

# Aplicação de Técnicas de Mineração de Dados para Identificação de Fatores que Influenciam o Desempenho Comercial e Industrial em uma Indústria de Beneficiamento de Malha.

Maxsuel Vieira Orti, Rafael Ballottin Martins

Laboratório de Inteligência Aplicada – Centro de Ciências Tecnológicas da Terra e do Mar – Universidade do Vale do Itajaí (UNIVALI)  
Rua Uruguai, 458 - 88302-202 – Itajaí– SC – Brasil  
maxsuel@edu.univali.br, ballottin@univali.br

**Abstract.** *The advances in globalization caused an increase in the competition between companies, where is of ultimate necessity to obtain a commercial advantage over a competitor. The commercial advantage can be obtained by reducing costs, improving the business process and by better understanding of the client's needs. The use of data mining tasks like association, classification and clustering can help by finding patterns and discovering knowledge previous unknown. In this context, a study was made with the objective of using data mining tasks in a Textile Industry, looking for patterns in both production and commercial process. The results of this analysis generated rules with potential to generate improvements in both industrial and commercial processes.*

**Resumo.** *O avanço na globalização provocou um aumento na competição entre empresas, gerando uma disputa onde é necessário obter uma vantagem comercial sobre seus concorrentes. A vantagem comercial pode ser obtida ao diminuir custos, melhorar processos e entender melhor as necessidades dos clientes. A aplicação de tarefas de mineração de dados como associação, classificação e clusterização pode extrair padrões em dados e gerar conhecimento. Neste contexto, foram aplicadas tarefas de mineração de dados em uma indústria têxtil, procurando por padrões nos processos industriais e comerciais. Os resultados desta análise geraram regras com potencial para gerar melhorias tanto nos processos industriais quanto comerciais.*

## 1. Introdução

A competição entre empresas está cada vez mais acirrada e para enfrentar os desafios da concorrência os gestores estão utilizando com maior frequência sistemas de informações que ajudam na tomada de decisões [Huang, Liu e Chang 2012]. Os consumidores exigem, das indústrias, uma diminuição no tempo de produção, aumento na complexidade de um produto e a possibilidade de customização [Stockton, Khalil e Mukhongo 2012]. Os processos de uma indústria são complexos, com várias etapas e muitas variáveis, fazendo com que engenheiros tenham dificuldades em controlar e identificar problemas [Çiflikli e Kahya-Özyirmidokuz 2010]. A informatização de todos os processos dentro de uma empresa gera uma grande quantidade de dados, fazendo com que empresas em

quase todas as áreas de atuação explorem os dados existentes para ganhar vantagem competitiva [Provost e Fawcett 2013].

Sistemas industriais estão ativamente armazenando dados relacionados a vários aspectos, como: estoque, design de produto, planejamento de processo industrial, sistemas e processos de produção, monitoramento e diagnóstico e previsão de mercado [Wang 2007]. Segundo Wang (2007) uma forma de transformar os dados armazenados em conhecimento é utilizar uma abordagem de mineração de dados com base em inteligência computacional, *machine learning* e estatística avançada. Os princípios que formalizam como aplicar essas técnicas e ferramentas é conhecido como descoberta de conhecimento em base de dados.

O processo de descoberta de conhecimento em base de dados (KDD) é utilizado em todo o processo de descoberta de conhecimento, que engloba, além da aplicação de técnicas de preparação e mineração de dados, a identificação do problema através de interação com especialistas no domínio da aplicação para o entendimento do contexto em que os dados estão inseridos [Augusto Junior 2015]. A metodologia CRISP-DM é uma formalização do KDD [Augusto Junior 2015] e pode ser dividida em seis etapas: entendimento do negócio; entendimento dos dados; preparação dos dados; modelagem; avaliação e desenvolvimento [Mariscal, Marbán e Fernández 2010].

As principais técnicas de mineração de dados incluem classificação, predição, clusterização, detecção de outliers, regras de associação, análise de sequência, análise de eventos em série, mineração de texto, e também algumas novas técnicas como mineração de opiniões e análise de redes sociais [Zhao 2012].

