

Algoritmo Apriori

Quando se trata de mineração de regras de associação, o algoritmo *Apriori* entra em cena. Ele encontra todos os conjuntos de itens frequentes, denominados “itemsets frequentes” (L_k).

A tarefa de associação é responsável por grande parte das soluções usadas para descoberta de padrões. Utiliza metodologia não supervisionada e busca encontrar relacionamentos significativos entre os itens de dados armazenados (KANTARDZIC, 2003).

O objetivo desta técnica é encontrar em grandes conjuntos de dados tendências que ajudem a compreender padrões. Os algoritmos de associação buscam encontrar relações entre os itens, verificando os eventos que ocorrem simultaneamente, possibilitando o entendimento de novos modelos (SERRA, 2002).

[...] algoritmo *Apriori* realiza a garimpagem em dois passos: geração e poda. No primeiro, é feita uma varredura sobre o arquivo, a fim de gerar todos os conjuntos de combinações de valores de colunas que aparecem no arquivo. No segundo, são considerados apenas aqueles conjuntos que aparecem no arquivo com uma frequência não menor que um valor mínimo pré-fixado, são os chamados *grandes conjuntos*. A medida da frequência de um conjunto X de valores é chamada de *suporte*. (CARVALHO, 2005)“.

Definição de Suporte:

$$\text{Suporte}(x) = \frac{\text{Nº de registros do arquivo que contêm os elementos do conjunto } X}{\text{Nº total de registros do arquivo}}$$

Quando se procura descobrir regras de associação, é necessário definir um suporte mínimo (*SupMin*) para as regras (CARVALHO, 2005).

Funcionamento do Algoritmo Apriori

O suporte mínimo (*SupMin*) garante a quantidade de vezes que aquele produto aparece nas vendas. Se for definido como suporte mínimo o valor “0,6”, isso quer dizer que o produto deve aparecer no mínimo em 60% do total das vendas para entrar nessa condição.

Digamos que o produto BORRACHA possua suporte mínimo de 0,9. Ele pode ser considerado um “conjunto grande”, porque sai em 90% das vendas (Tabela 01).

| Produto | Suporte |
|-------------------|---------|
| BORRACHA | 0,9 |
| FITA CREPE | 0,9 |
| PAPEL DE PRESENTE | 0,8 |

Tabela 01: Suporte Mínimo dos Produtos

Segundo Carvalho et. al. (2005), para cada grande conjunto de itens X , todas as regras com fator de confiança acima de um mínimo especificado ($Cmin$) são geradas como: $\forall Y \subset X$, se $\text{suporte}(X) / \text{suporte}(X - Y) \geq Cmin$. O fator de confiança é definido pela seguinte fórmula:

$$\text{Confiança}(R) = \frac{\text{Nº de registros com } X \text{ e } Y}{\text{Nº de registros com } X}$$

O fator de confiança determina a quantidade de vezes que os produtos aparecem juntos em uma venda. Se for definido como fator de confiança o valor 0,7, isso quer dizer que entraram nessa condição todos os produtos que apareceram juntos em 70% das vendas (Tabela 02).

| Regra | Fator de confiança |
|-------------------------------|--------------------|
| BORRACHA => FITA CREPE | 0,88 |
| FITA CREPE => BORRACHA | 0,88 |
| BORRACHA => PAPEL DE PRESENTE | 0,77 |
| PAPEL DE PRESENTE => BORRACHA | 0,87 |

Tabela 02: Fator de Confiança

Assim, percebe-se que quando se vende o produto BORRACHA em 88% das vendas, também é vendido o produto FITA CREPE. O mesmo acontece para os demais produtos apresentados.

| Produto 1 | Produto 2 | * Freq. P1 | ** Freq. P1 e P2 | *** % P1 e P2 |
|-------------------|-------------------|---------------|---------------------|------------------|
| BORRACHA | FITA CREPE | 0,9 | 0,88 | 97,77% |
| FITA CREPE | BORRACHA | 0,9 | 0,88 | 97,77% |
| BORRACHA | PAPEL DE PRESENTE | 0,9 | 0,77 | 85,55% |
| PAPEL DE PRESENTE | BORRACHA | 0,8 | 0,77 | 96.25% |

