

# Mineração de Dados para Identificar Aspectos que Influenciam o Desempenho nas Vendas de Produtos em uma Indústria de Confeção Do Vestuário

Um Estudo de Caso Sobre os Dados de uma Indústria de Confeção do Vestuário

Anderson D. E. Melchiorretto<sup>†</sup>  
Laboratório de Inteligência Aplicada  
Universidade do Vale do Itajaí/UNIVALI  
Itajaí, Santa Catarina, Brasil  
anderson.melchiorretto@edu.univali.br

Rafael B. Martins  
Laboratório de Inteligência Aplicada  
Universidade do Vale do Itajaí/UNIVALI  
Itajaí, Santa Catarina, Brasil  
ballottin@univali.br

## ABSTRACT

*In garment industry, the production process is composed of several steps, starting with product development, production until it reaches the customer. All these processes generate a large volume of data, about products, customers, sales, among others. The lack of knowledge about this data generated in the companies, can generate problems such as product returns and consequently the loss of customers. The application of data mining (DM) tasks can extract hidden patterns from this data. In this context, DM techniques following Knowledge Discovery in Database (KDD) steps were used to identify patterns that may influence product sales performance in an apparel manufacturing industry. The dataset considered were composed of 31 attributes, about customers, products, sales, from the years 2018 through 2020. The main results of this analysis identified the attributes that most influence product sales, which are: "product line", "quota item", "customer code", and "tax coupon number".*

## KEYWORDS

Garment Industry, KDD, Data Mining.

## 1. INTRODUÇÃO

A cadeia produtiva têxtil e confecção inicia-se na obtenção das fibras, transformando estas em fios, em seguida passa pela tecelagem ou malharia. Após a transformação dos fios em tecidos, é realizado o beneficiamento dos tecidos, e por fim vai para a etapa de confecção. A confecção é a etapa final da cadeia produtiva têxtil e muitas vezes faz a conexão com o consumidor, sendo a meta final deste processo [1].

O processo produtivo da confecção é uma sequência operacional que se inicia no planejamento da coleção e desenvolvimento do produto, passando por toda a produção até a expedição. Na indústria são necessários os dados de cada processo individual e parâmetro em um formato compilado para tomar uma decisão rápida e assertiva [2]. Todo processo de confecção até a venda final gera naturalmente grandes volumes de dados sobre os produtos, clientes, vendas, entre outros. Analisar e identificar

problemas dentre esses processos tornam-se uma tarefa complexa, que impossibilitam um ser humano realizar de forma manual.

Neste sentido, este trabalho visa aplicar técnicas de DM através do processo KDD em uma base de dados de uma indústria de confecção do vestuário. Teve-se como foco neste artigo, descobrir a frequência de produtos vendidos a cada coleção por cliente, e identificar a frequência de itens vendidos por linha de produto e artigo de cotas, das vendas B2B (*Business-to-business*, "De empresa para empresa", em português), e B2C (*Business-to-consumer*, "Empresa para consumidor", em português).

## 2. DESENVOLVIMENTO

Foram extraídos os dados de uma base de dados de uma indústria de confecção do vestuário, para o estudo de padrões que influenciam nas vendas de produtos. A empresa trabalha com foco em B2B, no entanto, também possui loja em anexo a fábrica, trabalhando assim com B2C. Neste estudo a ferramenta escolhida foi Weka<sup>1</sup> (*Waikato Environment for Knowledge Analysis*), devido conter todos os algoritmos necessários para a realização do DM e de partes das etapas de pré e pós-processamento, e a qual se destacou na literatura consultada [3].

Neste trabalho, foi adotado a metodologia CRISP-DM, como método para aplicação das técnicas de DM. O CRISP-DM foi um dos primeiros e mais bem aceitos modelos voltados à orientação do processo de KDD [4].

Os dados disponibilizados pela empresa de confecção do vestuário foram compostos por 31 diferentes atributos e por um total de 300.032 instâncias das coleções de inverno e verão entre os anos de 2018 até 2020. No total, existem registros de 2.262 clientes distintos, 1.725 produtos e 6 coleções. Cada registro presente nas coleções das vendas b2b são referentes a um número de nota fiscal contendo informações de clientes, produtos e da própria venda. Nas vendas b2c cada registro é referente a um número de cupom fiscal, contendo apenas informações do produto e da própria venda. A Tabela 1, apresenta cada atributo utilizado neste trabalho.

No total foram criadas mais de 45 versões de *datasets* através dos dados disponibilizados pela empresa. No entanto, serão apenas apresentadas as versões que obtiveram resultados relevantes.