Neste artigo, uma indústria de beneficiamento de malha é estudada e a metodologia CRISP-DM é utilizada como formalização do KDD, buscando utilizar as tarefas de mineração de dados para descobrir fatores que influenciam os fatores comerciais e industriais na indústria pesquisada.

A tarefa de associação foi utilizada para identificar os fatores comerciais através do algoritmo FP-GROWTH e CREATE ASSOCIATION RULES, para o mesmo objetivo foi utilizada a tarefa de classificação com os algoritmos DECISION TREE e TREE TO RULES. Para o setor industrial, foi utilizada a tarefa de classificação com o algoritmo RULE INDUCTION. A ferramenta RapidMiner foi utilizada para executar essas atividades.

## **2. Trabalhos Relacionados**

A escolha das tarefas e algoritmos utilizados, foi inicialmente realizada com base em alguns trabalhos similares, sendo ajustadas de acordo com as características do problema a medida que o processo foi sendo realizado. Estes trabalhos foram selecionados sob a premissa de apresentarem soluções para problemas relacionados a aplicação de técnicas de mineração de dados em indústrias, focando nos processos e na relação comercial com o cliente.

O trabalho de Kamsu-Foguem et al. (2013) procurou identificar o que causa atrasos no processo industrial e se assemelha a essa pesquisa por que as mesmas variáveis podem ser observadas nos dados explorados. A base de dados estudada nesse trabalho mantém informações sobre os tempos de entrada e saída de máquinas, além de apontamentos em todos os processos industriais.

Lee et al. (2013) buscaram a relação de defeitos dentro de um processo de confecção com o objetivo de identificar quais os defeitos acontecem juntos e quais as relações desse defeito com o processo. A similaridade existe por que os dados analisados por Lee et al. (2013) podem ser encontrados na base de dados analisada. Existe uma necessidade de identificar os processos e suas relações com os defeitos que acontecem durante as várias etapas.

Altintas e Trick (2014) pesquisaram a efetividade dos pedidos realizados por clientes em um caso onde esses clientes tinham a possibilidade de alterar a quantidade do pedido. O caso de Altintas e Trick (2014) é específico para a necessidade da empresa pesquisada, mas pode ser aproveitado no contexto de análise de pedidos de clientes.

### 3. Fatores Comerciais

A indústria pesquisada vende rolos de malhas de 20 kg de diferentes cores, tendo clientes distribuídos em 16 estados brasileiros. Hipóteses foram levantadas e buscou-se validar a relação de vendas de cores por feriado, temperatura no momento da aquisição, que podia apresentar valores diferentes das temperaturas comuns das estações devido a condições atípicas, e a tendência. Além disso, foi estabelecido uma relação entre as cores vendidas e identificado quando elas vendem juntas.

As informações de temperatura e tendência foram retiradas de um banco de dados externo ao da indústria. Para as informações de temperatura, foi utilizado o Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP). As informações de tendência de cores foram extraídas da análise de tendências feita pela empresa Pantone, que segundo os especialistas, definem as cores mais utilizadas em determinadas épocas. O período analisado nesse trabalho foi o de 01/01/2014 até 31/12/2016. Os atributos estudados podem ser visualizados na Tabela 1.

**Tabela 1. Atributos Comerciais**

Atributo	Descrição
COD_NOTA_FISCAL	Código de identificação de uma venda realizada
DESC_COR = [1...425]	Cada cor é uma coluna, se a cor está no pedido o valor é True.
ESTADO = [1...15]	Cada estado é uma coluna, se o pedido tiver origem em um estado o valor é True
TENDENCIA	Se o pedido tem uma cor tendência o valor é True.
FERIADO = “período”	Se o pedido não foi feito em um feriado “n_feriado”, ou foi feito em algum período comemorativo: “natal”, “pais”, “namorados”, “crianças” ou “mães”.
TEMP = [1...7]	Faixas de temperatura a partir do range1 até o range7.

### 4. Fatores Industriais

A indústria pesquisada tem 26 máquinas de tingimento, um armazém automatizado, uma cozinha industrial, três tanques de amaciante automatizado e quatro ramas. Uma peça é carregada em uma máquina de tingimento, sendo informado qual é a cor que ela deve ser tingida. O armazém automatizado busca pelo corante para o tingimento, a cozinha

industrial começa a seleção dos produtos químicos que serão enviados e a máquina de tingir começa os processos de tingimento. O tempo de tingimento varia conforme cor e a máquina. Após o tingimento, a peça é colocada em um equipamento chamado Abridor que tem como função abrir a malha para que ela seja utilizada em uma rama. A rama faz o processo de termo fixação da cor na malha.

