



Sistemática de seleção de variáveis para classificação de produtos em categorias de modelos de reposição

Proposition of a variable selection framework for product replenishment

César Augusto Mazzillo Júnior¹
Michel José Anzanello¹

Resumo: Empresas integradas a cadeias de suprimento constantemente buscam iniciativas para melhorar o desempenho global da cadeia, principalmente no que concerne ao fornecimento de produtos e gerenciamento dos canais diretos e indiretos de distribuição. O VMI (*Vendor Managed Inventory*) auxilia as empresas a melhor gerenciar e balancear o estoque ao longo da cadeia. Para o bom funcionamento do VMI, é necessário que parâmetros e modelos de cálculo alinhem-se às características dos produtos e locais de reposição, normalmente descritos por elevado número de variáveis. Este artigo propõe uma sistemática de seleção de variáveis para classificação de produtos em modelos de reposição. Para tanto, utiliza a Análise dos Componentes Principais (ACP) em conjunto com as ferramentas de classificação *k-Nearest Neighbor* (KNN) e Análise Discriminante Linear (ADL). Ao ser aplicado em um estudo prático do setor de consultoria em *Supply Chain*, o método proposto alcançou uma acurácia de classificação de 90% ao reter 55% das variáveis originais.

Palavras-chave: Gestão da Cadeia de Suprimentos (SCM). Estoque Gerenciado pelo Fornecedor (VMI). Análise dos Componentes Principais (PCA). Algoritmo do Vizinho mais próximo (KNN). Análise Discriminante Linear. Análise Multivariada de Dados.

Abstract: *Companies integrated in supply chains seek initiatives to improve the overall performance of their chains. Vendor Managed Inventory (VMI) enables better results when it comes to managing and balancing stocks along the chain. For that matter, VMI frameworks must rely on well-defined parameters and algorithms aimed at allocating products to replenishment local characteristics. This paper presents a method to classify products in replenishment categories based on Principal Component Analysis (PCA) along with two classification algorithms: k-Nearest Neighbor (KNN) and Linear Discriminant Analysis (LDA). The model seeks to identify the most relevant variables for assigning products to the most appropriate replenishment model. When applied to a real situation, the proposed method yielded 90% classification accuracy, retaining 55% of the original variables on average.*

Keywords: *Supply Chain Management (SCM). Vendor Managed Inventory (VMI). Principal Component Analysis (PCA). k-Nearest Neighbor (KNN). Linear Discriminant Analysis (LDA). Multivariate Data Analysis (MVA).*

1 Introdução

Com o crescimento da competição mundial, a eficiência dos processos internos das organizações não é mais fator diferencial de mercado. Empresas integradas em uma cadeia de suprimentos devem atentar para todos os seus elos e gerenciar os processos entre eles com eficácia e flexibilidade, assegurando elevados padrões de qualidade e custos reduzidos (SIMCHI-LEVI; KAMINSKY; SIMCHI-LEVI, 2000). Dentro deste contexto, a utilização de políticas apropriadas para gestão de modelos de reposição de estoque, aliadas a sistemas de transferência de informação ao longo da cadeia, tem oferecido potencial de diferenciação a diversas empresas do segmento industrial (HOBERG; THONEMANN, 2014).

A Análise Multivariada de Dados (*Multivariate Analysis - MVA*) é uma ferramenta estatística que consiste na análise simultânea de múltiplas variáveis, as quais possuem a capacidade de explicar diversos comportamentos do sistema. A MVA pode auxiliar as empresas a melhor entenderem seus processos, permitindo respostas mais rápidas diante de mudanças de mercado, produtos e serviços mais alinhados com as necessidades dos clientes, e redução de custos e tempo de operação. Além disso, a MVA permite transformar dados disponíveis em conhecimento para a tomada de decisão (HAIR JR et al., 2010; RENCHER; CHRISTENSEN, 2012).

¹ Departamento de Engenharia de Produção, Universidade Federal do Rio Grande do Sul – UFRGS, CEP 90450-001, Porto Alegre, RS, Brasil, e-mail: mazzillo.junior@ufrgs.br; anzanello@producao.ufrgs.br

Recebido em Out. 28, 2013 — Aceito em Ago. 20, 2014

Suporte financeiro: Nenhum.

A MVA possui aplicações em diversas áreas do mercado, como Engenharia, Medicina e Educação (RENCHER; CHRISTENSEN, 2012). Dentre as aplicações práticas, destaca-se o uso das técnicas de MVA para a operacionalização do *Data Mining*, ferramenta que permite extrair informações e conhecimento de grandes massas de dados para uso em análises de mercado, comportamento do consumidor e avaliação de produtos (HAN; KAMBER, 2006). As técnicas de MVA podem ser divididas entre as de relacionamento dependente e as de relacionamento interdependente, destacando-se dentre as independentes a Análise dos Componentes Principais (ACP) e a Análise Fatorial - tipicamente indicadas para estudos que envolvam elevado número de variáveis (HAIR JR et al., 2010). Neste contexto, Sahmer e Qannari (2008) indicam a importância de se trabalhar com um correto número de variáveis relevantes, reduzindo tempo de análise e esforço de analistas.

Por sua vez, o VMI (*Vendor Managed Inventory* – Estoque Gerenciado pelo Fornecedor) é uma iniciativa na qual o fornecedor passa a ser responsável pela reposição do estoque de produtos dos seus compradores, baseado em informações de vendas e nível de estoques dos elos a jusante da cadeia de suprimentos (ZANONI; JABER; ZAVANELLA, 2012). Empresas produtoras de bens de consumo, porém, possuem dificuldade na escolha do melhor modelo de reposição e dos melhores parâmetros para compô-lo, visto que tipicamente operam com uma gama de mais de 1500 SKUs (*Stock keeping unit* – Unidade de Manutenção de Estoque), demandas caracterizadas por comportamentos incertos. Dentro deste contexto, Mula et al. (2006) apontam a importância do desenvolvimento de sistemáticas estruturadas com vistas ao respaldo do processo de decisão em aplicações da gestão de produção, e apresentam uma agenda de potenciais temas a serem endereçados na área. Com enfoque semelhante, Ivanov e Solokov (2013) enfatizam a importância de estudos quantitativos para melhor compreensão e endereçamento de inconsistências decorrentes de incerteza na decisão de modelos de reposição. Tendo em vista que o VMI busca proporcionar incremento de vendas, redução de falta de estoques, melhoria na disponibilização dos produtos e redução global dos estoques da cadeia, a definição correta do modelo de reposição e dos seus parâmetros em função das diversas incertezas que caracterizam tais aplicações possui caráter fundamental para a correta implantação da metodologia.

Este artigo apresenta uma abordagem para identificar as variáveis (critérios quantitativos de venda e mercado) mais relevantes para classificação dos produtos em categorias de reposição (classes). Para tanto, A ACP é aplicada sobre os dados originais, e um

índice de importância das variáveis é derivado com base nos parâmetros gerados pela ACP. Na sequência, inicia-se um processo iterativo de classificação de observações (modelos de produtos) e remoção de variáveis. As ferramentas de classificação testadas são o *k-Nearest Neighbors* (KNN) e a Análise Discriminante Linear (ADL). Objetiva-se, com base em um conjunto reduzido de variáveis descritivas, aprimorar a acurácia de alocação dos produtos aos modelos de reposição, possibilitando uma melhor adequação à metodologia do VMI e, conseqüentemente, um melhor desempenho nos indicadores associados.

