

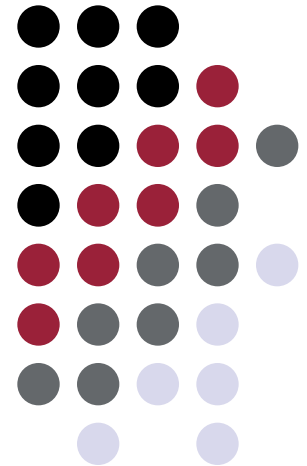
Mineração de Dados: Associação II



Universidade Federal
de Ouro Preto

CEA462 – Sistemas de Apoio à Decisão

Prof. MSc. George H. G. Fonseca
Universidade Federal de Ouro Preto



Algoritmo FP-Growth



- Codifica as transações na estrutura de dados FP-Tree
- Posteriormente, extrai conjuntos frequentes dessa estrutura

Algoritmo FP-Growth



- Cada transação representa um caminho na árvore FP
- Como diferentes transações podem ter itens em comum, seus caminhos na árvore podem se sobrepor
 - Quanto mais caminhos se sobrepõem, maior a compressão proporcionada pela árvore
- Com essa abordagem, temos que ler a base de dados **apenas duas vezes**

Algoritmo FP-Growth



- Procedimento para a geração da árvore FP
 - O conjunto de dados é varrido para determinar o contador de suporte de cada item e descartar itens infrequentes
 - Itens frequentes e transação são ordenados de acordo com seu suporte
 - Numa segunda leitura dos dados, a árvore FP é construída transação a transação
 - Se um item é inédito, é criado um novo nó para representá-lo
 - Senão, o mesmo se torna filho de um prefixo de outra transação

Algoritmo FP-Growth

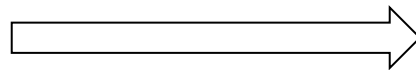


- Exemplo

- minSup = 0,2

TID	Itens
1	{b, a, f}
2	{b, d, c}
3	{a, c, e, d}
4	{a, d, e}
5	{b, c, a}
6	{a, b, c, d}
7	{a}
8	{a, c, b}
9	{a, b, d}
10	{b, e, c}

Item	Suporte
a	8
b	7
c	6
d	5
e	3
f	1



TID	Itens
1	{a, b}
2	{b, c, d}
3	{a, c, d, e}
4	{a, d, e}
5	{a, b, c}
6	{a, b, c, d}
7	{a}
8	{a, b, c}
9	{a, b, d}
10	{b, c, e}

