

# Aprendizado de Máquina

Prof. Aran Morales, UNISUL – Anima Educação  
Apostila 10: Regras de Associação

# Técnica: Regras de Associação



# Técnica: Regras de Associação



Visa descobrir **associações** importantes entre os itens (**k-itemsets**), tal que, a presença de um item em uma determinada transação implica na presença de outro item na mesma transação.

Cada registro corresponde a uma transação, com itens assumindo valores binários (sim/não), indicando se o cliente comprou ou não o respectivo item.

Uma regra de associação é uma **implicação** na forma

$$X \Rightarrow Y,$$

e possui dois parâmetros básicos: um **suporte** e uma **confiança**;

# Técnica: Regras de Associação



A função do **Suporte** é determinar a frequência (contagem ou em porcentagem) que ocorre um *itemset* entre todas as transações da Base de Dados.

A **confiança** mede a força da regra e determina a sua validade, isto é, quantifica a frequência do antecedente implicando o conseqüente.

A confiança e suporte são utilizadas como filtro para gerar menos regras.

# Técnica: Regras de Associação



| Id | Leite | Café | Cerveja | Pão | Manteiga | Arroz | Feijão |
|----|-------|------|---------|-----|----------|-------|--------|
| 1  | N     | S    | N       | S   | S        | N     | N      |
| 2  | S     | N    | S       | S   | S        | N     | N      |
| 3  | N     | S    | N       | S   | S        | N     | N      |
| 4  | S     | S    | N       | S   | S        | N     | N      |
| 5  | N     | N    | S       | N   | N        | N     | N      |
| 6  | N     | N    | N       | N   | S        | N     | N      |
| 7  | N     | N    | N       | S   | N        | N     | N      |
| 8  | N     | N    | N       | N   | N        | N     | S      |
| 9  | N     | N    | N       | N   | N        | S     | S      |
| 10 | N     | N    | N       | N   | N        | S     | N      |

Leite S = 0.2; Pão S=0.5

Café e Pão S= 0.3  
Pão/S(Café)

Se (Café) Então (Pão) C = 1.0 (S(Café e Pão)/S(Café))

Pão e Manteiga S = 0.4

Se (Pão) Então (Manteiga) C = 0.8

Se (Manteiga) Então (Pão) C = 0.8

Café, Pão e Manteiga S = 0.3

Se (Café e Pão) Então Manteiga C = 1.0

Se (Café) Então (Pão e Manteiga) C = 1.0

# Técnica: Regras de Associação



## Fases do algoritmo Apriori:

1. Geração dos conjuntos candidatos, com **suporte** acima do mínimo estabelecido;
2. Geração da regras de associação dos conjuntos candidatos gerados no passo anterior com **confiança** superior ao mínimo estabelecido;

# Algoritmo Apriori

Suporte mínimo = 2



| TID | Items      |
|-----|------------|
| 100 | A, C, D    |
| 200 | B, C, E    |
| 300 | A, B, C, E |
| 400 | B, E       |

Base de dados

**C1**

| Itemsets | So |
|----------|----|
| {A}      | 2  |
| {B}      | 3  |
| {C}      | 3  |
| {D}      | 1  |
| {E}      | 3  |

**L1**

| Itemsets | So |
|----------|----|
| {A}      | 2  |
| {B}      | 3  |
| {C}      | 3  |
| {E}      | 3  |

L1

**C3**

| Itemsets  | So |
|-----------|----|
| {A, B, C} | 1  |
| {A, B, E} | 1  |
| {A, C, E} | 1  |
| {B, C, E} | 2  |

**L2**

| Itemsets | So |
|----------|----|
| {A, C}   | 2  |
| {B, C}   | 2  |
| {B, E}   | 3  |
| {C, E}   | 2  |

L2

**C2**

| Itemsets | So |
|----------|----|
| {A, B}   | 1  |
| {A, C}   | 2  |
| {A, E}   | 1  |
| {B, C}   | 2  |
| {B, E}   | 3  |
| {C, E}   | 2  |

C2

**L3**

| Itemsets  | So |
|-----------|----|
| {B, C, E} | 2  |

**C4**

| Itemsets | So |
|----------|----|
| { Null } | 2  |

C4

**FIM (Fase I)**

# Algoritmo Apriori



Regras geradas com L2 (para  $s \geq 50\%$  e  $c \geq 60\%$ )