Este artigo está estruturado como segue, além desta introdução. Na segunda seção, uma revisão teórica apresenta os fundamentos da Análise de Componentes Principais, Análise Discriminante, KNN e Análise Multivariada, contextualizando sua aplicação nos conceitos de VMI. Na terceira seção, é apresentada a metodologia proposta para a definição das variáveis que melhor explicam a decisão pelos modelos de reposição e sua parametrização. A seção 4 apresenta um estudo de caso, ao passo que a quinta seção traz as conclusões.

2 Referencial teórico

2.1 Ferramentas multivariadas

Ferramentas multivariadas são métodos estatísticos e matemáticos que objetivam analisar a inter-relação entre um grande conjunto de variáveis ou sistemas complexos. Os diferentes métodos existem para serem aplicados em tipos e conjuntos específicos de dados, visando elucidar diferentes análises (MALINOWSKI, 2002; BARTHOLOMEW, 2010). Fundamentos de ACP, *k-Nearest Neighbor* (KNN) e Análise Discriminante Linear (ADL) são agora apresentados.

2.1.1 Análise dos Componentes Principais (ACP)

A ACP é uma técnica de ordenação multivariada que objetiva encontrar padrões de comportamento em dados interdependentes e não agrupados, reduzindo o número de dimensões e, então, exibindo a posição dos dados em variáveis latentes (componentes principais - CPs), os quais são não correlacionados. Dessa forma, a ACP busca reter a maior quantidade de informação e variabilidade existente nos dados, ordenando os CPs de forma que os primeiros contenham a maior parte da informação e variação presente nas variáveis originais (SYMS, 2008).

Matematicamente, a ACP consiste em uma transformação linear ortogonal, cuja aplicação transforma os dados em um novo sistema de coordenadas (JOLLIFFE, 2002; MOROPOULOU; POLIKRETI, 2009). O primeiro CP formado é a

combinação linear que representa a maior variação do conjunto de dados, enquanto o segundo, ortogonal ao primeiro, possui a segunda maior variação, e assim por diante. Os valores dessas novas variáveis são chamados de valores fatoriais e podem ser interpretados geometricamente (ABDI; WILLIAMS, 2010; RENCHER; CHRISTENSEN, 2012). O número de componentes principais formado é menor ou igual ao número de variáveis do conjunto de dados, sendo que os primeiros CPs explicam a maior parte da variabilidade do conjunto (MOROPOULOU; POLIKRETI, 2009).

Dado um conjunto de vetores de dados y_1, y_2, \dots, y_n com dimensões iguais a n , pode-se formar uma matriz $M_{p \times n}$ que represente o conjunto inicial dos dados para estudo e calcular-se o vetor médio \bar{y} por meio da Equação 1 (HE; WANG; COOK, 2011). Objetiva-se encontrar os eixos naturais do conjunto de dados, os quais possuem origem em \bar{y} . Isso é realizado ao se transladar a origem para \bar{y} e então rotacionar os eixos (Figura 1). Depois da rotação, as novas variáveis geradas (componentes principais) serão independentes (RENCHER; CHRISTENSEN, 2012).

$$\bar{y} = \frac{1}{n} * \sum_{i=1}^n y'_i \quad (1)$$

Conforme He, Wang e Cook (2011) e Rencher e Christensen (2012), para rotacionar os eixos, cada vetor observação y_i deve ser multiplicado por A . Como A é ortogonal, pode-se afirmar que a distância da origem não foi alterada (conforme a Equação 2), transformando y_i em um novo ponto t_i , o qual dista o mesmo valor da origem com os eixos rotacionados. Ao encontrar os novos eixos, e realizar operações algébricas de transposição e ortogonalização, encontra-se o vetor A (Equação 3)

para que os componentes principais t_1, t_2, \dots, t_n em $t = Ay$ sejam não relacionados. O vetor A traz os pesos (influência na variabilidade) de cada variável do conjunto de dados. Os CPs (variáveis latentes) consistem nas variáveis $t_1 = a_1' y, \dots, t_p = a_p' y$ em $z = Ay$. Análises baseadas nos coeficientes de importância do ACP são estruturadas nas variáveis latentes formadas, as quais seguem o exemplo da Equação 4, na qual a_{np} representa o peso da variável y_p na variabilidade do conjunto de dados.

$$t'_i t_i = (Ay_i)'(Ay_i) = y'_i A' A y_i = t'_i t_i \quad (2)$$

$$A = \begin{pmatrix} a'_1 \\ a'_2 \\ \vdots \\ a'_p \end{pmatrix} \quad (3)$$

$$t_1 = a_{11}y_1 + a_{12}y_2 + \dots + a_{1p}y_p \quad (4)$$

2.1.2 Algoritmo *k-Nearest Neighbor* (KNN)

O *k-Nearest Neighbor* (KNN) é um método supervisionado de classificação de dados baseado na proximidade de seus vizinhos em um espaço amostral (DAKHLAOU; BARGAOU; BÁRDOSY, 2012). Seu objetivo é formar uma generalização com base em um conjunto de treinamento, maximizando a acurácia da classificação de novos dados (GAO; GAO, 2010). O algoritmo pressupõe que o conjunto de treinamento é composto pelas variáveis descritivas e pela sua classificação; o KNN então utiliza tais variáveis para classificar um novo item (SU, 2011).

Com base em um conjunto de treinamento, formado por n observações previamente classificadas, como

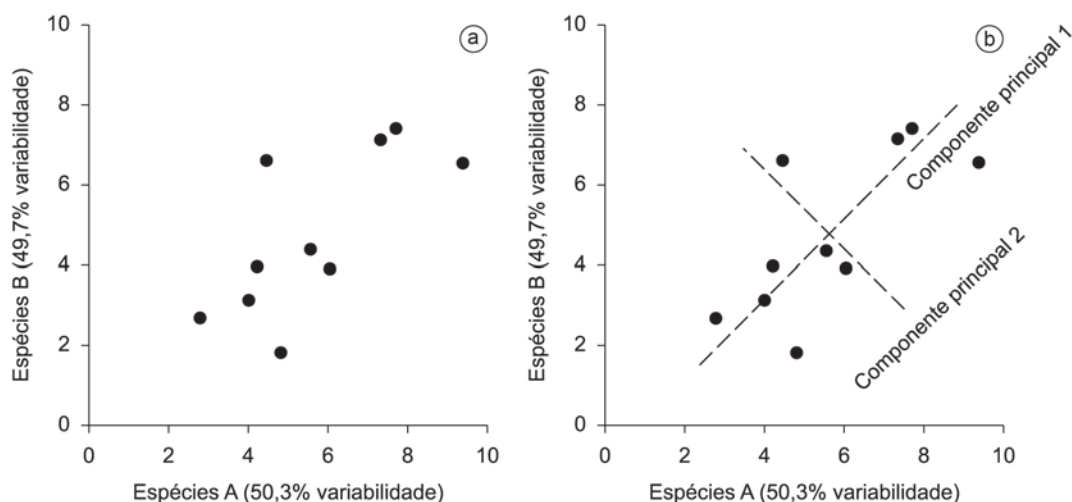


Figura 1. (a) Conjunto original de dados exemplo. (b) Novos eixos criados com centro em \bar{y} , obtendo os dois componentes principais não correlacionados. Fonte: SYMS (2008).

