Universidade de São Paulo
Faculdade de Economia, Administração e Contabilidade
Departamento de Administração

**Exploração de Dados Atomizados
para Previsão de Vendas no Varejo
Utilizando Redes Neurais**

Antonio Fabrizio Lima Passari

Orientador - Prof. Dr. Fernando Carvalho de Almeida

Dissertação apresentada ao Departamento de Administração da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, como requisito parcial para a obtenção do título de Mestre em Administração.

SÃO PAULO

2003

Índice

<i>Lista de Figuras</i>	<i>v</i>
<i>Lista de Tabelas</i>	<i>vi</i>
<i>Lista de Siglas</i>	<i>vii</i>
CAPÍTULO I – O PROBLEMA DE PESQUISA	1
1. <i>Introdução</i>	1
2. <i>Formulação da Situação Problema</i>	2
3. <i>Objetivos da Pesquisa</i>	7
4. <i>Questões de Pesquisa</i>	7
5. <i>Delimitações da Pesquisa</i>	8
CAPÍTULO II – FUNDAMENTAÇÃO TEÓRICA	9
1. <i>O Problema de Previsão de Vendas</i>	9
1.1. Estudos Empíricos Sobre o Problema de Previsão	10
1.1.1. Questões Metodológicas	11
Nível de agregação das previsões	11
Propósito, uso, frequência e horizonte de tempo das previsões	15
Analistas e Usuários das Previsões	16
Fontes de dados	18
1.1.2. Questões de Seleção e Especificação de Técnicas de Previsão	19
1.1.3. Questões de Avaliação	21
Medidas de Desempenho dos Modelos de Previsão	24
Comparação e Seleção de Técnicas de Previsão	26
1.2. Técnicas de Previsão	27
1.2.1. Técnicas Qualitativas de Previsão	27
Técnicas Exploratórias	29
Técnicas de Grupo	29
1.2.2. Modelos Bayesianos	30
1.2.3. Técnicas Quantitativas de Previsão	32
Modelos de Escolha Discreta	32
Análises de Séries Temporais	33
Técnicas Causais	40
1.3. Previsão de Vendas no Varejo	46
1.3.1. Composto de Marketing	47
1.3.2. Conjunto de Produtos Oferecidos	50
1.3.3. Gestão de Preços no Varejo	53
Elasticidades cruzadas	55
1.3.4. Propaganda e Promoção	57
2. <i>Metodologia de Mineração de Dados Aplicada à Previsão de Vendas</i>	58
2.1. Bases de Dados para Mineração de Dados	60
2.1.1. Data Warehouses	62
2.2. Usos da Mineração de Dados	64
2.3. O Processo de Mineração de Dados	65
2.4. Técnicas de Mineração de Dados	67

2.4.1.	Indução de Regras / Análise do Cesto de Compras	67
2.4.2.	Sistemas Especialistas	70
2.4.3.	Árvores de Regressão	70
2.4.4.	Memórias Associativas	71
2.5.	Seleção das Técnicas de Mineração de Dados	71
3.	<i>Redes Neurais Artificiais</i>	72
3.1.	Justificativa de utilização das Redes Neurais	74
3.2.	Desempenho e Limitações das Redes Neurais	76
CAPÍTULO III – METODOLOGIA DA PESQUISA		80
1.	<i>O Método de Pesquisa</i>	80
2.	<i>Desenvolvimento da Pesquisa</i>	81
2.1.	Modelo de Pesquisa	81
CAPÍTULO IV – PESQUISA EMPÍRICA		84
1.	<i>Contextualização da Pesquisa</i>	84
1.1.	Metodologia de Previsão Utilizada pela EMPRESA	84
2.	<i>Obtenção dos Dados</i>	85
3.	<i>Seleção e Limpeza dos Dados</i>	86
4.	<i>Análise do Cesto de Compras</i>	87
4.1.	Análise do Cesto de Compras por Transação	88
	Análise por Produto	89
	Análise por Linha de Produto	89
	Análise por Setor	90
	Análises por Taxonomias	91
4.1.1.	Conclusões da Análise do Cesto de Compras por Transação	92
4.2.	Análise do Cesto de Compras por Cliente	92
	Análise por Produto	93
	Análise por Linha de Produto	94
	Análise por Setores	95
	Análises por Taxonomias	95
4.2.1.	Conclusões da Análise do Cesto de Compras por Cliente	95
4.3.	Confirmação da Análise do Cesto de Compras	96
5.	<i>Tratamento dos Dados para Previsão</i>	98
5.1.	Transformações Realizadas	101
5.2.	Testes de Correlações	102
5.3.	Testes de Autocorrelação e Correlações com Lag de Tempo	103
6.	<i>Modelagem com Uso de Redes Neurais</i>	104
6.1.	Seleção de Variáveis de Entrada	105
6.2.	Incorporação de Variáveis de Entrada	107
6.3.	Incorporação de Informações do Centro de Interesse	110
6.4.	Conclusões da Modelagem Baseada em Redes Neurais	111
7.	<i>Comparação com Outras Modelagens</i>	114

7.1.	Comparação com a Modelagem Naive	114
7.2.	Comparação com a Regressão Linear	115
7.3.	Resumo do Desempenho das Redes Neurais	116
<i>CAPÍTULO V - CONCLUSÕES</i>		117
1.	<i>Resultados Alcançados</i>	117
2.	<i>Conclusões da Previsão Baseada em Redes Neurais</i>	119
3.	<i>Sugestões para Trabalhos Futuros</i>	120
<i>Referências Bibliográficas</i>		122
<i>ANEXO A — Questionário aplicado à EMPRESA</i>		129
<i>ANEXO B — Dicionário de Dados Originais</i>		130
<i>ANEXO C — Dicionário de Dados – Vendas Semanais</i>		133
<i>ANEXO D — Transformações nos Dados</i>		134
<i>ANEXO E — Correlações e Autocorrelações Verificadas</i>		139

Lista de Figuras

<i>Figura 1 – Abordagens de previsão de vendas com uso de dados agregados</i>	4
<i>Figura 2 – Abordagens de previsão de vendas com uso de dados individuais e modelagem de séries temporais</i>	5
<i>Figura 3 - Abordagens de previsão de vendas com uso de dados individuais e modelagem causal</i>	5
<i>Figura 4 - Abordagem da presente pesquisa</i>	6
<i>Figura 5 - Framework para análise de métodos de previsão (WINKLHOFER et al., 1996)</i>	10
<i>Figura 6 - Taxonomia das técnicas de previsão</i>	20
<i>Figura 7 - Modelo de inferência bayesiana</i>	31
<i>Figura 8 – Decisão de compra baseada nos conceitos de valor e compostos de marketing</i>	48
<i>Figura 9 – Classificação dos tipos de varejo segundo sua amplitude e profundidade (PARENTE, 2000)</i>	51
<i>Figura 10 – Amplitude, profundidade e atratividade das diferentes categorias (PARENTE, 2000)</i>	53
<i>Figura 11 – Relacionamentos entre produtos substitutos e complementares de um ou mais centros de interesse</i>	56
<i>Figura 12 – Estatística convencional X Mineração de Dados</i>	60
<i>Figura 13 – Consolidação de dados em um Data Warehouse</i>	62
<i>Figura 14 – Data Warehouse como repositório para análises de dados</i>	63
<i>Figura 15 – O ciclo virtuoso de data mining (BERRY & LINOFF, 1997, modificado)</i>	66
<i>Figura 16 – Esquema de funcionamento de cada nó em uma rede neural artificial</i>	72
<i>Figura 17 – Modelo da Pesquisa</i>	82
<i>Figura 18 – Metodologia conduzida para a Pesquisa</i>	83
<i>Figura 19 – Relacionamentos entre produtos detectados nas análises do cesto de compras</i>	97
<i>Figura 20 – Boxplot do total de vendas semanal por setor, indicando a presença de muitos outliers</i>	100
<i>Figura 21 – Médias de vendas semanais por setor, indicando forte sazonalidade e picos de vendas</i>	101
<i>Figura 22 – Histograma da dispersão dos erros, na escala da variável prevista</i>	113
<i>Figura 23 – Gráfico de vendas previstas X vendas reais</i>	114
<i>Figura 24 – Histograma da variável “Estoque” sem transformações</i>	134
<i>Figura 25 – Boxplot da variável “Estoque” sem transformações</i>	135
<i>Figura 26 - Histograma da variável “Estoque” após a transformação $Z = \log x$</i>	136
<i>Figura 27 - Boxplot da variável “Estoque” após a transformação $Z = \log x$</i>	136
<i>Figura 28 – Gráfico scatterplot entre “Estoque_log” e “Grade_log”</i>	141
<i>Figura 29 – Gráfico scatterplot entre “Mostruario_log” e “Preço_pot”</i>	141
<i>Figura 30 – Gráfico scatterplot entre “Classe_distribuicao” e “Classe”</i>	142
<i>Figura 31 – Gráfico scatterplot entre “Estoque_log” e “Vendas_log”</i>	142
<i>Figura 32 – Gráfico scatterplot entre “Pprazo” e “Vendas_log”</i>	143
<i>Figura 33 – Gráfico de autocorrelação da variável “Vendas_log”</i>	143

Lista de Tabelas

<i>Tabela 1 - Classificação das técnicas de previsão quantitativas segundo o nível de agregação dos dados (ROBERTS, 1998)</i>	11
<i>Tabela 2 – Exemplos de comparação de desempenho das técnicas de previsão na literatura</i>	27
<i>Tabela 3 – Comparação de desempenho da técnica de redes neurais artificiais frente a outras técnicas estatísticas</i>	79
<i>Tabela 4 – Resultados da análise do cesto de compras por produto, abordados por transação</i>	89
<i>Tabela 5 - Resultados da análise do cesto de compras por linha de produto, abordados por transação</i>	89
<i>Tabela 6 - Resultados da análise do cesto de compras por setor, abordados por transação</i>	90
<i>Tabela 7 – Exemplos de regras da análise do cesto de compras abordado por transação, com uso de taxonomias entre os produtos</i>	91
<i>Tabela 8 - Resultados da análise do cesto de compras por produto, abordados por cliente</i>	93
<i>Tabela 9 - Resultados da análise do cesto de compras por linha de produto, abordados por cliente</i>	94
<i>Tabela 10 – Lista de Linhas de Produtos relacionadas – resultados preliminares</i>	95
<i>Tabela 11 - Lista de Linhas de Produtos relacionadas</i>	96
<i>Tabela 12 – Centro de interesse detectado</i>	98
<i>Tabela 13 – Principais variáveis independentes usadas nos modelos de previsão semanal</i>	99
<i>Tabela 14 – Transformações aplicadas aos dados</i>	102
<i>Tabela 15 – Correlações significativas detectadas entre as variáveis independentes</i>	103
<i>Tabela 16 - Correlações significativas entre as variáveis independentes e a variável dependente</i>	103
<i>Tabela 17 – Autocorrelações e correlações parciais verificadas</i>	104
<i>Tabela 18 – Resultados das redes 1 e 2, comparando os efeitos das transformações nos dados</i>	105
<i>Tabela 19 - Resultados da rede neural 6, sem a presença da variável “Grade”</i>	106
<i>Tabela 20 - Resultados da rede neural 10, com as entradas numéricas normalizadas e padronizadas</i>	107
<i>Tabela 21 - Resultados da rede neural 26, com incorporação das vendas na semana anterior e missing values preenchidos</i>	110
<i>Tabela 22 - Resultados da rede neural 108, com incorporação do total de vendas de produtos substitutos (missing values preenchidos)</i>	111
<i>Tabela 23 – Resumo das redes neurais geradas e performance obtidas</i>	112
<i>Tabela 24 – Capacidade preditiva do Modelo Baseado em Redes Neurais (rede 108)</i>	112
<i>Tabela 25 – Capacidade preditiva da previsão naive não ajustada</i>	115
<i>Tabela 26 – Equações de regressão linear geradas</i>	115
<i>Tabela 27 – Capacidade preditiva das equações de regressão linear</i>	115
<i>Tabela 28 – Performance comparativa das redes neurais</i>	116

Lista de Siglas

ARIMA – *Autoregressive Integrated Moving Averages*, ou métodos auto-regressivos

MAPE – *Mean Absolute Percentage Error*, ou Erro Percentual Absoluto Médio

MSE – *Mean Square Error*, ou Erro Quadrado Médio

RMSE – *Root Mean Square Error*, ou Raiz do Erro Quadrado Médio

SKU – *Stock Keeping Unit*, ou Identificador Único do Produto

CAPÍTULO I – O PROBLEMA DE PESQUISA

1. Introdução

A gestão eficiente de qualquer corporação, seja ela do setor público ou privado, industrial ou varejista, requer planejamento. Para um planejamento efetivo é necessário que se tenha uma expectativa precisa das condições futuras em que a corporação irá operar, e de como se relacionam os elementos condicionantes desta expectativa.

O governo de uma cidade deve saber qual a expectativa de crescimento populacional — e qual o relacionamento deste com o aumento na demanda por serviços — para ampliar, entre outras coisas, sua infra-estrutura básica, como redes de água, luz e esgoto. Uma indústria deve antecipar qual a demanda de seus produtos para agendar sua produção, por exemplo. E deve saber também os principais fatores que afetam esta demanda, para tomar as decisões corretas quando necessário.

Tanto decisões estratégicas como operacionais de uma corporação requerem, portanto, explorar o relacionamento presente entre os elementos que compõem a realidade em que a corporação está inserida. Para apoiar decisões corporativas como as citadas acima, as empresas procuram criar sistemas e procedimentos a fim de explorar cenários, com base em informações quantitativas e/ou qualitativas. Com o avanço da tecnologia e da capacidade de armazenagem e processamento dos sistemas computacionais, diversos modelos e técnicas quantitativos de previsão têm sido pesquisados, complementando e aprimorando as análises qualitativas por uma série de fatores, incluindo maior precisão.

Modelos de previsão quantitativos utilizam-se basicamente de dados históricos para detectar padrões de comportamento e estimá-los no futuro. Tais modelos empregam ferramental matemático-estatístico para representar a realidade para a qual foram criados. Diversas técnicas estatísticas têm sido usadas na criação dos modelos, baseadas em diferentes pressupostos assumidos.

Este trabalho visa explorar o uso de técnicas relativamente novas para criação de modelos voltados à detecção de padrões e previsão de demanda no varejo. Estas técnicas avançadas, comumente chamadas de *data mining* ou mineração de dados, incluem métodos de inteligência artificial (sistemas especialistas, lógica fuzzy, entre outros), árvores de decisão, métodos de indução de regras, redes neurais, entre outras (GARGANO & RAGGAD, 1999).

Essas técnicas consistem na descoberta de novas e importantes informações, procurando em certos casos “aprender” a partir de relacionamentos escondidos, padrões, correlações e interdependências existentes em grandes bases de dados. A evolução do desempenho computacional para exploração e análise de dados através de pacotes estatísticos permitiu o desenvolvimento e uso da mineração de dados de maneira crescente, com sucesso em muitas empresas (BERRY & LINOFF, 1997).

2. Formulação da Situação Problema

Toda previsão é uma tentativa de prognosticar o futuro através do exame do passado (HANKE & REITSCH, 1995). Consiste em gerar previsões não enviesadas da magnitude de alguma variável, como vendas, com base no conhecimento presente e passado acumulado em bases de dados e na experiência dos gestores e outros profissionais envolvidos. Muitas variáveis são comumente previstas por gestores, analistas, acadêmicos, instituições governamentais e outros, como renda, nível de emprego, inflação, produção, criminalidade, custo de vida, gastos, vendas, e muitas outras.

Segundo os autores, dois tipos de reação às previsões podem ser identificados:

1. Preparar-se para as conseqüências advindas das mudanças previstas. Por exemplo, se o gestor de uma empresa prevê queda nas vendas devido a problemas macroeconômicos, ele deve preparar a organização para enfrentar a contração.
2. Tomar medidas visando modificar as previsões. Se o cenário previsto é de queda, o gestor pode decidir enfrentá-la, incrementando ações de venda ou marketing, por exemplo.

O planejamento de ações é comumente realizado de uma forma interativa: planejamentos são utilizados como base para previsões, que por sua vez mostram os resultados prováveis se as ações planejadas forem tomadas. Com base nos resultados previstos, o gestor pode refazer seu planejamento, considerando novas decisões, que por sua vez geram novos resultados, e assim sucessivamente, até estabelecer uma decisão final ideal. Ainda segundo os autores, a visão oposta — planejar as ações após as previsões, como, por exemplo, planejar ações de marketing para enfrentar adversidades — não é realista, uma vez que os resultados dependem das ações tomadas. Assim, o planejamento das ações – como o

nível de preços ou ações de marketing planejadas – devem ser levadas em consideração nas previsões.

As previsões variam quanto à sua aplicação em níveis agregados (como na economia) ou para um componente específico (como para uma companhia individual). Também diferem quanto ao horizonte de previsão – curto, médio ou longo prazo são conceitos flexíveis, que variam conforme a aplicação. Previsões de vendas podem ser construídas para toda a companhia, para produtos ou grupo de produtos particulares, para regiões geográficas ou territórios de vendas específicos, para clientes particulares, ou para combinações dos elementos citados.

Este trabalho está focado no processo acima descrito, de geração de previsões de demanda para tomada de decisões no varejo. A preocupação central desta pesquisa é na exploração de uma metodologia baseada em técnicas de mineração de dados para identificação de relacionamentos entre a demanda de diversos produtos e os fatores de maior influência destas. Vislumbrando estes relacionamentos com maior clareza, o gestor fica mais habilitado à tomada de decisões acertadas no gerenciamento de vendas e outras atividades conseqüentes, como gerenciamento de estoques, de fluxo de caixa e outros.

Para que o gestor possa tomar decisões bastante específicas, os dados utilizados devem ser o mais atomizados (individuais) possível, gerando previsões com grande nível de detalhamento, porém pequeno horizonte futuro. A metodologia explorada busca ser útil, portanto, para tomada de decisões operacionais no varejo — e não estratégicas, de longo prazo. Situações típicas do setor varejista como: “o que acontece com a demanda e o lucro de minha loja se aumentar o preço da linha de produtos X?”, ou “meu concorrente cortou preços da ordem de 30%; qual a melhor medida a ser tomada?” — podem ser vislumbradas e as possíveis soluções endereçadas com o uso de uma metodologia de prospecção de cenários como a aqui proposta.

Previsões de curto prazo são empregadas para auxiliar decisões que devem ser feitas num futuro próximo, como a determinação do nível de inventário, número de funcionários e produção necessários, capital de giro, suprimentos a serem adquiridos, entre outras. Por outro lado, em situações onde o impacto das decisões tomadas pode influenciar a posição da companhia nos próximos anos, previsões de longo prazo são requeridas. Tais situações podem envolver decisões a respeito da construção de uma nova fábrica, a entrada em novos mercados, exportar ou não seus produtos, modernizar as operações, entre outras.

Inúmeros autores já pesquisaram sobre a adoção de modelos quantitativos de previsão de vendas, utilizando-se de diferentes técnicas e abordagens. Alguns autores, como TANG et al. (1991), LUXHOJ et al. (1996), ANSUIJ et al. (1996), KUO & XUE (1998), e SPEDDING & CHAN (2000), pesquisaram a adoção de modelos de previsão de vendas agregadas – para toda uma corporação ou loja, por exemplo. Estes modelos foram desenvolvidos a partir de duas abordagens clássicas:

(1) através da modelagem de séries temporais, onde se considera como variável de entrada os valores históricos da variável a ser prevista (a demanda agregada, no caso); ou

(2) com o uso de variáveis explicativas daquela a ser prevista como entrada de dados — os prováveis causadores da demanda. Esta abordagem é conhecida como modelagem causal.

A figura abaixo ilustra o desenvolvimento desses modelos.

Figura 1 – Abordagens de previsão de vendas com uso de dados agregados

Outros autores exploraram a previsão da demanda de produtos de maneira individualizada, a partir de séries históricas das vendas de cada produto. Esses estudos, como o de BARASH & MITCHELL (1998), GORDON (1998), e ALIBAIG & LILLY (1999), utilizaram técnicas de modelagem de séries temporais, tomando como entrada dos modelos os valores históricos das demandas de cada produto no tempo, construindo assim um modelo distinto para cada produto.

A figura a seguir mostra o processo de desenvolvimento destes modelos.

Figura 2 – Abordagens de previsão de vendas com uso de dados individuais e modelagem de séries temporais

Ainda outros autores exploraram o impacto de diversas variáveis na previsão de demanda de um produto isolado, realizando, portanto, o que se costuma referir como previsão causal, ou seja, a previsão realizada tendo-se como base as variáveis que possam explicar a previsão. Esses autores, como MENTZER & KENT (1999) e SUBRAHMANYAN (2000), buscaram relacionar as variáveis que possam estar relacionadas com a demanda individual de cada produto, desenvolvendo modelos de previsão que tomam como entrada os valores destas variáveis explicativas da demanda no tempo.