Se A Então C (  $s = 50\%$ ,  $c = 100\%$  )

Se C Então A (  $s = 50\%$ ,  $c = 66.7\%$  )

Se B Então C (  $s = 50\%$ ,  $c = 66.7\%$  )

Se C Então B (  $s = 50\%$ ,  $c = 66.7\%$  )

Se B Então E (  $s = 75\%$ ,  $c = 100\%$  )

Se E Então B (  $s = 75\%$ ,  $c = 100\%$  )

Se C Então E (  $s = 50\%$ ,  $c = 66.7\%$  )

Se E Então C (  $s = 50\%$ ,  $c = 66.7\%$  )



# Algoritmo Apriori



Regras geradas com L3 (para  $s \geq 50\%$  e  $c \geq 60\%$ )

Se B e C Então E (  $s = 50\%$ ,  $c = 100\%$  )

Se B e E Então C (  $s = 50\%$ ,  $c = 66.7\%$  )

Se C e E Então B (  $s = 50\%$ ,  $c = 100\%$  )

Se B Então C e E (  $s = 50\%$ ,  $c = 66.7\%$  )

Se C Então B e E (  $s = 50\%$ ,  $c = 66.7\%$  )

Se E Então B e C (  $s = 50\%$ ,  $c = 66.7\%$  )

# Algoritmo Apriori



| Id | Leite | Café | Cerveja | Pão | Manteiga | Arroz | Feijão |
|----|-------|------|---------|-----|----------|-------|--------|
| 1  | N     | S    | N       | S   | S        | N     | N      |
| 2  | S     | N    | S       | S   | S        | N     | N      |
| 3  | N     | S    | N       | S   | S        | N     | N      |
| 4  | S     | S    | N       | S   | S        | N     | N      |
| 5  | N     | N    | S       | N   | N        | N     | N      |
| 6  | N     | N    | N       | N   | S        | N     | N      |
| 7  | N     | N    | N       | S   | N        | N     | N      |
| 8  | N     | N    | N       | N   | N        | N     | S      |
| 9  | N     | N    | N       | N   | N        | S     | S      |
| 10 | N     | N    | N       | N   | N        | S     | N      |

# Algoritmo Apriori



Item sets com suporte 25%

The screenshot displays the Orange3 interface for the Apriori algorithm. On the left, a workflow diagram shows a 'File' widget connected to 'Frequent Itemsets (1)', 'Association Rules (1)', and 'Data Table (1)'. The 'Frequent Itemsets (1)' widget is selected, showing its configuration and results.

**\*\*\* Frequent Itemsets (1)**

Info  
Number of itemsets: 7  
Selected itemsets: 0  
Selected examples: 0

Expand all Collapse all

Find itemsets  
Minimal support: 25%  
Max. number of itemsets: 10000

Filter itemsets  
Contains:   
Min. items: 1 Max. items: 999

Find Itemsets

| Itemsets   | Support | %  |
|------------|---------|----|
| ▼ Café=1   | 3       | 30 |
| ▼ Pao=1    | 3       | 30 |
| Manteiga=1 | 3       | 30 |
| Manteiga=1 | 3       | 30 |
| ▼ Pao=1    | 5       | 50 |
| Manteiga=1 | 4       | 40 |
| Manteiga=1 | 5       | 50 |

**Data Table (1)**

Info  
10 instances  
7 features (70.0 % missing data)  
No target variable.  
No meta attributes

Variables  
 Show variable labels (if present)  
 Visualize numeric values  
 Color by instance classes