$X = \{x_1, x_2, \dots, x_n\}$, e uma observação a ser classificada, formada por $Y = \{y_1, y_2, \dots, y_n\}$ em um espaço n -dimensional, tem-se que a distância Euclidiana, na qual o algoritmo baseia-se, pode ser calculada para cada ponto do conjunto de treinamento pela Equação 5. Com base nesta distância, os k vizinhos mais próximos são identificados e, com base em um k escolhido, uma nova observação é atribuída à classe com maior número de observações em k , conforme ilustrado na Figura 2 (SU, 2011; DAKHLAOU; BARGAOU; BÁRDOSSY, 2012; DUDA; HART; STORK, 2012).

$$\text{dist}(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (5)$$

2.1.3 Análise Discriminante Linear (ADL)

A análise discriminante linear (*Linear Discriminant Analysis*) é um dos métodos estatísticos supervisionados mais conhecidos para classificação de dados (XU; BROCK; PARRISH, 2008). Seu objetivo é, com base em um conjunto de treinamento de dados, buscar uma combinação linear de variáveis que melhor explique aqueles dados (FISHER, 1936; MARTINEZ; KAK, 2001). Proposto por Fisher (1936), o método baseia-se em um conjunto de aprendizado com n observações e K classes $(x_i, y_i), i = 1, \dots, n$. $x_i \in \mathbb{R}^p$ e $y_i \in \{1, \dots, K\}$, o qual formará duas matrizes, a de dispersão intraclasses S_b (que computa a variância dos padrões em relação à classe que pertence) e a de dispersão entre as classes S_w (computa a variância entre as classes), as quais são definidas pelas Equações 6 e 7, respectivamente:

$$S_b = \sum_{k=1}^K n_k * (\bar{x}_k - \bar{x}_{..}) * (\bar{x}_k - \bar{x}_{..})^T \quad (6)$$

$$S_w = \sum_{k=1}^K \sum_{i=1}^{n_k} n_k * (\bar{x}_{ki} - \bar{x}_k) * (\bar{x}_{ki} - \bar{x}_k)^T \quad (7)$$

nas quais x_{ki} é a observação i em um espaço p -dimensional da classe k , n_k é o número de amostras

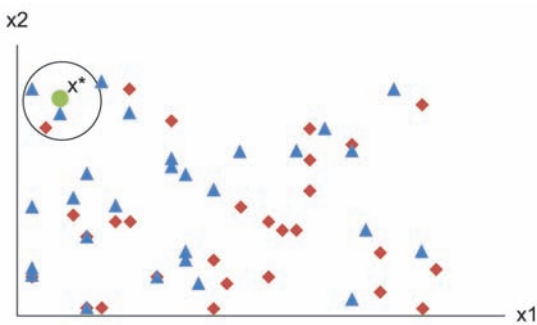


Figura 2. Uma observação x é testada em um conjunto treinamento com $k = 3$, formado pelas observações com menor distância euclidiana do ponto x^* . No caso apresentado, a observação x é alocada à categoria dos pontos representados pelos triângulos azuis. Fonte: Adaptado de Duda, Hart e Stork (2012).

do conjunto de aprendizado da classe k e \bar{x}_k , $\bar{x}_{..}$ representam o vetor médio da classe k e a média geral, respectivamente (PARK; LEE, 2008; KITANI; THOMAZ, 2007; XU; BROCK; PARRISH, 2008). Segundo Härdle e Simar (2006) e Kitani e Thomaz (2007), o objetivo é encontrar a razão entre o determinante da matriz S_b e a matriz S_w , conhecido como critério de Fischer (Equação 8).

$$\text{Critério de Fischer} = \max \frac{|S_b|}{|S_w|} \quad (8)$$

O vetor de projeção responsável pelo máximo critério de Fischer denota a maior separação entre as classes.

2.2 Métodos para seleção de variáveis para classificação de dados

Sistemáticas para seleção de variáveis em aplicações industriais e corporativas têm recebido crescente atenção do meio acadêmico por possibilitarem classificações e predições mais acuradas. Wold, Sjöström e Eriksson (2001) foram pioneiros na proposição de técnicas de PLSR (*Partial least squares projection regression* – Regressão dos mínimos quadrados parciais) aliadas a outras técnicas multivariadas (ACP, KNN e LDA) com vistas à criação de índices de importância de variáveis para predição (VIP – *Variable importance for the projection*). Choi et al. (2012) ressaltam que tais métodos buscam descartar variáveis irrelevantes para aperfeiçoar a classificação de dados e poupar esforço computacional no processamento de dados.

Anzanello, Albin e Chaovalitwongse (2009) propõem a criação de uma série de diferentes índices de importância de variáveis (IIV), um deles baseado nos pesos das variáveis das combinações lineares independentes originadas pela regressão PLS, e posterior classificação pelo algoritmo de KNN para remover as variáveis menos relevantes e aumentar a acurácia do conjunto de treinamento (Figura 3). Choi et al. (2012) também apresentam um método baseado em pesos das variáveis oriundos da ACP e posterior aplicação de ADL para classificação dos dados. Por sua vez, Gertheiss e Tutz (2009) propõem a aplicação de um conjunto de métodos baseados no algoritmo de vizinhos mais próximos para estimar a probabilidade de um determinado dado pertencer a uma classe e depois classificá-lo de acordo com tal análise. Os métodos avaliados consideram o algoritmo clássico de KNN e um algoritmo cujas variáveis possuem pesos de acordo com sua influência.

Ao abranger outros métodos de MVA, Ballabio et al. (2010) propõem a construção da Medida Canônica de Correlação (Índice CMC), calculado com base em dois conjuntos de dados, o das variáveis independentes e o da matriz das classes existentes. Os índices são

calculados por variável e um *ranking* é gerado com base na contribuição das variáveis em discriminar as classes. Raghuraj e Lakshminarayanan (2009) apresentam um método para classificação baseado na matriz dos coeficientes de correlação parciais e no coeficiente de Pearson, medida que define a associação entre variáveis contínuas, comparando-o com uma aplicação de ADL na mesma base de dados. Por fim, Shreve, Schneider e Soysal (2011) propõem uma sistemática baseada em simulação de Monte Carlo para possibilitar a comparação de métodos de classificação com base na estabilidade dos modelos e validade da seleção de variáveis; os autores concluem que um conjunto de dados amostral suficientemente grande permite a comparação dos métodos.

Ao final desta seção de revisão, é importante mencionar que, com base na literatura consultada, os autores não identificaram nenhuma abordagem estruturada para seleção de variáveis no contexto do VMI. De tal forma, tal endereçamento constitui-se na principal contribuição deste artigo.

3 Procedimentos metodológicos

Esta seção descreve a classificação da pesquisa quanto ao seu método, definindo a natureza do estudo, sua abordagem, objetivos e procedimentos. Em seguida, descreve-se o método aplicado no trabalho, detalhando os passos para o desenvolvimento da sistemática de seleção de variáveis para inserção de produtos em categorias de reposição de estoques.

O presente estudo é definido como de natureza aplicada, sendo a sistemática desenvolvida com base na solução de VMI da empresa. Quanto à sua

abordagem, a pesquisa é de caráter quantitativo, pois busca o desenvolvimento de uma sistemática com base em dados históricos de vendas e de demanda para classificar os produtos em métodos de reposição automática de estoques. Os objetivos do estudo são definidos como exploratórios e explicativos, visto que parte de uma massa de informações, formulando hipóteses para alcançar os seus resultados. Por fim, quanto ao procedimento, o estudo é caracterizado como experimental, visto que se baseia na manipulação de fatores e variáveis para observar um resultado (YIN, 2003).