Tabela 03: Resultado da Aplicação do Algoritmo

* Freq. P1 – Frequência do Produto 1;

** Freq. P1 e P2 – Frequência em que Produto 1 e Produto 2 aparecem juntos;

*** % P1 e P2 - Percentagem da probabilidade de compra do Produto 1 atrelado ao produto 2.

Analisando a Tabela 03, entende-se que existe uma probabilidade na compra de um determinado produto associado a outro. Esse fato auxilia na realocação dos produtos. Assim, pode-se colocar os produtos que são vendidos juntos com maior frequência numa mesma prateleira.

Relatório Gerencial Baseado no Algoritmo Apriori

O algoritmo Apriori está armazenado no banco de dados MySQL como uma *Stored Procedure*.

```
CREATE DEFINER=`root`@`localhost` PROCEDURE `sp_apriori`(IN sup_min INTEGER,IN conf INTEGER,
in mes varchar(2), in ano varchar(4))
BEGIN
    DROP TABLE IF EXISTS `l_item`;
    DELETE FROM regras;
    SET @total := (SELECT COUNT(DISTINCT c_venda_codigo ) AS total
    FROM d_venda
    INNER JOIN c_venda ON c_venda.codigo = d_venda.c_venda_codigo
    WHERE MONTH(c_venda.dataVenda) = mes
    AND YEAR(c_venda.dataVenda) = ano
    );

    -- criacao da tabela temporaria para um item
    CREATE TEMPORARY TABLE l_item(
    id INT NOT NULL AUTO_INCREMENT PRIMARY KEY,
    prdno VARCHAR(13) NOT NULL,
    freq FLOAT NOT NULL);

    -- inserindo na tabela temporaria os itens de acordo com o
    -- suporte minimo definido no parametro de entrada
    INSERT INTO l_item(prdno,freq) SELECT
    produto_codigo,ROUND(100*COUNT(produto_codigo)/@total,2) AS freq
    FROM d_venda
    INNER JOIN c_venda ON c_venda.codigo = d_venda.c_venda_codigo
    WHERE MONTH(c_venda.dataVenda) = mes
    AND YEAR(c_venda.dataVenda) = ano
    GROUP BY produto_codigo
    HAVING freq >= sup_min;

    SET @total_aux:= (SELECT COUNT(*) FROM l_item);
    SET @i:=1;

    -- while onde são verificadas as associacoes dos itens classificados acima
    WHILE @i <= @total_aux DO
        SET @cod:= (SELECT prdno FROM l_item WHERE id=@i);
        SET @freqx:= (SELECT freq FROM l_item WHERE id=@i);
        INSERT INTO regras
        SELECT @cod,produto_codigo,@freqx,ROUND(100*COUNT(produto_codigo)/@total,2) AS freq
        FROM d_venda
        INNER JOIN c_venda ON c_venda.codigo = d_venda.c_venda_codigo
        WHERE c_venda_codigo IN(SELECT c_venda_codigo FROM d_venda WHERE
        produto_codigo=@cod)

    -- filtro dos produtos que são vendidos juntos
        AND produto_codigo <> @cod
        AND MONTH(c_venda.dataVenda) = mes
        AND YEAR(c_venda.dataVenda) = ano
        GROUP BY produto_codigo HAVING freq >=sup_min;
        SET @i:=@i+1;
    END WHILE;

    -- seleciona as regras que estao dentro da confianca passada como parametro, ]
    -- ou seja, deleta aquelas que nao interessam
    DELETE FROM regras WHERE (freq_xy/freq_x) < (conf/100);

    SELECT p1.descricao AS produto1,
    p2.descricao AS produto2,
    freq_x,
    freq_xy,
    SUBSTRING(((freq_xy / freq_x) * 100 ),1,5) AS porcetagem
```

```

FROM regras
LEFT JOIN produto AS p1 ON p1.codigo = regras.x
LEFT JOIN produto AS p2 ON p2.codigo = regras.y
;
END

