

## Técnicas escolhidas: KMeans e Affinity Propagation

### 1. Explicação básica

KMeans é um algoritmo de agrupamento que particiona os dados em k grupos, minimizando a soma das distâncias quadradas entre os pontos e o centroide de seu grupo.

Parâmetros utilizados:

n\_clusters=5: define que o número de grupos desejado é 5.

random\_state=42: fixa a semente do gerador de números aleatórios para garantir reprodutibilidade dos resultados.

Affinity Propagation é um algoritmo de clustering baseado na troca de mensagens entre pares de amostras, que automaticamente determina o número de grupos a partir da similaridade entre os pontos.

Parâmetros utilizados:

random\_state=42: fixa a semente do gerador de números aleatórios para garantir reprodutibilidade dos resultados.

preference=-50: controla a probabilidade de cada ponto ser escolhido como "exemplar" (valor mais baixo → menos grupos; valor mais alto → mais grupos).

Testando, o valor -50 gerou 5 grupos, permaneci com ele para se igualar ao KMeans e facilitar comparação.

### 2. Comparação entre os grupos

Abaixo uma tabela comparando a similaridade entre os grupos gerados. Podemos observar um certo alinhamento entre os grupos 0, 2, 3 do Affinity Propagation e 3, 0, 2 do KMeans. Pode-se entender, por exemplo, que 84.21% dos pontos do grupo 2 do Affinity também estavam presentes no grupo 0 do KMeans. Isso indica que ambos algoritmos convergem para uma similaridade entre os grupos. No entanto, o desalinhamento de outros grupos, como Affinity 4, que está amplamente disperso entre os clusters do KMeans, aponta para possíveis instabilidades ou ambiguidade nos dados.

|            | KMeans 0 | KMeans 1 | KMeans 2 | KMeans 3 | KMeans 4 |
|------------|----------|----------|----------|----------|----------|
| Affinity 0 | 4.55%    | 9.09%    | 4.55%    | 72.73%   | 9.09%    |
| Affinity 1 | 0.00%    | 31.25%   | 31.25%   | 12.50%   | 25.00%   |
| Affinity 2 | 84.21%   | 0.00%    | 10.53%   | 5.26%    | 0.00%    |
| Affinity 3 | 0.00%    | 20.00%   | 66.67%   | 13.33%   | 0.00%    |
| Affinity 4 | 30.00%   | 25.00%   | 15.00%   | 5.00%    | 25.00%   |

No entanto, ao comparar os Silhouette Score dos dois algoritmos, temos o seguinte:

- KMeans: 0.0633
- Affinity Propagation: 0.0573

Mesmo que leve, Affinity Propagation criou grupos menos distintos que o KMeans, o que indica uma performance pior.

3. Usando o resultado, indicar um gênero musical e estilo de jogo que uma pessoa que gosta de Bandinha e Jogo de cartas poderia gostar.

Estudando o Affinity Propagation percebi que os “centros” gerados por ele são compostos por pontos significativos, que existem dentro do grupo. Desse modo, é possível ver quais gostos uma pessoa que representa um grupo teria. Infelizmente a tabela possui colunas demais para incluir no texto, mas enviarei como anexo.

Nela é possível perceber que o grupo 4 cumpre o requisito de gostar de Bandinha e Jogo de cartas, então, boas recomendações seriam outros gêneros também presentes naquele grupo, para os exemplos tentei selecionar categorias que estavam presentes principalmente no grupo 4, e poucas vezes nos outros grupos. Alguns exemplos são: Sertanejo, Brega, Reggaeton, Battle Royale, Esporte e Realidade Virtual.

Andrei Carlesso Camilotto