As etapas propostas, ilustradas na Figura 3, apoiam-se nas técnicas ACP, ADL e KNN, e são detalhadas na sequência.

3.1 Coleta de dados

Para a montagem da base de dados, foi estruturada uma matriz nos moldes da Tabela 1, contendo as variáveis pertinentes ao processo de distribuição disponíveis no sistema VMI e que poderiam influenciar a escolha do modelo de reposição para uma relação produto/local de reposição (no caso estudado, centros de distribuição e lojas). Cada observação $(y_i, x_i), i = 1, 2, \dots, n$ representa um produto/local de reposição, descrito por j variáveis e um único modelo de reposição; as variáveis x descrevem as características de cada produto, enquanto a variável y denota o modelo de reposição mais adequado (indicado por meio de uma classe).

Nas proposições deste artigo, foram levantadas variáveis quantitativas contínuas e discretas (que incluíam, dentre outras, dados de venda, dados

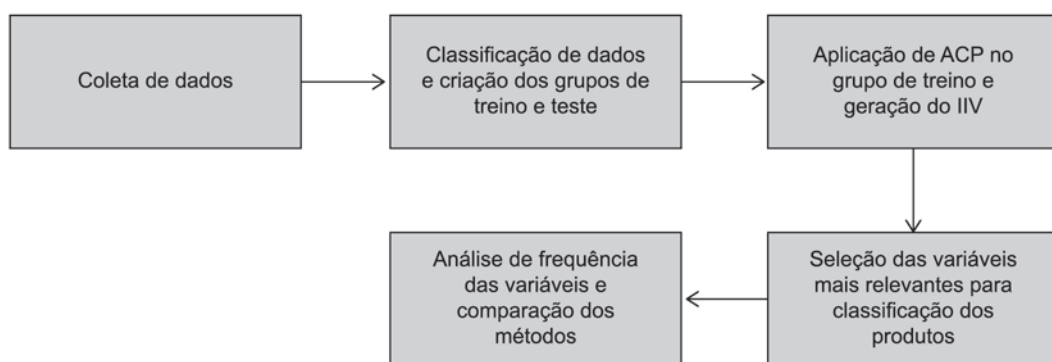


Figura 3. Fluxograma das etapas para o desenvolvimento do método. Fonte: Elaborado pelos autores.

Tabela 1. Estrutura da Matriz de Dados.

Observações	Variável 1	Variável 2	...	Variável J	Modelo de Reposição
Produto 1/Loja 1	-	-	-	-	-
...	-	-	-	-	-
Produto n/Loja 2	-	-	-	-	-

Fonte: Elaborado pelos autores.

logísticos, dados de previsão de demanda e demais dados de mercado presentes na ferramenta), e qualitativas (as quais incluíam nível de serviço desejado, estoque máximo parametrizado, algoritmo do estoque de segurança, entre outros dados que impactam a gestão de estoque e distribuição). Em relação aos produtos, analisa-se parte da gama de produtos de uma empresa de bens de consumo (produtos de higiene pessoal e doméstica, limpeza e alimentos), além dos produtos voltados a consumidores finais de uma empresa fabricante de pneus.

3.2 Classificação de dados e criação dos grupos de treino e teste

Com o auxílio da equipe de consultores especializada em soluções de distribuição, o comportamento e características específicas dos produtos foram vinculados ao modelo de reposição mais adequado. Para tal estudo, foram considerados produtos e locais de reposição já consolidados na empresa, objetivando uma base mais sólida para o aprendizado do método.

Na sequência, os dados foram divididos em um grupo de treino (T_{tr}), contendo n_{Tr} observações, e um grupo de teste (T_{s}), contendo n_{Ts} observações, de forma que $n_{Tr} + n_{Ts} = n$. Conforme Chong, Albin e Jun (2007), é adequado manter 60% das observações no grupo de treino.

3.3 Aplicação de ACP no grupo de treino e geração do Índice de Importância de Variáveis (IIV)

A fim de caracterizar a relação entre as variáveis e os métodos de reposição de cada produto/local, foi aplicada ACP no grupo de treino, gerando pesos “ w ” para cada variável. O peso está relacionado com a importância da variável na explicação da variabilidade presente nos dados. Nas proposições deste trabalho, assume-se, conforme Anzanello, Albin e Chaovalitwongse (2009), que variáveis com maior peso absoluto denotam variáveis com maior importância na explicação de variância dos dados.

Com base nos pesos gerados pela ACP, foram estimados os Índices de Importância das Variáveis (IIV) de acordo com a Equação 9, adaptada de Rossini, Anzanello e Fogliatto (2012). O módulo dos pesos w de cada variável foi somado para os Δ 's componentes retidos, a fim de verificar quais variáveis mantinham a maior variância, visto que essas são mais adequadas para aplicações de classificação de dados (ANZANELLO; ALBIN; CHAOVALITWONGSE, 2009). Tais índices não foram normalizados, visto que têm como função indicar uma ordem de eliminação das variáveis, logo se entende que a magnitude relativa dos coeficientes não interfere no propósito com que foram gerados.

$$IIV_j = \sum_{i=1}^n |w_{ij}|, i=1, \dots, n \quad (9)$$

3.4 Seleção das variáveis mais relevantes para classificação dos produtos

Nessa etapa, objetivou-se definir as variáveis mais relevantes para classificação dos produtos em categorias de reposição. Para tanto, as variáveis foram ordenadas em ordem decrescente de IIV, as observações descritas pela totalidade de variáveis classificadas utilizando KNN e ADL, e a acurácia de classificação (razão entre o número de observações classificadas corretamente e o número total de observações) do grupo de treino calculada. Depois do cálculo da acurácia, a variável menos relevante (com menor IIV) foi retirada da base de dados e o processo de classificação reiniciado, até restar apenas uma variável. Tal sistemática gera um gráfico associando o número de variáveis retidas à acurácia para cada método de classificação.

Com base no gráfico gerado (ilustração genérica na Figura 4), determinou-se o conjunto de variáveis responsável pela maior acurácia de classificação no grupo de treino. Tal conjunto foi escolhido com base na máxima acurácia.

Objetivando não ancorar o algoritmo de classificação e os seus resultados em um único grupo de treino, visto que existem produtos e locais com características distintas, os dados foram aleatoriamente embaralhados após a coleta de indicadores e os passos anteriormente descritos executados novamente. O processo foi repetido 100 vezes e, ao final de cada iteração, a acurácia e número de variáveis retidas foram armazenados.

3.5 Análise de frequência das variáveis e comparação dos métodos

Apoiando-se no conjunto de variáveis retidas para cada método de classificação em cada iteração, foram criados histogramas para representar a frequência com que cada variável foi retirada. Dessa forma, foi

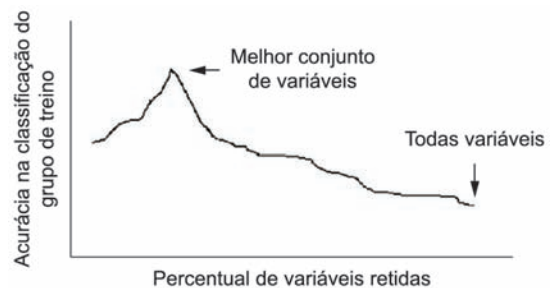


Figura 4. Perfil hipotético da acurácia do conjunto de treinamento após a eliminação de variáveis.

possível identificar as variáveis que oferecem maior contribuição na alocação de produtos aos modelos de reposição.

