Ciência de Dados
(Big Data Processing and Analytics)

Big Data Analytics – Mineração e Análise de Dados

Professor curador
Prof. Dr. Rogério de Oliveira

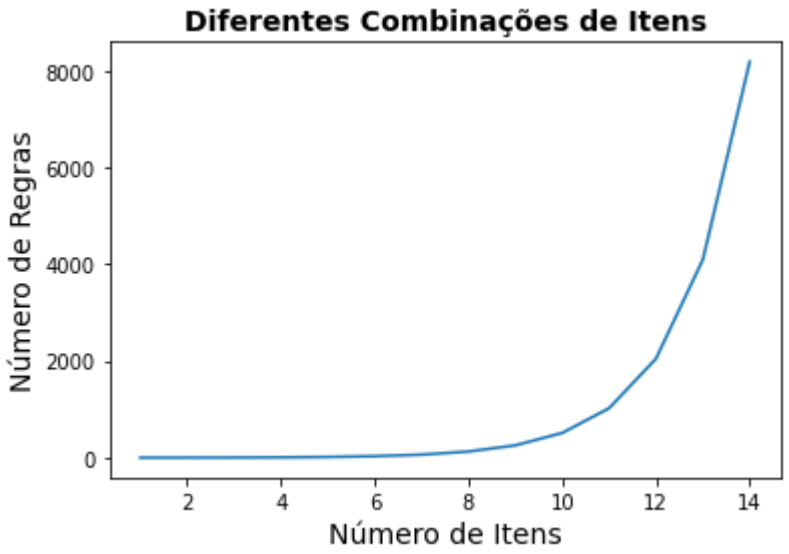
TRILHA 7
Aprendizado não Supervisionado: Regras de Associação e Filtros de Conteúdo

Parte A

Minerando Regras de Associação

Beer and Nappies. Baby love drinking beers?

Número Exponencial de Regras de Associação



$$N_{rules} = 3^n - 2^{n+1} + 1$$

Métricas e Poda

Suporte

A métrica mais simples é o *Suporte* e o compartilhamento de transações que contêm um conjunto de itens.

$$Support(X) = \frac{freq(X)}{N}$$

$$Support(X \to Y) = \frac{freq(X \cap Y)}{N}$$

Ela é uma medida de quão frequente a regra é no domínio das transações.

Elevação ou Lift

$$Lift(X \to Y) = \frac{Support(X \to Y)}{Support(X) \times Support(Y)}$$

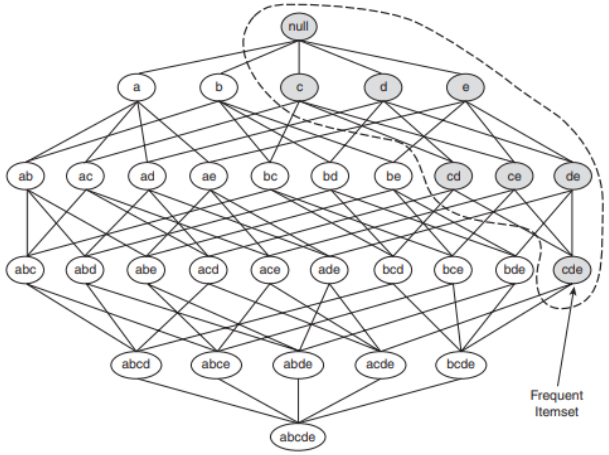
Confiança

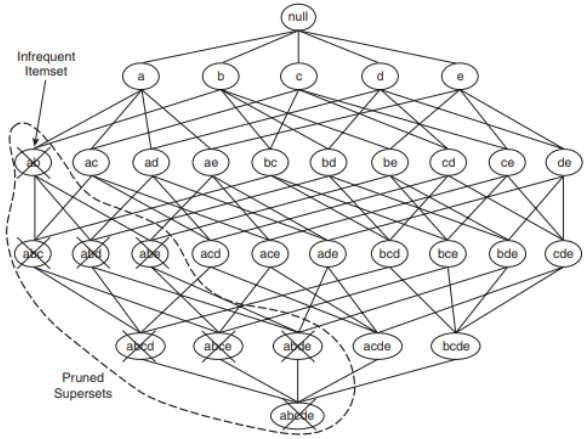
Esta métrica mede a frequência com que os itens em Y aparecem em transações que contêm X e é dado pela fórmula.

$$Confidence(X \to Y) = \frac{Support(X \to Y)}{Support(X)}$$

E agora já podemos entender que as regras abaixo são de fato diferentes:

Métricas e Poda: *Apriori*





Candidate 1-Itemsets

Item	Count
Beer	3
Bread	4
Cola	2
Diapers	4
Milk	4
Eggs	1

Minimum support count = 3

Candidate 2-Itemsets

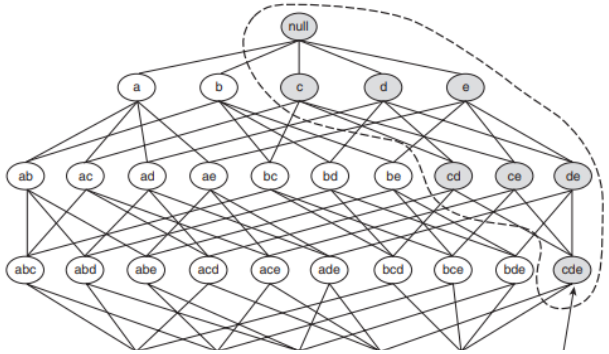
Itemset	Count
{Beer, Bread}	2
{Beer, Diapers}	3
{Beer, Milk}	2
{Bread, Diapers}	3
{Bread, Milk}	3
{Diapers, Milk}	3