<sup>1</sup> <https://www.cs.waikato.ac.nz/ml/weka/>

Tabela 1: Atributos das vendas.

Atributo	Descrição
COLECAO = [13-18]	Código da coleção
LINHA_PRODUTO = [1-12]	Código da linha do produto.
ARTIGO_COTAS = [1-21]	Código do tipo de produto.
CD_CLIENTE = [1-2262]	Código do cliente
NUM_NOTA_FISCAL = [ > 0]	Número da nota fiscal
NUMNOT = [ > 0]	Número de cupom fiscal
DATA_HR_VENDA	Data e hora da venda
QTD_REF_VENDA = [ > 0]	Quantidade de itens comprados por produto
VAL_VENDA = [ > 0 ]	Valor da compra
VAL_DESCONTO [ > 0]	Valor de desconto
COND_PGT = [D; CD; CC]	Tipo de pagamento realizado. Ex: Dinheiro; Débito; Crédito

## 2.1. Algoritmos de DM

Na geração e avaliação do modelo de classificação foram utilizados o algoritmo “J48” e “JRip”. O algoritmo “J48” é uma implementação do algoritmo de árvores de decisão “C4.5” presente na ferramenta Weka. A qualidade dos modelos gerados pelos algoritmos de classificação é medida pelos valores apresentados na matriz de confusão gerada pela ferramenta. Para validação dos modelos gerados nos algoritmos de classificação foi utilizado a técnica de *10-fold-cross-validation*. As medidas de desempenho consideradas neste trabalho foram a estatística Kappa e o F1-Score.

Na aplicação do algoritmo Apriori para a tarefa de associação, foram utilizados alguns parâmetros disponíveis na ferramenta, a qual foram configurados conforme demonstrados na Tabela 2.

Tabela 2: Configuração dos datasets do algoritmo Apriori.

Nome Dataset	Total de atributos	Total registros	Suporte Mín.	Conf.
PF_B2B_P1	6	2.262	0.05	0.85
PF_B2B_COL_14_P3	10	866	0.4	0.9
PF_B2B_COL_16_P3	10	980	0.46	0.9
PF_B2B_COL_18_P3	10	777	0.46	0.9

## 3. RESULTADOS

### 3.1. Regras de Associação B2B

No dataset “PF\_B2B\_P1”, foram impostos o suporte mínimo de 0.05 e a métrica de confiança de 0.85, para encontrar quais clientes compravam com frequência em determinadas coleções. Na Tabela 3 são apresentados os principais resultados extraídos das regras com a confiança mínima esperada. Nas regras 1, 4, 5, 6 e 8, observou-se que os clientes compraram em cinco das seis coleções, variando de 125 a 176 clientes dos 2.262 presentes no dataset. Nas regras 18, 22 e 29 os clientes representaram aproximadamente 9% do total de clientes que compraram em pelo menos 4 das 6 coleções. Por fim foram identificados 121 clientes que compraram em todas as coleções, representando um total de 5,35% do total de clientes existentes na base de dados. Conclui-se que 82,3% dos 686 clientes que compraram na coleção 13, não compraram em uma das outras coleções ou em nenhuma.

Tabela 3: Regras de associação para o dataset “PF\_B2B\_P1”.

Nº	Regra	Conf.
1	13=True 15=True 16=True 17=True 146 ==> 14=True 142	0.97
2	13=True 15=True 16=True 17=True 18=True 125 ==> 14=True 121	0.97
4	13=True 15=True 17=True 18=True 137 ==> 14=True 130	0.95
5	13=True 14=True 16=True 18=True 153 ==> 15=True 144	0.94
6	13=True 16=True 17=True 18=True 134 ==> 14=True 126	0.94
8	13=True 16=True 17=True 18=True 134 ==> 15=True 125	0.94
12	15=True 16=True 17=True 18=True 192 ==> 14=True 176	0.92
18	13=True 14=True 16=True 242 ==> 15=True 219	0.9
22	15=True 16=True 17=True 237 ==> 14=True 212	0.89
29	15=True 16=True 18=True 249 ==> 14=True 220	0.88

Os datasets apresentados na Tabela 4 até a Tabela 6, procurou-se encontrar 5 atributos em um conjunto de regras, que representassem aproximadamente 50% do total de linha de produtos vendidos com frequência nas coleções de verão. A Tabela 4 apresenta as principais regras para a coleção 14 – Verão 2018/19. Nota-se que foram encontrados 6 conjuntos de regras com 5 atributos. Observou-se que as mesmas regras 1 e 7, foram encontrados nos datasets das coleções 16 e 18.