Para a comparação dos métodos, foram utilizados três critérios de avaliação, conforme a Tabela 2. O indicador de acurácia foi considerado como “maior é melhor”, pois quanto mais elevada a acurácia da classificação, melhor o desempenho que o modelo trará para os produtos/locais de reposição. O critério de número de variáveis com frequência igual a 100% e o número médio de variáveis retidas foram considerados como “menor é melhor”, visto que quanto menos variáveis, menor o esforço de coleta e menor o trabalho computacional de manipulação de dados. Além disso, são contemplados dois indicadores de desvio, do tipo “menor é melhor”, para avaliar a variação no comportamento dos dados e métodos ao longo das 100 iterações.

4 Resultados numéricos

Esta seção aplica o método proposto aos produtos gerenciados por uma ferramenta de VMI de uma empresa de consultoria em *Supply Chain*. A empresa foco do estudo atua na área de tecnologia da informação aplicada à gestão da cadeia de suprimentos e demanda, possuindo quatro grandes áreas de atuação: planejamento avançado e sequenciamento de produção, gestão da demanda, gestão da distribuição e gestão da visibilidade, nas quais oferece *softwares* e consultoria de negócios para auxiliar as empresas a aperfeiçoarem a gestão e operacionalização de seus processos. Está presente em todo o território nacional, com sede em São Leopoldo/RS e escritório de negócios em São Paulo/SP. Ao longo de nove anos de atuação, realizou mais de 250 projetos na área de *Supply Chain*, os quais apresentam *softwares* operando em três continentes.

Na área de gestão de distribuição, possui uma de suas soluções mais avançadas, o VMI (*Vendor Managed Inventory* – Estoque Gerenciado pelo Fornecedor). Nessa solução colaborativa, empresas das áreas de cosméticos, bens de consumo, alimentação humana e pet, pneumáticos e produtos farmacêuticos realizam a gestão de abastecimento em parceria com todo o seu canal indireto e parte do seu canal direto no Brasil, reabastecendo automaticamente, em

média, 10000 diferentes SKUs em 150 cidades nos 26 estados brasileiros. Frente a essa complexidade, a solução busca a melhor adequação dos cálculos a cada tipo de produto, empresa e região. Uma das principais dificuldades das empresas nesse contexto é determinar, junto à consultoria especializada em distribuição, quais os melhores algoritmos, métodos e parâmetros para cada um de seus produtos. Isso se deve ao fato de cada produto possuir diferentes públicos alvos, comportamentos de demanda e estratégias de comercialização e distribuição em cada região do Brasil.

A etapa inicial do método proposto define as variáveis a serem coletadas para inclusão dos produtos em categorias associadas a modelos específicos de reposição de estoque. Para tal, foram levantadas variáveis quantitativas (dados de venda, dados logísticos, dados de previsão de demanda e demais dados de mercado presentes na ferramenta) e qualitativas (nível de serviço desejado, estoque máximo parametrizado, algoritmo do estoque de segurança, entre outros dados que impactam a gestão de estoque e distribuição), com base nos estudos de Seth, Deshnuh e Vrat (2006), Zhu, Dou e Sarkis (2010), Liu et al. (2008), Chae (2009) e Kim, Kumar e Kumar (2010). Tais variáveis foram apresentadas aos consultores e gestores da área de Gestão de Distribuição – responsáveis pelo sistema VMI – que retiraram da análise, por meio de uma reunião expositiva, variáveis com pouco histórico ou baixa consistência de informações. As 24 variáveis remanescentes são apresentadas na Tabela 3. Quanto aos produtos, foram escolhidos os de uma empresa de bens de consumo (produtos de higiene pessoal e doméstica, limpeza e alimentos), além dos produtos voltados a consumidores finais de uma empresa fabricante de pneus. Esse conjunto de análise gerou uma base de dados com 2000 registros (produtos a serem classificados pelo método proposto). A coleta dos dados foi realizada diretamente no banco de dados do sistema, o qual já possui as informações organizadas por produto e local de reposição, cabendo apenas uma extração simples e um filtro nos locais e produtos desejados.

Os modelos de reposição (classes) nos quais os produtos foram classificados (Tabela 4) foram

Tabela 2. Avaliação comparativa dos métodos de classificação.

Critério	Tipo	KNN	ADL
Acurácia média de classificação (%)	Maior é melhor	-	-
Desvio da Acurácia de classificação (%)	Menor é melhor		
Nº médio de variáveis retidas	Menor é melhor		
Desvio do Nº médio de variáveis retidas	Menor é melhor	-	-
Nº de variáveis com frequência = 100%	Menor é melhor	-	-

Fonte: Elaborado pelos autores.

Tabela 3. Variáveis selecionadas para análise.

Nº	Variável	Descrição da variável
1	<i>Sell-out</i> (R\$)	Representa o resultado financeiro de um determinado produto em uma loja.
2	<i>Sell-out</i> (Un.)	Representa o número de unidades transacionadas de um determinado produto em uma loja.
3	Preço Médio	Representa o preço de venda de um determinado produto em uma loja.
4	Cancelamento (Un.)	Representa o número de vendas canceladas de um determinado produto em uma loja.
5	Bonificação (Un.)	Representa o número de unidades de produto que foram “vendidas” sem custo em uma loja.
6	Devolução (Un.)	Representa o número de devoluções de um determinado produto em uma loja.
7	<i>Sell-out</i> Médio (Un.)	Representa, em média, o número de unidades comercializadas de um produto quando existe uma venda para ele.
8	Dias com venda	Representa o número de dias em que um produto foi comercializado em uma loja.
9	Dias úteis sem venda	Representa o número de dias em que um produto não foi comercializado em uma loja.
10	Frequência (%)	Representa a frequência de vendas de um determinado produto em uma loja.
11	Lote Múltiplo (Un.)	Representa o lote de ressuprimento de estoque de um determinado produto para uma loja.
12	Venda Máxima (Un.)	Representa a maior quantidade vendida de um produto em um único dia em uma loja.
13	Número Transações	Representa o número de compras, realizadas em uma loja, que continham um determinado produto.
14	Base Clientes	Representa o número de clientes finais que adquirem o produto em uma determinada loja.
15	ABC	Representa a classificação ABC de um produto de acordo com seu volume de vendas.
16	VMD	Representa o número médio de unidades vendidas dentro de um período.
17	Desvio	Representa o desvio médio de unidades vendidas dentro de um período.
18	CV	Representa o coeficiente de variação (desvio/média) de um produto em uma loja.
19	Cobertura Objetivo	Representa o número de dias em que se objetiva manter estoques para cada produto em uma loja.
20	Nível de Serviço	Representa o coeficiente k da curva normal que representa o nível de serviço que se objetiva manter para cada produto em uma loja.
21	<i>Lead Time</i>	Representa o número de dias que um determinado produto demora a chegar a uma loja.
22	Desvio LT	Representa o desvio médio em dias que um determinado produto demora a chegar a uma loja.
23	Dias de Ruptura	Representa o número de dias que um produto manteve-se sem estoques em uma loja.
24	Cobertura Média (Dias)	Representa o número médio de dias de estoque para um produto em uma loja.

Fonte: Sistema VMI da empresa em estudo.

Tabela 4. Modelos de Reposição de Estoques (classes) considerados.