```

Tabela 04: Código Fonte do Algoritmo Apriori

Na Tabela 05 pode-se observar alguns dados armazenados no banco, na tabela de cabeçalho das vendas (c_venda).

| codigo | dataVenda | valor | desconto | total |
|--------|------------|-------|----------|-------|
| 131 | 2009-10-12 | 16 | 0 | 16 |
| 132 | 2009-10-12 | 16 | 0 | 16 |
| 133 | 2009-10-12 | 16 | 0 | 16 |
| 134 | 2009-10-12 | 16 | 0 | 16 |
| 135 | 2009-10-12 | 3,2 | 0 | 3,2 |
| 136 | 2009-10-12 | 3,2 | 0 | 3,2 |
| 137 | 2009-10-12 | 3,2 | 0 | 3,2 |
| 138 | 2009-10-12 | 3,2 | 0 | 3,2 |
| 139 | 2009-10-12 | 12,8 | 0 | 12,8 |
| 140 | 2009-10-12 | 12,8 | 0 | 12,8 |
| 141 | 2009-10-12 | 12,8 | 0 | 12,8 |
| 142 | 2009-10-12 | 12,8 | 0 | 12,8 |
| 143 | 2009-10-12 | 12,8 | 0 | 12,8 |

Tabela 05: Alguns Dados da Tabela de Cabeçalho das Vendas

Na Tabela 06 pode-se observar alguns dados armazenados no banco, na tabela de detalhes das vendas (d_venda), relacionados com os dados da Tabela 05.

| c_venda_codigo | produto_codigo | quantidade | vlrUnitario | vlrTotal |
|----------------|----------------|------------|-------------|----------|
| 131 | 0001190000009 | 1 | 1,5 | 1,5 |
| 131 | 0001234567898 | 1 | 5,8 | 5,8 |
| 131 | 1201011747598 | 1 | 1,7 | 1,7 |
| 131 | 1201011759164 | 1 | 7 | 7 |
| 132 | 0001190000009 | 1 | 1,5 | 1,5 |
| 132 | 0001234567898 | 1 | 5,8 | 5,8 |
| 132 | 1201011747598 | 1 | 1,7 | 1,7 |

| | | | | |
|-----|---------------|---|-----|-----|
| 132 | 1201011759164 | 1 | 7 | 7 |
| 133 | 0001190000009 | 1 | 1,5 | 1,5 |
| 133 | 0001234567898 | 1 | 5,8 | 5,8 |
| 133 | 1201011747598 | 1 | 1,7 | 1,7 |
| 133 | 1201011759164 | 1 | 7 | 7 |
| 134 | 0001190000009 | 1 | 1,5 | 1,5 |
| 134 | 0001234567898 | 1 | 5,8 | 5,8 |
| 134 | 1201011747598 | 1 | 1,7 | 1,7 |
| 134 | 1201011759164 | 1 | 7 | 7 |
| 135 | 0001190000009 | 1 | 1,5 | 1,5 |
| 135 | 1201011747598 | 1 | 1,7 | 1,7 |
| 136 | 0001190000009 | 1 | 1,5 | 1,5 |
| 136 | 1201011747598 | 1 | 1,7 | 1,7 |
| 137 | 0001190000009 | 1 | 1,5 | 1,5 |
| 137 | 1201011747598 | 1 | 1,7 | 1,7 |
| 138 | 0001190000009 | 1 | 1,5 | 1,5 |
| 138 | 1201011747598 | 1 | 1,7 | 1,7 |
| 139 | 0001234567898 | 1 | 5,8 | 5,8 |
| 139 | 1201011759164 | 1 | 7 | 7 |
| 140 | 0001234567898 | 1 | 5,8 | 5,8 |
| 140 | 1201011759164 | 1 | 7 | 7 |
| 141 | 0001234567898 | 1 | 5,8 | 5,8 |
| 141 | 1201011759164 | 1 | 7 | 7 |
| 142 | 0001234567898 | 1 | 5,8 | 5,8 |
| 142 | 1201011759164 | 1 | 7 | 7 |
| 143 | 0001234567898 | 1 | 5,8 | 5,8 |
| 143 | 1201011759164 | 1 | 7 | 7 |