A pesquisa busca associar as falhas com as etapas do processo de tingimento, avaliando os tempos entre os processos, a cor e as máquinas utilizadas durante a produção de um lote. O período estudado é de 01/01/2016 até 31/12/2016 e os atributos utilizados são apresentados na Tabela 2.

**Tabela 2. Atributos Industriais**

Atributo	Descrição
COD_OB_ITEM	Item de um Lote
MAQUINA_TINGIMENTO	Máquina que a peça foi tingida
RAMA	Rama que termofixou a peça
CONFORMIDADE	A falha apresentada pela peça
TEMPO_REAL	Tempo que demorou para tingir uma peça
TEMPO_TEORICO	Tempo teórico, calculado por sistema, que deveria ter demorado
TIEMPOPARO	Tempo que a máquina ficou parada
TEMPO RAMA – TEMPO ABRIDOR	Tempo entre que de espera entre a Rama e o Abridor
TEMPO REAL – TEMPO TEORICO	Diferença do tempo real para o tempo teórico.
DESC_COR	Descrição da Cor produzida

## 5. Resultados

Nesta seção são apresentados os resultados obtidos através da execução dos atributos selecionados com os algoritmos de associação e classificação. Os principais resultados e regras são expostos e é explicado a relação de importância deles para a indústria pesquisada.

### 5.1. Análise Comercial

Duas tarefas de mineração de dados foram utilizadas na análise dos dados do comercial, a primeira busca associar atributos com uma cor e usa os algoritmos FP-Growth e Create Association Rules. As regras de associação foram criadas considerando um suporte mínimo de 0.1 e uma confiança superior a 0.6. Esses números foram definidos com base em testes e análises dos resultados na perspectiva do especialista. As principais regras podem ser observadas na Tabela 3.

**Tabela 3. Regras de Associação**

Premissas	Conclusão	Suporte	Confiança
BRANCO, PICADILLY	PRETO	0.143	0.84
BRANCO, PINK	PRETO	0.109	0.83

BRANCO, JADE	PRETO	0.107	0.83
PICADILLY, JADE	PRETO	0.106	0.79
PICADILLY, PINK	PRETO	0.102	0.77
N_FERIADO, BRANCO	PRETO	0.166	0.77
BRANCO	PRETO	0.285	0.76
Temperatura[22.143 – 25.429], BRANCO	PRETO	0.107	0.75
N_FERIADO, PICADILLY	PRETO	0.134	0.74
AZULCHIC	PRETO	0.138	0.73
PICADILLY	PRETO	0.225	0.73
MARINHOBLACK	PRETO	0.133	0.69
JADE	PRETO	0.163	0.69
MARFIM	PRETO	0.137	0.69
PINK	PRETO	0.163	0.69
INKBLUE	PRETO	0.110	0.67
PRETO, PINK	BRANCO	0.109	0.67
PRETO, JADE	BRANCO	0.107	0.65
PRETO, JADE	PICADILLY	0.106	0.65
ESTADO_SP	PRETO	0.188	0.64
PRETO, PICADILLY	BRANCO	0.143	0.63
N_FERIADO, ESTADO_SP	PRETO	0.107	0.62
PRETO, PINK	PICADILLY	0.102	0.62
Temperatura[18.857 – 22.143]	PRETO	0.158	0.61

Duas regras chamaram atenção em relação a quantidade presente no conjunto de dados e o suporte obtido nas regras de associação. A Temperatura no range4 [18.86 – 22.14] está presente em 26% dos pedidos analisados e com um suporte de 0.158 ela foi relacionada com a cor PRETO em uma confiança de 0.61. Da mesma forma, a Temperatura no range5 [22.15 – 25.43] se associa com a cor BRANCO e tem como conclusão a cor PRETO. A conclusão ser PRETO era algo esperado, existir uma relação entre BRANCO e PRETO também era esperado, o range de temperatura não era esperado nessa regra.