A figura a seguir ilustra o desenvolvimento desses modelos.

Figura 3 - Abordagens de previsão de vendas com uso de dados individuais e modelagem causal

Poucos autores, contudo, realizaram estudos a fim de desenvolver modelos que incorporassem o relacionamento existente entre os produtos. PARENTE (1997) mostra a importância do conceito de centros de interesse – grupo de produtos que têm alguma relação, seja concorrendo entre si ou complementando-se mutuamente – e da sua utilização na gestão da loja. Segundo o autor, a gestão por categorias de produtos (que representam centros de

interesse dos consumidores) é fundamental para o sucesso de um empreendimento de varejo, e costuma influenciar diretamente a disposição dos produtos na loja.

Alguns autores, como BEN-AKIVA & GERSHENFELD (1998) e GUADAGNI & LITTLE (1998) investigaram sobre o relacionamento entre produtos na previsão de vendas, mas nestes estudos pouco se utilizaram dados históricos sobre os produtos, baseando-se mais em análises exploratórias e pesquisas de grupo. Vê-se então uma lacuna na bibliografia sobre previsão existente, que é o estudo do relacionamento entre os produtos na previsão de vendas, com a construção de modelos quantitativos que possam incorporar e utilizar-se de informações acerca de mais de um produto simultaneamente.

CHAKRABORTY et al. (1992) mostraram, com o uso de redes neurais artificiais, ser possível melhorar a previsão de séries temporais de variáveis afins ao incorporá-las em um único modelo, e não realizar as previsões isoladamente. A exemplo da pesquisa de CHAKRABORTY et al. (1992), pretende-se neste estudo explorar uma metodologia de previsão de vendas que, a partir de dados históricos da demanda individual de cada produto pertencente a um centro de interesse, levar em consideração o impacto da demanda (ou das ações sobre a demanda) de cada produto sobre a demanda dos demais. O modelo deve, portanto, incorporar os relacionamentos existentes entre os produtos de um centro de interesse, utilizando para tanto a técnica de redes neurais artificiais.

A figura a seguir descreve a abordagem a ser adotada neste trabalho.

Figura 4 - Abordagem da presente pesquisa

Como mostrado na figura acima, este trabalho visa investigar o relacionamento entre produtos afins na previsão de vendas individual de produtos, e, através da adoção de um modelo causal, averiguar como esta previsão pode ser utilizada para auxílio à tomada de decisões no varejo.

O problema de investigação pode ser centrado na seguinte questão:

“Como utilizar o processo de mineração de dados para a descoberta de relacionamentos entre produtos e previsão de vendas individual de produtos para apoio à tomada de decisões operacionais no varejo?”

3. Objetivos da Pesquisa

O objetivo geral desta pesquisa é explorar a possibilidade de usar uma metodologia capaz de identificar padrões de relacionamento úteis na previsão de vendas individual no varejo, com o uso do processo de mineração de dados. Essas previsões devem abordar grande parte das decisões de curto prazo enfrentadas no cotidiano do gestor da loja, num nível aprofundado – detalhado quanto a produtos – de decisões. O objetivo é explorar um modelo de previsão de demanda para os produtos visando identificar um composto de marketing adequado (preços, produtos e promoções).

Os seguintes objetivos específicos podem ser destacados:

- Explorar a técnica de redes neurais artificiais para a criação de modelos de previsão de vendas individualizadas no varejo.
- Desenvolver uma rede neural artificial para previsão de vendas capaz de detectar relacionamentos entre variáveis que impactam no volume de vendas de uma loja ou cadeia de lojas, de forma detalhada (por produto).

4. Questões de Pesquisa

Durante o desenvolvimento da metodologia de criação de modelos para previsão de demanda, pretende-se explorar as seguintes questões:

- Como tratar os dados de vendas de maneira atomizada – sem agregações – e de forma sistemática para serem utilizados em um processo de mineração de dados. Em outras palavras, como realizar a coleta, filtragem, limpeza e transformações necessárias a um adequado pré-processamento de dados para criação de modelos, em se tratando de dados não agregados?
- É possível utilizar as informações a respeito dos centros de interesse de um grupo de produtos em modelos de previsão? Em caso afirmativo, esse

procedimento, incorporando os relacionamentos entre as demandas de produtos afins melhora os resultados dos modelos de previsão construídos?

- A técnica de redes neurais artificiais é adequada para geração de modelos de previsão no varejo?

5. Delimitações da Pesquisa

Esta pesquisa está focada em análises de dados de vendas no varejo, e, portanto, não trata de problemas referentes à demanda em serviços, na indústria primária ou mesmo outros elementos da cadeia de distribuição que não seja aquele de distribuição ao consumidor final.

Outra delimitação importante diz respeito ao horizonte de previsão: como os modelos gerados tratam de dados extremamente atomizados, sua utilização em horizontes amplos torna-se inviável. A princípio, apenas decisões operacionais do composto de marketing – num horizonte de curto prazo – são consideradas.

Não é preocupação deste estudo, também, a comparação direta entre modelos de previsão de vendas. Os modelos gerados foram confrontados com outras modelagens apenas para efeito de confirmação de sua capacidade preditiva mínima. A preocupação central deste trabalho é com a metodologia utilizada para detecção de relacionamentos entre as variáveis que afetam a demanda no varejo, a fim de dar uma melhor compreensão da realidade em que a loja está sujeita por meio de uma previsão de vendas causal mais acertada.

CAPÍTULO II – FUNDAMENTAÇÃO TEÓRICA

Esta revisão bibliográfica está dividida em três partes. Na primeira parte, o problema de previsão de vendas é abordado; são mostradas as diversas iniciativas tomadas ao longo da evolução do tema na tentativa de solucionar o problema, e a abordagem desta pesquisa é situada perante as demais abordagens encontradas na literatura. Na segunda parte, o processo de mineração de dados é abordado, mostrando algumas das técnicas desenvolvidas até o presente momento e a metodologia de análise envolvida em projetos de mineração de dados. Finalmente, a terceira parte traz uma revisão da literatura sobre a técnica de redes neurais artificiais e sua aplicação para problemas de previsão.

1. O Problema de Previsão de Vendas

“These hieroglyphics have evidently a meaning. If it is a purely arbitrary one, it may be impossible for us to solve it. If, on the other hand, it is systematic, I have no doubt that we shall get to the bottom of it” (DOYLE, 2002).

Para aumentar a vantagem competitiva comercial em um ambiente de constantes mudanças, os gestores de uma organização devem tomar a decisão correta no momento certo, com as informações que tiverem em mãos. Uma previsão acertada tem papel fundamental nesse processo. Compreendendo melhor o comportamento do consumidor e sua resposta às alterações nos fatores controláveis que o varejista articula (o composto de marketing), os gestores podem prever os resultados de suas ações e obter dados de diagnóstico para se suprir de informações que permitam ações acertadas no futuro.

Previsão, ou em inglês *forecasting*, pode ser definida como uma sequência de passos que o tomador de decisões realiza, seja implícita ou explicitamente, para antever satisfatoriamente um valor futuro. Entretanto, devido às complexidades do processo decisório em ambientes não estruturados, não é uma tarefa fácil saber quais passos tomar (LUXHOJ et al., 1996).

MAKRIDAKIS et al. (1983) destacam os fatores que fizeram com que a importância da previsão aumentasse:

- Aumento na complexidade das organizações (como o aumento do número de nichos de mercado e de produtos oferecidos) e de seu ambiente (mudanças tecnológicas e na

estrutura da demanda), tornando mais difícil levar em consideração todos os fatores relacionados ao desenvolvimento futuro da organização.

- As organizações passaram a adotar procedimentos decisórios mais sistemáticos, que envolvem justificativas explícitas para cada ação tomada. Ter uma previsão de vendas formal é uma forma de suportar tais procedimentos.
- O desenvolvimento contínuo das técnicas de previsão e suas aplicações, permitindo que não só analistas especializados como também gerentes e outros tomadores de decisão entendessem e utilizassem tais técnicas.

Apesar do reconhecimento da importância das previsões entre executivos de todos os setores, sua implementação não é universal – existe uma lacuna entre o desenvolvimento teórico das técnicas de previsão e sua aplicação prática nas organizações. De acordo com o que foi levantado nesta revisão bibliográfica, o uso prático das técnicas de previsão ainda está muito distante dos avanços teóricos, fato comprovado pelo uso maciço de técnicas subjetivas ou de técnicas quantitativas rudimentares.

1.1. Estudos Empíricos Sobre o Problema de Previsão

WINKLHOFER et al. (1996) propõem um *framework* (estrutura) para organizar os estudos empíricos relativos à previsão, que é apresentado abaixo.

Figura 5 - *Framework* para análise de métodos de previsão (WINKLHOFER et al., 1996)

O *framework* distingue três diferentes grupos de problemas:

1. **Questões metodológicas** (*design issues*) compreendem o propósito e o tipo de previsão, os recursos envolvidos, as características dos analistas e usuários envolvidos e as fontes de dados utilizadas;
2. **Questões de seleção e especificação** (*selection/specification issues*) dizem respeito às técnicas de previsão e respondem as questões de familiaridade, seleção e uso de métodos de previsão alternativos;
3. **Questões de avaliação** (*evaluation issues*) focam nos resultados das atividades de previsão, como a apresentação e revisão das previsões, medição da performance da previsão e as forças que afetam sua acurácia.

Os três grupos são interligados de tal forma que cada um deles tem implicações nos demais. As seções seguintes abordam o problema de previsão segundo o *framework* proposto.

1.1.1. Questões Metodológicas

Nível de agregação das previsões

Para analisar o nível de agregação dos dados, ROBERTS (1998) propõe que as técnicas de previsão quantitativas existentes podem ser classificadas dentro de quatro categorias básicas, que são:

	Nível individual	Nível agregado
Produtos existentes	Modelos de varredura de dados do comportamento individual (I)	Modelos econométricos de comportamento do mercado (II)
Novos produtos	Modelos de teste e pré-teste de mercado (III)	Modelos de difusão da inovação (IV)

Tabela 1 - Classificação das técnicas de previsão quantitativas segundo o nível de agregação dos dados (ROBERTS, 1998)

As técnicas de modelagem voltadas para previsão em marketing podem ser divididas, de acordo com o autor, segundo duas dimensões principais: (1) granularidade (individual ou agregada) e (2) se estão dirigidos ao estudo de novos produtos ou a sensibilidades a produtos já existentes. Segundo o próprio autor, as divisões entre as categorias não são perfeitamente claras e precisas, mas elas de fato formam grupos bem distintos de modelos de previsão.

A seguir são apontadas as principais características dos modelos de previsão de cada um dos quadrantes apresentados por ROBERTS (1998), e é destacada a abordagem foco desta pesquisa (tipo I).

Modelagens Tipo I - Modelos de Varredura de Dados

Há muito tempo os pesquisadores estudam séries históricas de dados de produtos existentes para compreender melhor as respostas às variáveis do composto de marketing, particularmente preço e propaganda. Para realizar estudos sobre dados atomizados (individuais), faz-se necessária uma análise longitudinal de uma grande quantidade de dados capazes de representar o comportamento dos consumidores, caracterizando o tipo de estudo que costuma ser referido como varredura de dados, ou *scanner data analysis*.

Este tipo de análise pode ser utilizado a fim de obter o composto de marketing otimizado (que resulta em maior faturamento ou lucro, por exemplo) em ambientes altamente competitivos, entender fontes de heterogeneidade de consumidores, e para estudar as relações entre o ponto de venda e a indústria em diferentes segmentos (ROBERTS, 1998).

Diversas técnicas têm sido utilizadas neste tipo de modelagem, entre elas modelos de escolha discreta, árvore de decisões, técnicas de regressão linear múltipla e análises de séries temporais. Todas elas demandam um grande esforço do analista para que possam ser adaptadas para trabalharem com dados individualizados, e os resultados nem sempre são satisfatórios (ROBERTS, 1998).

Uma das técnicas para previsão desagregada que vem sendo utilizada com sucesso é a modelagem com uso de redes neurais artificiais, tendo mostrado superioridade sobre várias outras modelagens, em diversas áreas do conhecimento (ANSUJ et al., 1996).

Dentro da classificação proposta por ROBERTS (1998), este trabalho localiza-se nesta categoria de previsão — modelagens tipo I — pois está focado no desenvolvimento e avaliação de modelos para o estudo individual das sensibilidades de mercado às variáveis do composto de marketing para cada produto. Ou seja, dentro de um ambiente de varejo, este trabalho propõe uma análise longitudinal de grande quantidade de dados históricos, a fim de desenvolver um modelo quantitativo capaz de prever a demanda futura de um grupo de produtos de forma individualizada (não agregada).

A fim de melhor localizar o presente projeto perante as demais abordagens de previsão segundo o nível de agregação dos dados, as demais tipologias são brevemente apresentadas.

Modelagens Tipo II - Modelos Econométricos de Comportamento do Mercado

Modelos econométricos e de análise de séries de tempo são enquadrados nesta categoria. Apesar de os modelos de comportamento individual serem considerados modelos econométricos, e os modelos de difusão de inovações serem baseados em séries de tempo, esta categoria é voltada àqueles modelos mais genéricos, desenvolvidos principalmente no campo da econometria e estatística econômica, e então utilizados na área de marketing.

Os modelos econométricos surgiram efetivamente para prever índices econômicos, como o próprio nome diz, e foram posteriormente adaptados para previsões microeconômicas, e baseiam-se principalmente em técnicas estatísticas de regressão. O uso deste tipo de modelo supõe a construção das equações matemáticas que o regem, o que envolve a seleção *a priori* de todas as variáveis a ser consideradas e do tipo de relacionamento existente entre elas.

Já a análise de séries temporais é o processo de avaliar os relacionamentos históricos entre uma variável e o tempo. O propósito básico é identificar comportamentos periódicos e utilizá-los para prever séries futuras.

Exemplos típicos da utilização destes modelos são a previsão de tendências de mercado como a taxa de retração de uma indústria ou setor; de índices macroeconômicos como inflação e crescimento do PIB; de vendas agregadas em uma empresa, como o total de demanda anual ou mensal; dos efeitos de variáveis do composto de marketing na demanda total, exame de canais de distribuição, entre outros.

Esta classe de modelagem foi revista por HANSSENS et al. (1990), e aplicações e exemplos das técnicas podem ser encontrados em BROWN (1959), KLEIN & YOUNG (1980) e GROSS & PETERSON (1983).

Modelagens Tipo III - Modelos de Teste e Pré-teste de Mercado

São modelos calibrados com dados de pré-lançamento, utilizando para tanto dados individualizados de respondentes selecionados. Os modelos assim desenvolvidos são geralmente voltados ao estudo da participação e penetração no mercado para novos produtos, e se utilizam basicamente da análise de escolhas discretas. Outras possibilidades são a estimação do tamanho de um novo mercado ou de sua transformação com a entrada de um novo produto ou marca.

O uso típico destes modelos é a aplicação em experimentos com a finalidade de criar uma função-utilidade para situações de escolhas baseadas em menus, ou seja, em seleções

pré-definidas de itens ou benefícios em um produto. Podem ser utilizadas técnicas baseadas em julgamento, em técnicas estatísticas como análise *conjoint* ou modelos de escolha discreta econométricos tipo *multinomial logit* ou *probit*, desde que utilizem dados desagregados.

Observam-se aplicações destes modelos em diversas áreas - além da análise de penetração de mercado de novos produtos -, como na seleção de cesta de produtos para venda casada (BEN-AKIVA & GERSHENFELD, 1998), determinação de visitação a um evento (POWERS, 1998), e ainda na previsão de demanda de transporte, para, por exemplo, selecionar qual infra-estrutura viária deve ser construída ou qual meio de transporte coletivo implantar (DAGANZO, 1979), entre outros.

Modelagens Tipo IV - Modelos de Difusão de Inovação

Os modelos para previsão de difusão de inovações surgiram na década de 60, com FORT & WOODLOCK (1960), MANSFIELD (1961) e principalmente com o modelo proposto por BASS (1969), o mais conhecido e explorado modelo de difusão de inovações. Esses modelos foram criados com base na analogia da idéia de propagação de doenças aplicada ao lançamento de novos produtos ou tecnologias: a taxa de adoção inicialmente cresce rapidamente com o efeito de contágio provocado por aqueles que já adotaram a inovação, e posteriormente decresce com o efeito de saturação e diminuição do mercado disponível.

O processo de difusão consiste de quatro elementos básicos: inovação, canais de comunicação, tempo e sistema social. Como uma teoria de comunicação, a teoria de difusão foca nos canais de comunicação. Estes consistem em dois meios fundamentais e distintos: (1) mídia de massa e (2) comunicações pessoais, ou “boca-a-boca”. O modelo de Bass assume então que os adotantes estão subdivididos em dois grupos: aqueles influenciados apenas pela mídia e aqueles influenciados apenas pelo boca-a-boca — os “inovadores” e os “imitadores”, respectivamente.

O modelo de Bass e suas formas revisadas têm sido usados para prever a difusão de inovações no varejo, tecnologia industrial, agricultura, educação, farmácia e mercados de bens duráveis. O objetivo de um modelo de difusão é prever o nível de uso de uma inovação ao longo do tempo em uma população de possíveis adotantes, e serve basicamente para prever as vendas de inovações consideradas “primeira compra”.

A principal equação do modelo de Bass pode ser escrita como:

$$\text{Equação I: } n(t) = \frac{dN(t)}{dt} = p.[m - N(t)] + \frac{q}{m}.N(t).[m - N(t)]$$

onde: $n(t)$ é o número de novos adotantes no período;
 p é o coeficiente inovação ou coeficiente de influência externa;
 q é o coeficiente de imitação, ou coeficiente de influência interna;
 m é o potencial de adotantes totais.

Na equação, nota-se que um dos termos representa a adoção por consumidores não influenciados pela imitação — os inovadores — e um segundo termo representando a adoção dos imitadores.

Além de permitir previsões de venda antes do lançamento para os primeiros estágios do ciclo de vida — quando as compras de reposição ainda não ocorreram — estes modelos também são usados após o lançamento da inovação, sendo este o maior desafio identificado na literatura. Quando calibrados com base em dados históricos, os modelos de difusão só são estáveis após o ponto de máximo absoluto da taxa de vendas (HEELER & HUSTAD, 1980). Entretanto, quando ajustados com dados externos, a calibração pré-lançamento pode ser alcançada (LAWRENCE & LAWTON, 1981).

É interessante notar que, mesmo depois de mais de trinta anos de pesquisas e desenvolvimento dos modelos de inovação, um dos primeiros modelos — aquele proposto por BASS (1969) — continua a ser extensamente utilizado como referencial teórico, e os resultados práticos de sua aplicação em previsões têm-se mostrado com desempenho igual ou superior aos modelos mais recentes e sofisticados (PUTSIS, 1998; FILDES & BOTTOMLEY, 1998). MAHAJAN et al. (1990) trazem uma revisão completa dos modelos de inovação, seus avanços e conclusões.

Propósito, uso, frequência e horizonte de tempo das previsões

Diversos estudos analisam o porquê de se realizar previsões e seu uso. Em WHITE (1986), 64% dos respondentes afirmaram que o principal propósito das previsões é formular um objetivo, uma performance a ser alcançada, enquanto apenas 30% gostariam de obter uma medida realista do potencial do mercado. MENTZER & COX (1984) pesquisaram sobre as principais áreas de aplicação das previsões, e os resultados indicam planejamento de produção, inventários e orçamento como as áreas decisórias mais influenciadas pelas previsões. Este resultado é coerente com aqueles encontrados por MCHUGH & SPARKES (1983) e PETERSON (1993). Em muitos casos, os gestores confundem previsão com metas a

serem alcançadas, ou seja, previsão de vendas com planejamento de vendas (FILDES & HASTINGS, 1994), o que gera freqüentemente mudanças nos resultados das previsões por motivos externos, como motivações políticas e estratégicas da organização (PETERSON, 1993).

O estudo de PETERSON (1993) vai mais além, concluindo que grandes firmas de varejo utilizam técnicas de previsão de vendas com maior freqüência que os pequenos varejistas, resultado confirmado por PETERSON & JUN (1999), que mostraram que entre as empresas grandes, 100% afirma preparar previsões ao menos uma vez por ano, número este que cai para 93,2% entre as empresas pequenas. As freqüências de preparação das previsões são segundo os autores, nesta ordem: anual (a mais comum), quadrimestral, semestral e mensal (a menos comum), resultado compatível com aqueles obtidos por CERULLO & AVILA (1975) e WHITE (1986).