Tabela 4: Regras de associação – dataset PF\_B2B\_COL\_14\_P3.

Nº	Regra	Conf.
1	M_Bebe=True F_Bebe=True F_Juvenil=True M_Infantil=True 443 ==> F_Infantil=True 441	100
7	M_Bebe=True F_Bebe=True F_Juvenil=True M_Juvenil=True 402 ==> F_Infantil=True 400	100
8	F_Infantil=True F_Bebe=True F_Adulto=True M_Juvenil=True 402 ==> F_Juvenil=True 400	100
15	F_Infantil=True F_Bebe=True M_Infantil=True M_Juvenil=True 422 ==> F_Juvenil=True 419	99
25	F_Infantil=True M_Bebe=True M_Infantil=True M_Juvenil=True 406 ==> F_Juvenil=True 402	99
30	F_Bebe=True F_Juvenil=True F_Adulto=True M_Infantil=True 432 ==> F_Infantil=True 427	99

Tabela 5: Regras de associação – dataset PF\_B2B\_COL\_16\_P3.

Nº	Regra	Conf.
4	M_Bebe=True F_Bebe=True F_Juvenil=True M_Infantil=True 492 ==> F_Infantil=True 490	100
6	F_Bebe=True F_Juvenil=True M_Infantil=True M_Juvenil=True 465 ==> F_Infantil=True 463	100
7	M_Bebe=True F_Bebe=True M_Infantil=True M_Juvenil=True 450 ==> F_Infantil=True 448	100
12	M_Bebe=True F_Bebe=True F_Juvenil=True M_Juvenil=True 457 ==> F_Infantil=True 453	99
13	M_Bebe=True F_Juvenil=True M_Infantil=True M_Juvenil=True 455 ==> F_Infantil=True 451	99
23	M_Bebe=True F_Bebe=True M_Infantil=True M_Juvenil=True 450 ==> F_Juvenil=True 444	99
8	M_Bebe=True F_Bebe=True F_Juvenil=True M_Infantil=True M_Juvenil=True 444 ==> F_Infantil=True 442	100
94	F_Juvenil=True F_Plus=True 457 ==> F_Adulto=True 439	96

Na Tabela 5 foram apresentadas as principais regras para a coleção 16–Verão 2019/20. Nota-se que foram encontrados os mesmos conjuntos de regras com 5 atributos, e 1 regra com 6 atributos em comparação com a coleção 18, a qual é apresentada na Tabela 6. Já na regra 94 identificou-se que 44% dos clientes compraram a linha feminina: juvenil, adulto e plus.

A tabela 6 apresenta as principais regras encontradas para a coleção 18, com a confiança mínima esperada.

**Tabela 6: Regras de associação – dataset PF\_B2B\_COL\_18\_P3.**

Nº	Regra	Conf.
3	M_Bebe=True F_Bebe=True F_Juvenil=True M_Infantil=True 428 ==> F_Infantil=True 427	100
5	M_Bebe=True F_Bebe=True F_Juvenil=True M_Juvenil=True 414 ==> F_Infantil=True 413	100
6	M_Bebe=True F_Bebe=True M_Infantil=True M_Juvenil=True 412 ==> F_Infantil=True 411	100
14	F_Bebe=True F_Juvenil=True M_Infantil=True M_Juvenil=True 415 ==> F_Infantil=True 413	100
19	M_Bebe=True F_Juvenil=True M_Infantil=True M_Juvenil=True 420 ==> F_Infantil=True 417	99
22	M_Bebe=True F_Bebe=True M_Infantil=True M_Juvenil=True 412 ==> F_Juvenil=True 409	99
7	M_Bebe=True F_Bebe=True F_Juvenil=True M_Infantil=True M_Juvenil=True 409 ==> F_Infantil=True 408	100

Todas as regras apresentadas tiveram a confiança maior que 96%. As linhas de produtos Bebê, Infantil e Juvenil para os sexos Feminino e Masculino, foram comprados em conjuntos por mais de 42% dos clientes a cada coleção nas vendas b2b. As informações obtidas através das regras ajudaram a compreender qual o perfil por linha de produto grande parte dos clientes costumam comprar nas coleções de verão.