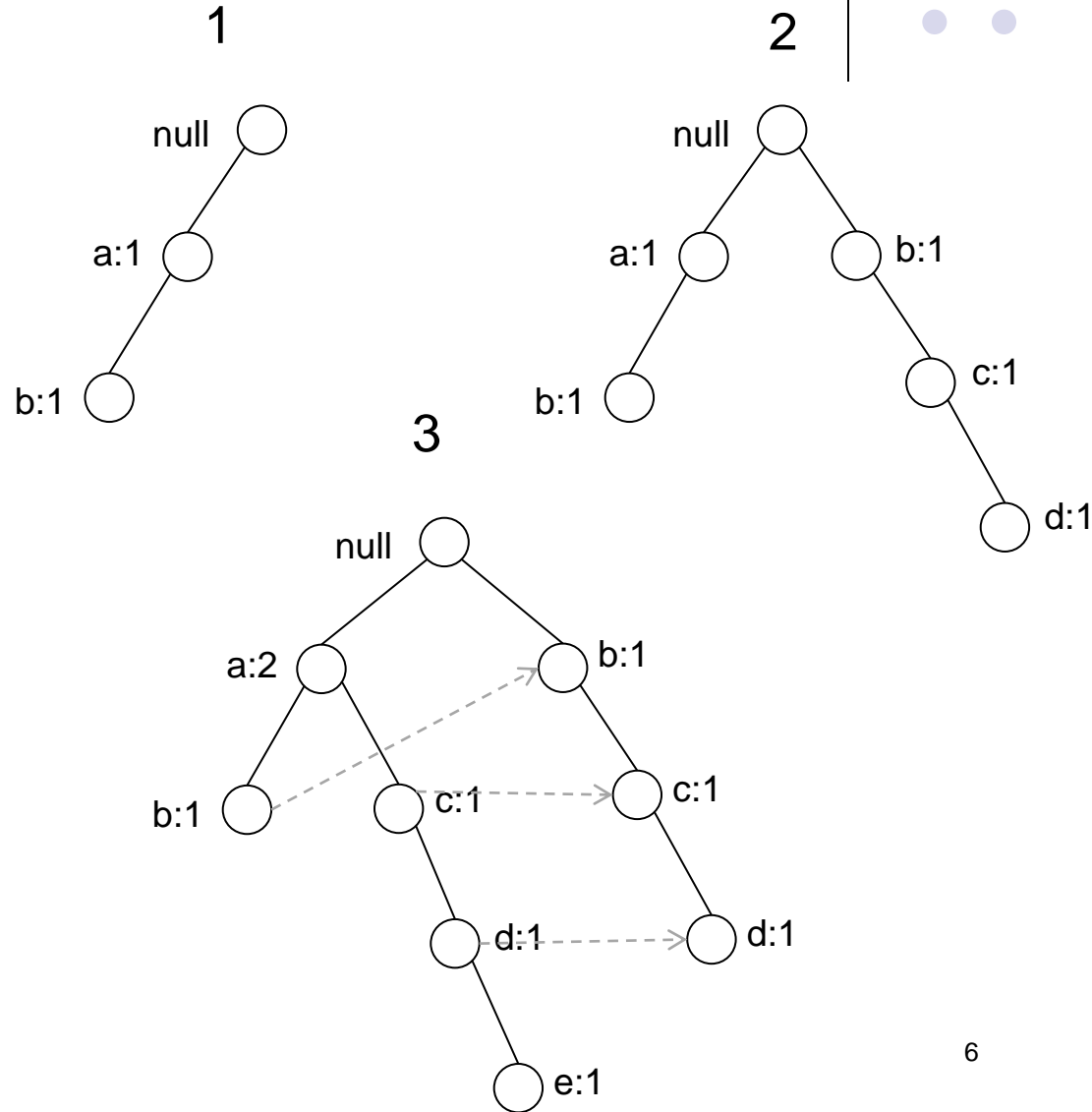
Algoritmo FP-Growth



● Exemplo

● minSup = 0,2

TID	Itens
1	{a, b}
2	{b, c, d}
3	{a, c, d, e}
4	{a, d, e}
5	{a, b, c}
6	{a, b, c, d}
7	{a}
8	{a, b, c}
9	{a, b, d}
10	{b, c, e}