Quanto ao horizonte de tempo utilizado nas previsões, há pouco consenso entre os autores, tanto no que se refere ao desempenho das técnicas quanto às próprias definições de curto, médio e longo prazos. Estudos já foram conduzidos para previsões com horizonte de tempo de um dia (SAKAI et al., 1999) até 25 anos (NAYLOR, 1981). Os estudos relacionados ao horizonte de tempo e freqüência das previsões confirmam, no entanto, a proposição de WHITE (1986), de que as “companhias utilizam a freqüência que melhor se adapta aos seus produtos, mercado e método de operação. Não há uma freqüência de previsão ideal”.

Analistas e Usuários das Previsões

Parece não haver um consenso sobre a responsabilidade na preparação das previsões, isto é, de maneira geral não há nas organizações um cargo de analista de previsões fixo. De acordo com DRURY (1990), apenas uma em cada cinco empresas possui equipe específica para planejamento e previsão; em metade das empresas essa função está a cargo da alta administração, e em cerca de 15% das companhias a responsabilidade não é definida. Essa função — de analista, ou seja, quem prepara as previsões — combina as disciplinas de economia, matemática, estatística, contabilidade e, mais recentemente, computação; seu objetivo é tentar explicitar através de equações ou modelos matemático-computacionais seus pressupostos (teorias) e processos lógicos sobre a realidade, tendo a vista a obtenção de previsões sobre determinadas grandezas.

Há observações que indicam falta de comunicação entre usuários e analistas, e uma falta de preparo na elaboração de previsões efetivas, mesmo com a crescente facilidade na obtenção de análises oferecida por pacotes estatísticos comerciais à disposição atualmente (WINKLHOFFER et al., 1996). Há ainda grande disparidade na percepção do status e necessidades das previsões entre os analistas e os usuários (WHEELWRIGHT & CLARKE, 1976), o que causa análises desconexas da necessidade real do gestor e, portanto, freqüentemente ignoradas.

SANDERS (1992) conclui que os gestores responsáveis pelas previsões em empresas de serviço e varejo tendem a ter formação educacional mais baixa que aqueles de empresas industriais. O autor afirma ainda que os gestores de empresas industriais estão mais familiarizados com técnicas quantitativas relativamente complexas que os gestores de empresas de serviço e varejo. Outra conclusão relacionada diz respeito à correlação positiva encontrada entre o nível de escolaridade dos analistas e o uso de técnicas de previsão mais sofisticadas (CERULLO & AVILA, 1975), o que ajuda a explicar a maior importância e aplicação dada pelas empresas industriais à previsão de vendas, em comparação às empresas de varejo.

PAN et al. (1977) concluem que grandes empresas industriais reconhecem a importância da previsão de vendas e empregam recursos na elaboração de previsões formais e regulares. Ainda segundo PAN et al. (1997), as pequenas empresas também costumam utilizar com maior freqüência técnicas subjetivas, em oposição às grandes companhias, que utilizam técnicas quantitativas mais sofisticadas com maior freqüência.

De acordo com CHASE (1999) os responsáveis pelas previsões tendem a ser analistas de negócios, ou seja, analistas com uma visão muito mais ampla do que simplesmente prover os gestores com números relativos a previsões acuradas. A combinação de apreciação estatística e conhecimento do mercado é necessária para tal análise, o que resulta em uma previsão de vendas que pode ser utilizada na confecção de planos de vendas, de marketing, de produção, de compras e finalmente de negócios.

Essas conclusões fortalecem a carência latente dos segmentos ligados ao varejo em conhecimento sobre as técnicas de planejamento e previsão. Também corroboram para a idéia de que o analista tem a função de definir e medir os elementos do mercado que direcionam o negócio — e não de simplesmente tentar adivinhar o futuro — provendo a alta administração com análises de suporte à decisão acionáveis, capazes de influenciar os objetivos de vendas.

Fontes de dados

Uma expressão comum entre os analistas e estatísticos é, em inglês, *Garbage In, Garbage Out*, ou “entra lixo, sai lixo”. Ela expressa bem o sentido de que nenhum modelo é capaz de gerar boas previsões sem a presença de dados confiáveis e, principalmente, que representem o problema em questão. O problema reside no fato de que os sistemas voltados à operação da empresa nem sempre enfatizam a presença de dados “limpos” e confiáveis. Algumas vezes, os dados disponíveis não são adequados ao que se quer prever; por exemplo, na previsão de demanda, muitas vezes só se dispõe de dados de entregas, e estas não são uma boa representação do fenômeno de interesse – a demanda.

A presença de dados confiáveis e que representem o fenômeno estudado é, segundo BROWN (1963), um dos fatores de maior ganho em acurácia. Normalmente, demanda-se um grande esforço de modelagem para um pequeno ganho de acurácia, porém dados que melhor representem o problema geram grandes ganhos, muitas vezes com pouco esforço.

Para a preparação de previsões – tanto agregadas como individuais – diversas fontes de dados podem ser utilizadas. Normalmente, elas são divididas em dois grandes grupos de fontes: internas e externas (WINKLHOFFER et al., 1996).

Fontes internas de dados são aquelas, como o próprio nome diz, internas à companhia, e muitas vezes estão armazenadas em diversas bases de dados nos departamentos ou setores da empresa. Para ter uma fonte interna de dados maciça, é necessário manter um registro de todas as atividades do composto de marketing, incluindo modificações realizadas nos produtos, promoções, políticas internas, mudanças ambientais e da concorrência, já que modelos quantitativos podem utilizar todos esses dados, muitas vezes com ganhos de acurácia aliados a pouco esforço adicional para coleta de tais informações. Os dados internos podem ter origem nos canais de venda da companhia, como dados históricos de vendas; em informações coletadas com a força de vendas (quantitativas e qualitativas); em dados dos centros de distribuição e dos pontos de venda final. Naturalmente, estes últimos são os mais indicados para previsões de venda, já que refletem a demanda real dos produtos, sem distorções devido a reposições de estoque, devoluções, diferenças de tempo entre distribuição e demanda final, entre outras.

As fontes externas de dados mais utilizadas são aquelas que dizem respeito aos fatores macroeconômicos que possam influenciar nas previsões, como índices governamentais, tendências político-econômicas, estatísticas financeiras e sócio-econômicas. Também é

comum a compra de bases de dados externas para se obter dados da concorrência, dados econômicos e do setor, promoções, tendências de preços, de lançamento de produtos e tecnológicas, entre outros.

É freqüente a presença de distorções nos dados disponíveis, que podem ter várias origens, entre elas (GEURTS & WHITLARK, 1996):

- Presença de *outliers*. Podem ser resultado de ações de marketing extraordinárias, ou fenômenos externos com efeito apenas sobre alguns pontos da série (de curta duração). Normalmente, a ação de *outliers* é de difícil detecção pelos modelos.
- Presença de cotas de venda. Os vendedores manipulam suas vendas para atingir as cotas, e muitas vezes prorrogam as vendas para ficar mais fácil atingir as cotas do próximo período, ou realizam promoções agressivas para atingir as metas do período. Este fenômeno é conhecido como *game playing*.
- Séries de dados contaminadas com dados de fenômenos concorrentes, como o caso da demanda representada pelas entregas ou produção, ao invés de vendas propriamente ditas. Nestes casos as previsões terão acurácia invariavelmente baixa, e a única solução é tentar aproximar os dados do que se quer estudar de fato. Por exemplo, pode-se inserir o valor de pedidos atrasados ou adiantados a fim de aproximar os dados de entregas da demanda verdadeira.
- Vendas gerenciadas por “contas”. As contas podem não representar a demanda real, uma vez que também estão sujeitas ao *game playing*.
- Estratégias para fugir de taxas e impostos, como a produção acelerada no último mês do ano. Estas devem ser detectadas e amenizadas.
- Políticas empresariais diversas, por questões legais, de marketing ou outras.

1.1.2. Questões de Seleção e Especificação de Técnicas de Previsão

Inúmeras técnicas de previsão já foram desenvolvidas na tentativa de melhor elucidar o problema de previsão de vendas. Tais técnicas podem ser divididas em dois grupos principais de abordagem: quantitativo e qualitativo. Dentro de cada grupo, várias subdivisões distintas já foram propostas na literatura para classificar as técnicas de previsão; neste trabalho é proposta uma nova taxonomia das técnicas de previsão, buscando classificá-las

primeiramente de acordo com sua abordagem conceitual, em seguida por sua metodologia de coleta e tratamento dos dados, e finalmente distinguindo as diferentes técnicas de análise de dados dentro de cada abordagem.

A figura a seguir mostra a taxonomia das técnicas de previsão proposta (a caixa em destaque sinaliza a categoria de modelos foco desta pesquisa):

Figura 6 - Taxonomia das técnicas de previsão

Conforme mostra a figura, os métodos quantitativos podem ser divididos em: 1) análise de séries temporais, baseada na descoberta de padrões nos dados históricos de uma

variável; 2) métodos causais, baseados em relacionamentos de causa e efeito, tentando descrever matematicamente o relacionamento de uma variável dependente com uma ou mais variáveis independentes; e 3) métodos de escolha discreta, utilizados essencialmente em problemas de classificação. Já os métodos qualitativos podem ser divididos em: 1) exploratórios, como as pesquisas de mercado e teoria de utilidade; e 2) técnicas de grupo, como as técnicas Delphi e composição da força de vendas.

1.1.3. Questões de Avaliação

O processo de previsão pode ser decomposto em três fases principais (LUXHOJ et al., 1996):

1. Identificação do problema;
2. Desenvolvimento de alternativas;
3. Seleção da melhor alternativa baseada em um critério definido.

A partir do problema identificado – como previsão de vendas no varejo com uso de dados atomizados – as alternativas de técnicas são consideradas, como aquelas mostradas nas seções anteriores. A fase final do desenvolvimento de um modelo de previsão passa pela avaliação dos modelos e seleção daquele mais apropriado ao problema, a partir de critérios bem definidos.

Os critérios utilizados para seleção da melhor técnica de previsão são muito variados, e passam por:

- Acurácia. É o critério mais utilizado para seleção da melhor técnica de previsão, e muitas vezes o único. Sem dúvida é um critério primordial, dado que previsões com grandes erros não suportam o gestor na tomada de decisões; pelo contrário, levam a decisões danosas à corporação.
- Custo. Envolve o custo de obtenção, tratamento e preparação dos dados, o custo no desenvolvimento e avaliação dos modelos, e o custo de implantação do modelo na operação da empresa. Relaciona-se sempre aos elementos humano, de tempo e computacional.
- Habilidade do analista. Muitas vezes a falta de know-how sobre uma determinada técnica de previsão impede que ela seja sequer considerada. Uma familiaridade maior na aplicação de alguma técnica influencia decisivamente a

escolha da mesma, até porque o analista tende a gerar modelos melhores do que ao utilizar técnicas sobre as quais detém pouco conhecimento.

- Características desejadas na previsão, como explicabilidade, incorporação de determinadas variáveis, capacidade de realizar simulações, entre outras.
- Características específicas do problema. Envolve todo o sistema investigado: número e características das variáveis relevantes, como elas se comportam – alguns métodos só podem ser aplicados a dados com determinados padrões de comportamento, horizonte de tempo considerado, número de itens previstos, periodicidade das previsões, entre outras.
- Facilidade de uso. Refere-se à facilidade de uso da técnica em si, na geração do modelo, e da utilização posterior deste modelo no contexto da organização. Envolve: complexidade, tempo de desenvolvimento, nível de conhecimento exigido, e base conceitual aplicada.
- Requerimento de dados. A disponibilidade de dados dentro dos padrões esperados por determinadas técnicas pode pressionar a escolha das mesmas.
- Disponibilidade de softwares. A exploração das técnicas de previsão depende diretamente da disponibilidade de pacotes estatísticos que as suportem, sobretudo nas previsões quantitativas.
- Velocidade. Refere-se à agilidade e flexibilidade no desenvolvimento e aplicação de novos modelos.
- Facilidade de interpretação. Esta é uma característica fundamental de qualquer previsão, uma vez que ela deve ser bem compreendida pelo gestor. Está relacionada com o poder explicativo dado pela técnica, e além de influenciar o modo como o administrador irá tomar decisões a partir da previsão, afeta também sua credibilidade.
- Eficiência. Capacidade de melhorar a performance com o passar do tempo, conforme assimila maior quantidade de dados. Também conhecida como taxa de “aprendizado” do modelo de previsão.
- Enviesamento. Ausência de respostas enviesadas; que tendem sempre para mais ou para menos.
- Capacidade de incorporar a experiência do gestor. Este tem sido o critério mais usado para justificar o uso de métodos qualitativos de previsão (MCHUGH & SPARKES, 1983).

Não há consenso na literatura sobre qual o critério ou conjunto de critérios mais importante na seleção das técnicas de previsão. WHEELWRIGHT & CLARKE (1976) citam custo, habilidade do analista, características do problema e desejadas no método como os critérios mais relevantes. Já SANDERS (1992) obteve como respostas de gestores de empresas de manufatura: acurácia, requerimento de dados e facilidade de uso. Facilidade de uso também foi destacada por MCHUGH & SPARKES (1983), que adicionalmente encontraram a (falta de) habilidade do analista como o fator predominante para o não uso de técnicas quantitativas mais formais. Finalmente, no estudo de NAYLOR (1981) os maiores obstáculos para uso de técnicas quantitativas foram enviesamento, falta de know-how (habilidade do analista) e tempo (velocidade) para construção dos modelos.

Os resultados de WILSON & DAUBECK (1989), entretanto, mostram que há uma associação positiva entre a importância percebida dos métodos de previsão e sua acurácia. Isto significa que quanto maior a acurácia de um modelo, maior a importância percebida dele. De fato, acurácia tem sido relatada como o critério mais importante na seleção das técnicas de previsão (BROWN, 1963; LAWRENCE et al., 2000).

É importante distinguir os critérios de seleção de técnicas com aqueles de avaliação das previsões. Durante a seleção, a preocupação é com a escolha da técnica mais adequada ao problema, aquela capaz de modelar o fenômeno com maior precisão ou atingir outro critério de seleção especificado. Já durante o processo de avaliação, está-se preocupado com a performance do modelo criado a partir da técnica selecionada, ou seja, com a medida da qualidade das previsões efetuadas. Naturalmente, uma escolha acertada da técnica de previsão deve levar a uma melhor capacidade preditiva do modelo gerado.

Acurácia é, também no caso de avaliação de performance, o critério mais utilizado (MENTZER & COX, 1984; SANDERS, 1992). Tal resultado é natural que seja esperado, uma vez que a acurácia das previsões de vendas tem um papel fundamental na tomada de decisões, especialmente no que diz respeito ao planejamento financeiro e reposição de inventários. Previsões mal feitas podem levar a excesso de inventário ou perda de vendas, problemas no agendamento da produção, decisões de preço incorretas, falhas no atendimento ao cliente, entre outros – todas situações de perda financeira, ocasionando diminuição da lucratividade.

Quanto aos resultados das previsões, as empresas de bens de consumo (varejo) atingem previsões otimistas (enviesadas para mais) com maior frequência que as empresas

industriais, e também consideram a previsão de vendas não importante com maior frequência (SANDERS & MANRODT, 1994).

Medidas de Desempenho dos Modelos de Previsão

Uma vez que acurácia é o critério mais importante na seleção e avaliação das técnicas de previsão, ela deve ser definida matematicamente. A fim de avaliar o desempenho (acurácia) absoluto e relativo dos diversos modelos de previsão, diversas ferramentas matemáticas foram desenvolvidas com o tempo. Os principais mecanismos são (ACZEL, 1993):

- ME. *Mean Error*, ou Erro Médio, é a medida mais simples de erro das previsões, definida como o somatório dos erros (erro total) dividido pelo número de observações realizadas.

$$\text{Equação II: } ME = \frac{\sum_{i=1}^n e_i}{n}$$

- MAE. *Mean Absolute Error*, ou Erro Absoluto Médio, é o erro médio tomado em termos absolutos, para que um erro positivo não seja anulado por outro negativo.

$$\text{Equação III: } MAE = \frac{\sum_{i=1}^n |e_i|}{n}$$

- MSE. *Mean Square Error*, ou Erro Quadrado Médio, foi definido com a mesma finalidade de não anular os erros durante sua somatória.

$$\text{Equação IV: } ME = \frac{\sum_{i=1}^n e_i^2}{n}$$

- MAPE. *Mean Absolute Percentage Error*, ou Erro Percentual Absoluto Médio, é uma medida do erro absoluto médio em termos percentuais, para que se tenha uma visão do erro comparado com o valor previsto, e também para permitir comparações entre modelos que utilizam dados diferentes.

$$\text{Equação V: } MAPE = 100 \cdot \frac{\sum_{i=1}^n \left| \frac{X_i - F_i}{X_i} \right|}{n}$$

Outras medidas de acurácia podem ainda ser definidas, apesar de não serem tão comuns na prática, como:

- AE. *Absolute Error*, ou Erro Absoluto, é o erro total acumulado, somado de forma absoluta (com o uso do módulo do erro).
- SSE. *Sum of Squared Erros*, ou Soma Quadrada dos Erros, é o total quadrado dos erros verificados. Ou seja, representa o MSE sem a medida da média.
- PE. *Percentage Error*, ou Erro Percentual, é o erro total medido em porcentagem.
- MPE. *Mean Percentage Error*, ou Erro Percentual Médio, medida percentual do erro médio, ou seja, o PE medido em termo médio.
- APE. *Absolute Percentage Error*, ou Erro Percentual Absoluto, é o erro absoluto total medido em porcentagem. Em outras palavras, representa o MAPE sem a medida da média.
- RMSE. *Root Mean Square Error*, ou Raiz do Erro Quadrado Médio, tomado a partir da raiz do MSE, para que a medida volte à unidade original.
- SMAPE. *Symetric Mean Absolute Percentage Error*, ou Erro Percentual Absoluto Médio Simétrico, é uma adaptação do MAPE, evitando que grandes erros pontuais tenham peso exagerado na medida da acurácia. É recomendado quando existirem falhas ou picos repentinos na demanda.
- YTD MAPE. É a medida do MAPE para o último ano corrido, isto é, para as últimas 52 semanas ou 12 meses até o momento, dependendo da frequência tomada.
- CumRAE. *Cumulative Relative Absolute Error*, ou Erro Absoluto Relativo Acumulado, mede a capacidade preditiva de dois modelos, comparativamente. É definido como a relação entre os erros totais verificados por dois modelos.
- MdCumRAE. *Median Cumulative Relative Absolute Error*, ou Erro Absoluto Relativo Acumulado Mediano, é uma forma de usar o CumRAE na comparação de mais de dois modelos.

A escolha do critério a ser utilizado para medir a acurácia dos modelos não é arbitrária – ela deve ser feita baseada nas características do problema e das medidas citadas. Um dos métodos mais utilizados é o MSE, por ser um estimador não enviesado. Porém este – e também sua raiz, RMSE – são largamente afetados por *outliers*: grandes e possivelmente raros erros. Caso ocorram picos de erros, e o objetivo seja ter um modelo buscando acurácia ao longo do tempo, este critério deve ser desconsiderado. A medida do MAPE pode ser a melhor medida nestes casos, por ser computada por medidas absolutas (e não quadradas) e em porcentagem do valor previsto. Outra vantagem da MAPE é ter uma visão da amplitude do

erro: por exemplo, dizer que um modelo teve erro médio de 8% (MAPE) fornece mais informações do que dizer que seu MSE foi de 361, ou que seu MAE foi 147. No caso de se ter como objetivo um modelo de previsão que respeite limites de erro máximo, a medida do MAE pode ser a mais indicada, assim como as medidas de AE e APE (HARDIE et al., 1998).

Comparação e Seleção de Técnicas de Previsão

Não há, de maneira geral, uma conclusão uniforme na literatura sobre o desempenho comparativo das técnicas de previsão. Geralmente as técnicas são comparadas apenas com relação à sua capacidade preditiva, e as conclusões são díspares – não há consenso sobre qual técnica é a mais apropriada para cada tipo de problema. Porém, é fato que nenhuma técnica supera as demais em todas as aplicações. Algumas, entretanto, têm baixa performance sempre, como os métodos de suavização através de médias – estes são citados mais por questões históricas, e raramente empregados na prática, a não ser para comparar o ganho de acurácia na adoção de uma nova técnica de previsão (MAKRIDAKIS et al., 1983).

Outra conclusão que parece ser consenso na literatura é quanto à baixa acurácia das técnicas qualitativas (LUXHOJ et al., 1996; ALIBAIG & LILLY, 1999; SUBRAHMANYAN, 1998; SUBRAHMANYAN, 2000; LAWRENCE et al., 2000). Segundo MAKRIDAKIS et al. (1983), tais técnicas só apresentam resultados equivalentes às quantitativas no longo prazo (mais que dois anos), e são freqüentemente mais caras dado que demandam muito tempo dos analistas e da alta administração. Já as técnicas quantitativas podem ser aplicadas a qualquer horizonte de tempo, desde que os padrões não se modifiquem.