As primeiras 16 regras têm como conclusão a cor Preto. A partir da regra 17, a conclusão muda brevemente para Branco e Picadilly, mas a cor Preto continua sendo fator dominante.

Foram realizados testes com o suporte em 0.05 e suporte em 0.01. Mesmo com essas configurações, não foi possível observar os feriados e nem as tendências em nenhuma das análises realizadas.

A segunda análise comercial buscou utilizar o algoritmo Decision Tree e o Tree Rules para criar uma árvore de decisões e depois gerar regras com base na árvore. Novamente, buscou-se validar a influência ou não dos feriados, temperatura e tendência. Nessa análise, também é considerado o peso de um pedido. A árvore foi configurada para obter um ganho mínimo de informação de 0.06 e confiança de 0.6. Os valores foram

determinados após testes indicarem que a árvore não seria montada se o ganho de informação fosse inferior a 0.06. A Tabela 4 apresenta as regras geradas.

**Tabela 4. Árvore de Decisão como Regras**

if Tendência = $\tilde{N}$ _TENDENCIA and Estado = SP and (peso_item = range4 [201.265 - 405.815] or range5 [405.815 - $\infty$ ]) then PRETO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = SP and peso_item = (range1 [- $\infty$ - 62.35] or range2 [62.35 - 147.535] or range3 [147.535 - 201.265]) then BRANCO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = (SC or RS or RN or RJ) then PRETO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = PR and peso_item = range2 [62.35 - 147.535] then BRANCO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = PR and peso_item = range1 [- $\infty$ - 62.35] then PICADILLY
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = PE and peso_item = range5 [405.815 - $\infty$ ] then PRETO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = PE and peso_item = (range2 [62.35 - 147.535] or range3 [147.535 - 201.265] or range4 [201.265 - 405.815]) then PICADILLY
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = PE and peso_item = range1 [- $\infty$ - 62.35] then BRANCO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = MG and peso_item = (range2 [62.35 - 147.535] or range3 [147.535 - 201.265] or range4 [201.265 - 405.815] or range5 [405.815 - $\infty$ ]) then PRETO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = MG and peso_item = range1 [- $\infty$ - 62.35] then PICADILLY
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = MA then BRANCO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = GO and peso_item = (range4 [201.265 - 405.815] or range5 [405.815 - $\infty$ ]) then PRETO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = GO and peso_item = (range2 [62.35 - 147.535] or range3 [147.535 - 201.265]) then PICADILLY
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = GO and peso_item = range1 [- $\infty$ - 62.35] then JADE
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = ES and peso_item = range5 [405.815 - $\infty$ ] then PRETO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = ES and peso_item = range4 [201.265 - 405.815] then PICADILLY
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = ES and peso_item = range2 [62.35 - 147.535] then BRANCO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = ES and peso_item = range1 [- $\infty$ - 62.35] then VANILLA
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = DF and peso_item = range3 [147.535 - 201.265] then PRETO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = DF and peso_item = range1 [- $\infty$ - 62.35] then BRANCO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = CE then PRETO