Classe	Modelo
1	Reposição Contínua Clássica
2	Reposição <i>Push & Pull</i> (Método do Mínimo/Máximo)
3	Reposição Periódica com Segurança Cadastrada
4	Reposição Contínua com Segurança Cadastrada
5	Reposição Periódica baseada na previsão

Fonte: Sistema VMI da empresa em estudo.

determinados em conjunto com os consultores em consistência com as ferramentas e algoritmos já existentes no sistema de VMI em estudo. Uma breve descrição das características de cada classe é apresentada na sequência.

Classe 1 - Reposição Contínua Clássica: modelo de reposição com revisão diária, respeitando os parâmetros calculados de ponto de pedido, estoque objetivo e estoque de segurança; é calculado com base nas médias e desvios das vendas e *lead times*. Considera também um coeficiente de segurança subjetivamente ajustado de acordo com o nível de serviço desejado (tipicamente utilizando a distribuição normal).

Classe 2 - Reposição *Push/Pull* (Mínimo e Máximo): reposição de revisão diária baseada em duas curvas de estoque, mínimo e máximo (ambas cadastradas). A reposição é realizada sempre que o estoque alcançar o parâmetro mínimo, repondo a quantidade necessária para atingir o nível máximo.

Classe 3 - Reposição Periódica com segurança cadastrada: modelo de reposição realizado em

intervalos fixos pré-determinados. Utiliza o estoque de segurança cadastrado em dias de cobertura e um estoque objetivo também cadastrado em dias. A quantidade necessária é sempre a diferença da posição atual de estoque para a soma dos parâmetros de segurança e objetivo

Classe 4 - Reposição Contínua com segurança cadastrada: modelo de reposição com revisão diária, respeitando os parâmetros calculados de ponto de pedido e estoque objetivo considerando o estoque de segurança como um parâmetro cadastrado em dias. Considera também um coeficiente de segurança ajustado de acordo com o nível de serviço desejado utilizando a distribuição normal.

Classe 5 - Reposição periódica com base em forecast: modelo de reposição realizado em intervalos fixos pré-determinados, levando em consideração a previsão de vendas diária para cada produto em cada local de reposição. A quantidade ressuprida respeita a previsão de vendas agregada nos dias de *lead time*, sempre as antecipando no tempo de acordo com o *lead time* de entrega.

Na sequência, aplicou-se a ACP para definição dos pesos de cada variável em cada componente principal gerado. Foram retidos dois componentes principais com base no percentual de variância explicada pelos mesmos (83%) e análise do *Scree Graph* (RENCHER; CHRISTENSEN, 2012). O IIV calculado através da Equação 9 é apresentado na Tabela 5; com base nesses valores, as variáveis foram ordenadas em ordem decrescente de IIV, de modo a facilitar o processo iterativo de remoção de variáveis.

Na sequência, partiu-se para um processo iterativo de remoção de variáveis e classificação de produtos utilizando os métodos KNN e ADL. A cada variável retirada de acordo com a ordem estabelecida pelo IIV, uma nova classificação era realizada (sempre com base nas variáveis restantes) e a acurácia calculada. Esse procedimento gerou o gráfico da Figura 5 para uma das 100 iterações realizadas, na qual se observa que 20 variáveis devem ser retidas ao utilizar-se a ferramenta KNN. Ressalta-se, porém, que a acurácia obtida com 20 variáveis (88%) é muito similar à

obtida com 14 variáveis (87%), por vezes preferível por necessitar um menor esforço de coleta de dados e menor desempenho computacional. Destaca-se, também, a redução significativa da acurácia a partir de 11 variáveis restantes, o que denota a dificuldade de classificação dos produtos com menos de 50% das variáveis originais. Tal procedimento foi repetido 100 vezes para cada ferramenta de classificação (KNN e ADL).

A Figura 6 traz a frequência de retenção das variáveis para o KNN e ADL. Percebe-se que um número significativo de variáveis (mais de 50%) foi

Tabela 5. Índices de Importância das Variáveis (IIV).

Variável	IIV
Sell-out Médio (Un.)	3,625
Sell-out (Un.)	3,625
VMD	3,560
Desvio	3,464
Dias com venda	3,447
Frequência (%)	3,447
Dias úteis sem venda	3,446
Sell-out (R\$)	3,347
Devolução (Un.)	3,324
Venda Máxima (Un.)	3,167
Desvio LT	3,047
Base Clientes	3,018
Nível de Serviço	3,008
Bonificação (Un.)	2,951
Número Transações	2,546
Cobertura Objetivo	2,372
Preço Médio	2,177
Lote Múltiplo (Un.)	1,942
ABC	1,722
CV	1,468
Lead Time	1,431
Cancelamento (Un.)	0,714
Dias de Ruptura	0,499
Cobertura Média (Dias)	0,230

Fonte: Elaborado pelos autores.

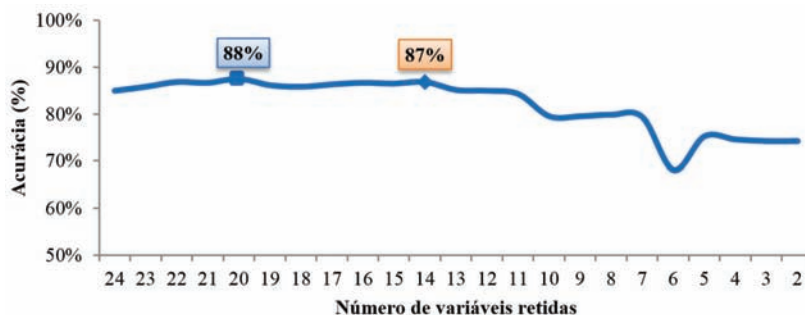


Figura 5. Acurácia x número de variáveis retidas. Fonte: Elaborado pelos autores.