Uma importante ressalva deve ser feita quanto ao padrão dos dados: muitas técnicas, principalmente as de séries temporais, são aplicáveis apenas a alguns padrões, mostrando grandes erros se aplicadas em outros padrões. De fato, a maior parte das técnicas quantitativas, incluindo as técnicas causais, requer uma análise prévia do padrão geral dos dados, a fim de formular um modelo adequado.

Além das diferenças estruturais nos inúmeros problemas já estudados, outro fator que pode ter influenciado na disparidade das conclusões acerca do desempenho das técnicas é a qualidade (representatividade e quantidade suficiente) e natureza (seleção das variáveis) dos dados apresentados, independentemente do modelo utilizado. Os resultados dos estudos de CHURCH & CURRAM (1996) e GEURTS & WHITLARK (1996) confirmam que a qualidade dos dados é fator preponderante na acurácia das previsões.

A tabela a seguir demonstra a disparidade nos resultados alcançados por diversos autores a respeito da acurácia das diferentes técnicas de previsão, quando comparadas entre si para um mesmo conjunto de dados:

<i>Autor</i>	<i>Técnicas</i>	<i>Estudo</i>	<i>Conclusões</i>
(LAWRENCE et al., 2000)	Técnicas qualitativas X <i>naive</i> não ajustada	Compararam a acurácia dos métodos de previsão de 13 empresas de manufatura, todas utilizando métodos de previsão subjetivos. Os resultados foram comparados com uma previsão <i>naive</i> não ajustada.	Os resultados mostraram que as previsões qualitativas das empresas não foram consistentemente melhores que a (simples) técnica <i>naive</i> . Este resultado poderia ser ainda pior se tivesse sido utilizada a técnica <i>naive</i> ajustada, igualmente simples, mas que leva em consideração a sazonalidade.
(SPEDDING & CHAN, 2000)	Modelagem bayesiana X ARIMA	Desenvolveram modelos de previsão de vendas para controle de inventários. Um modelo bayesiano baseado em séries de tempo dinâmicas lineares é proposto para a previsão de demanda (agregada), e é comparado com uma modelagem ARIMA.	Os resultados obtidos indicam que a melhor técnica para previsão de demanda agregada é a modelagem bayesiana. O modelo bayesiano também apresentou melhor resultado com a inserção de maior número de séries temporais, indicando maior eficiência no aprendizado.
(BIDARKOTA, 1998)	Diferentes métodos auto-regressivos, uni e bivariados	Compara a performance de dois modelos de séries temporais na previsão de taxas de juros: um deles com o uso de uma série de dados composta por valores de taxas de juros reais, e outro composto por dados de juros nominais e inflação. Duas variações da modelagem ARIMA foram utilizadas.	O modelo bivariado mostrou maior acurácia para dados fora da amostra, apesar de ambos os modelos apresentarem grandes erros no longo prazo. Prova, até certo ponto, que prever através dos componentes teóricos da série (como uma técnica causal) tem resultado igual ou melhor que a previsão a partir da série temporal isolada.
(NEWBOLD & GRANGER, 1977)	Séries temporais e ARIMA	Compararam a acurácia de diversos modelos baseados nas técnicas de suavização exponencial (método de Holt), ARIMA (método Box-Jenkins) e regressão <i>stepwise</i> .	Os modelos gerados a partir da técnica Box-Jenkins (ARIMA) mostraram-se superior em 73% dos casos. Com o aumento do horizonte de tempo, os modelos ARIMA continuaram superiores, porém em apenas 57% dos casos.
(MAKRIDAKIS, 1982)	Séries temporais	Comparou a acurácia de três técnicas de séries temporais: suavização através de médias, suavização exponencial e ARIMA. Para tanto, utilizou 1001 séries de dados de vendas reais.	A suavização exponencial simples mostrou-se ser a técnica mais acurada, seguida da ARIMA e de técnicas de suavização exponencial mais complexas.

Tabela 2 – Exemplos de comparação de desempenho das técnicas de previsão na literatura

1.2. Técnicas de Previsão

1.2.1. Técnicas Qualitativas de Previsão

As técnicas de previsão qualitativas, também chamadas de técnicas subjetivas ou baseadas em critérios de juízo, são aquelas que utilizam primordialmente a capacidade humana de estabelecer generalizações e extrapolações. Estas técnicas pouco se utilizam – ou simplesmente não se utilizam – de séries de dados históricos, quantitativos.

Uma conclusão presente em vários estudos indica que, em geral, as empresas parecem estar mais à vontade com essas técnicas que com métodos quantitativos (WHEELWRIGHT & CLARKE, 1976; MCHUGH & SPARKES, 1983; MENTZER & COX, 1984; FILDES & HASTINGS, 1994). Contudo, um estudo recente conduzido por SANDERS & MANRODT (1994) sinaliza que a familiaridade com métodos quantitativos é crescente — em sua pesquisa 76% dos respondentes afirmaram possuir familiaridade com pelo menos uma técnica quantitativa para previsão de vendas, contra 61% obtidos no estudo de MENTZER & COX (1984).

Talvez essa familiaridade com métodos qualitativos seja devido ao fato de eles serem mais simples, e algumas vezes até intuitivos; fazendo com que sejam os métodos mais utilizados nas empresas (SANDERS & MANDRODT, 1994; PETERSON & JUN, 1999). De fato, facilidade de uso e capacidade de incorporar a experiência do gestor são os principais argumentos utilizados para justificar o uso intensivo de técnicas de julgamento subjetivo nas previsões, em detrimento de técnicas quantitativas (WINKLHOFFER et al., 1996).

Entretanto, diversos estudos mostraram que os métodos qualitativos oferecem baixa acurácia, e não raro, as previsões subjetivas geram grandes erros, ocasionando distúrbios no planejamento e resultado final das operações corporativas (LUXHOJ et al., 1996; ALIBAIG & LILLY, 1999; SUBRAHMANYAN, 1998; SUBRAHMANYAN, 2000; LAWRENCE et al., 2000).

LAWRENCE et al. (2000) sugerem alguns motivos pelos quais as previsões subjetivas, apesar de serem muitas vezes realizadas por analistas experientes e com informações contextuais do mercado, não possuem boa performance:

- As informações contextuais podem não ter valor preditivo.
- Enviesamento e ineficiência na interpretação dos relacionamentos podem mascarar as informações contextuais levadas em consideração.
- Excesso de dados a serem consideradas pelo analista, fazendo com que ele ignore ou dê pesos errados às informações contextuais recebidas.
- Perturbações recentes na série temporal podem ser entendidas pelo analista como um sinal de tendência, confundindo a interpretação das informações contextuais.
- Acurácia não é o critério mais importante nas previsões das empresas.

Dada a baixa capacidade preditiva das técnicas qualitativas, esta pesquisa está focada em métodos quantitativos de previsão, e, portanto, apenas apresenta brevemente algumas das técnicas qualitativas mais utilizadas, sem detalhar seu funcionamento.

Técnicas Exploratórias

Aqui estão classificadas todas as técnicas qualitativas de exploração de dados e desenvolvimento de cenários. Não há grande formalidade na aplicação destes métodos, por isso as variações são incontáveis.

Algumas das técnicas mais comuns e sistematizadas são:

- Pesquisas de mercado. Pesquisas e testes de mercado são instrumentos de previsão úteis especialmente quando dados históricos não estiverem disponíveis ou não forem confiáveis. O lançamento de um novo produto é o exemplo clássico. As desvantagens incluem baixa acurácia, alto custo e tempo dispensado.
- Método Pert-modificado. Método baseado em estimativas otimista, pessimista e realista dos acontecimentos futuros. Às estimativas são associadas probabilidades de acontecimento, que são posteriormente avaliadas. Fundamentado na Teoria das Decisões (teoria estatística para avaliação de probabilidades consecutivas), porém simplificado quanto à parte quantitativa.
- Teoria da utilidade. Refere-se à otimização de um parâmetro ou função utilidade, com base em probabilidades previstas por especialistas.
- Métodos de extrapolação de gráficos. A partir de gráficos relacionados à variável de interesse, são traçadas extrapolações à mão livre, com base no comportamento esperado da curva mais informações externas, como crescimentos ou retrações esperados no setor, informações da força de vendas, entre outras.

Técnicas de Grupo

Constituem técnicas de previsão onde as conclusões são essencialmente providas de análises de mais de um especialista ou analista. As contribuições podem provir de: executivos da empresa, equipe de vendas, clientes e outros especialistas da área. A maior parte desses

métodos utiliza técnicas estatísticas, com maior ou menor intensidade, para combinar as informações providas dos diversos especialistas.

As técnicas mais utilizadas são:

- Técnica Delphi. Baseada em questionários enviados para especialistas, que são revisados até que se chegue em consensos.
- Diagnóstico de expectativas. Baseado em pesquisas de opiniões de um júri executivo; utilizado principalmente para detecção de novas tendências, e para combinar opiniões sobre previsões quantitativas previamente calculadas.
- Composição da força de vendas. Uma fonte importante de dados qualitativos é a força de vendas de uma empresa. A percepção de futuro da equipe de vendas é de particular importância quando a demanda tiver um padrão com muitos picos e vales (MOON & MENTZER, 1999). Existem duas condições primárias onde o uso da força de vendas faz sentido na previsão de vendas: (1) os vendedores têm bom discernimento das mudanças no padrão da demanda e conhecimento profundo de seus clientes; e (2) os vendedores conhecem bem as probabilidades de garantir grandes vendas. A demanda de armamentos de guerra é um bom exemplo de situação onde a modelagem matemática é virtualmente impossível, mas a equipe de vendas tem boa expectativa do comportamento futuro da demanda.

1.2.2. Modelos Bayesianos

São modelos baseados na Teoria Bayesiana, combinando opiniões subjetivas com procedimentos quantitativos para revisar as probabilidades previstas – e por isso foram colocados em uma categoria própria, dentro da classificação aqui proposta.

O Teorema de Bayes deriva da Lei de Probabilidade Total, que afirma que dados dois eventos A e B, pode-se sempre dizer que a probabilidade de A é igual à probabilidade da intersecção de A e B mais a probabilidade da intersecção de A e o complemento de B. Ou seja: $P(A) = P(A \cap B) + P(A \cap \bar{B})$

Ou, expressando em termos de probabilidade condicional:

$$P(A) = P(A | B)P(B) + P(A | \bar{B})P(\bar{B})$$

O Teorema de Bayes permite reverter a condicionalidade dos eventos, ou, em outras palavras, obter a probabilidade de B *dado* A a partir da probabilidade de A *dado* B . Segundo a definição de probabilidade condicional: $P(B | A) = \frac{P(A | B)P(B)}{P(A)}$

O que, combinado com a Lei de Probabilidade Total, resulta em:

$$\text{Equação VI: } P(B | A) = \frac{P(A | B)P(B)}{P(A | B) + P(A | \overline{B})P(\overline{B})}$$

a qual é o Teorema de Bayes. Ele pode ser entendido como uma forma de transformar uma probabilidade *a priori* — $P(B)$ — em uma probabilidade *a posteriori* — $P(B|A)$. Convencionou-se, então, em denominar as abordagens que utilizam informação *a priori* sobre um problema particular, combinando-a com resultados estatísticos, de bayesianas.

Figura 7 - Modelo de inferência bayesiana

A informação *a priori* a ser utilizada pode ser um resultado estatístico anterior, ou, mais freqüentemente, reflexo do conhecimento do especialista ou analista.

Uma das grandes vantagens de se utilizar a abordagem bayesiana, além da possibilidade de melhorar a acurácia com o uso de informação *a priori*, é a análise de forma seqüencial que ela permite. Ou seja, informação obtida em uma análise pode ser utilizada como informação *a priori* em uma nova análise, quando novos dados estiverem disponíveis. Assim, a segunda amostragem é utilizada como dados, enquanto os resultados da primeira análise são utilizados como informação *a priori*. Esta metodologia é bastante utilizada em problemas de previsão: os dados atuais são combinados com a previsão anterior (esta última como informação *a priori*) para obter novas previsões. Também o conhecimento de especialistas pode ser utilizado como informação *a priori* nas previsões.

1.2.3. Técnicas Quantitativas de Previsão

Técnicas de previsão quantitativas são aquelas que usam dados históricos para calcular matematicamente extrapolações dos dados no futuro. A previsão com uso de técnicas quantitativas pode ser aplicada quando (MAKRIDAKIS et al., 1983):

1. Informação sobre o passado esteja disponível.
2. Esta informação possa ser quantificada em termos matemáticos.
3. Seja possível assumir que alguns aspectos do padrão verificado no passado continuarão no futuro. Esta colocação é também chamada de pressuposto da continuidade.

O procedimento geral para estimar um padrão de relacionamento, seja causal ou de série temporal, é através da aderência a uma forma funcional matemática qualquer, de forma a minimizar o componente de erro. Exceção deve ser feita à modelagem baseada em redes neurais artificiais, que pode ser usada tanto para modelos causais como de séries temporais, porém não buscam a aderência a um modelo matemático explícito (ANSUJ et al., 1996).

A previsão causal, também chamada de explanatória, assume um relacionamento de causa e efeito entre as entradas e saídas de um sistema. O sistema pode ser visto como a economia nacional, o mercado de uma empresa, e assim por diante. Já a previsão com uso de séries temporais trata o sistema como uma “caixa preta”, sem tentar descobrir os fatores que causam os comportamentos observados.

A seguir as principais técnicas quantitativas de previsão são abordadas.

Modelos de Escolha Discreta

Uma das técnicas possíveis para previsões quantitativas é a utilização de modelos de escolha discreta, os quais têm sido aplicados principalmente na análise de participação de mercado e também na área de transportes, para planejamento viário e de operações de transporte. Entretanto, estes são, a rigor, modelos de classificação, sendo que sua aplicabilidade para previsão de vendas é restrita. Por este motivo, esta modelagem não será abordada neste trabalho.

Análises de Séries Temporais

Uma série temporal é uma sequência de valores, ordenados no tempo, de uma variável de interesse particular (NEWBOLD, 1995). Modelos de séries temporais realizam previsões baseadas em uma série de dados observados em intervalos de tempo regulares, buscando padrões no passado para prever o futuro. Este tipo de modelagem é especialmente útil quando há pouco conhecimento da base teórica sobre o processo em que os dados foram gerados. As previsões são muitas vezes confiáveis no curto prazo, sendo a classe de técnicas quantitativas há mais tempo em desenvolvimento.

Segundo MAKRIDAKIS et al. (1983), existem duas razões básicas para tratar um sistema como uma caixa preta:

1. O sistema não pode ser compreendido, ou, mesmo que possa, é extremamente difícil medir o relacionamento entre as variáveis que governam seu comportamento.
2. A preocupação pode ser simplesmente prever com algum grau de precisão o que vai acontecer, e não por que.

É importante observar que as técnicas de análise de séries temporais desenvolveram-se de forma paralela à estatística convencional. Isto porque praticamente todas as técnicas estatísticas são baseadas no pressuposto de amostragem aleatória, ou seja, no pressuposto de que os dados disponíveis são observações independentes do fenômeno de interesse. Este pressuposto raramente é verdadeiro para séries temporais, demandando portanto o desenvolvimento de novas e específicas técnicas estatísticas.

Existem duas razões básicas para desconfiar do pressuposto de observações independentes em séries temporais (NEWBOLD, 1995):

1. Os fatores econômicos que contribuíram para a geração de um valor não mudam repentinamente, tendendo-se a manter um nível próximo em períodos adjacentes. Por exemplo, se as vendas de um mês foram altas, muito provavelmente as do mês subsequente também o serão, estando portanto os dados de vendas relacionados de alguma forma.
2. Há geralmente a presença de sazonalidade, ou seja, de padrões de longo prazo (usualmente anuais) constantes, repetitivos.

Assim, o pressuposto de amostragem aleatória não é válido para séries temporais, e o uso de técnicas convencionais leva a grandes erros neste tipo de análise. Pelo contrário, as

técnicas de previsão a partir de séries temporais valem-se justamente do fato de as observações serem bastante dependentes, possibilitando a inferência de valores futuros a partir de dados históricos.

Apesar de sua grande aceitação, as técnicas de séries temporais possuem sérias limitações. A mais visível delas é o fato de as causas que agem sobre as variáveis previstas serem completamente ignoradas. Todas as forças externas, como fatores econômicos, esforços de marketing, ações dos competidores, e assim por diante, são desprezadas. Outra deficiência é que os padrões históricos que geraram as séries mudam com o tempo, e as técnicas podem não detectar tais mudanças. Suas deficiências resultam em previsões com baixa acurácia, especialmente no longo prazo (GROSS & PETERSON, 1983).

Como nessa classe de modelagem a preocupação é apenas com o comportamento da variável de interesse no tempo – desprezando todos os fatores exógenos condicionantes de seu comportamento – as técnicas quantitativas de previsão mais simples, e que demandam menor esforço matemático-computacional, inserem-se nessa categoria. No entanto, dado o desenvolvimento constante das técnicas, algumas delas são de fato bastante complexas, e de implementação difícil.

Apesar de todos os problemas verificados com as técnicas de análise de séries temporais, elas continuam a ser extensivamente usadas e estudadas, e portanto são brevemente abordadas neste trabalho.

Métodos de Suavização Através de Médias

O nome desta classe de métodos – *smooth* em inglês – significa homogeneizar, tornar uniforme, aperfeiçoar, suavizar. E é justamente o que propiciam: uma suavização no gráfico da variável estudada, através de processos matemáticos relativamente simples.

Técnicas “Naive”

Talvez o mais simples método de suavização – e também de previsão – seja o chamado método *naive* (“ingênuo”) não ajustado, que consiste simplesmente em considerar a última observação conhecida da série como sua melhor expectativa para o futuro. Ou seja, parte da idéia de que amanhã será, com grande chance, como hoje. Por mais simplória que pareça, esta metodologia é freqüentemente usada para comparar o ganho em capacidade preditiva ao se adotar uma técnica mais elaborada; e já foi mostrado que sua aplicação oferece

acurácia igual ou significativamente melhor que técnicas qualitativas (LAWRENCE et al., 2000).

Uma forma ligeiramente melhorada de técnica *naive* é a chamada *naive* ajustada, onde também é utilizada a última observação para efeito de previsão, porém ela é antes desazonalizada: o efeito da sazonalidade é minimizado através de técnicas de decomposição – mostradas adiante – melhorando assim sua acurácia. Este método também é usado para fins de comparação entre capacidades preditivas.

Técnicas “Averaging”

São as técnicas usualmente referenciadas como de suavização através de médias – ou *averaging* – propriamente ditas, aplicando procedimentos matemáticos – especialmente o cálculo sucessivo de médias – para extrapolar a série temporal no futuro.

O mais simples método dessa classe, conhecido como *simple average* é o cálculo da média das últimas observações como uma aproximação de seu valor no futuro. De fato, se considerarmos um fenômeno que gere saídas constantes estacionárias, flutuando ao redor de um ponto, a média é uma boa previsão. Entretanto, se a série contiver tendências ou sazonalidades, a média deixará de ser efetiva. Mas outras técnicas utilizam mecanismos de suavização dos dados através de médias buscando incorporar tais características dos dados em seus resultados.

A mais conhecida delas – está presente em praticamente todos os livros de estatística aplicada – é a de médias flutuantes, ou em inglês *moving averages*. A técnica consiste em tomar uma média de um número fixo de observações que se move conforme se progride na série de dados. Assim, para ter a primeira observação em um *moving average* que utiliza a média de k observações é necessário passar pelos k primeiros números, realizando sua média. A segunda observação seria composta também pela média de k valores da série, mas iniciando pelo segundo número, e assim por diante.

Quanto maior o número de pontos tomados, mais estáveis as séries obtidas. Maior estabilidade também pode ser conseguida através da aplicação do algoritmo de *averaging* mais de uma vez, fazendo a média das médias. Surgiram assim as chamadas *double moving averaging* e *triple moving averaging*, com a aplicação das médias duas, três ou mais vezes.

As técnicas *averaging* não são muito utilizadas para previsão dada sua baixa acurácia. De fato, seu uso mais freqüente é o utilizado na desazonalização dos dados, ou seja, na

tentativa de minimizar os efeitos da sazonalidade em dados que servirão de entrada a outras técnicas.

Métodos de Suavização Exponencial

As técnicas de *moving averages* possuem alta estabilidade nas respostas fornecidas, característica esta que é bastante desejável num sistema de previsão, para superar o problema das grandes oscilações devido a alterações puramente aleatórias. Porém a taxa de resposta às variações é de difícil modificação — baseia-se inerentemente no número de observações tomadas. Outro problema grave dessas técnicas é sua baixa eficiência, isto é, baixo nível de “aprendizado”, ou correção dos erros com o tempo.

Na tentativa de sanar os problemas citados, foram desenvolvidas técnicas de suavização exponencial, que envolvem a aplicação de pesos distintos aos dados históricos, para dar maior importância às observações mais recentes. Nestas técnicas a taxa de resposta pode ser facilmente corrigida, dando maior flexibilidade aos modelos produzidos. Outra vantagem é que elas utilizam menor número de dados da série, e portanto a necessidade de registro histórico é menor.