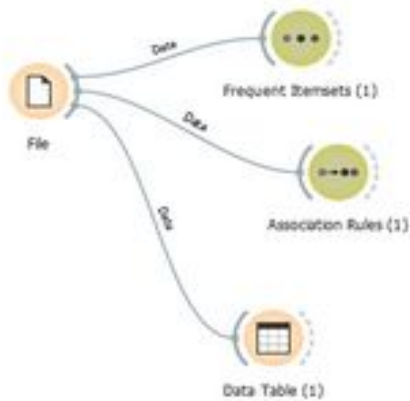
Selection  
 Select full rows

|    | Leite | Café | Cerveja | Pao | Manteiga | Arroz | Feijao |
|----|-------|------|---------|-----|----------|-------|--------|
| 1  | ?     | 1    | ?       | 1   | 1        | ?     | ?      |
| 2  | 1     | ?    | 1       | 1   | 1        | ?     | ?      |
| 3  | ?     | 1    | ?       | 1   | 1        | ?     | ?      |
| 4  | 1     | 1    | ?       | 1   | 1        | ?     | ?      |
| 5  | ?     | ?    | 1       | ?   | ?        | ?     | ?      |
| 6  | ?     | ?    | ?       | ?   | 1        | ?     | ?      |
| 7  | ?     | ?    | ?       | 1   | ?        | ?     | ?      |
| 8  | ?     | ?    | ?       | ?   | ?        | ?     | 1      |
| 9  | ?     | ?    | ?       | ?   | ?        | 1     | 1      |
| 10 | ?     | ?    | ?       | ?   | ?        | 1     | ?      |

# Algoritmo Apriori



Regras com suporte de 25% e confiança de 75%



Association Rules (1)

Info  
Rules: 8 (shown 8)

Find association rules  
Min. supp.: 25 %  
Min. conf.: 75 %  
Max. rules: 10k  
 Induce only classification rules  
 Restrict search by below filters  
Find Rules

Filter by Antecedent  
Contains:   
Items, min: 1 max: 999

Filter by Consequent  
Contains:   
Items, min: 1 max: 999

Send selection

| Supp  | Conf  | Covr  | Strg  | Lift  | Levr  | Antecedent         | Consequent        |
|-------|-------|-------|-------|-------|-------|--------------------|-------------------|
| 0.400 | 0.800 | 0.500 | 1.000 | 1.600 | 0.150 | Manteiga=1         | Pao=1             |
| 0.400 | 0.800 | 0.500 | 1.000 | 1.600 | 0.150 | Pao=1              | Manteiga=1        |
| 0.300 | 1.000 | 0.300 | 1.667 | 2.000 | 0.150 | Café=1             | Pao=1             |
| 0.300 | 1.000 | 0.300 | 1.667 | 2.000 | 0.150 | Café=1             | Manteiga=1        |
| 0.300 | 0.750 | 0.400 | 0.750 | 2.500 | 0.180 | Pao=1, Manteiga=1  | Café=1            |
| 0.300 | 1.000 | 0.300 | 1.667 | 2.000 | 0.150 | Café=1, Manteiga=1 | Pao=1             |
| 0.300 | 1.000 | 0.300 | 1.667 | 2.000 | 0.150 | Café=1, Pao=1      | Manteiga=1        |
| 0.300 | 1.000 | 0.300 | 1.333 | 2.500 | 0.180 | Café=1             | Pao=1, Manteiga=1 |

# Algoritmo Apriori



Exemplo com mais de 100 mil instancias

The screenshot displays the Orange Data Mining software interface. The workflow diagram shows a 'Data Table' connected to 'Select Columns' and 'Frequent Itemsets'. The 'Frequent Itemsets' window shows a list of itemsets with their support and percentage. The 'Association Rules' window shows a table of rules with columns for Support, Confidence, Coverage, Strength, Lift, and Leverage, along with antecedent and consequent itemsets.

**Frequent Itemsets**

| Itemsets                | Support | %     |
|-------------------------|---------|-------|
| Doces=1                 | 33696   | 31.16 |
| Carnes=1                | 11219   | 10.38 |
| Farinhas_Graos_Massas=1 | 17457   | 16.14 |
| Carnes=1                | 35401   | 32.74 |
| Alimentos_Congelados=1  | 11139   | 10.3  |
| Condimentos=1           | 11629   | 10.75 |
| Bebidas_alcoolicas=1    | 11529   | 10.66 |
| Frutas=1                | 17573   | 16.25 |
| Aguas_Sucos=1           | 23747   | 21.96 |
| Farinhas_Graos_Massas=1 | 34869   | 32.25 |
| Bebidas_alcoolicas=1    | 12391   | 11.46 |
| Frutas=1                | 10999   | 10.17 |
| Alimentos_Congelados=1  | 35616   | 32.94 |
| Condimentos=1           | 20349   | 18.82 |
| Bebidas_alcoolicas=1    | 16748   | 15.49 |
| Bebidas_alcoolicas=1    | 20005   | 18.5  |
| Frutas=1                | 11621   | 10.75 |
| Condimentos=1           | 36537   | 33.79 |
| Bebidas_alcoolicas=1    | 20609   | 19.06 |
| Frutas=1                | 11248   | 10.4  |
| Bebidas_alcoolicas=1    | 36606   | 33.85 |
| Frutas=1                | 11221   | 10.38 |
| Laticinios=1            | 23128   | 21.39 |
| Padaria=1               | 23649   | 21.87 |
| Frutas=1                | 36179   | 33.46 |