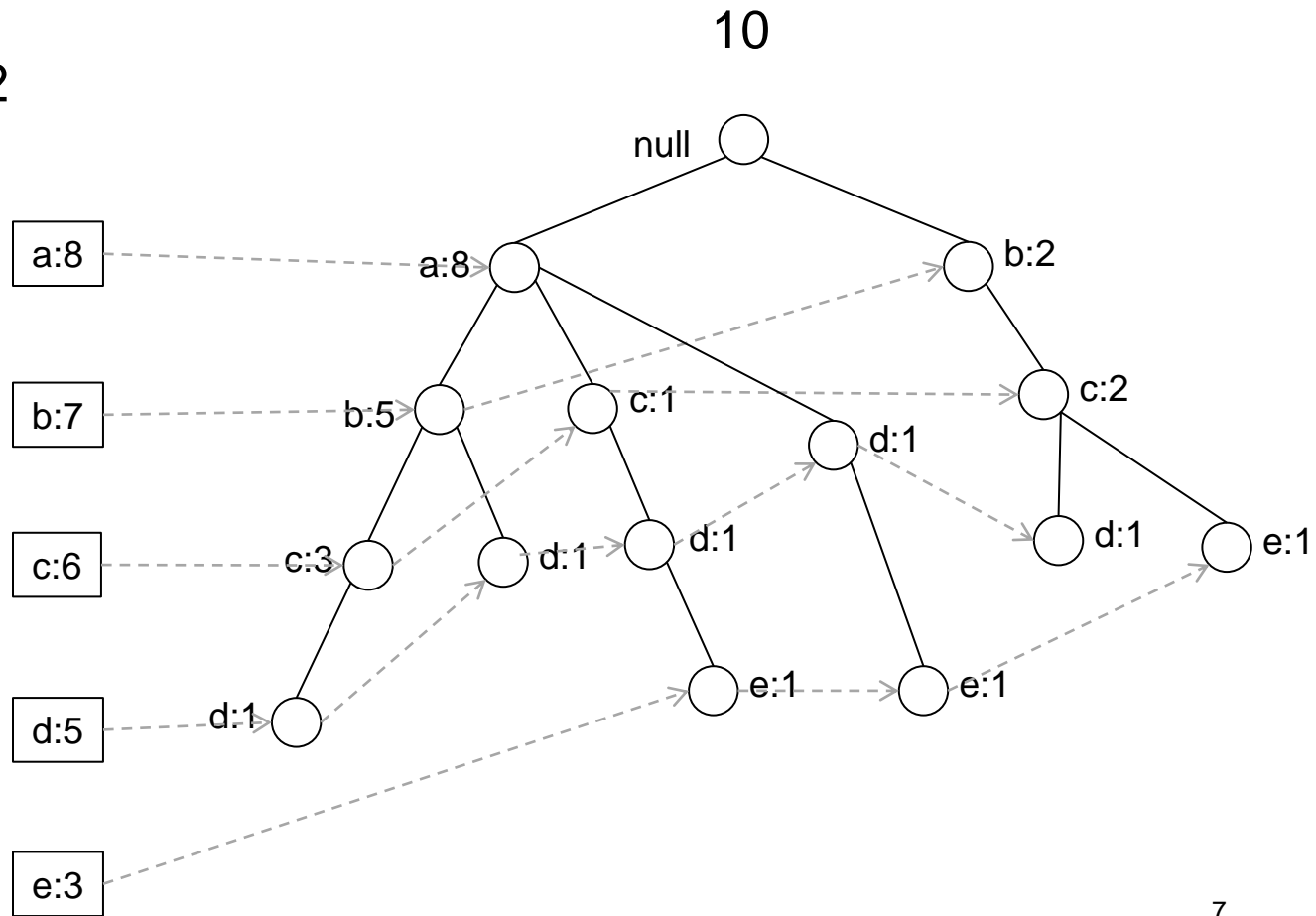
Algoritmo FP-Growth



- Exemplo

- minSup = 0,2

TID	Itens
1	{a, b}
2	{b, c, d}
3	{a, c, d, e}
4	{a, d, e}
5	{a, b, c}
6	{a, b, c, d}
7	{a}
8	{a, b, c}
9	{a, b, d}
10	{b, c, e}



Algoritmo FP-Growth

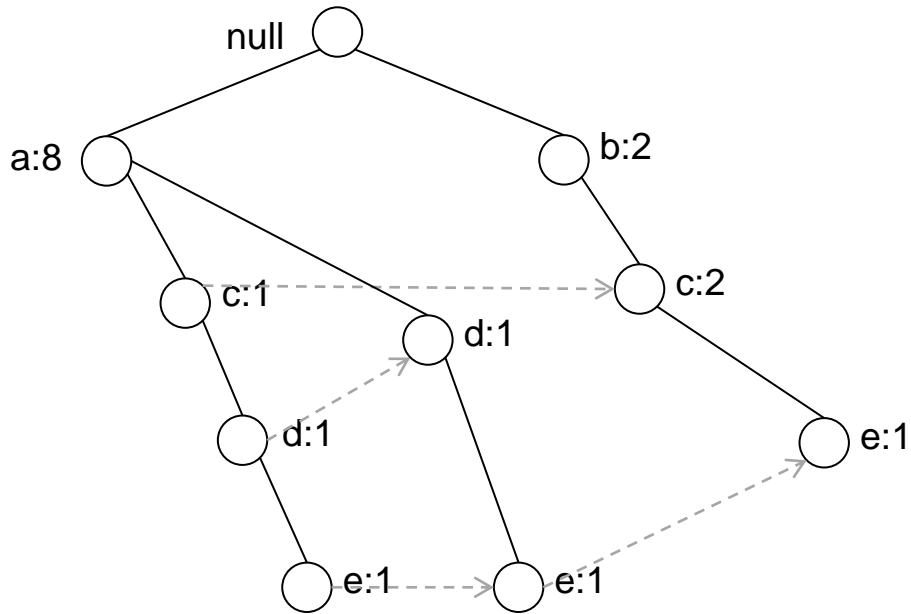


- Gerando itemsets frequentes
 - Estratégia bottom-up
 - Procura por **caminhos** terminando em 'e', depois 'd', ... até 'a'
 - Gerar **árvore condicional** para e
 - Árvore que contém os caminhos que levam a 'e' com suportes recalculados
 - Verifica-se quais **caminhos atendem ao suporte mínimo**
- Atenção: há implementações mais elaboradas!