A regra geral utilizada no desenvolvimento das técnicas de suavização exponencial é: para ter uma estimativa da demanda, adicione à estimativa do mês anterior uma fração do erro verificado. Ou seja:

Nova estimativa = antiga estimativa + α (demanda – antiga estimativa).

Rearranjando:

$$\text{Equação VII: } P_t = \alpha \cdot Q + (1 - \alpha) \cdot P_{t-1}$$

onde P é o valor previsto, Q é a observação da demanda atual e α é o fator de peso. Quanto maior o fator de peso, mais rápida a resposta das previsões às mudanças observadas; ao contrário, quanto menor o parâmetro α , mais estáveis serão as previsões.

As principais técnicas desenvolvidas com base na idéia de suavização exponencial são:

- Single exponential smoothing. Aplicação da Equação VII uma única vez para cada período previsto.

- Double exponential smooting. Aplicação da suavização exponencial duas vezes sucessivas, utilizando o resultado da primeira iteração como entrada na segunda. Pode utilizar o mesmo parâmetro (Método de Brown), usando, portanto, a mesma equação duas vezes; ou utilizar dois parâmetros distintos nas iterações (método de Holt).
- Triple exponential smooting. Aplicação da suavização exponencial por três vezes sucessivas. De forma semelhante ao *double exponential smooting*, pode-se utilizar um (método de Brown quadrático) ou três parâmetros distintos (método de Winter).
- Adaptative parameter exponential smoothing (ARRSES). É na verdade um caso típico de *single exponential smooting*, porém o fator de peso é definido matematicamente a partir dos valores de P e Q , e desta forma passa a ter certa liberdade para variar conforme se progride na série de dados. Esta característica aumenta a eficiência dos modelos, fazendo com que possam adaptar-se melhor aos dados e diminuir os erros com o passar do tempo.

Outras equações podem ser geradas a partir da equação geral de suavização exponencial (Equação VII), como aquelas escritas a partir do modelo de classificação de Pegel (MAKRIDAKIS et al., 1983). Tais extrapolações não são, entretanto, muito utilizadas na prática.

As maiores vantagens das técnicas de suavização exponencial são sua simplicidade e baixo custo de desenvolvimento — elas podem facilmente ser aplicadas através de planilhas de cálculo. Com dados estacionários, pode-se utilizar com sucesso as técnicas simples de suavização exponencial. O método de Brown é indicado para dados não-estacionários sem sazonalidade, principalmente por requerer apenas um parâmetro. O método de Brown quadrático também é bastante utilizado por sua simplicidade e boa capacidade de detectar pontos de inflexão na tendência. Seu ponto falho é sua rápida reação (baixa estabilidade), podendo indicar pontos de inflexão onde na verdade há apenas pequenas variações aleatórias. Para dados com sazonalidade, a única técnica largamente utilizada é a de Winter (MAKRIDAKIS et al., 1983). Estas técnicas são normalmente preferidas em detrimento das técnicas *averaging* para previsão, por oferecerem maior acurácia e necessitarem de menor número de dados históricos armazenados e computados.

Métodos de Decomposição

Ao contrário das técnicas de suavização, que tentam distinguir entre os padrões e os erros nos dados através de uma medida média de valores anteriores, as técnicas de decomposição buscam separar os componentes que possam estar presentes nos dados. Os componentes usualmente citados na literatura são:

- Tendência (T). Envolve a expectativa de crescimento ou decaimento em um período longo de tempo. Por exemplo, apesar de apresentar altos e baixos, o índice do Produto Interno Bruto (PIB) de um país desenvolvido tende a aumentar com o tempo.
- Sazonalidade (I). Flutuações periódicas de comprimento constante. Muitas variáveis econômicas possuem forte padrão anual, semestral ou trimestral, e outras variáveis podem ter sazonalidades quaisquer. Pode-se verificar facilmente um aquecimento da economia próximo ao fim do ano, por exemplo.
- Componente cíclico (C). Padrões oscilatórios, desconectados de comportamentos sazonais repetitivos. Não são necessariamente regulares, mas seguem um certo padrão ao longo do tempo, geralmente relacionado ao comportamento macroeconômico.
- Componente irregular (ϵ). Agrega a multitude de fatores que influenciam o comportamento de uma série real, e cujo padrão parece ser imprevisível.

Os modelos de decomposição supõem que se pode escrever o padrão dos dados como:

$$\text{Equação VIII: } Y = F(\text{tendência, ciclo, sazonalidade}) + \epsilon$$

Tal relação pode ser aditiva, do tipo $Y = (T + C + I) + \epsilon$

Ou multiplicativa, como $Y = TCI + \epsilon$

Todas as técnicas deste gênero, como a Decomposição Clássica, ou a Decomposição Censo II, envolvem a decomposição dos dados nos fatores acima (GROSS & PETERSON, 1983).

Métodos Auto-Regressivos (ARIMA)

Os métodos Auto-Regressivos (*Autoregressive Integrated Moving Average* – ARIMA) constituem uma classe genérica de métodos capazes de gerar modelos através da combinação

de três técnicas matemático-estatísticas: auto-regressão¹ (regressão com base no tempo), *moving averages* (para suavizar e sazonalizar), e diferenciação (para incluir processos não estacionários). BOX & JENKINS (1976) uniram à teoria de modelagem auto-regressiva a capacidade de tratar dados não-estacionários, através de um processo de diferenciação, criando a classe genérica de modelos ARIMA, que também passaram a ser conhecidos como modelos *Box-Jenkins*.

As equações utilizadas no método ARIMA são normalmente referenciadas por ARIMA (a,b,c), onde: a representa o grau de auto-regressão utilizado, b representa o grau de diferenciação, e c representa o grau de *moving average* usado. Desta forma, são possíveis inúmeras combinações, gerando um conjunto possível de equações que são escolhidas conforme as características dos dados².

A metodologia proposta por BOX & JENKINS (1976) define três passos para a construção do modelo de previsão:

1. Identificação de um ou mais modelos que descrevam a série temporal adequadamente. É feita utilizando ferramental estatístico, testando a estrutura de correlação da série, a partir da classe genérica ARIMA (a,b,c).
2. Estimação dos parâmetros dos modelos. É similar à estimação de parâmetros de modelos de regressão, porém com uso de técnicas de estimação não-lineares.
3. Condução do diagnóstico dos modelos, através da medição dos erros provocados por cada modelo, e seleção daquele com maior acurácia.

Apesar de seu apelo estatístico e teórico, estes modelos nunca foram muito utilizados na prática principalmente devido à sua complexidade matemática, que nem sempre se traduz em melhoria na acurácia (MAKRIDAKIS et al., 1983).

¹ Auto-regressão refere-se à utilização da técnica estatística de regressão sem o uso de variáveis exógenas, mas sim da própria variável dependente em diferentes períodos de tempo: $Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_k Y_{t-k} + e_t$. Assim, o pressuposto de independência dos termos de erro formulado na teoria estatística é imediatamente violado, por isso a obtenção de resultados satisfatórios não é garantida, e só é feita em termos empíricos.

² Há uma certa confusão ocasionada pela nomenclatura utilizada no método ARIMA. Duas são as possíveis ambigüidades: 1) o termo *Integrated* no nome da técnica na verdade refere-se ao processo de diferenciação da série de dados (e não integração); 2) o termo *moving averages* tem neste método um significado diferente daquele usado nos métodos de suavização – aqui ele refere-se à série de termos de erro em diferentes períodos de tempo considerados ($Y_t = a + b_1 e_{t-1} + b_2 e_{t-2} + \dots + b_k e_{t-k} + e_t$).

Técnicas Causais

As técnicas de previsão baseadas em análises de séries temporais são especialmente úteis quando há pouco conhecimento sobre a teoria envolvida no processo investigado. Por exemplo, se houver total desconhecimento dos fatores que influenciam a demanda de um determinado produto, ou se estes fatores não puderem ser mensurados, então o uso de uma técnica de análise de séries temporais pode ser a mais adequada.

Porém o uso dessa abordagem resulta em previsões não explicativas, e conseqüentemente o gestor não é capaz de realizar simulações sobre o comportamento da variável de interesse com mudanças em outros fatores, especialmente sobre aqueles aos quais ele tem controle. Por exemplo, a demanda de um produto é claramente influenciada pelo preço praticado, e este preço é um fator que está sob controle do gestor. Habilitar o gestor a simular o comportamento da demanda com mudanças no preço é fornecer a ele uma ferramenta de apoio à decisão, o que é um passo além de uma simples ferramenta de previsão de vendas.

As técnicas de previsão causais buscam descrever matematicamente as relações de causa e efeito entre a variável que está sendo medida e seus fatores constituintes (explicativos). Um modelo causal, como o descrito por uma equação de regressão, tem a vantagem de permitir ao tomador de decisões explorar como mudanças nas variáveis explicativas alteram a saída prevista.

Esta pesquisa está focada na geração de uma ferramenta de apoio à decisão no varejo, e, portanto, as técnicas de previsão causais – que habilitam a geração de cenários de vendas a partir de mudanças no composto de marketing – são aquelas de maior preocupação neste trabalho.

Modelos de Regressão

Freqüentemente, o relacionamento entre duas ou mais variáveis de interesse pode ser bem modelado através de uma equação matemática. O processo de adaptação de uma equação matemática a um fenômeno envolvendo duas ou mais variáveis, de forma que uma delas seja dependente das demais, é chamado de regressão. Ele gera modelos matemáticos que buscam representar a relação de causa e efeito entre as variáveis consideradas.

Se existir uma associação entre duas variáveis quaisquer, diz-se que há covariância positiva (ou negativa). Se não há associação entre as variáveis, a covariância é nula.

Entretanto, este valor depende da unidade de medida das variáveis. Para superar esta restrição, foi definida a correlação, que é uma medida pura, independente de escala. O coeficiente de correlação é definido como a covariância dividida pelo produto dos desvios padrões individuais:

$$\text{Equação IX: } \rho = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sqrt{E[(X - \mu_x)^2 (Y - \mu_y)^2]}}$$

onde: $-1 \leq \rho \leq 1$

Quanto maior o coeficiente de correlação (em módulo), mais forte a associação entre as variáveis. Todas as técnicas de regressão são baseadas na idéia de associação entre as variáveis, demonstrada estatisticamente através da correlação.

A técnica mais simples de regressão é a Regressão Linear, que modela o relacionamento entre uma variável dependente Y e uma variável independente X de forma linear. Assim:

$$\text{Equação X: } Y = \alpha + \beta X + \varepsilon$$

onde ε é uma variável aleatória com média zero.

Por exemplo, se demanda for a variável dependente, e ela tiver uma correlação linear com o preço, escreve-se:

$$Q = \alpha + \beta P + \varepsilon; \text{ onde } Q \text{ é a demanda e } P \text{ é o preço ofertado.}$$

Na verdade, a regressão linear é um caso específico da técnica geral de análise de Regressão Múltipla; que estabelece matematicamente a relação entre uma variável dependente e uma ou mais variáveis independentes. Assim:

$$\text{Equação XI: } Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

No caso de previsão de vendas, por exemplo, pode-se relacionar a demanda do produto (variável dependente) com a renda dos consumidores, o tamanho da população, o preço do produto, o preço de seus substitutos e complementares, o nível de promoção, o crescimento macroeconômico, entre outras.

Para que as equações possam ser efetivamente empregadas como modelos de previsão, é necessária a determinação dos parâmetros α e β_k . A principal técnica para determinação dos parâmetros de um modelo de regressão é a técnica de mínimos quadrados, que se baseia no

teorema de Gauss-Markov. NEWBOLD (1995) mostra que a técnica de mínimos quadrados é não-enviesada e ótima para modelos lineares, e muito bem aplicada para modelos de regressão múltipla. A qualidade de aderência do modelo aos dados é calculada em termo do erro verificado, que é minimizado através da técnica de mínimos quadrados.

Um termo relevante nas análises de regressão é o coeficiente de determinação (R^2), que denota a proporção de variabilidade da variável dependente explicada pela relação com as variáveis independentes. Em outras palavras, traduz o poder explanatório da regressão. Quanto maior o R^2 (assumindo sempre valores entre 0 e 1), diz-se que a curva está mais ajustada aos dados, e portanto o modelo representa bem os dados utilizados para construí-lo.

Comparativamente, os modelos de regressão requerem um esforço maior para construção do que as técnicas de séries temporais (excetuando-se possivelmente os modelos ARIMA). Além disso, eles requerem uma série de testes estatísticos para a seleção das variáveis de entrada mais relevantes, e exigem um conhecimento prévio da forma funcional do relacionamento entre as variáveis. Em condições normais, tal determinação a priori é difícil de ser obtida. Uma técnica usual é tentar diversas formas funcionais, acarretando em uma família de equações, optando-se pela equação com maior aderência aos dados. Mesmo neste caso, deve-se decidir as formas funcionais a serem consideradas.

Freqüentemente, os analistas assumem hipóteses simplificadoras de linearidade na estrutura dos dados, para que os modelos sejam construídos mais facilmente. Modelos lineares, entretanto, têm performance ruim na determinação de pontos de inflexão. Como os problemas em marketing sempre lidam com dados como vendas e preços, as séries de dados estão inseparavelmente ligadas a pontos de inflexão, tendências e não-linearidades, e muitas vezes podem ser até mesmo caóticas (THALL, 1992).

MAKRIDAKIS et al. (1983) desenvolveram uma metodologia para a escolha das variáveis a serem utilizadas nos modelos de regressão múltipla. Ela envolve:

1. Determinação de uma “lista longa” de variáveis. Uma lista com todas as variáveis potenciais que possam ter algum efeito sobre a variável dependente analisada. Baseada na opinião de especialistas, na disponibilidade dos dados e no esforço e custo associado à aquisição dos dados.
2. Redução para uma “lista curta”. Vários métodos podem ser usados na eliminação de algumas das possíveis variáveis da lista longa. Os mais simples (e menos recomendados) são: a) plotar cada variável contra Y, observando se há alguma relação

visível; b) observar as correlações entre todas as variáveis independentes potenciais, eliminando as de grande correlação para evitar multicolineariedade; e c) efetuar uma regressão múltipla com todas as variáveis, e eliminar aquelas com baixo t . Os métodos mais sofisticados (e recomendados pelos autores) são: a) regressão *stepwise* (método iterativo de retirada de variáveis e testes de significância); b) análise de componentes principais; e c) fazer uma análise horizontal no tempo (*lag analysis*).

Já (ACZEL, 1993) sugere as seguintes abordagens:

- ❑ Testar todas as regressões possíveis. Assim, tendo k variáveis independentes, tem-se 2^k possibilidades, já que cada variável pode estar ou não incluída no modelo. Aplica-se então um critério de performance para seleção da melhor opção, que pode ser maior R^2 ou menor MSE (Erro Quadrado Médio).
- ❑ Seleção “para frente”. Começa com um modelo sem variáveis, testando uma de cada vez e selecionando aquele com maior F estatístico (da análise de componentes principais). O processo recomeça, adicionando mais uma variável, e mantendo aquela pré-selecionada.
- ❑ Eliminação “para trás”. Funciona de maneira oposta à seleção para frente: começa com um modelo com todas as possíveis k variáveis e é retirada aquela com menor F , recomeçando o processo.
- ❑ Regressão *stepwise*. É na verdade uma mistura da seleção para frente e eliminação para trás, que reavalia o nível de significância de cada variável a cada passada — ao contrário dos demais que, uma vez tomada a decisão de incluir ou excluir uma determinada variável, ela não mais é revista.

Apesar das dificuldades na seleção das variáveis e construção dos modelos, HUSSEY & HOOLEY (1995) mostraram que a análise de regressão é a técnica estatística mais utilizada nas empresas européias, sendo utilizada por cerca de 18% dos respondentes. A popularidade das técnicas de regressão deve-se a pelo menos dois fatores, segundo GROSS & PETERSON (1983):

1. São de fácil compreensão, mesmo para aqueles sem profundos conhecimentos matemáticos ou estatísticos.
2. Resultam em boa acurácia para a maior parte das aplicações.

Certamente, os modelos de regressão são os que oferecem maior capacidade explicativa para as variações nos dados, já que é possível enxergar a relação matemática entre as variáveis explicitamente. Mas, de acordo com ZHOU (1999), eles nem sempre oferecem as melhores previsões. A maior parte das limitações das regressões está relacionada aos pressupostos que foram assumidos para a formulação das equações e determinação dos parâmetros ideais. Os pressupostos assumidos ao se desenvolver um modelo de regressão qualquer, com n pares de observações disponíveis, são (NEWBOLD, 1995):

1. Os valores de x_i são números fixos (determinados, por exemplo, por um experimento) ou realizações de variáveis aleatórias X_i , que são independentes do erro ε .
2. Os termos de erro ε são variáveis aleatórias com média zero.
3. As variáveis ε têm a mesma variância.
4. As variáveis ε não estão correlacionadas uma com a outra.
5. Não é possível determinar um grupo de números $c_0 \dots c_k$ de forma que:

$$c + c_1x_1 + c_2x_2 + \dots + c_kx_k = 0$$

(ou seja, as variáveis independentes não estão correlacionadas).

As principais limitações da análise de regressão são decorrentes dos pressupostos acima, e podem ser listadas como sendo:

1. Relacionamento constante. As técnicas de regressão assumem um relacionamento constante entre as variáveis, durante o período histórico compreendido pelos dados utilizados na criação do modelo. Ou seja, todos as observações possuem o mesmo peso durante a regressão. Isto pode ser um problema sério ao se utilizar grande horizonte de dados históricos, ou ao analisar mercados altamente dinâmicos.
2. Autocorrelação. Refere-se à condição onde o valor da variável prevista está relacionado ao seu valor no período anterior. A análise de regressão assume valores aleatórios, e não ordenados (correlacionados) das variáveis previstas. Este problema gera previsões enviesadas – sub ou sobre estimadas constantemente.
3. Multicolinearidade. Presente quando uma ou mais, e até mesmo todas as variáveis independentes estão relacionadas entre si. A existência desta condição teoricamente invalida as medidas de correlação e o procedimento de *stepwise*. Se existir multicolineariedade perfeita numa regressão, a solução ótima baseada no método dos Mínimos Quadrados não pode ser encontrada. No caso de multicolineariedades quase

perfeitas, esta solução será afetada por problemas de arredondamento nos computadores.

4. Causalidade. A regressão assume uma relação de causalidade entre a variável dependente e as variáveis independentes. Porém esta relação não pode ser tida como certa. Pode-se afirmar que duas variáveis estão estatisticamente correlacionadas (a mudança em uma é acompanhada de uma mudança proporcional na segunda), porém nada se pode afirmar em respeito à relação de causalidade entre as duas.
5. Linearidade. Refere-se à linearidade dos coeficientes, e relata diretamente ao desenvolvimento dos testes F e T. Em outras palavras, se for violada os testes estatísticos não são mais válidos.
6. Independência dos residuais. Também se reflete nos testes F e T, tornando os resultados destes testes não válidos. O teste Durbin-Watson é uma forma de examinar se este pressuposto está sendo respeitado.
7. Homocedasticidade. Refere-se à variância dos erros, que deve ser constante. Mais uma vez, o impacto está na validade dos testes F e T.
8. Normalidade dos residuais. Se há forte violação da normalidade nos erros, os testes estatísticos não são recomendados.
9. Intervalo de confiança. Há um intervalo de confiança para o qual o modelo é aplicável, que pode ser determinado matematicamente. O desrespeito ao intervalo de confiança gera previsões errôneas.

Modelos Econométricos

Assim como a regressão linear simples é um caso especial da regressão múltipla, esta última pode ser vista como um caso especial da modelagem econométrica. Os modelos econométricos envolvem várias equações de regressão, com mais de uma variável dependente. Vale ressaltar que alguns autores usam o termo “econométrico” para qualquer modelo de regressão, seja simples, múltiplo ou de várias equações. Na verdade, as técnicas econométricas surgiram efetivamente para prever índices econômicos, como o próprio nome diz, e foram posteriormente adaptadas para previsões microeconômicas.

A análise de regressão assume que cada uma das variáveis independentes seja determinada por fatores externos, exógenos ao sistema. Este pressuposto nem sempre é

verdadeiro na análise macroeconômica, onde as variáveis exibem comportamentos interdependentes. A maior vantagem dos modelos econométricos é sua habilidade em lidar com interdependências, ou seja, em levar em consideração variáveis com efeito mútuo, onde não é possível determinar uma relação de causa e efeito clara. Sua principal desvantagem é a inexistência de um conjunto de regras que possam ser aplicadas em situações distintas, o que torna o desenvolvimento dos modelos altamente dependente do problema específico, e requer sempre a presença de um analista experiente.