if Tendência = $\tilde{N}$ _TENDENCIA and Estado = BA and Temp Comp Media = range6 [25.429 - 28.714] then PRETO
if Tendência = $\tilde{N}$ _TENDENCIA and Estado = BA and Temp Comp Media = range5 [22.143 - 25.429] then MARAJOARA
if Tendência = $\acute{E}$ _TENDENCIA and feriado = $\tilde{N}$ _FERIADO and Temp Comp Media = range7 [28.714 - $\infty$ ] then ROSABARROCO
if Tendência = $\acute{E}$ _TENDENCIA and feriado = $\tilde{N}$ _FERIADO and Temp Comp Media = (range5 [22.143 - 25.429], range6 [25.429 - 28.714]) then EMPIRE
if Tendência = $\acute{E}$ _TENDENCIA and feriado = $\tilde{N}$ _FERIADO and Temp Comp Media = (range1 [ $-\infty$ - 12.286] or range2 [12.286 - 15.571] or range3 [15.571 - 18.857] or range4 [18.857 - 22.143]) then PARIS
if Tendência = $\acute{E}$ _TENDENCIA and feriado = pais and Temp Comp Media = range6 [25.429 - 28.714] then PARIS
if Tendência = $\acute{E}$ _TENDENCIA and feriado = pais and Temp Comp Media = (range2 [12.286 - 15.571] or range3 [15.571 - 18.857] or range4 [18.857 - 22.143] or range5 [22.143 - 25.429]) then CORAL
if Tendência = $\acute{E}$ _TENDENCIA and feriado = pais and Temp Comp Media = range1 [ $-\infty$ - 12.286] then GOLD
if Tendência = $\acute{E}$ _TENDENCIA and feriado = natal then PARIS
if Tendência = $\acute{E}$ _TENDENCIA and feriado = namorados and Estado = (SP or PR or GO or CE)
if Tendência = $\acute{E}$ _TENDENCIA and feriado = namorados and Estado = SC then CELESTIAL
if Tendência = $\acute{E}$ _TENDENCIA and feriado = namorados and Estado = RJ then CAVALI
if Tendência = $\acute{E}$ _TENDENCIA and feriado = namorados and Estado = PE then ATLANTIS
if Tendência = $\acute{E}$ _TENDENCIA and feriado = namorados and Estado = MG then EMPIRE
if Tendência = $\acute{E}$ _TENDENCIA and feriado = namorados and Estado = ES then GOLD
if Tendência = $\acute{E}$ _TENDENCIA and feriado = mães and Estado = SP then GOLD
if Tendência = $\acute{E}$ _TENDENCIA and feriado = mães and Estado = SC then ROSABARROCO
if Tendência = $\acute{E}$ _TENDENCIA and feriado = mães and Estado = RS then EMPIRE
if Tendência = $\acute{E}$ _TENDENCIA and feriado = mães and Estado = (RJ or PR) THEN NUUK
if Tendência = $\acute{E}$ _TENDENCIA and feriado = mães and Estado = PE then ROSABARROCO
if Tendência = $\acute{E}$ _TENDENCIA and feriado = mães and Estado = (MG or GO) THEN CORAL
if Tendência = $\acute{E}$ _TENDENCIA and feriado = mães and Estado = ES then CAQUI
if Tendência = $\acute{E}$ _TENDENCIA and feriado = mães and Estado = CE then ATLANTIS

if Tendência = É_TENDENCIA and feriado = crianças and Estado =(SP or PE or CE) THEN PARIS
if Tendência = É_TENDENCIA and feriado = crianças and Estado =(SC or PR or MG) THEN NUUK
if Tendência = É_TENDENCIA and feriado = crianças and Estado =(RJ or GO) THEN LIGHTBLUE

Nesta análise é possível perceber uma definição de cores por estado, tendência e feriado. Em alguns estados foi possível identificar a relação entre os pesos e o pedido, como o estado de SP que é responsável por 25,8%, e as primeiras regras definem o padrão de quantidade de vendas e quais são as cores dominantes dentro dessa quantidade. A cor PRETO é a responsável pelos pedidos maiores que 200kg, enquanto a cor BRANCA é responsável pela maioria dos pedidos entre 62kg e 200kg. A cor PRETO continua em um fator dominante, a terceira regra define ela como a cor dominante em todos os ranges de quantidade nos estados SC, RS, RN e RJ. Esses estados representam 20,3% das vendas. Pernambuco é outro estado grande em vendas, sendo responsável por 17,1% dos itens vendidos. Nesse estado já existe uma divisão mais clara entre Branco, Preto e Picadilly. Goiás é o maior estado, em relação a vendas, no período analisado, detendo 28,6% das vendas e suas regras surpreendem por não conter a cor Branco e sim Jade, Picadilly e Preto.

## 5.2. Análise Industrial

A análise industrial foi feita utilizando a tarefa de classificação com o algoritmo de Rule Induction. Buscou-se mapear os padrões que geram falhas e isso foi transformado em regras e apresentado aos especialistas da empresa.