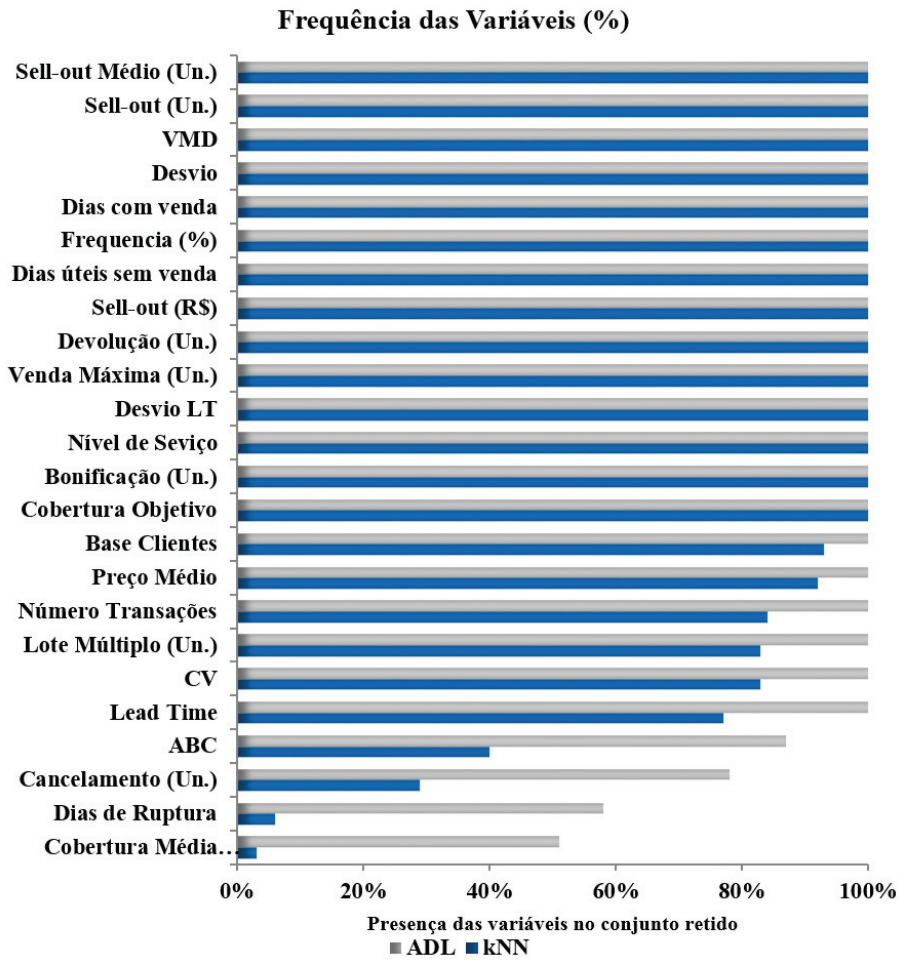


Figura 6. Histograma de presença de variáveis no conjunto retido. Fonte: Elaborado pelos autores.

retido em todas as replicações do método proposto. Dentre essas, algumas possuem relação direta com os modelos de reposição de estoque, sendo utilizadas diretamente nos cálculos e análises do sistema. Outras, porém, relacionam-se ao perfil de mercado do produto nos locais, impactando de forma indireta, mas significativa, na escolha de cada modelo de reposição. Consideraram-se as variáveis com frequência igual a 100% como fundamentais na escolha do modelo de reposição de um determinado produto, devendo ser controladas e estudadas em maior profundidade para definições de novos produtos e locais de comercialização.

Quanto aos métodos de classificação, percebe-se um comportamento mais conservador do método ADL na eliminação de variáveis, retendo aproximadamente 85% das variáveis presentes em todas as iterações – o método de KNN armazenou aproximadamente 55%. Das variáveis retidas, todas também estavam presentes no método ADL, evidenciando uma consistência

entre os dois métodos. De tal forma, evidencia-se um desempenho superior do KNN para a classificação de dados, o qual apresentou maior acurácia e menor desvio desse indicador. Além de obter um melhor grau de acurácia, o KNN reteve menor número de variáveis nas medições e também possui menor quantidade de atributos com presença constante nas classificações. Isso permite que se opere com uma base de dados reduzida e, mesmo assim, alcance-se um resultado satisfatório em termos de acurácia de classificação. A Tabela 6 resume os resultados obtidos por cada um dos métodos.

Por fim, ressalta-se que o número elevado de variáveis retidas pode estar relacionado à significativa diferença de perfil de vendas e negócio entre os produtos de higiene, limpeza e alimentos e os produtos da empresa fabricante de pneus. Com a acurácia obtida, porém, o método mostra-se robusto o suficiente para ambas as realidades de negócio.

Tabela 6. Avaliação comparativa dos métodos de classificação.

Critério	Tipo	KNN	ADL
Acurácia média de classificação (%)	Maior é melhor	89,41%	87,1%
Desvio da Acurácia de classificação (%)	Menor é melhor	0,94%	1,31%
Nº médio de variáveis retidas	Menor é melhor	19,8	22,58
Desvio do Nº médio de variáveis retidas	Menor é melhor	2,45	1,47
Nº de variáveis com frequência = 100%	Menor é melhor	14	20

Fonte: Elaborado pelos autores.

5 Conclusões

Empresas cujas estruturas estão dispostas em cadeias de suprimento não podem mais apenas buscar um desempenho satisfatório dos seus processos internos. O desempenho de cada elo presente em uma cadeia é fundamental para seu bom desempenho global. O VMI busca passar a responsabilidade do gerenciamento de estoque dos clientes para o fornecedor, otimizando os níveis de estoque dentro da cadeia e aperfeiçoando o nível de serviço. Para o correto funcionamento dessa sistemática, é preciso que os modelos de reposição dos estoques para cada produto e local estejam bem definidos e parametrizados. Dentro desse contexto, a MVA, por meio de métodos com ACP, KNN e ADL auxilia na identificação das variáveis mais relevantes na identificação de um modelo de reposição para os produtos em análise.

Este artigo apresentou uma sistemática para classificação de produtos em modelos de reposição, identificando quais variáveis eram mais relevantes para tal, utilizando como base o *software* de VMI de uma empresa de consultoria especializada em *Supply Chain*. Para o estudo, foi necessária a montagem de uma base de dados de produtos em locais específicos de reposição. Isso permitiu a aplicação da ACP e levantamento da importância para cada variável. Depois dessa ponderação, os métodos de classificação foram executados, atribuindo classes às observações e medindo a acurácia e número de variáveis retidas no conjunto. As variáveis retidas foram organizadas em um histograma de acordo com sua frequência, o que determinou sua importância na definição dos modelos de reposição.

Quando aplicado em um cenário prático, a ferramenta de classificação KNN obteve um melhor desempenho frente à ADL, gerando uma acurácia média de 90% – com um desvio de 0,9% – e retendo cerca de 80% das variáveis (55% em todas as 100 iterações). Já a ADL alcançou 87% de acurácia média – com desvio de 1,3% – porém retendo 95% das variáveis (83% em todas as iterações). Com tais resultados, foi possível identificar que 14 das 24 variáveis são fundamentais na determinação de um modelo de reposição de estoques para um produto em um local específico de reposição.

O estudo realizado utilizou, como base, apenas variáveis quantitativas referentes a duas empresas

multinacionais de segmentos distintos (higiene pessoal e doméstica, alimentos e pneumáticos). Sugere-se, para estudos futuros, a consideração de variáveis qualitativas e a análise segregada por empresa, objetivando um viés estratégico e uma acurácia mais voltada à realidade e produtos de cada empresa.

Referências

- ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 2, n. 4, p. 433-459, 2010. <http://dx.doi.org/10.1002/wics.101>
- ANZANELLO, M.; ALBIN, L. S.; CHAOVALITWONGSE, A. W. Selecting the best variables for classifying production batches into two quality levels. **Chemometrics and Intelligent Laboratory Systems**, v. 97, n. 2, p. 111-117, 2009. <http://dx.doi.org/10.1016/j.chemolab.2009.03.004>
- BALLABIO, D. et al. Canonical Measure of Correlation (CMC) and Canonical Measure of Distance (CMD) between sets of data. Part 3. Variable selection in classification. **Analytica Chimica Acta**, v. 657, n. 2, p. 116-122, 2010. PMID:20005322. <http://dx.doi.org/10.1016/j.aca.2009.10.033>
- BARTHOLOMEW, D. J. The interpretation of multivariate data. In: BAKER, E.; PETERSON, P.; MCGAW, B. (Ed.). **International Encyclopedia of Education**. 3rd ed. New York: Elsevier, 2010. p. 12-17. <http://dx.doi.org/10.1016/B978-0-08-044894-7.01303-8>
- CHAE, B. Developing key performance indicators for supply chain: an industry perspective. **Supply Chain Management: An International Journal**, v. 14, n. 6, p. 422-428, 2009. <http://dx.doi.org/10.1108/13598540910995192>
- CHOI, S-II. et al. Input variable selection for feature extraction in classification problems. **Signal Processing**, v. 92, n. 3, p. 636-648, 2012.
- CHONG, I.; ALBIN, S.; JUN, C. A data mining approach to process optimization without an explicit quality function. **IIE Transactions**, v. 39, n. 8, p. 795-804, 2007. <http://dx.doi.org/10.1080/07408170601142668>
- DAKHLAOU, H.; BARGAOU, Z.; BÁRDOSSY, A. Toward a more efficient Calibration Schema for HBV rainfall-runoff model. **Journal of Hydrology**, v. 444-445, p. 161-179, 2012. <http://dx.doi.org/10.1016/j.jhydrol.2012.04.015>
- DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern classification**. New York: John Wiley & Sons, 2012.
- FISHER, R. A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**,