Por exemplo, se um governo quisesse saber qual o impacto de uma redução de impostos em uma economia em recessão, uma modelagem econométrica seria bem apropriada. Tal redução afetaria a renda das pessoas e empresas, e influenciaria toda a economia, incluindo nível de preços, desemprego, capital circulante aplicado, e assim por diante. Todos estes fatores são inter-relacionados, e uma equação de regressão múltipla não seria capaz de modelar esse fenômeno eficazmente — tal problema seria mais bem expresso com o uso de um conjunto de equações econométricas.

1.3. *Previsão de Vendas no Varejo*

PARENTE (2000, pág. 22) define varejo da seguinte forma:

“Varejo consiste em todas as atividades que englobam o processo de venda de produtos e serviços para atender a uma necessidade pessoal do consumidor final”. O varejista desempenha o elo de ligação entre o consumidor e o produtor ou atacadista.

O setor varejista vem atravessando intensas transformações ao longo das últimas décadas, assumindo importância relativa crescente no Brasil e no Mundo (PARENTE, 2000). Uma das transformações mais visíveis é o aumento do poder do varejo observado no Mundo Ocidental. Viu-se surgirem grandes conglomerados varejistas, muitas vezes com atuação multinacional. Prova disso é o número crescente de varejistas que figuram na relação de maiores empresas do Brasil. Tal fenômeno mudou a relação de poder fornecedor-varejista, fazendo com que o varejista ganhe poder de barganha. Grandes cadeias de varejo têm hoje força suficiente para impor condições de fornecimento (quantidades e frequências), forma de abastecimento e pagamento, e até mesmo procedimentos de gestão logística.

O surgimento e fortalecimento das marcas próprias – produtos comercializados com a marca do varejista, exclusivamente em suas lojas – é outro exemplo da mudança no relacionamento varejo-indústria. Além disso, é crescente a polarização entre grandes e

pequenos varejistas. Os grandes varejistas assumiram uma posição de varejo de massa, que operam em grande parte do território nacional, desenvolvem sofisticados sistemas logísticos e de informações de marketing, e exercem forte poder de barganha junto aos fornecedores. Já os pequenos varejistas, que não detêm o poder de barganha dos grandes, não são capazes de oferecer preços tão competitivos, e para sobreviver atuam em mercados restritos ou optam por uma estratégia de especialização, aprofundando-se no atendimento de necessidades específicas em certos segmentos de mercado.

O aumento da concorrência no varejo acompanhou aquele verificado em outros setores, porém com uma característica peculiar: ele agora se dá não somente entre empresas que operam no mesmo formato, mas também entre diferentes tipos de varejistas. Ou seja, diferentes tipos de varejo vendem as mesmas categorias de produto, aumentando as opções de compra do consumidor final. As padarias, por exemplo, hoje têm como concorrentes não só outras padarias, mas também hipermercados, supermercados e lojas de conveniência.

Finalmente, o surgimento e ampliação das vendas realizadas sem o uso de lojas mudaram a realidade do varejo no mundo todo. Vendas através da televisão, por catálogo, através de máquinas automáticas e principalmente pela Internet registram contínua expansão, e oferecem desafios crescentes de planejamento e execução aos gestores do varejo.

1.3.1. Composto de Marketing

Em qualquer mercado, a população desenvolve padrões de consumo que se refletem em gastos realizados nos diversos setores varejistas. Estes gastos representam a demanda de mercado, que é dividida entre todos os varejistas que atuam no setor. Conseqüentemente, cada varejista é detentor de uma parcela da demanda total do mercado; esta parcela é comumente referida como sendo sua fatia de mercado.

PARENTE (2000, pág. 61) define composto de marketing como sendo “o conjunto de variáveis que compõem o esforço de marketing do varejista, e engloba todos os fatores controláveis que o varejista articula (...) para conquistar as preferências dos consumidores”. Ainda de acordo com o autor, a fatia de mercado de uma loja depende principalmente das características de seu composto de marketing, frente ao composto de marketing de seus concorrentes.

Segundo MASON et al. (1993), o composto de marketing do varejo pode ser classificado a partir de seis macro-fatores (chamados “seis P’s”):

1. Conjunto (*mix*) de **Produtos** – linhas de produtos comercializados, caracterizados por sua variedade, qualidade e quantidade.
2. **Preços** – preços de venda e créditos oferecidos.
3. **Promoção** – nível de propaganda, ofertas, sinalização da loja, programas de fidelização entre outros.
4. **Apresentação** – Layout da loja, disposição dos produtos, espaço físico, limpeza da loja, conforto e decoração.
5. **Pessoal** – qualidade do atendimento, rapidez, cortesia, serviços.
6. **Ponto** – localização da loja, facilidade de acesso, visibilidade, estacionamento, complementaridade com outras lojas.

Ainda segundo o autor, o comportamento de compra do consumidor perante uma loja é influenciado basicamente por três conjuntos de variáveis: variáveis ambientais (economia, clima, demografia, época do ano, tendências de mercado), o composto de marketing da loja e o composto de marketing da concorrência. Influenciado por estes três grupos de forças, o consumidor decide o quanto gastar no setor varejista e em cada loja especificamente.

A decisão final do consumidor é tomada com base na relação custo-benefício percebida para cada loja, definida por KOTLER (1999) como *valor*:

“Os clientes são maximizadores de valor, e agem dentro de seu repertório de alternativas, limitados pela sua mobilidade, conhecimento, custos e renda. Eles formam uma expectativa de valor que determina seu comportamento. O grau em que essas expectativas são atendidas define tanto sua satisfação como sua possibilidade de recompra”.

Figura 8 – Decisão de compra baseada nos conceitos de valor e compostos de marketing

A Figura 8 ilustra as forças e variáveis que influenciam a decisão de compra do consumidor, enquanto maximizador de valor.

Alguns modelos de comportamento de compra foram desenvolvidos para tentar explicar de forma simplificada o processo de aquisição de uma mercadoria e os fatores que o influenciam. O modelo mais citado tem sido aquele desenvolvido por ENGEL et al. (1993), que identifica cinco estágios no processo de compra:

1. Reconhecimento do problema. O consumidor identifica uma necessidade não atendida, e reconhece que poderá atendê-la através da compra de um produto ou serviço.
2. Busca de informação. Pesquisa de alternativas de compra, pesquisas de preço e busca de informações sobre os produtos e serviços ofertados. Este processo varia conforme o consumidor e o tipo de produto adquirido (conveniência, compra comparada ou especialidade).
3. Análise das alternativas. Os consumidores comparam e avaliam as alternativas para tomar sua decisão de compra. Os atributos utilizados para a tomada de decisão também variam conforme o consumidor e o tipo de produto, mas podem ser relacionados ao composto de marketing (seis P's).
4. Decisão de compra. Finalmente, após analisar as alternativas consideradas para compra, o consumidor decide o que e onde comprar. Além das variáveis do composto de marketing e ambientais, outros fatores podem influenciar o processo de compra, como a fidelidade a determinadas marcas, tempo e recursos disponíveis para compra, e os pesos dados aos atributos dos produtos e das lojas consideradas.
5. Avaliação pós-compra. Após a compra, o consumidor reavalia o acerto de sua decisão, e desenvolve sentimentos de satisfação ou insatisfação.

Dentro da classificação proposta por MASON et al. (1993), este trabalho está focado na exploração do impacto na demanda proporcionado por alterações nos três primeiros componentes do composto de marketing: produto, preço e promoção. Esta escolha — de restringir as análises a variáveis pertencentes apenas a estes componentes — deveu-se basicamente à dificuldade observada ao se tentar quantificar e comparar as variáveis dos demais componentes: apresentação, pessoal e ponto. Tal tentativa de quantificação — necessária para fins de modelagem para mineração de dados — fugiria do escopo deste trabalho.

A seguir são tecidos alguns comentários a respeito dos três primeiros componentes de marketing citados acima, já que são os componentes mais facilmente computados de forma quantitativa, e que podem, portanto, ser incorporados nos modelos de previsão de vendas baseados em técnicas quantitativas.

1.3.2. Conjunto de Produtos Oferecidos

A composição do conjunto de produtos oferecido — o *mix* de produtos — é uma das tarefas mais importantes de uma empresa de varejo, já que determina o ramo ou setor de atividade da empresa, e quais necessidades de seus consumidores estará atendendo. Na composição do conjunto de produtos, o varejista enfrenta dois objetivos que podem ser conflitantes. Por um lado, procura atender ao máximo as diferentes necessidades de seus (vários) segmentos de consumidores, ampliando assim seu leque de produtos. Por outro lado, dispor de uma variedade muito ampla de linhas de produtos acarreta alguns problemas como: maior investimento em estoques, incorporação de produtos com baixo giro de estoque, maior custo operacional com atividades como armazenagem, controle de estoque e compras, enfim, menor retorno sobre o investimento.

O número de itens (produtos distintos) comercializados por lojas de diferentes ramos pode ser muito variado. Normalmente varia entre aproximadamente mil — lojas de conveniência, por exemplo — a 50 mil itens ou mais — grandes hipermercados (PARENTE, 2000). Para entender e gerenciar melhor tal universo de produtos, os profissionais do varejo geralmente utilizam a subdivisão dos produtos em classificações hierárquicas (LEWISON, 1997). Não existe, entretanto, consenso sobre o número de níveis a serem utilizados e nem mesmo sobre a estrutura de classificação de produtos. Cada empresa busca classificar seus produtos conforme mais lhe convém, dependendo do foco de mercado, produtos oferecidos e grau de especialização. Vários autores sugerem níveis hierárquicos semelhantes, como: departamento, seção, categoria, subcategoria, segmento, subsegmento, linhas de produtos entre outros.

A subdivisão dos produtos em níveis hierárquicos reflete, em última análise, as diferentes necessidades dos consumidores, ou seja, seus diversos interesses. Por este motivo, os profissionais do setor costumam referir-se a seus principais grupos de produtos pela expressão “centro de interesse”. Pode-se definir um centro de interesse como sendo um grupo de produtos que mantém uma sinergia entre si, do ponto de vista do consumidor. Em outras

palavras, são produtos que orbitam ao redor de algum interesse específico do consumidor. Por definição, todo produto está necessariamente dentro de um ou mais centros de interesse, e, existindo um mínimo de dois produtos em um centro de interesse, eles são necessariamente relacionados de alguma forma.

A problemática existente na determinação do conjunto de produtos a ser oferecido reflete as duas dimensões fundamentais que as norteiam: 1) a amplitude ou extensão, que se refere à quantidade de centros de interesse distintos atendidos pela loja, e 2) a profundidade, que se refere ao número de produtos diferentes existentes dentro de cada centro de interesse: quanto maior a profundidade, maior a sortidade de produtos disponíveis. Tomando por exemplo o caso do centro de interesse que inclui queijos e derivados, uma loja especializada em queijos possui conjunto de produtos pouco extenso (pode atender somente a este centro de interesse), porém aprofundado, já que dispõe de tipos variados de queijos. Já um supermercado tem a característica de ser generalista, ou seja, com conjunto de produtos extenso, mas pouco profundo — sua seção de queijos será provavelmente menor que a loja citada, porém ele atende a muitos outros centros de interesse. Finalmente, um hipermercado pode ter conjunto de produtos ao mesmo tempo extenso e profundo, e no exemplo citado oferecer grande variedade de queijos, além de muitos outros grupos de produtos.

A figura a seguir exemplifica alguns tipos de varejo de acordo com a amplitude e profundidade do conjunto de produtos usualmente ofertado.

Figura 9 – Classificação dos tipos de varejo segundo sua amplitude e profundidade (PARENTE, 2000)

A literatura sobre varejo no Brasil costuma referir-se ao conceito de centro de interesse como sendo um dos níveis hierárquicos da classificação de produtos: a categoria (PARENTE, 2000; PARENTE, 1997). PARENTE (1997, pág 230) define gerenciamento de categorias como “o processo de administrar categorias como unidades estratégicas de negócio, visando obter não só uma melhor satisfação do consumidor, mas também melhores indicadores de desempenho e lucratividade”. Para um gerenciamento por categorias eficiente, uma hierarquia adequada deve ser criada para que cada produto seja devidamente classificado.

Em geral, existe uma relação direta entre a estrutura de classificação dos produtos e sua apresentação na loja: o *layout* da loja geralmente reflete perfeitamente a classificação dos produtos, agrupando-os sempre por categorias com o objetivo de estimular a compra de produtos complementares e facilitar a escolha dos substitutos (PARENTE, 2000).

As categorias exercem diferentes papéis para os varejistas. Fatores como localização, segmentos de mercado atingidos, perfis dos consumidores, concorrência, política de preços, amplitude e profundidade do conjunto de produtos, entre outros, determinam a importância relativa de cada categoria para uma determinada loja.

A importância de cada categoria deve ser definida de acordo com sua capacidade de atrair clientes para a loja e de definir a imagem do varejista (PARENTE, 1997). Uma classificação usual dos papéis de cada categoria segundo esta definição de importância é:

1. Categorias de destino. São os centros de interesse que definem o foco da loja, contendo os grupos de produtos de maior atração de clientes para a loja.
2. Categorias de rotina. Apesar de não serem o foco principal da loja, possuem grande poder de atração de consumidores, que desenvolvem a rotina de comprar tais produtos na loja.
3. Categorias ocasionais. Possuem baixo poder de atração de clientes, que compram os produtos destas categorias apenas ocasionalmente.
4. Categorias de conveniência. São as categorias de menor poder de atração de clientes. Normalmente contém produtos que os consumidores sequer esperavam encontrar na loja, e que podem eventualmente adquiri-los apenas por conveniência.

Naturalmente, o gestor deve tratar os produtos de categorias diferentes de forma distinta, dado que sua importância relativa para a loja varia. PARENTE (2000) sugere que a amplitude e a profundidade a serem adotadas para cada categoria dentro de uma loja específica varie conforme sua classificação: categorias de destino, por terem uma atratividade de clientes para a loja maior, devem ter maior amplitude e profundidade, e assim sucessivamente, conforme ilustra a figura a seguir.

Figura 10 – Amplitude, profundidade e atratividade das diferentes categorias (PARENTE, 2000)

1.3.3. Gestão de Preços no Varejo

De todas as variáveis do composto de marketing, a decisão de preço é aquela que mais rapidamente afeta a competitividade, o volume de vendas, as margens e a lucratividade das empresas varejistas (PARENTE, 2000). As políticas de preço podem ser alteradas em curtíssimo prazo, e são imediatamente percebidas pelos consumidores; isto é, os consumidores demonstram ter bastante sensibilidade às mudanças de preço.

Conforme PARENTE (2000), o varejista normalmente segue uma política de preço mais ou menos uniforme para todos os produtos da loja. Esta política pode ser de 1) preços acima do mercado; 2) preços médios de mercado; e 3) preços abaixo do mercado.

Na política de preços acima do mercado, o varejista declara explicitamente sua intenção de não concorrer com base em preço, procurando maximizar os lucros não através de

alto volume de vendas, mas sim por meio da elevada margem bruta por produto que pratica. Esta política é normalmente adotada quando:

- A concorrência é pouco intensa.
- Os consumidores são menos sensíveis a preço, dando maior valor aos benefícios e serviços prestados;
- O varejista consegue diferenciar-se no mercado através de um ou mais componentes de marketing (ponto, serviços, produtos, etc.) superior.

A política de preços médios de mercado é a adotada pela maioria dos varejistas, fazendo com que o preço não constitua vantagem nem desvantagem competitiva.

Já com a política de preços abaixo do mercado os varejistas pretendem ter no preço sua principal arma competitiva. Em geral, esta política está associada a uma forte pressão para baixar os custos, que inclui a oferta de menos benefícios aos clientes. Para ter sucesso adotando-se esta política de preços, o varejista precisa ter alto volume de vendas e giro rápido dos estoques.

Um dos conceitos mais importantes com relação à definição dos preços de venda dos produtos é o de elasticidade a preços. A elasticidade a preços, ou simplesmente elasticidade, reflete a sensibilidade dos clientes às alterações de preço. Ou seja, é a relação entre a variação da quantidade demandada e a variação do preço:

$$\text{Elasticidade} = \frac{\Delta Q}{\Delta P} \quad (\text{Equação XII})$$

Situações de alta elasticidade normalmente ocorrem quando há forte substitutibilidade, ou seja, quando existem no mercado muitos produtos substitutos para atender à mesma necessidade. Assim por exemplo se existirem muitas marcas de cerveja, o aumento de preço de apenas uma delas deve ocasionar queda aguda em sua demanda. O mesmo ocorre no caso de alta substitutibilidade de lojas: se muitos estabelecimentos concorrem nas mesmas categorias de produtos, a variação de preço de apenas um deles deve ocasionar forte variação em sua demanda. Produtos que não são vistos como sendo de primeira necessidade (considerados supérfluos) também têm forte elasticidade.

A situação oposta – de baixa elasticidade ou inelasticidade – ocorre basicamente em duas situações: quando há urgência em adquirir o produto (bens de primeira necessidade, por exemplo), ou quando há baixo grau de substitutibilidade. Em situações de monopólio, ou

quando todas as alternativas de compra sofrem reajustes simultâneos (como no caso de aumentos no preço de combustíveis, por exemplo), verifica-se pouca alteração na demanda.

Elasticidades cruzadas

A elasticidade cruzada reflete como a variação do preço de um produto afeta o comportamento de outros produtos (PARENTE, 2000). É portanto a relação entre a variação da quantidade demandada do produto *A* e a variação do preço do produto *B* (correlação preço-demanda):

$$\text{Elasticidade cruzada} = \frac{\Delta Q_a}{\Delta P_b} \text{ (Equação XIII)}$$

Este conceito reflete como variações no preço de um produto afetam a demanda de outros produtos. Dependendo do resultado da elasticidade cruzada, pode-se identificar os seguintes tipos de relação entre dois produtos ou categorias:

1. Produtos complementares. Apresentam elasticidade cruzada negativa, isto é, quando um aumento no preço do produto A afeta negativamente as vendas do produto B, e vice-versa. Por exemplo, aumentos no preço de bebidas refrigerantes e/ou cerveja acarretam uma diminuição na demanda de aperitivos.
2. Produtos substitutos. Apresentam elasticidade cruzada positiva, ou seja, uma variação no preço do produto A acarreta variação no mesmo sentido na demanda do produto B. Utilizando exemplo semelhante: aumento no preço de Coca-cola causa aumento na demanda por Guaraná.

Os conceitos de elasticidade cruzada, produtos substitutos e produtos complementares estão intimamente ligados com aquele de centros de interesse, ou categorias de produtos. De fato, um centro de interesse é, de acordo com a definição adotada neste texto (pág. 50), um grupo de produtos que estão de alguma forma relacionados; este relacionamento nada mais é do que uma elasticidade cruzada. Portanto, havendo relacionamento na forma de elasticidade cruzada entre um grupo de produtos, conclui-se que tais produtos formam um centro de interesse, e que dois produtos quaisquer desse grupo devem ser, necessariamente, complementares ou substitutos. Em outras palavras, deve existir sempre uma correlação entre o preço e a demanda de dois produtos pertencentes à mesma categoria; se esta correlação for

negativa, os produtos são complementares; sendo a correlação positiva, os produtos são substitutos.

Na Figura 11, é mostrado um exemplo de relacionamento entre produtos substitutos e complementares de dois centros de interesse. Os produtos substitutos são considerados pelo consumidor como sendo altamente semelhantes, já que atendem a uma mesma necessidade. Por isso, podem ser vistos como constituintes de grupos uniformes de produtos (S1, S2, S3 e S4, na figura). Cada centro de interesse possui um ou mais grupos de substitutos; produtos que pertencem a um mesmo centro de interesse mas não fazem parte do mesmo grupo de substitutos são, necessariamente, complementares (P1 e P3 na figura, por exemplo).

Figura 11 – Relacionamentos entre produtos substitutos e complementares de um ou mais centros de interesse

Como ilustrado na figura, alguns produtos podem fazer parte de mais de um centro de interesse simultaneamente, já que atendem a necessidades distintas para o consumidor (P6, P7 e P8). Entretanto, tais produtos continuam a formar um grupo coeso de substitutos (S3). Uma cadeira de pano, por exemplo, pode servir como cadeira de praia ou cadeira de pesca, dois interesses bem distintos. Para cada um desses centros de interesse, os produtos complementares são diferentes: por exemplo guarda-sol no primeiro e vara de pescar no segundo. Todas as cadeiras de pano vistas como equivalentes para o consumidor, contudo, formam um mesmo grupo de produtos substitutos.

1.3.4. Propaganda e Promoção

O composto promocional é o elemento que o varejista utiliza para atrair os clientes para a loja e motivá-los para a compra (PARENTE, 2000). O composto promocional deve estar integrado às demais decisões acerca do composto de marketing, tendo como objetivos básicos informar (fornecendo dados sobre a loja, produtos e serviços oferecidos), persuadir (influenciando o comportamento de compra do consumidor) e relembrar a audiência-alvo, firmando marcas e qualidades (LEWISON, 1997).