**Tabela 5. Rules Induction**

if MAQUINA_TINGIMENTO = M08 and COR = (PINK, CORAL, FLAN, AZUL CHIC, MARAJOARA, BATOR, SAARA, SUEDE) then MANCHAS DE TINGIMENTO
if RAMA = RAMA 04 and MAQUINA_TINGIMENTO = M12 then LARGURA
if RAMA = RAMA 01 and MAQUINA_TINGIMENTO = M13 then MANCHAS DE TINGIMENTO
if Tempo Real - Tempo Teorico (min) = range5 [39.500 - 55.500] and MAQUINA_TINGIMENTO = M14 then MANCHAS DE TINGIMENTO
if MAQUINA_TINGIMENTO = M15 and COR = (FRASER, FLAN) then MANCHAS DE TINGIMENTO
if MAQUINA_TINGIMENTO = M16 and COR = SUEDE then MANCHAS DE TINGIMENTO
if Tempo Real - Tempo Teorico (min) = range12 [146.500 - ∞] and MAQUINA_TINGIMENTO = M18 then MANCHAS DE TINGIMENTO
if MAQUINA_TINGIMENTO = M25 and Tempo Rama - Abridor (min) = range5 [727.500 - 911.500] then LARGURA
if Tempo Rama - Abridor (min) = range3 [370.500 - 526.500] and MAQUINA_TINGIMENTO = M25 then MANCHAS DE TINGIMENTO
if MAQUINA_TINGIMENTO = M25 and COR = BRANCO then LARGURA
if MAQUINA_TINGIMENTO = M26 and Tempo Rama - Abridor (min) = range10 [1601.500 - 2002.500] then MANCHAS DE TINGIMENTO

```
if COR = PINK and MAQUINA TINGIMENTO = M29 then BARRAMENTO
```

```
if COR = PRETO and RAMA = RAMA 01 then LARGURA
```

As regras estabeleceram uma relação de cor, as máquinas e os problemas que são observados com a combinação dessas variáveis. A máquina 08 apresentou problemas com várias cores diferentes e todos concluíram um problema de mancha de tingimento. Foi possível observar diferentes problemas envolvendo diferentes máquinas e concluindo em problemas não esperados, sendo um exemplo a regra que diz que se a Cor for Preto, a Rama for a 01, então o problema é de largura. A princípio, isso foi considerado como uma coincidência pelo gerente de produção da empresa, mas após uma análise do gerente de manutenção, foi identificado que o sensor de cor da Rama 01 é diferente das outras ramas, e ele tem dificuldades de detectar a presença dessa cor em específico.

Além das regras geradas, foi observado que o atributo “Tempo Real – Tempo Teórico” influencia diretamente na quantidade de falhas observada no conjunto de dados. A Figura 1 apresenta a relação de defeitos e o tempo e os números apresentados são em relação a todas as peças produzidas em um determinado range. Percebe-se que existe um aumento de falhas conforme o tempo real se distancia do tempo teórico. O ponto de menor falha é o range2[-16.50 - -1.50] e representa 5,96%, enquanto o range12[134.5 até infinito] apresenta 10,77% de falhas.

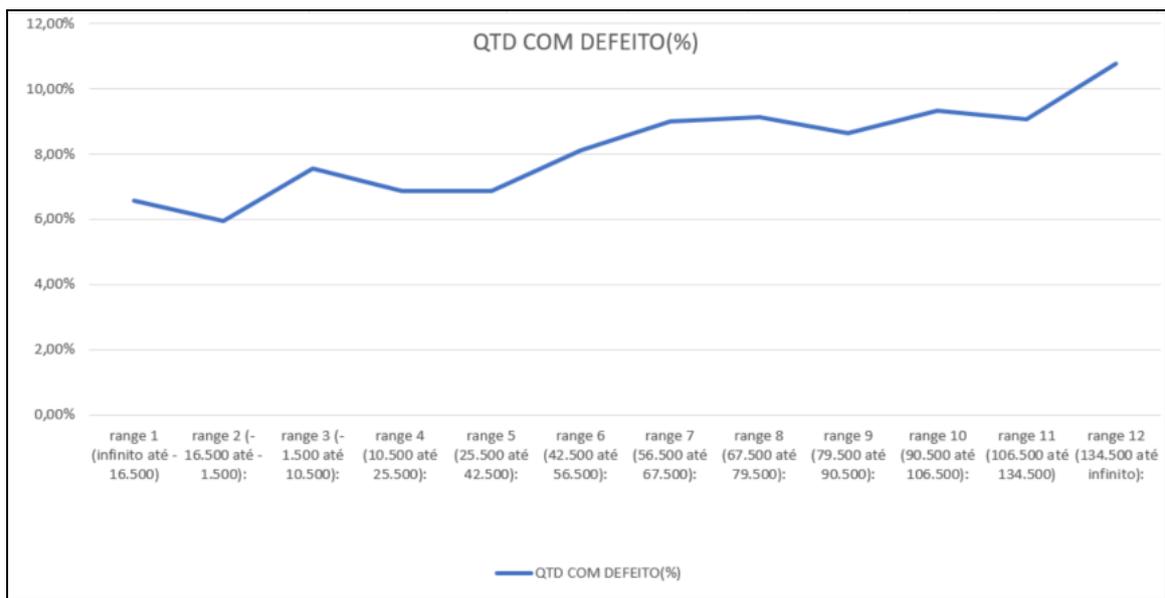


Figura 1. Defeitos (%) em relação ao Tempo Real – Tempo Teórico

## 6. Conclusões

Os conjuntos de dados gerados para o setor Comercial e Industrial foram gerados utilizando a metodologia CRISP-DM. Esses dados foram analisados utilizando as tarefas de Associação e Classificação. Os resultados obtidos foram analisados com auxílio de especialistas da área Comercial e Industrial.

No setor comercial, as regras geradas não foram suficientes para definir se o feriado, a tendência e a temperatura são relevantes para as vendas, mas foi possível identificar para quais cores elas influenciam. De forma geral, as regras podem ser

utilizadas para melhorarias na organização do estoque, ao colocar as cores que vendem juntas próximas e diminuir o tempo que demora para separar um pedido.

No setor industrial, as regras geradas foram suficientes para definir que o fator de tempo é um atributo que influencia no aumento de defeitos de produção. Além disso foi possível verificar que as regras relacionam os defeitos as máquinas e as cores. Essa informação permite que a programação de produção seja feita para evitar que os defeitos aconteçam.

## Referências

- Altintas, N., Trick, M. A data mining approach to forecast behavior. *Annals of Operations Research*, v. 216, n. 1, p. 3-22, March 2014.
- Augusto Junior, R. G., Descoberta de conhecimento para identificação de fatores que influenciam o desempenho discente: diferença entre cursos de direito e engenharia civil. 158 f. Dissertação (Mestrado em Computação Aplicada) – Universidade do Vale do Itajaí (UNIVALI), Itajaí, 2015.
- Çiflikli, C., Kahya-Özyirmidokuz, E. Implementing a data mining solution for enhancing carpet manufacturing productivity. In: *Knowledge-Based Systems*, pages 501-251, 2010
- Huang, T. C. -K., Liu, C. -C. and Chang, D. C. An empirical investigation of factors influencing the adoption of data mining tools. In: *International Journal of Information Management*, pages 257-270, 2012.
- Kamsu-Foguem, B., Rigal, F., Mauget, F. Mining association rules for the quality improvement of the production process. **Expert Systems with Applications**, v. 40, n. 4, p. 1034–1045, 2013.
- Lee, C. K. H., Choy, K. L., Ho, G. T. S., Chin, K. S., Law, K. M. Y., & Tse, Y. K., “A hybrid OLAP-association rule mining based quality management system for extracting defect patterns in the garment industry,” *Expert Systems with Applications*, vol.40, pp.2435-2446, 2013.
- Mariscal, G., Marbán, O., Fernández, C. A survey of data mining and knowledge discovery process models and methodologies. In: *The Knowledge Engineering Review*, pages 137-166, 2010.
- Provost, F. Fawcett, T. *Data Science for Business: what you need to know about data mining and data-analytic thinking*. 1. ed. Sebastopol: O’Reilly Media, 2013.
- Stockton, D. J., Khalil, R. A., Mukhongo, M. L, Cost Model development using virtual manufacturing and data mining: part II – comparison of data mining algorithms. In: *The International Journal of Advanced Manufacturing Technologi*, pages 1389-1396, 2012.
- Wang, K., *Applying data mining to manufacturing: The nature and implications*. In: *Journal of Intelligent Manufacturing*, pages 487-495, 2007.
- Zhao, Y. R and *Data Mining*. 1 ed. Academic Press, 2012.