- v. 7, n. 2, p. 179-188, 1936. <http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- GAO Y.; GAO, F. Edited AdaBoost by weighted KNN. **Neurocomputing**, v. 73, n. 16-18, p. 3079-3088, 2010.
- GERTHEISS, J.; TUTZ, G. Feature selection and weighting by nearest neighbor ensembles. **Chemometrics and Intelligent Laboratory Systems**, v. 99, n. 1, p. 30-38, 2009. <http://dx.doi.org/10.1016/j.chemolab.2009.07.004>
- HAIR JR., J. F. et al. **Multivariate data analysis**. New Jersey: Pearson, 2010.
- HAN, J.; KAMBER, M. **Data mining: concepts and techniques**. San Francisco: Elsevier, 2006.
- HÄRDLE, W.; SIMAR, L. Applied multivariate statistical analysis. **Metrika**, v. 64, n. 1, p. 121-122, 2006.
- HE, S.-G.; WANG, G. A.; COOK, D. F. Multivariate measurement system analysis in multisite testing: An online technique using principal component analysis. **Expert Systems with Applications**, v. 38, n. 12, p. 14602-14608, 2011. <http://dx.doi.org/10.1016/j.eswa.2011.05.022>
- HOBERG, K.; THONEMANN, U. Modeling and analyzing information delays in supply chains using transfer functions. **International Journal of Production Economics**, v. 156, p. 132-145, 2014. <http://dx.doi.org/10.1016/j.ijpe.2014.05.019>
- IVANOV, D.; SOKOLOV, B. Control and system-theoretic identification of the supply chain dynamics domain for planning, analysis and adaptation of performance under uncertainty. **European Journal of Operational Research**, v. 224, n. 2, p. 313-323, 2013. <http://dx.doi.org/10.1016/j.ejor.2012.08.021>
- JOLLIFFE, I. T. **Principal component analysis**. 2nd ed. New York: Springer, 2002.
- KIM, D.; KUMAR, V.; KUMAR, U. Performance assessment framework for supply chain partnership, **Supply Chain Management: An International Journal**, v. 15, n. 3, p. 187-195, 2010.
- KITANI, E. C.; THOMAZ, C. E. Análise de discriminantes lineares para modelagem e reconstrução de imagens de face. In: CONGRESSO DA SBC, 27., 2007, Rio de Janeiro. **Anais...**
- LIU, L. B. et al. Applying the analytic hierarchy process to the offshore outsourcing location decision. **Supply Chain Management: An International Journal**, v. 13, n. 6, p. 435-449, 2008.
- MALINOWSKI, E. R. **Factor analysis in chemistry**. New York: John Wiley & Sons, 2002.
- MARTINEZ, A. M.; KAK, A. C. PCA versus LDA. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 23, n. 2, p. 228-233, 2001. <http://dx.doi.org/10.1109/34.908974>
- MOROPOULOU, A.; POLIKRETI, K. Principal component analysis in monument conservation: three application examples. **Journal of Cultural Heritage**, v. 10, n. 1, p. 73-81, 2009. <http://dx.doi.org/10.1016/j.culher.2008.03.007>
- MULA, J. et al. Models for production planning under uncertainty: a review. **International Journal of Production Economics**, v. 103, n. 1, p. 271-285, 2006. <http://dx.doi.org/10.1016/j.ijpe.2005.09.001>
- PARK, C. H.; LEE, M. On applying linear discriminant analysis for multi-labeled problems. **Pattern Recognition Letters**, v. 29, n. 7, p. 878-887, 2008. <http://dx.doi.org/10.1016/j.patrec.2008.01.003>
- RAGHURAJ, R.; LAKSHMINARAYANAN, S. Partial correlation based variable selection approach for multivariate data classification methods. **Pattern Recognition**, v. 42, n. 1, p. 7-16, 2009. <http://dx.doi.org/10.1016/j.patcog.2008.07.005>
- RENCHER, A. C.; CHRISTENSEN, W. F. **Methods of multivariate analysis**. New Jersey: Wiley, 2012. <http://dx.doi.org/10.1002/9781118391686>
- ROSSINI, K.; ANZANELLO, M.; FOGLIATTO, F. Seleção de atributos em avaliações sensoriais descritivas. **Produção**, v. 22, n. 3, p. 380-390, 2012. <http://dx.doi.org/10.1590/S0103-65132012005000032>
- SAHMER, K.; QANNARI, E. M. Procedures for the selection of a subset of attributes in sensory profiling. **Food Quality and Preference**, v. 19, n. 2, p. 141-145, 2008. <http://dx.doi.org/10.1016/j.foodqual.2007.03.007>
- SETH, N.; DESHNUKH, S. G.; VRAT, P. A framework for measurement of quality of service in supply chains, **Supply Chain Management: An International Journal**, v. 11, n. 1, p. 82-94, 2006.
- SHREVE, J.; SCHNEIDER, H.; SOYSAL, O. A methodology for comparing classification methods through the assessment of model stability and validity in variable selection. **Decision Support Systems**, v. 52, n. 1, p. 247-257, 2011. <http://dx.doi.org/10.1016/j.dss.2011.08.001>
- SIMCHI-LEVI, D.; KAMINSKY, P.; SIMCHI-LEVI, E. **Designing and managing the supply-chain: concepts, strategies, and case studies**. New York: McGraw-Hill, 2000.
- SU, M.-Y. Using clustering to improve the KNN-based classifiers for online anomaly network traffic identification. **Journal of Network and Computer Applications**, v. 34, n. 2, p. 722-730, 2011. <http://dx.doi.org/10.1016/j.jnca.2010.10.009>
- SYMS, C. **Principal component analysis**. Oxford: Academic Press, 2008.
- WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109-130, 2001. [http://dx.doi.org/10.1016/S0169-7439\(01\)00155-1](http://dx.doi.org/10.1016/S0169-7439(01)00155-1)
- XU, P.; BROCK, N.; PARRISH, R. S. Modified linear discriminant analysis approaches for classification of high-dimensional microarray data. **Computational Statistics & Data Analysis**, v. 53, n. 5, p. 1674-1687, 2008. <http://dx.doi.org/10.1016/j.csda.2008.02.005>
- YIN, R. K. **Estudo de caso planejamento e métodos**. 3. ed. Porto Alegre: Bookman, 2003.
- ZANONI, S.; JABER, M.; ZAVANELLA, L. Vendor managed inventory (VMI) with consignment considering learning and forgetting effects. **International Journal of Production Economics**, v. 140, n. 2, p. 721-730, 2012. <http://dx.doi.org/10.1016/j.ijpe.2011.08.018>
- ZHU, Q.; DOU, Y.; SARKIS, J. A portfolio-based analysis for green supplier management using the analytical network process. **Supply Chain Management: An International Journal**, v. 15, n. 4, p. 306-319, 2010. <http://dx.doi.org/10.1108/13598541011054670>