A literatura existente sobre este tema é vasta, sendo usual encontrar diversas subdivisões para o composto promocional. Uma das divisões mais aceita é aquela proposta por KOTLER (1999), que divide as ações possíveis em propaganda, promoção de vendas e relações públicas.

KOTLER (1999) define propaganda como “qualquer forma paga de apresentação e promoção de idéias, bens ou serviços por um patrocinador identificado”. Segundo o autor, a propaganda pode ser uma forma efetiva de transmitir mensagens, e para tanto é necessário que se siga uma metodologia coerente na aplicação dos recursos em propaganda, que deve passar por:

- 1) Definição dos objetivos da propaganda.
- 2) Decisão sobre o orçamento a ser empregado.
- 3) Definição da mensagem e do(s) meio(s) a ser(em) utilizado(s).
- 4) Medição do impacto da propaganda.

Já a promoção de vendas pode ser definida como uma “ferramenta de comunicação que oferece um valor extra ao consumidor, que pode ou não envolver o uso de mídia” (PARENTE, 2000). Em outras palavras, enquanto a propaganda justifica *porque* comprar, a promoção de vendas *incentiva* a compra. De acordo com KOTLER (1999), a promoção de vendas pode consistir em: promoção ao consumidor (amostras, descontos, prêmios, garantias, entre outros); promoção mercadológica (distribuição de bens, combinação com propaganda, entre outros); e promoção de negócios e força de vendas (feiras e eventos, premiações a representantes de vendas, entre outros). Dentre os três componentes da propaganda e promoção, a promoção de vendas vem crescendo em participação relativa dos recursos empregados – representando cerca de 65-75% do orçamento médio nas empresas americanas – em detrimento da propaganda, cuja participação no orçamento do composto promocional vem caindo constantemente (KOTLER, 1999).

O terceiro item do composto promocional é representado pelas relações públicas, ou simplesmente publicidade, e envolve os programas corporativos designados para promover ou proteger a imagem de uma companhia ou de seus produtos. Muitas empresas, incluindo as de varejo, operam ou subcontratam departamentos de relações públicas, a fim de monitorar as atitudes públicas de uma organização e distribuir informações à imprensa, seus clientes e comunidade em geral.

2. Metodologia de Mineração de Dados Aplicada à Previsão de Vendas

Mineração de Dados, ou *Data Mining* em inglês, consiste na geração de conhecimento a partir de dados acumulados. Ela permite estruturar o conhecimento que está escondido nos bancos de dados corporativos, utilizando para tanto de tecnologias de banco de dados, reconhecimento de padrões, aprendizado automático, estatística entre outras (HAND, 1998). Entre as técnicas de mineração de dados hoje empregadas, citam-se, entre outras: redes neurais, árvores de decisão, regras de associação, métodos de indução de regras, técnicas de segmentação, métodos para análise de cesta de compras, e dedução baseada em memória.

O termo *data mining* é definido em um escopo estreito por alguns autores, e mais amplo por outros. Na sua definição estreita, mineração de dados é a descoberta automatizada de padrões interessantes escondidos nas bases de dados, e agrega métodos de aprendizado automático, muitos com origem na inteligência artificial (PEACOCK, 1998a). Uma definição mais ampla abrange também a confirmação dos relacionamentos descobertos; ou seja, os métodos de teste de hipóteses — incluindo análises estatísticas tradicionais — também são considerados mineração de dados. Esta definição também é conhecida como o processo de *Knowledge Discovery in Databases* (KDD – descoberta de conhecimento em bases de dados). Finalmente, no escopo mais abrangente, a mineração de dados é vista como um processo – e não um método ou conjunto de técnicas – de descoberta de informações valiosas e não-óbvias a partir de grandes quantidades de dados (BERRY & LINOFF, 1997). Incorpora, assim, atividades de aquisição de dados, limpeza, formatação e pré-processamento dos mesmos, análise e validação dos dados, construção e implementação de ferramentas de apoio à decisão, e acompanhamento (recalibração e manutenção dos modelos) dos sistemas de apoio à decisão.

As raízes da mineração de dados estão na metodologia criada por J. W. Tukey nos anos 70 e 80, referenciada como Análise Exploratória de Dados (TUKEY et al., 1983). As

forças que pressionaram o avanço desta tecnologia podem ser divididas em (PEACOCK, 1998a):

- Supply-Side Factors. São os efeitos do avanço tecnológico. Inclui avanços no armazenamento de dados, na tecnologia de processamento, redução de custos de comunicação eletrônica, nas técnicas de análise (como redes neurais, algoritmos genéticos, árvores de decisão, indução de regras), na arquitetura cliente-servidor dos computadores e no advento dos repositórios de dados.
- Demand-Side Factors. Aumento nas necessidades de análises rápidas em ambientes cada vez mais competitivos. Colabora também a mudança na hierarquia organizacional, onde geralmente a figura do analista já não existe, e o gerente de marketing é o responsável pelas análises. Por fim, a atenção dada pela mídia e livros sobre mineração de dados e relacionamento com o cliente pressionam a “corrida tecnológica”.

BERRY & LINOFF (1997) confirmam a visão acima, afirmando que o processo de mineração de dados só pode amadurecer durante os anos 90 devido aos seguintes fatores:

- Os dados estão sendo produzidos. Mineração de dados só faz sentido quando existir grandes volumes de dados.
- Os dados estão sendo guardados, e freqüentemente em bases separadas dos dados operacionais, permitindo análises para suporte à decisão.
- O poder computacional necessário é acessível.
- A pressão competitiva é forte.
- Pacotes comerciais dedicados a análises de dados estão disponíveis.

As principais aplicações de mineração de dados estão nas áreas de: aquisição de novos clientes, manutenção dos clientes atuais, abandono de clientes não rentáveis, priorização de clientes (para mala direta ou outros fins), análise de cesto de compras, previsão de vendas, segmentação de mercado, risco de crédito, previsão de insolvência, detecção de fraude, entre outras.

SAS (1996) frisa que essas aplicações raramente podem ser automatizadas, apesar da crescente disponibilidade de pacotes comerciais que implementem o processo de mineração de dados.

Figura 12 – Estatística convencional X Mineração de Dados

Com relação às técnicas estatísticas, a mineração de dados distancia-se delas na medida em que (HAND, 1998):

- A quantidade de dados é muito maior.
- Os dados estão “contaminados” com frequência, isto é, não estão limpos – livres de erros, de lacunas, inconsistências etc.
- Os dados costumam estar fora da distribuição esperada (normal).
- Há observações não-estacionárias, enviesadas e dependentes. Ou seja, a amostragem não é aleatória.
- Envolve a “descoberta” de padrões interessantes, muitas vezes de maneira indireta e com pouca interferência humana. Isto significa que o processo pode ser aplicado mesmo sem ter um modelo imposto *a priori* para o fenômeno estudado.
- Pode envolver dados não numéricos, qualitativos.

2.1. Bases de Dados para Mineração de Dados

Para dar início ao processo de mineração de dados, faz-se necessária a constituição de um banco de dados a ser explorado, o que levanta algumas questões como: Quais as variáveis relevantes? Qual o volume de dados necessário? Onde estão os dados? Qual o custo para obtê-los?

De maneira geral, os dados podem ser coletados interna ou externamente à organização. Dados externos são de difícil obtenção e de qualidade suspeita, uma vez que não

podem ser auditados pela própria organização. Portanto, uma origem importante dos dados — senão a única — é a coleta feita pelos sistemas operacionais da empresa — aqueles que controlam sua operação diária, como sistemas de ponto de venda, sistemas transacionais, e geradores de relatórios operacionais. Tais sistemas podem acumular dados diversos como informações dos clientes, dos produtos, de vendas, e mantê-los em formato altamente atomizado, guardando todos os detalhes sobre as operações realizadas.

Formam-se assim grandes bases de dados históricas, que geralmente têm de ser trabalhadas a fim de que possam ser exploradas convenientemente. Normalmente, os dados existentes não foram coletados com o propósito específico de realizar análises exploratórias e de mineração, mas apenas para suportar os processos de negócio operacionais. Isto pode levar a dados de baixa qualidade, valores ausentes e dados enviesados, que reduzem a aplicabilidade dos resultados encontrados. Um dos objetivos da aplicação de técnicas avançadas de mineração de dados é justamente desenvolver um modelo válido mesmo se utilizando dados “sujos” (HAND, 1998).

Uma questão comum diz respeito à quantidade de dados necessária para a realização de mineração de dados. Uma resposta freqüente para esta questão é a de que as atividades de mineração de dados requerem enormes quantidades de dados. FEELDERS et al. (2000) lembram, entretanto, que:

- Uma vez que o problema foi adequadamente especificado, apenas uma pequena parte de uma base de dados altamente heterogênea é necessária.
- Mesmo se a quantidade de dados resultante for grande, normalmente apenas uma amostra é suficiente para a construção de modelos suficientemente precisos.

Quanto à seleção de dados, a abordagem da mineração de dados difere consideravelmente daquela usual na estatística. Ao contrário de técnicas de seleção de variáveis, como a *setpwise*, a experiência até o momento sugere que esta deve ser feita da maneira mais ampla possível, já que o ponto forte das técnicas de mineração de dados é justamente sua capacidade de fazer com que os dados falem por si mesmo, sem restringir a análise a hipóteses previamente formuladas (GARGANO & RAGGAD, 1999). As técnicas devem, idealmente, ser capazes de identificar as variáveis relevantes e desprezar as irrelevantes.

2.1.1. Data Warehouses

É normal que os sistemas operacionais de uma empresa sejam fruto de vários esforços no seu processo de informatização. Sendo assim, é freqüente a existência de sistemas operando em ambientes (arquiteturas de *hardware*, versões de *software* etc.) distintos. Para que a mineração de dados possa ocorrer, estes dados devem ser consolidados em um mesmo ambiente, o *Data Warehouse* (DW).

Figura 13 – Consolidação de dados em um *Data Warehouse*

A Figura 13 ilustra o processo de consolidação de dados em um *Data Warehouse* para fins analíticos. Pode-se descrever um *Data Warehouse* como sendo um banco de dados especial, contendo as seguintes características básicas (GRAY, 1997; IMMON, 1997):

1. Armazena os dados separadamente daqueles utilizados em aplicações operacionais.
2. Seu foco está no apoio à tomada de decisão.
3. Cobre um horizonte de tempo muito maior que sistemas transacionais, guardando séries temporais dos dados.
4. Inclui dados provindos de várias bases de dados operacionais, unificando-os e consolidando-os em um único local. Os dados são processados e validados antes de serem inseridos no *Data Warehouse*, para que, uma vez no DW, passam a representar uma fonte segura, “oficial” dos dados. Isso implica também que as definições dos atributos existentes no DW sejam únicos — por exemplo, o atributo *lucro* pode ter

diferentes definições, conforme o propósito da base de dados; esta definição têm de ser única no ambiente DW.

5. É otimizado para chamadas complexas e com grande volume de dados, vindas de analistas e administradores. Isso quer dizer que antigos paradigmas não são mais válidos neste ambiente, e normas como a simplificação de tabelas complexas em várias tabelas e relacionamentos, e a regra máxima de se evitar a todo custo qualquer tipo de redundância ficam enfraquecidas, sendo utilizadas apenas seletivamente.
6. Da mesma forma, é projetado para poucas inclusões ou alterações — a grande maioria das chamadas são para recuperação de dados. Ou, resumindo, seus dados são não voláteis.
7. É integrado — une vários pontos de vista, não sendo específico a uma dada aplicação ou comunidade de usuários.
8. É orientado por assunto — os dados são organizados de forma que as análises sejam feitas por um determinado assunto do negócio.
9. Possuem, além de dados transacionais (atomizados), dados agregados, para agilizar consultas de agregação frequentes. Esta propriedade é chamada de granularidade do DW.

Figura 14 – *Data Warehouse* como repositório para análises de dados

2.2. Usos da Mineração de Dados

Existem na literatura diversas taxonomias para classificar os usos e objetivos genéricos dos processos de mineração de dados. Os objetivos mais citados são: sumarização (descrição e visualização), previsão, classificação, segmentação e análise de ligações ou associações (PEACOCK, 1998a; BERRY & LINOFF, 1997). Algumas destas tarefas são melhor abordadas de maneira *top-down*, através de teste de hipóteses. Outras são melhor abordadas de maneira *botton-up*, através de técnicas de descoberta de conhecimento (KDD).

Os principais objetivos visados com a mineração de dados são:

1. **Previsão**. Lida com comportamentos futuros, conforme detalhado na seção “O Problema de Previsão de Vendas”. Todas as técnicas utilizadas para classificação podem ser usadas com sucesso para previsões, formulando-se os modelos com base em dados históricos, onde a variável a ser prevista é conhecida (BERRY & LINOFF, 1997). Alguns autores separam esta tarefa daquela de “estimação”, onde o elemento tempo não está necessariamente presente – pode-se estimar variáveis como renda, altura, número de filhos, balanço da conta corrente ou a probabilidade de responder a uma campanha publicitária.
2. **Sumarização e Descrição**. Visa aumentar o grau de compreensão sobre um fenômeno complexo, representado por grande quantidade de dados de difícil compreensão, através da sumarização destes dados. Utiliza basicamente técnicas estatísticas descritivas, e ferramentas de visualização gráfica. Estas últimas – de visualização gráfica automática a partir de dados e análises simples – são consideradas essenciais, dadas as características inerentes ao ser humano de dificuldade em interpretar dados numéricos aliada à enorme facilidade em interpretar informações visuais (GARGANO & RAGGAD, 1999).
3. **Classificação**. É, segundo BERRY & LINOFF (1997), o trabalho mais comum. Consiste em examinar as características de um objeto e tentar colocá-lo dentro de uma série de classes pré-definidas. A construção de um modelo de classificação consiste na análise das características de elementos previamente classificados, para que se possa “aprender” a classificar novas observações.
4. **Segmentação**. Segmentação é a atividade de separar em grupos homogêneos uma população heterogênea. Os grupos de registros semelhantes são chamados de *clusters*,

por isto estas técnicas também são conhecidas por técnicas de *clustering*. A diferença entre segmentação e classificação é que, nesta última, as classes são pré-definidas, enquanto durante uma segmentação elas são dinamicamente criadas com base na similaridade entre os elementos. O objetivo é identificar grupos semelhantes de registros, para tratar um grupo populacional como grupos com qualidades distintas. Assim, as ações sobre o mercado podem ser realizadas de maneira dirigida, focando esforços em pontos onde a resposta é mais favorável.

5. **Agrupamento por Afinidade.** É usado para determinar afinidades ou ligações entre objetos. A aplicação clássica é a determinação de quais produtos vendem em conjunto em um supermercado – também chamada de análise de cesto de compras – para dispor ou promover itens de maneira a incentivar a compra de produtos afins. Consiste basicamente na geração de probabilidades de compra conjunta: quem compra o produto *A* tem *X%* de chances de também comprar o produto *B*.

2.3. O Processo de Mineração de Dados

A simples aplicação de uma técnica de análise de dados avançada a fim de encontrar algum padrão nos dados não é suficiente para caracterizar um processo de mineração de dados. Para tanto deve-se estar apto a responder aos padrões, a agir com base neles, transformando dados em informação, e informação em ação, e conseqüentemente em valor para a empresa.

SPSS (1999) frisa a importância de se ter uma estratégia para completar um processo de mineração de dados com sucesso. A estratégia proposta pelo autor foi chamada de 5 A's, e corresponde às seguintes etapas:

1. **Avaliar.** Avaliação da situação do negócio, e definição do problema a ser resolvido — dos objetivos da análise.
2. **Acessar.** Acesso aos dados disponíveis e sua organização física. Algum tempo deve ser gasto conhecendo e preparando os dados — o significado de cada campo e como eles são medidos. Inclui também o pré-processamento dos dados, etapa fundamental para a construção de modelos representativos do problema, e que envolve tratamentos (limpeza, agrupamentos, agregações etc.) e transformações (normalização, incremento ou redução de grandezas etc.) no dados. Muitas vezes, o pré-processamento dos dados geralmente é uma tarefa mais demorada que a construção do modelo em si.

3. Analisar. Uso das técnicas de mineração de dados. O desenvolvimento do modelo é um processo cíclico de ajuste e ação recíproca entre a definição do problema, definição / representação dos dados, definição do modelo e da acurácia desejada dos resultados.
4. Atuar. Planejamento e execução de mudanças a partir do conhecimento gerado.
5. Automatizar. Incorporar o processo às práticas usuais, através da criação de aplicações específicas.

Metodologias semelhantes são propostas na literatura, mantendo-se sempre a estrutura Identificação do problema → Análise / construção do modelo → Ação sobre o problema. BERRY & LINOFF (1997), por exemplo, identificam um ciclo semelhante, porém acrescentam a importância de uma última etapa – a de medição dos resultados das ações: “A chave para o sucesso de um projeto de mineração de dados é incorporá-lo aos processos de negócio corporativos. Os estágios são interativos, e sempre dirigidos a resultados”. Os autores definem o processo como o “ciclo virtuoso de *data mining*”, conforme mostra a Figura 15.

Esse processo é altamente interativo e exploratório (FEELDERS et al., 2000). Segundo os autores, conforme novos conhecimentos são gerados, novas hipóteses podem ser formuladas e testadas, fazendo com que se passe pelas diferentes etapas de maneira não linear.

Figura 15 – O ciclo virtuoso de *data mining* (BERRY & LINOFF, 1997, modificado)

2.4. Técnicas de Mineração de Dados

BERRY & LINOFF (1997) definem análise de dados como o processo de construção de um modelo apropriado aos dados. De maneira concreta, as técnicas de mineração de dados são algoritmos (procedimentos matemáticos repetitivos, programáveis em computadores) que modelam relações ou padrões não-aleatórios (estatisticamente significantes) em grandes bases de dados históricas.

Cada técnica de mineração de dados tem sua própria abordagem, porém elas compartilham algumas características em comum. Dentre elas destacam-se:

- Elas melhoram sua performance gradativamente, conforme “aprendem” a partir dos dados de treinamento coletados.
- Há sempre uma fase de treinamento, quando o modelo “aprende” os padrões e relacionamentos; seguida de uma fase de implementação, quando o modelo é posto à prova.

De maneira geral, os dados disponíveis afetam a escolha da modelagem mais apropriada. Por exemplo, quando há muitas variáveis categóricas, as árvores de decisão são bem apropriadas. Quando o relacionamento entre as variáveis de entrada e saída é difícil de ser decifrada, as redes neurais são a melhor escolha (BERRY & LINOFF, 1997).

A seguir, uma breve descrição de algumas das técnicas de mineração de dados mais utilizadas é exposta. A técnica de redes neurais é explanada em capítulo separado, dada sua importância para esta pesquisa.

2.4.1. Indução de Regras / Análise do Cesto de Compras

As técnicas de indução de regras consistem no uso de ferramental matemático e estatístico — como frequência de correlação, taxa de acurácia, e acurácia das previsões — para desenvolver relacionamentos a partir dos dados apresentados. Tipicamente, são criadas correspondências do tipo “se-então”, baseadas em relações causais detectadas nas variáveis. Cada relacionamento “se-então” extraído é chamado de “regra”.

A metodologia para a geração de regras é não-supervisionada, e tende a consumir grande capacidade de processamento. Um efeito negativo muito comum é a geração de padrões em demasia — muitos deles são triviais, contraditórios ou inúteis. Ela também requer validação das regras por parte de um especialista, já que ela obtém apenas correlações, e não

relações de causa e efeito propriamente ditas. Outro problema visível desta técnica é que o número de possibilidades cresce exponencialmente com o aumento no número de itens.

Um dos principais usos das técnicas de indução de regras é o agrupamento de produtos por afinidade, ou, em sua denominação mais comum, a análise do cesto de compras. A análise do cesto de compras é utilizada para encontrar grupos de itens que ocorrem simultaneamente (e em que ordem) em transações. As informações coletadas são usualmente utilizadas no planejamento do *layout* de lojas, na elaboração de promoções de produtos conjuntos, agrupamento de produtos, comparação de vendas entre lojas ou entre períodos com promoções, detectar sazonalidades etc. Quando as transações não são anônimas, as análises podem utilizar o fator tempo como um elemento, para agrupar produtos adquiridos por um mesmo consumidor ao longo do tempo.