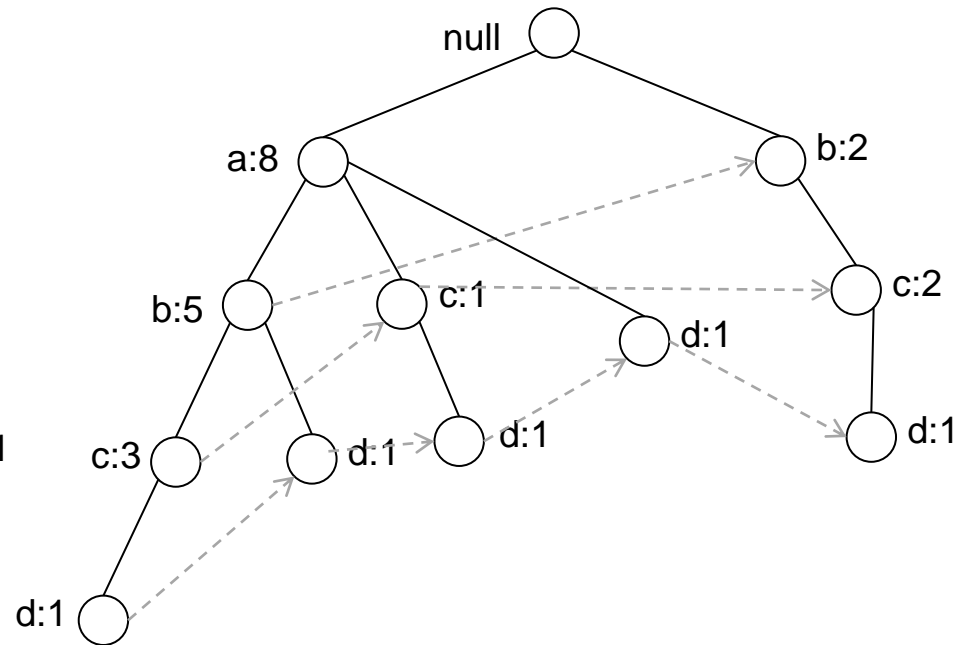
Algoritmo FP-Growth



● Exemplo



Caminhos contendo e

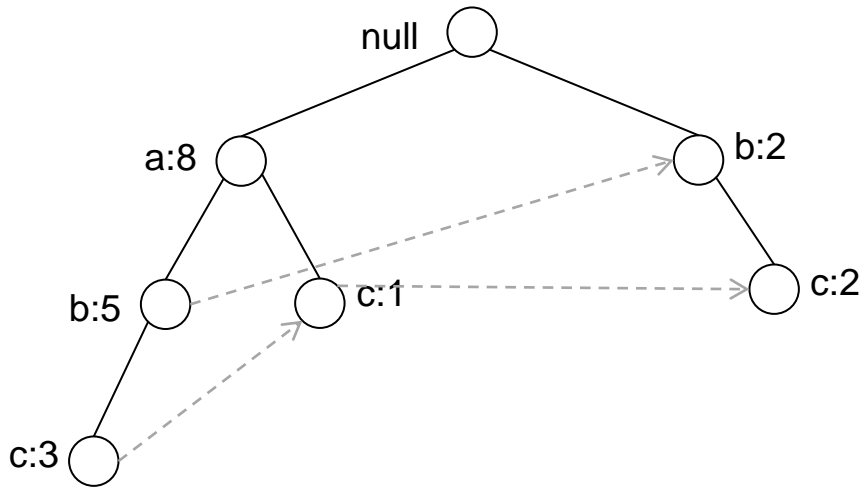


Caminhos contendo d

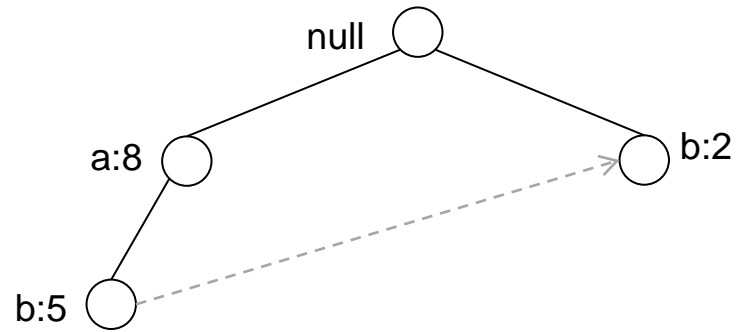
Algoritmo FP-Growth



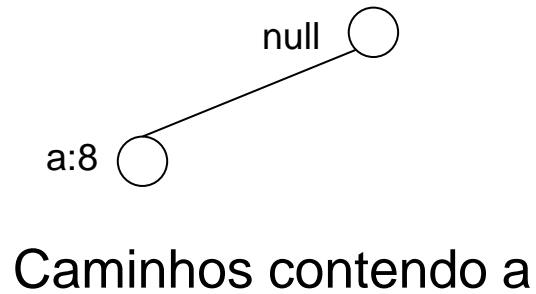
● Exemplo



Caminhos contendo c



Caminhos contendo b

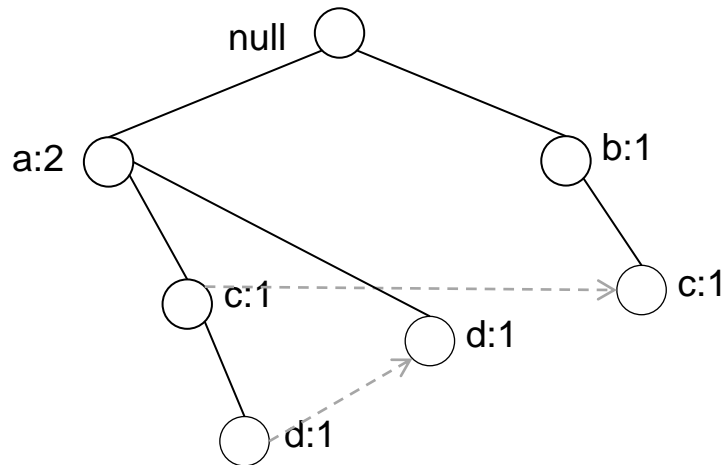


Caminhos contendo a

Algoritmo FP-Growth



● Exemplo



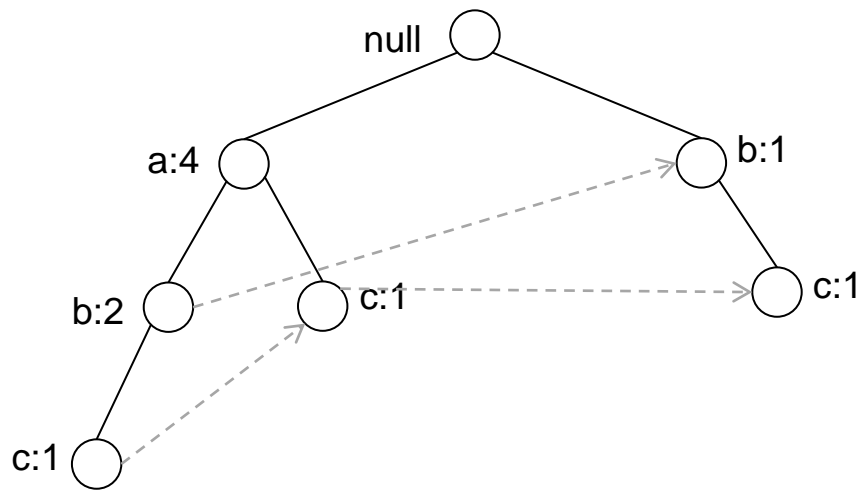
Árvore FP condicional para e

Itemset	Suporte
{d, e}	2
{c, e}	2
{b, e}	4
{a, e}	2
{a, c, e}	4
{a, d, e}	2
{b, c, e}	4
{a, c, d, e}	4