Tabela 06: Alguns Dados da Tabela de Detalhe das Vendas

Na Tabela 07 pode-se observar o resultado da aplicação do Algoritmo Apriori nos dados das Tabelas 05 e 06.

| Produto1 | Produto2 | P1 | P1_P2 | Porcentagem |
|---------------|---------------|-------|-------|-------------|
| 0001190000009 | 0001234567898 | 61,54 | 30,77 | 50 |
| 0001190000009 | 1201011747598 | 61,54 | 61,54 | 100 |
| 0001190000009 | 1201011759164 | 61,54 | 30,77 | 50 |
| 0001234567898 | 0001190000009 | 69,23 | 30,77 | 44,44 |
| 0001234567898 | 1201011747598 | 69,23 | 30,77 | 44,44 |
| 0001234567898 | 1201011759164 | 69,23 | 69,23 | 100 |
| 1201011747598 | 0001190000009 | 61,54 | 61,54 | 100 |

| | | | | |
|---------------|---------------|-------|-------|-------|
| 1201011747598 | 0001234567898 | 61,54 | 30,77 | 50 |
| 1201011747598 | 1201011759164 | 61,54 | 30,77 | 50 |
| 1201011759164 | 0001190000009 | 69,23 | 30,77 | 44,44 |
| 1201011759164 | 0001234567898 | 69,23 | 69,23 | 100 |
| 1201011759164 | 1201011747598 | 69,23 | 30,77 | 44,44 |

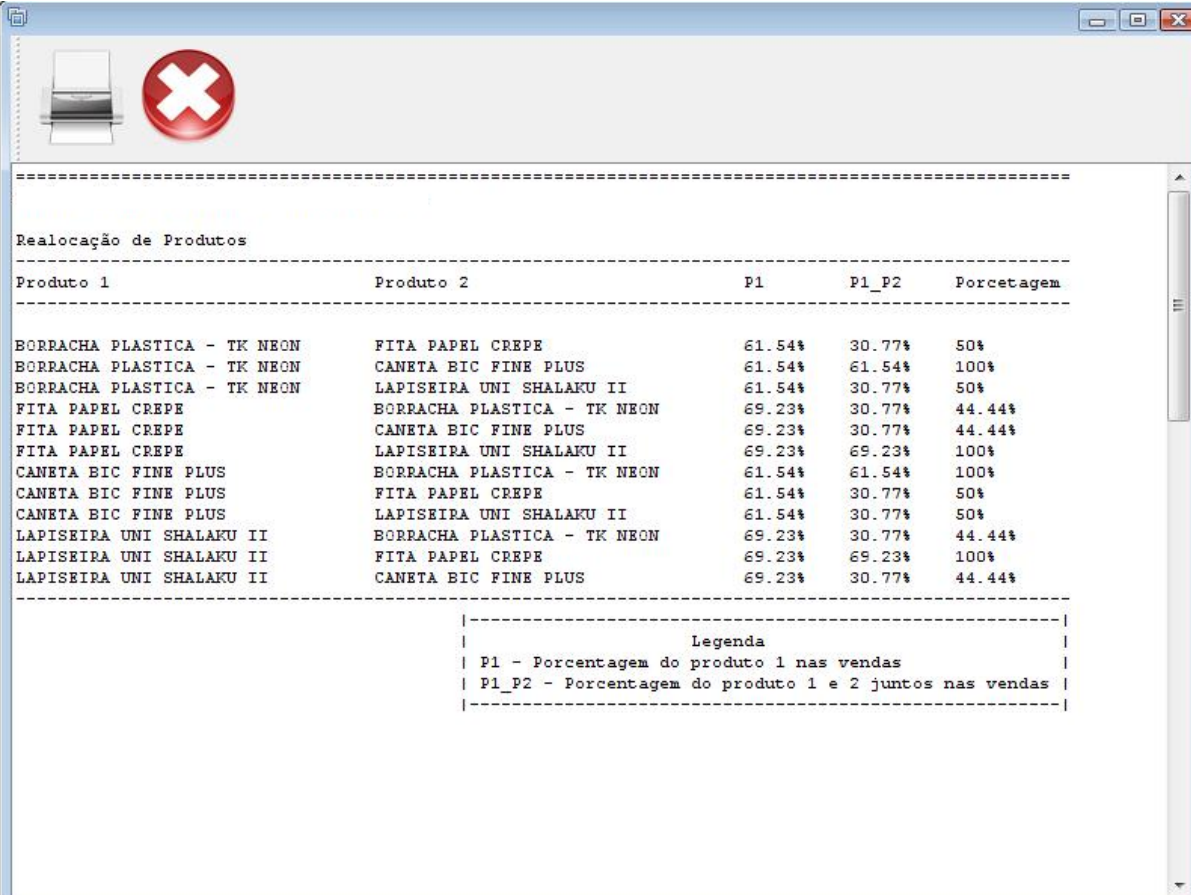
Tabela 07: Resultado da Aplicação do Algoritmo Apriori