Itemsets removed because of low support

Candidate 3-Itemsets

Itemset	Count
{Bread, Diapers, Milk}	2

Métricas e Poda: *Apriori*



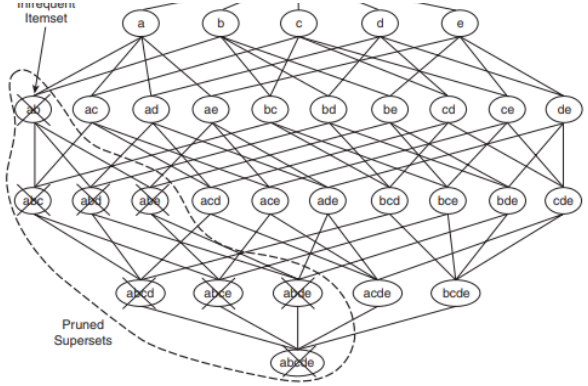
Candidate
1-Itemsets

Item	Count
Beer	3
Bread	4
Cola	2
Diapers	4

Minimum support count = 3

Candidate
2-Itemsets

Regras de Associação *not in* `scikit-learn`



Itemsets removed
because of low
support

{Bread, Milk}	3
{Diapers, Milk}	3

Candidate
3-Itemsets

Itemset	Count
{Bread, Diapers, Milk}	2

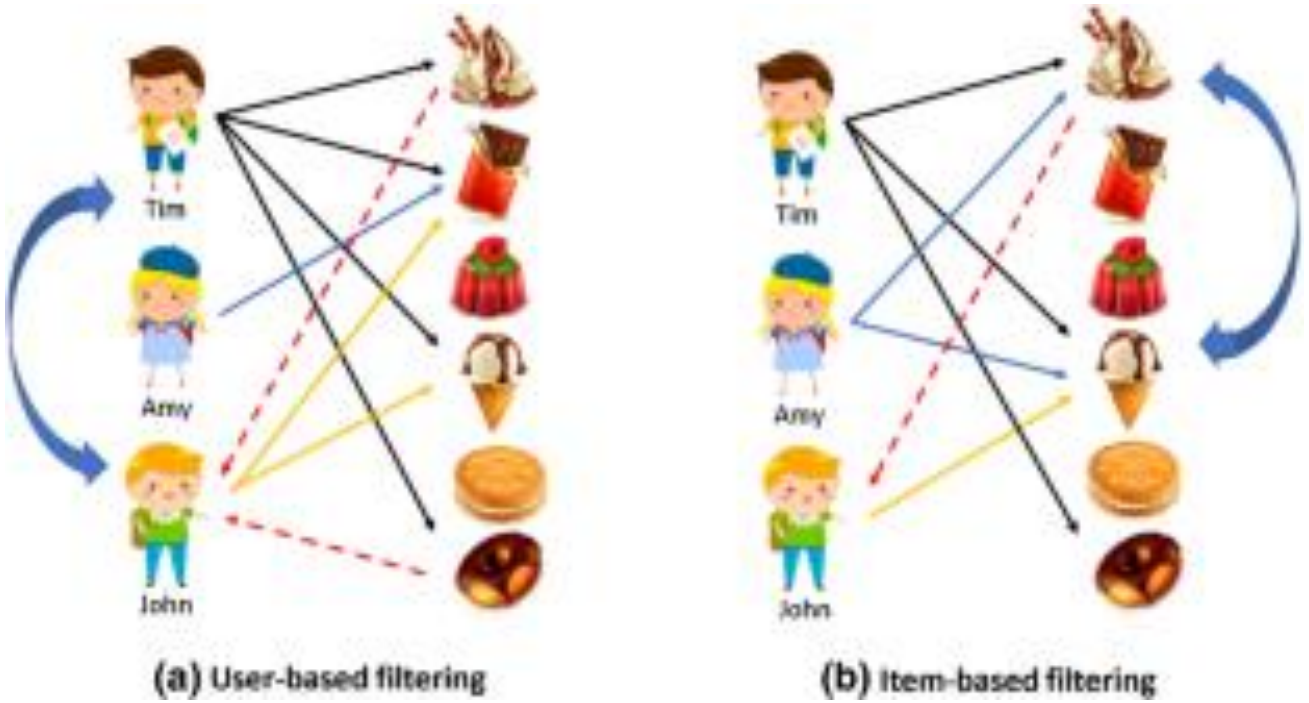
TRILHA 7
Aprendizado não Supervisionado: Regras de Associação e Filtros de Conteúdo

Parte B

Filtros de Conteúdo para Sistemas de Recomendação

- Sistemas baseados em conteúdo (Content-based filtering)
- Sistemas de filtragem colaborativa (Collaborative filtering)
- Sistemas híbridos (que usam uma combinação dos outros dois)

Filtros Baseados em Conteúdo:



Vizinhos Mais Próximos: User ou Item?

Filtros Baseados em Conteúdo:

Vizinhos Mais Próximos Não Supervisionado!

```
from sklearn.neighbors import NearestNeighbors

# Fit k-nearest neighbors
X = users.drop(columns='ID')

n_neighbors = 3

knn = NearestNeighbors(n_neighbors=n_neighbors+1)
knn.fit(X)
```

Universidade Presbiteriana
Mackenzie