**Association Rules**

| Supp  | Conf  | Covr  | Strg  | Lift  | Levr  | Antecedent                                   | Consequent              |
|-------|-------|-------|-------|-------|-------|--|-------------------------|
| 0.191 | 0.563 | 0.339 | 0.998 | 1.666 | 0.076 | Bebidas_alcoolicas=1                         | Condimentos=1           |
| 0.191 | 0.564 | 0.338 | 1.002 | 1.666 | 0.076 | Condimentos=1                                | Bebidas_alcoolicas=1    |
| 0.188 | 0.557 | 0.338 | 0.975 | 1.691 | 0.077 | Condimentos=1                                | Alimentos_Congelados=1  |
| 0.188 | 0.571 | 0.329 | 1.026 | 1.691 | 0.077 | Alimentos_Congelados=1                       | Condimentos=1           |
| 0.185 | 0.546 | 0.339 | 0.973 | 1.659 | 0.074 | Bebidas_alcoolicas=1                         | Alimentos_Congelados=1  |
| 0.185 | 0.562 | 0.329 | 1.028 | 1.659 | 0.074 | Alimentos_Congelados=1                       | Bebidas_alcoolicas=1    |
| 0.161 | 0.501 | 0.322 | 0.966 | 1.607 | 0.061 | Farinhas_Graos_Massas=1                      | Doces=1                 |
| 0.161 | 0.518 | 0.312 | 1.035 | 1.607 | 0.061 | Doces=1                                      | Farinhas_Graos_Massas=1 |
| 0.155 | 0.813 | 0.191 | 1.728 | 2.467 | 0.092 | Condimentos=1, Bebidas_alcoolicas=1          | Alimentos_Congelados=1  |
| 0.155 | 0.837 | 0.185 | 1.826 | 2.478 | 0.092 | Alimentos_Congelados=1, Bebidas_alcoolicas=1 | Condimentos=1           |
| 0.155 | 0.823 | 0.188 | 1.799 | 2.431 | 0.091 | Alimentos_Congelados=1, Condimentos=1        | Bebidas_alcoolicas=1    |

Orange Update Available  
Current version: 3.27.1  
Latest version: 3.38.0  
[Download](#) [Skip this Version](#)

# Algoritmo Apriori



Apriori\_Exemplo\_1.ipynb ☆

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda [Todas as alterações foram s](#)

+ Código + Texto



✓ 10s

```
!pip install efficient_apriori
from efficient_apriori import apriori

# Dados de exemplo
transactions = [
    ("cerveja", "amendoim", "fralda"),
    ("cerveja", "fralda"),
    ("cerveja", "amendoim"),
    ("cerveja"),
    ("amendoim", "fralda")
]

# Executando o algoritmo Apriori
itemsets, rules = apriori(transactions, min_support=0.3, min_confidence=0.2)

# Exibindo os resultados
print("Itemsets frequentes:")
for itemset in itemsets:
    print(itemset)

print("\nRegras de associação:")
```

{x}



<>



✓ 10s conclusão: 15:00

# Algoritmo Apriori



efficient\_aprioriAran.ipynb ☆

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda [Todas as alterações foram sal](#)

+ Código + Texto

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
```

```
[3] import io
from google.colab import files
uploaded_train = files.upload()
df = pd.read_csv(io.BytesIO(uploaded_train['groceries.csv']))
df.head()
```

Escolher arquivos groceries.csv

- **groceries.csv**(text/csv) - 802870 bytes, last modified: 21/05/2023 - 100% done  
Saving groceries.csv to groceries.csv

| Item(s) | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 | Item 7 | Item 8 | Item 9 | ... |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-----|
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-----|



# Aprendizado de Máquina

Prof. Aran Morales, UNISUL – Anima Educação

Apostila 10: Regas de Associação