Legenda

P1 - Porcentagem do produto 1 nas vendas

P1_P2 - Porcentagem do produto 1 e 2 juntos nas vendas

Para executar o algoritmo, deve-se chamar a *stored procedure* de dentro da aplicação construída passando os devidos parâmetros. Pode-se criar um relatório como visto na imagem abaixo. A criação desse relatório ficará como um exercício. Vejamos a chamada da *stored procedure* na prática.



| Produto 1 | Produto 2 | P1 | P1_P2 | Porcetagem |
|-----------------------------|-----------------------------|--------|--------|------------|
| BORRACHA PLASTICA - TK NEON | FITA PAPEL CREPE | 61.54% | 30.77% | 50% |
| BORRACHA PLASTICA - TK NEON | CANETA BIC FINE PLUS | 61.54% | 61.54% | 100% |
| BORRACHA PLASTICA - TK NEON | LAPISEIRA UNI SHALAKU II | 61.54% | 30.77% | 50% |
| FITA PAPEL CREPE | BORRACHA PLASTICA - TK NEON | 69.23% | 30.77% | 44.44% |
| FITA PAPEL CREPE | CANETA BIC FINE PLUS | 69.23% | 30.77% | 44.44% |
| FITA PAPEL CREPE | LAPISEIRA UNI SHALAKU II | 69.23% | 69.23% | 100% |
| CANETA BIC FINE PLUS | BORRACHA PLASTICA - TK NEON | 61.54% | 61.54% | 100% |
| CANETA BIC FINE PLUS | FITA PAPEL CREPE | 61.54% | 30.77% | 50% |
| CANETA BIC FINE PLUS | LAPISEIRA UNI SHALAKU II | 61.54% | 30.77% | 50% |
| LAPISEIRA UNI SHALAKU II | BORRACHA PLASTICA - TK NEON | 69.23% | 30.77% | 44.44% |
| LAPISEIRA UNI SHALAKU II | FITA PAPEL CREPE | 69.23% | 69.23% | 100% |
| LAPISEIRA UNI SHALAKU II | CANETA BIC FINE PLUS | 69.23% | 30.77% | 44.44% |

| Legenda | |
|---------|--|
| P1 | - Porcentagem do produto 1 nas vendas |
| P1_P2 | - Porcentagem do produto 1 e 2 juntos nas vendas |