

◇ Análise do Resultado do SimpleKMeans

1. Configuração do Algoritmo

- Foi utilizado **SimpleKMeans** com **N=10** , mas na execução foram gerados **5 clusters**.
- Distância usada: **Euclidiana**.
- Número máximo de iterações: **500** .

2. Características dos Dados

- Total de instâncias:** 4.773
- Atributos:** Idade, Atraso, Valor, Contato, Efetivo, Acordo.

3. Centroides dos Clusters

Os valores médios de cada atributo dentro de cada cluster apresentam padrões diferenciados nos dados.

4. Distribuição dos Clusters

- Cluster 0** → 663 instâncias (14%)
- Cluster 1** → 834 instâncias (17%)
- Cluster 2** → 269 instâncias (6%)
- Cluster 3** → 873 instâncias (18%)
- Cluster 4** → 2.134 instâncias (45%)

O cluster 4 é o maior, representando **quase metade dos dados** .

◇ Relação com o Apriori

Agora, para comparar com as **regras de associação geradas pelo Apriori** , podemos refletir:

1. Clusterização identifica grupos de perfis semelhantes

- Exemplo: Pode ter um cluster de clientes que pagam em dia, outro de inadimplentes, outro de novos clientes, etc.

2. A priori descobrimos padrões de comportamento dentro desses grupos

- Se aplicarmos ou a priori dentro de cada cluster, podemos identificar padrões mais específicos.
- Exemplo: "Clientes no cluster 4 têm maior probabilidade de fechar um acordo se tiverem um contato telefônico ativo".

3. Interpretação Complementar

- A clusterização mostra **quem são os grupos**.
- A priori revela **quais padrões existem dentro desses grupos** .

Clusterer

Choose SimpleKMeans -N 10 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10

Cluster mode

☒ Use training set☐ Supplied test set

Set...

☐ Percentage split

% 66

☐ Classes to clusters evaluation

(Num) Acordo

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

15:48:24 - SimpleKMeans

15:56:29 - SimpleKMeans

15:57:13 - SimpleKMeans

Clusterer output

=== Run information ===

Scheme:weka.clusterers.SimpleKMeans -N 5 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10

Relation: A4

Instances: 4773

Attributes: 6

Idade

Atraso

Valor

CONTATO

EFETIVO

Acordo

Test mode:evaluate on training data

=== Model and evaluation on training set ===

kMeans

=====

Number of iterations: 16

Within cluster sum of squared errors: 909.7776150855439

Missing values globally replaced with mean/mode

Cluster centroids:

Attribute	Full Data (4773)	Cluster#				
		0 (663)	1 (834)	2 (269)	3 (873)	4 (2134)
Idade	1.7461	4	1.8321	4	1.0332	1.0197
Atraso	2.4588	1.181	1.8885	6	6	1.1837
Valor	1.4718	1.9593	2.0731	0.4498	0.3963	1.6542
CONTATO	4.7591	4.3922	5.2842	4.0595	4.7205	4.7718
EFETIVO	1.1672	1.1825	1.1091	1.223	1.1775	1.1739
Acordo	0.8184	1	0	0.8773	1	1

Time taken to build model (full training data) : 0.19 seconds

=== Model and evaluation on training set ===

Clustered Instances

```


0      663 ( 14%)
1      834 ( 17%)
2      269 (  6%)
3      873 ( 18%)
4     2134 ( 45%)

```

Status

OK

Log

 x 0