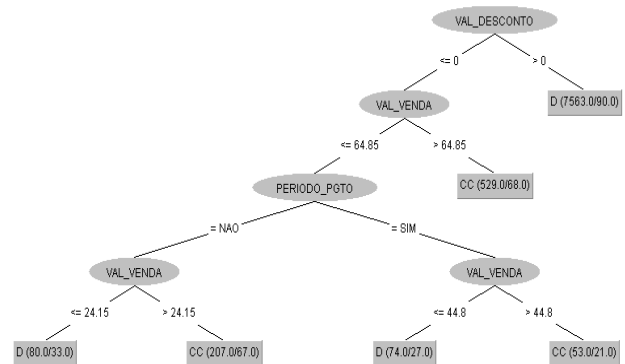
### 3.2. Classificação

Na tarefa de classificação, foi utilizado o dataset “PF\_B2C\_P8” que possui os dados das vendas B2C. Neste dataset buscou-se identificar o comportamento por tipo de pagamento das vendas. Foram realizados vários testes para se chegar no melhor modelo para a realização desta tarefa. O parâmetro “minNumObj” foi definido no valor de 50, por ter apresentado uma árvore simples de ser compreendida. Foram aplicados dois algoritmos na realização desta tarefa, sendo: J48 e JRip, no qual a Tabela 7 apresenta a matriz de confusão e os resultados da avaliação dos modelos.

**Tabela 7: Resultado do modelo – dataset PF\_B2C\_P8.**

Algoritmos	J48		JRip	
	<b>Matriz de confusão</b>	<b>7533</b>	<b>190</b>	<b>7499</b>
	139	<b>644</b>	109	<b>674</b>
<b>Pagamento (D OU CC)</b>	<b>D</b>	<b>CC</b>	<b>D</b>	<b>CC</b>
<b>Precisão (%)</b>	98,2	77,2	98,6	75,1
<b>Recall (%)</b>	97,5	82,2	97,1	86,1
<b>F1-Score</b>	0,978	0,796	0,978	0,802
<b>Kappa</b>	0,7752		0,7763	

Os resultados apresentados na Tabela 17 mostraram que os algoritmos classificadores conseguiram criar modelos com a métrica de avaliação F1-Score acima de 0,75. Por fim o modelo obtido para o algoritmo J48 encontram-se na Figura 1.



**Figura 1: Resultado da árvore de decisão.**

Nota-se que quando é aplicado algum desconto na compra, geralmente a venda é pago à vista. No entanto, quando não se aplica desconto, é analisado se o valor total da venda é maior que R\$ 64,85 frequentemente são efetuados o pagamento por cartão de crédito, no entanto valor abaixo deste, é analisado o período de pagamento, que é entre os dias 03 á 15 de cada mês. Notou-se que quando é dentro do período de pagamento as vendas abaixo de R\$ 44,80 são pagas em dinheiro. Já quando não é período de pagamento, compras acima de R\$ 24,15 são parcelados no cartão.

### 4. CONCLUSÕES

Os resultados obtidos na tarefa de associação em relação aos atributos “linha\_produto”, permitiu avaliar e evidenciar que alguns fatores podem apresentar maior influência nas vendas de produtos. Das regras encontradas nas vendas b2b por linha de produto, concluiu-se que as linhas Bebê, Infantil e Juvenil em ambos os sexos, são comprados em conjunto por mais de 42% dos clientes a cada coleção. Identificou-se que 5,35% dos clientes compram recorrentes em todas as coleções, isso tende a mostrar que parte dos clientes compram em apenas em uma estação do ano, seja verão ou inverno, ou ainda não estão satisfeitos com o produto ou atendimento. De forma geral, as regras encontradas podem ser utilizadas para melhorar o processo de comercialização dos produtos e ajuda na relação e retenção dos clientes.

### REFERÊNCIAS

- [1] Patricia L. Gutierrez. 2006. Aprender a Empreender- Têxtil e Confeção.
- [2] Prof., S. K. Tyagi and Dr., B. K Sharma. 2011. Data Mining Tools and Techniques to Manage the Textile Quality Control Data for Strategic Decision Making. International Journal of Computer Applications (0975 – 8887) Vol. 13, n 4, pp. 26-29, 2011.
- [3] JOVIC, A.; BRKIC, K.; BOGUNOVIC, N.. *An overview of free software tools for general data mining*. 2014 37th International Convention On Information And Communication Technology, Electronics And Microelectronics (Mipro), [S.L.], p. 1112-1117, maio 2014. <http://dx.doi.org/10.1109/mipro.2014.6859735>.
- [4] Ronaldo Goldschmidt, Emmanuel Passos, and Eduardo Bezerra. 2015. *Data Mining: Conceitos, Técnicas, Algoritmos, Orientações e Aplicações*.