Algoritmo FP-Growth



- Exemplo



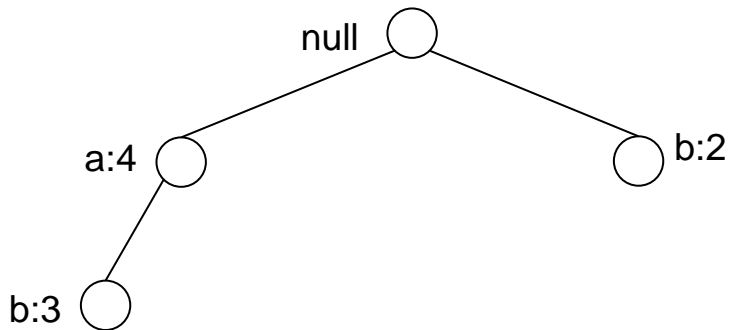
Árvore FP condicional para d

Itemset	Suporte
{c, d}	3
{b, d}	3
{a, d}	4
{b, c, d}	2
{a, c, d}	2
{a, b, d}	2
{a, b, c, d}	4

Algoritmo FP-Growth



- Exemplo



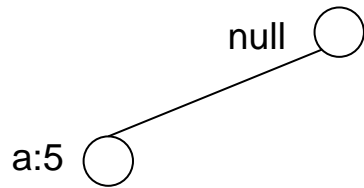
Itemset	Suporte
{b, c}	5
{a, c}	4
{a, b, c}	3

Árvore FP condicional para c

Algoritmo FP-Growth



- Exemplo



Itemset	Suporte
{a, b}	5

Árvore FP condicional para b

Algoritmo FP-Growth



- Exemplo

null ○

Itemset	Suporte
---------	---------

Árvore FP condicional para a

Algoritmo FP-Growth



- E para gerar as regras!?
 - Para cada conjunto frequente $f \in F$, gere todas as possibilidades de regras e avalie sua confiança
 - Valores de suporte, previamente armazenados são utilizados para tal, assim não é necessário ler a base de dados novamente!
 - Conjuntos frequentes de tamanho 1 são descartados

Algoritmo FP-Growth



- E para gerar as regras!?
- Seguindo o exemplo, suponha $\text{minConf} = 0.6$ (α)

Conjunto de itens frequentes (F)
{a} $\sigma = 8$
{b} $\sigma = 7$
{c} $\sigma = 6$
{d} $\sigma = 5$
{e} $\sigma = 3$

Conjunto de itens frequentes (F) cont.
{d, e} $\sigma = 2$
{a, d, e} $\sigma = 2$
{c, e} $\sigma = 2$
{a, e} $\sigma = 2$
{c, d} $\sigma = 3$
{b, c, d} $\sigma = 2$
{a, c, d} $\sigma = 2$
{b, d} $\sigma = 3$
{a, b, d} $\sigma = 2$
{a, d} $\sigma = 4$
{b, c} $\sigma = 5$
{a, b, c} $\sigma = 3$
{a, c} $\sigma = 4$
{a, b} $\sigma = 5$

Regras
{d} \rightarrow {e} $\alpha = 0.40$
{e} \rightarrow {d} $\alpha = 0.66$
{a, d} \rightarrow {e} $\alpha = 0.50$
{a, e} \rightarrow {d} $\alpha = 1.00$
{d, e} \rightarrow {a} $\alpha = 1.00$
{a} \rightarrow {d, e} $\alpha = 0.25$
{d} \rightarrow {a, e} $\alpha = 0.40$
{e} \rightarrow {a, d} $\alpha = 0.66$
{c} \rightarrow {e} $\alpha = 0.33$
{e} \rightarrow {c} $\alpha = 0.66$
...

Algoritmo FP-Growth



- Base de dados para o trabalho
 - Registros de votos no congresso americano
 - <http://archive.ics.uci.edu/ml/datasets/Congressional+Voting+Records>
 - Encontrar regras entre convicções e posicionamento partidário e/ou convicções que são correlacionadas





- Gere a Árvore FP para a seguinte base de dados considerando
 - $\text{minSup} = 0.2$

Registros
{Imigracao_S, AjudaELSlavador_S, ReligiaoEscolas_S, Democrata}
{Imigracao_S, AjudaELSlavador_S, ReligiaoEscolas_N, Democrata}
{Imigracao_N, AjudaELSlavador_N, ReligiaoEscolas_S, Republicano}
{Imigracao_N, AjudaELSlavador_N, ReligiaoEscolas_N, Republicano}
{Imigracao_S, AjudaELSlavador_N, ReligiaoEscolas_N, Democrata}



- *Introdução ao Data Mining*. Steinbach, Michael; Kumar, Vipin; Tan, Pang-ning, Rio de Janeiro: Ed. Ciência Moderna, 2009. Capítulo 6.