O processo de análise do cesto de compras consiste em:

1. Escolher os itens. Para tanto é comum realizar uma taxonomia dos itens a fim de generalizá-los propriamente. Usualmente, são analisados inicialmente os itens genéricos — em hierarquia superior, como segmentos ou linhas de produtos. Posteriormente, podem ser analisados os itens mais específicos (atomizados), geralmente com o uso de apenas uma amostra dos dados. O nível de generalização entre os produtos, dentro da taxonomia existente, pode variar. A frequência com que os itens aparecem também é indicativa da generalização necessária — produtos que raramente vendem devem ser agrupados para que sua frequência aumente, e vice-versa. Por exemplo, se há muitos registros de “coca-cola”, este produto pode ser analisado de forma atomizada, enquanto a baixa demanda por outras marcas concorrentes pode forçar seu agrupamento na categoria “refrigerantes de cola”.
2. Gerar as regras através da contagem de ocorrência simultânea dos produtos.
3. Analisar as regras interessantes, uma vez que podem ser geradas milhares de combinações diferentes, muitas delas inexplicáveis ou óbvias.

Algumas vezes faz-se necessária a geração de itens “virtuais”, ou seja, que não aparecem na taxonomia original. Por exemplo, uma marca específica pode aparecer em produtos bem distintos, e um item virtual indicando sua presença pode ser útil para detectar relacionamentos nas vendas dos produtos dessa marca. Outros exemplos típicos são informações sobre as transações — se foi paga em dinheiro, cheque ou cartão, qual era o dia

da semana, entre outros. A inclusão de itens “virtuais” deve ser limitada, contudo, a itens acionáveis — aqueles que geram regras úteis à corporação.

Outro tipo de regra que pode ser gerada são as regras de dissociação, que funcionam como as regras de associação, porém possuem o conector “não”, além do conector “e”, como por exemplo a regra “se A e não B então C”. Corre-se, contudo, o risco de gerar inúmeras regras do tipo “se não A e não B então não C”, que não são regras acionáveis. Assim, é comum testar regras dissociativas apenas com os itens mais frequentes.

Durante o processo de análise, os fatores mais importantes para monitorar a qualidade das regras criadas são (BERRY & LINOFF, 1997):

- ❑ Suporte à regra. Número de observações da regra em relação ao número de observações total. Por exemplo, se tivermos 5 transações e 2 delas incluírem “soda” e “suco”, então o suporte para a regra “se soda então suco” é $2 / 5$ ou 40%.
- ❑ Confiança. Razão entre o número de transações que suportam a regra e o número de transações onde o elemento condicional da regra ocorre. Para o mesmo exemplo acima, se em todos os casos onde ocorre soda também ocorrer suco, a confiança é de 100%. Entretanto se 4 transações contiverem suco, com apenas 2 delas contendo também soda, a regra “se suco então soda” possui confiança de 50%.
- ❑ Melhoria, também chamada de *lift* ou *improvement*. Mede o quanto uma regra melhora a previsão de um resultado do que simplesmente assumindo o resultado. Por exemplo, se um produto ocorre 45% das vezes, uma regra com confiança igual a 45% tem *lift* igual a 1. A melhoria é definida matematicamente como a frequência observada para uma regra dividida pela frequência esperada, dadas as frequências de cada um dos itens.

A análise do cesto de compras é de particular interesse nesta pesquisa, pois se trata da técnica utilizada para a detecção do grupo de produtos com maior grau de inter-relacionamento dentro do conjunto de dados coletados, definindo assim um centro de interesse. A aglutinação dos produtos relacionados serviu para a separação dos produtos a serem analisados (cuja demanda foi prevista) e também como ferramental de auxílio na qualidade das previsões, conforme será mostrado adiante.

2.4.2. Sistemas Especialistas

Um sistema especialista consiste de uma base de conhecimentos composta por regras extraídas de especialistas, aliada a um sistema de inferência lógica. Um sistema especialista é capaz de simular o comportamento de especialistas com conhecimentos em um domínio muito estreito e profundo. O objetivo é chegar nas mesmas conclusões que um especialista chegaria, passadas as informações a respeito do problema.

Sistemas especialistas são altamente supervisionados, ou seja, o treinamento não é automático e exige grande esforço do analista. Eles possuem performance comparável a de humanos especializados em um domínio de problema, consistindo de conhecimento cognitivo específico. Os domínios de conhecimento ideais são os de escopo bastante estreito, que permitam que o especialista resolva o problema num curto espaço de tempo. O conhecimento deve ser facilmente capturável, consistente, de explanação simples, de representação honesta (geralmente através de afirmações do tipo “se-então”) e que não dependa de bom senso (GARGANO & RAGGAD, 1999).

2.4.3. Árvores de Regressão

Árvores de Regressão são modelos simples baseados na anatomia de uma árvore, onde cada galho particiona estrategicamente os dados em classes e subclasses sucessivamente. A cada divisão, é escolhida a melhor forma de separar e classificar os dados, utilizando-se da característica que mais os distingue (maior poder de separação), através de medidas estatísticas. Em cada separação (ou nó), pode-se medir: o número de registros presentes no nó; a forma como eles foram separados; o número de registros em cada galho (separação); e a porcentagem de registros classificados corretamente (de acordo com a variável dependente).

As árvores de regressão possuem algoritmos não-supervisionados, ou seja, são capazes de processar automaticamente os dados. Com o modelo criado, cada partição (ou galho) representa uma regra clara, constituída durante o treinamento. Estas regras são então usadas para classificar o comportamento de novas situações.

A grande atratividade das árvores de regressão está no fato de que suas respostas representam regras facilmente acionáveis e com grande poder explicativo. Outra característica positiva é o fato de que fornecem indicações claras dos campos mais importantes para previsão ou classificação (SAS, 1998).

2.4.4. Memórias Associativas

Memórias associativas, também conhecidas por dedução baseada em memória ou *Memory Based Reasoning* (MBR), são pares ou grupos de itens associados através de uma memória de longo prazo, usando instâncias conhecidas como um modelo para fazer previsões sobre novas instâncias. Consiste na procura de vizinhanças nas instâncias conhecidas e combinação de seus valores para prever novas instâncias. Os dois elementos chaves são a função distância utilizada para encontrar os vizinhos mais próximos e a função de combinação para calcular os valores das previsões (BERRY & LINOFF, 1997).

O processo segue os seguintes passos:

1. Escolher os dados históricos apropriados.
2. Determinar a forma mais eficiente de representar os dados históricos.
3. Determinar as funções distância e combinação, e o número de vizinhos. Ou seja, são identificados os casos semelhantes nos dados históricos, e então a informação associada a estes casos é aplicada através de equações matemáticas convenientemente selecionadas.

A técnica é apropriada para tarefas de classificação e previsão. Técnicas estatísticas podem ser combinadas às técnicas de memórias associativas para auxiliar na construção da função combinação.

2.5. Seleção das Técnicas de Mineração de Dados

Não há consenso na literatura sobre a técnica ideal para cada aplicação, nem sequer de quais são os critérios para avaliação das diferentes técnicas de mineração de dados. GARGANO & RAGGAD (1999) citam como critérios para avaliação das técnicas: robustez, grau de automação, velocidade, poder explanatório, acurácia, quantidade de pré-processamento necessário, escalabilidade, facilidade de integração e habilidade para lidar com muitos atributos (variáveis). Já BERRY & LINOFF (1997) citam: facilidade de compreensão do modelo, facilidade de treinamento, facilidade de aplicação, capacidade de generalização, utilidade e disponibilidade. Ainda de acordo com os autores, os principais fatores que determinam a escolha da técnica a ser utilizada para determinado fim são: preponderância de variáveis categóricas ou numéricas, número de campos, número de variáveis dependentes, orientação no tempo, e presença de dados textuais.

Conclui-se portanto que não há critérios universais, aplicáveis a todas as técnicas de mineração de dados — cada técnica requer critérios específicos de medição. É extremamente difícil comparar os modelos entre si, já que operam de maneira distinta. Então, a única forma de avaliá-los é através da medição de sua habilidade em desempenhar a tarefa para a qual foram construídos, ou seja, classificação, previsão, etc.

Neste trabalho, são de interesse as técnicas voltadas para previsão, e, portanto, o principal critério de avaliação deve ser a habilidade do modelo em desempenhar sua função: prever a demanda futura. Assim sendo, a medida de maior interesse a acurácia do modelo.

3. Redes Neurais Artificiais

Redes neurais artificiais são sistemas de processamento de informações distribuídos, compostas por muitos elementos computacionais simples que interagem através de conexões com pesos distintos. Inspiradas na arquitetura do cérebro humano, elas exibem características como a habilidade de aprender padrões complexos de informação e generalizar a informação aprendida (BAETS & VENUGOPAL, 1994).

Cada elemento computacional não-linear é chamado nó, e é densamente interconectado através de conexões diretas. Os nós operam em passos discretos, de forma análoga a uma função de dois estágios: o primeiro estágio calcula a soma dos sinais de entrada, atribuindo pesos aos sinais; o segundo estágio consiste da aplicação de uma função de saída, chamada de ativação (PETRON, 1999).

Figura 16 – Esquema de funcionamento de cada nó em uma rede neural artificial

Como ilustrado na Figura 16, cada nó recebe um ou mais valores de entrada ($E_1 \dots E_n$), que são combinados em um único valor com o uso de diferentes pesos para cada entrada; e transformam-no em um valor de saída (S) através de uma função de ativação. Uma das funções de ativação mais usadas é a função logística (PETRON, 1999), que pode ser escrita como:

$$\text{Equação XIV: } F(x) = \frac{1}{1 + e^{-x}}$$

Este funcionamento aparentemente simples de cada neurônio resulta, após o processamento coletivo de todos os nós, em uma capacidade de execução de diversas tarefas com eficiência. Segundo BAETS & VENUGOPAL (1994), as redes neurais podem ser caracterizadas pelas seguintes propriedades: a estrutura da rede (topologia), como e o que a rede processa (propriedade computacional), e como a rede aprende (propriedade de treinamento).

A topologia da rede refere-se ao número de camadas e nós utilizados. Uma rede neural artificial deve ser composta por pelo menos duas camadas: uma contendo as entradas e uma contendo as saídas da rede. Usualmente, são utilizadas também uma ou mais camadas intermediárias, caracterizando assim as chamadas redes neurais multicamadas, compostas por (KUO & XUE, 1998):

- Camada de entrada: consiste das variáveis de entrada do modelo, devem ser sempre conhecidas;
- Camada de saída: contém um ou mais nós representando os resultados finais do processamento – para uma dada entrada, a rede fornece uma saída correspondente;
- Camadas intermediárias ou camadas ocultas: podem existir uma ou mais camadas de processamento intermediário, que tornam o modelo mais refinado e não-linear, com a capacidade de aprender padrões mais complexos.

A propriedade computacional de uma rede neural artificial refere-se ao modo pelo qual os elementos computacionais (nós) são ativados e processados (BAETS & VENUGOPAL, 1994). Diversas funções de ativação podem ser empregadas, de forma linear e não-linear.

Já o aprendizado é o processo no qual uma série de valores de entrada é apresentada sequencialmente, e os pesos da rede são ajustados até que ela reflita a saída desejada. As estratégias de treinamento são divididas em treinamento supervisionado e não-supervisionado.

O treinamento supervisionado requer a presença de valores de entrada e a correspondente saída, a partir da qual é calculado o erro, utilizado para corrigir o peso das conexões da rede. No treinamento não-supervisionado, são apresentados à rede apenas vetores de entrada. A saída é construída automaticamente pela rede, sendo capaz de capturar regularidades nos valores de entrada, sem receber nenhuma informação adicional (ALMEIDA, 1993).

O processo de montagem e treinamento de uma rede neural pode ser descrito pelos seguintes passos (PETRON, 1999):

- É definida a topologia da rede.
- A rede sofre um processo de aprendizado, através de informações de entrada e saída conhecidas.
- As conexões entre os nós são modificadas num processo iterativo, até que a rede forneça saída compatível com as informações apresentadas. Por isso diz-se que as redes neurais adquirem conhecimento através da experiência.
- A rede está pronta para fornecer respostas para novos problemas.
- O aprendizado pode continuar indefinidamente. O conhecimento reside na “força” entre as ligações entre neurônios, ou seja, no peso dado a cada conexão.

3.1. *Justificativa de utilização das Redes Neurais*

Segundo BERRY & LINOFF (1997), a modelagem através de redes neurais artificiais é uma das técnicas de mineração de dados mais utilizadas, largamente empregada em tarefas de classificação e previsão, além de poderem ser aplicadas à geração de conhecimento indireto (na forma de mapas auto-organizáveis). Neste trabalho, utilizaram-se as redes neurais para previsão de vendas; sua escolha como técnica de modelagem baseou-se em algumas características decisivas reportadas na literatura, mostradas a seguir.

De acordo com BAETS & VENUGOPAL (1994), as redes neurais são robustas e tolerantes a falhas, graças ao seu paralelismo (a contribuição dada por um único elemento de processamento não é tão importante). Isto a torna uma técnica bastante indicada para uso na mineração de dados, onde nem sempre os dados disponíveis são confiáveis.

Diversos autores mostraram a capacidade das redes neurais em lidar com estruturas de dados não-lineares (LUXHOJ et al., 1996; ZHANG et al., 1998). Esta é uma característica primordial, já que, de acordo com BAETS & VENUGOPAL (1994), os problemas em

marketing sempre lidam com dados como vendas e preços, onde as séries de dados estão inseparavelmente ligadas a pontos de inflexão, tendências e não-linearidades. Muitas vezes os dados podem ser até mesmo caóticos (THALL, 1992; HIBBERT & WILKINSON, 1994) — e mesmo nesta condição podem ser previstos com precisão por uma rede neural, como demonstraram LAPEDES & FARBER (1987).

Com o uso de redes neurais, um conhecimento a priori da relação entre as variáveis não é requerido, já que elas são auto-adaptativas (BAETS & VENUGOPAL, 1994; ZHANG et al., 1998; WRAY et al., 1994). Ou seja, elas são capazes de gerar modelos não lineares sem um conhecimento prévio sobre os relacionamentos entre as variáveis de entrada e saída, podendo ser portanto mais genéricas e flexíveis que metodologias estatísticas tradicionais. Tome-se como exemplo uma modelagem feita através de regressão: a estipulação da estrutura dos dados (de segundo ou terceiro grau, por exemplo), assim como a determinação de uma variável dependente e de uma ou mais variáveis independentes, são necessários a priori, para que o modelo matemático possa ser gerado.

Já com o uso de redes neurais, os relacionamentos (relação de dependência) entre as variáveis são detectados automaticamente. Podem até mesmo ignorar variáveis que não contribuem nas previsões, o que permite que algumas experimentações possam ser feitas sem prejuízo para o modelo resultante (CHURCH & CURRAM, 1996). Como consequência, um melhor resultado é esperado quando o relacionamento entre as variáveis não é aderente ao modelo assumido, e quando há forte interdependência entre os fatores.

Pode-se provar também que uma rede neural com uma camada oculta é capaz de aproximar qualquer função contínua, com a precisão que se desejar (ZHANG et al., 1998). Uma rede neural pode também mimificar uma série temporal qualquer, como mostraram ANSUI et al. (1996).

Além disso, relações de multicolinearidade são automaticamente descobertas e assimiladas pela rede de relacionamentos (WRAY et al., 1994). Segundo os autores, a performance das redes neurais não é depreciada por relações de multicolinearidade, como acontece com a regressão. Condições fora da normalidade, violação de hipóteses, pontos de alta influência e transformações também são bem assimiladas pelas redes neurais.

As vantagens das redes neurais que levaram à sua escolha como método de modelagem para previsão de vendas, conforme reportadas na literatura citada acima, podem ser resumidas da seguinte forma:

- São robustas e tolerantes a falhas. Dados imprecisos ou faltantes, relações de multicolinearidade e violações de hipóteses são automaticamente descobertas e assimiladas, não interferindo fortemente no resultado final das previsões.
- São inerentemente não-lineares.
- O conhecimento a priori da relação entre as variáveis não é requerido. Portanto, elas comportam-se bem em problemas onde o conhecimento a respeito é de difícil estruturação.
- São capazes de aproximar qualquer função contínua ou série temporal, inclusive séries caóticas.
- Possuem boa capacidade de generalização. Após a fase de aprendizado, onde a rede é submetida a dados históricos conhecidos, ela é capaz de prever novos registros com considerável sucesso, mesmo na presença de ruído e variáveis expúrias ao problema.

3.2. *Desempenho e Limitações das Redes Neurais*

Vários autores buscaram comparar o desempenho de modelos baseados em redes neurais com metodologias estatísticas tradicionais, porém os resultados ainda não são conclusivos. Segundo ZHANG et al. (1998), a principal explicação para o fato parece ser a ausência de uma metodologia sistemática para o desenvolvimento de modelos baseados em redes neurais, consistindo basicamente de tentativa e erro, o que favorece o surgimento de resultados inconsistentes. O número de fatores envolvidos na criação de modelos neurais — entre eles sua topologia, método de treinamento, e dados fornecidos para o aprendizado — torna o processo de desenvolvimento de tais modelos ainda mais complexo e sujeito a grandes variações no resultado final.

Os modelos construídos com o uso de redes neurais artificiais já mostraram ter boa aplicabilidade em problemas de previsão, tendo superado significativamente várias outras modelagens em diversas áreas do conhecimento (ANSUJ et al., 1996). Em alguns casos, entretanto, as redes neurais podem mesmo ter poder preditivo menor que metodologias tradicionais, seja simplesmente porque os dados analisados são lineares ou porque o desenvolvimento do modelo baseado em redes neurais não foi adequado (ZHANG et al., 1998).

Vários autores compararam as capacidades preditivas de redes neurais com técnicas estatísticas tradicionais, muitos deles concluindo que as redes neurais possuem performance superior (TANG et al., 1991; ANSUJ et al., 1996).

CHURCH & CURRAM (1996) compararam quatro modelos econométricos entre si, e então utilizam os mesmos dados na geração de modelos baseados em redes neurais, para comparação de suas previsões. Os resultados sugeriram pouca ou nenhuma melhoria nas previsões quando da utilização de redes neurais. De acordo com os autores, um ponto crítico durante a modelagem de redes neurais foi a seleção de camadas ocultas (*hidden layers*). Quanto mais complexa a inter-relação entre as variáveis, maior o número de camadas ocultas necessárias. Camadas a menos fazem com que o modelo perca a riqueza das relações entre as variáveis — perdendo em qualidade nas previsões; camadas em demasia provocam o efeito de *overfitting*, que é o excessivo ajuste aos dados de treinamento, perdendo sua capacidade de generalização. Segundo os autores, apenas uma camada oculta é geralmente suficiente para problemas de previsão; o número ideal, entretanto, deve ser encontrado por tentativa e erro.

Para BERRY & LINOFF (1997), as maiores dificuldades relacionadas à modelagem com uso de redes neurais são:

- a dificuldade em compreender os modelos produzidos (baixa explicabilidade);
- a sensibilidade ao formato dos dados de entrada — diferentes representações para os mesmos dados podem resultar em modelos distintos. Além disso, requerem entrada entre 0 e 1 — o que significa que necessitam muitas vezes de grande pré-processamento.
- Podem convergir prematuramente, resultando em soluções não ideais.

A tabela a seguir mostra alguns exemplos na literatura onde o desempenho das redes neurais foi melhor e/ou pior que outras técnicas analisadas:

Autor	Estudo	Conclusões
Houve melhora significativa com o uso de redes neurais		
PARK & SANDBERG, 1991	Aplicação de redes neurais na previsão de demanda por energia elétrica. A rede construída toma como entradas apenas informações sobre a temperatura ambiente, e produz como saídas as previsões de consumo por hora, pico e consumo total de energia elétrica de uma determinada região geográfica.	Os autores mostraram que a rede neural, apesar de simples, possui desempenho (previsão) muito superior às técnicas anteriormente adotadas, baseadas em regressão múltipla.
TANG et al., 1991	Analisaram diversas séries temporais de demanda agregada de diversas empresas, comparando as redes neurais à técnica de previsão ARIMA.	As redes neurais tiveram, no geral, desempenho igual ou superior à técnica de séries temporais. De forma geral, para séries com grandes irregularidades, as redes neurais foram superiores; para séries mais regulares as performances não foram significativamente diferentes. Outras conclusões sugerem: que as redes neurais são melhores para maiores horizontes de tempo; e que também são superiores quando o número de nós de entrada aumenta.
ANSUJ et al., 1996	Mostraram uma aplicação de modelagem ARIMA com intervenções e, comparativamente, de redes neurais retro-propagadas, para análise do comportamento das vendas de uma empresa média localizada no Brasil, no período de Janeiro de 1979 a Dezembro de 1989.	O modelo baseado em redes neurais apresentou menor variação residual e menor MAE para prever os últimos 12 meses da série temporal. Este resultado demonstra que as redes neurais retropropagadas podem ser usadas para previsão com base no conceito de séries temporais, com desempenho igual ou melhor que métodos altamente sofisticados.
Não houve melhora significativa com o uso de redes neurais		
LUXHOJ et al., 1996	Desenvolveram um modelo de previsão misto séries temporais-regressão-redes neurais, tendo como objetivo a previsão de vendas mensal total. As saídas dessas técnicas foram então combinadas em uma rede neural. