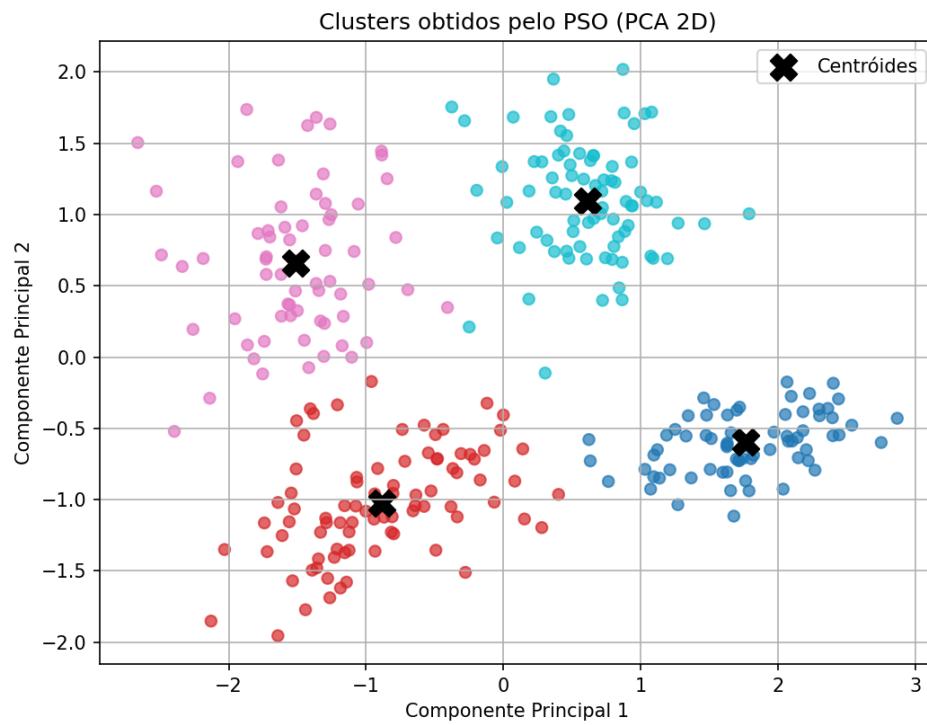
**Tabela 1. Resumo dos clusters e médias RFM.**

| Cluster | N (%)      | Recency | Frequency | Monetary |
|---------|------------|---------|-----------|----------|
| 0       | 71 (23.67) | 11.15   | 13.14     | 114.98   |
| 1       | 85 (28.33) | 39.05   | 3.07      | 299.31   |
| 2       | 65 (21.67) | 35.19   | 1.93      | 919.32   |
| 3       | 79 (26.33) | 10.21   | 9.36      | 788.45   |

A partir da Tabela 1, o Cluster 3 pode ser interpretado como *premium frequente* (baixa recency, alta frequência e alto gasto), enquanto o Cluster 0 representa *frequentes de baixo ticket* (baixa recency, alta frequência e baixo gasto). O Cluster 2 é caracterizado como *premium ocasional* (alto gasto, baixa frequência e recency elevada), e o Cluster 1 apresenta *baixo engajamento* (recency elevada e baixos valores de frequência e gasto).

### 3.3. Visualização por PCA

A Figura 3 apresenta a projeção bidimensional via PCA. Os eixos “Componente Principal 1” e “Componente Principal 2” correspondem a combinações lineares dos atributos RFM que preservam a maior variância possível na projeção, sendo utilizados apenas para visualização.



**Figura 3. Visualização dos clusters gerados pelo PSO após projeção PCA 2D.**

## 4. Conclusão

Este trabalho aplicou PSO para segmentação de clientes baseada no modelo RFM, formulando o agrupamento como um problema de otimização de centróides com minimização

de SSE. Os resultados indicaram quatro perfis de clientes com interpretações consistentes e boa separação visual na projeção PCA. Como trabalhos futuros, pode-se investigar dados reais e incluir variáveis adicionais de comportamento.

## Referências

- [1] Chan, C. C. H.; Hwang, Y. R.; Wu, H. C. (2016). Marketing segmentation using the particle swarm optimization algorithm: a case study. *Journal of Ambient Intelligence and Humanized Computing*, 7, 855–863.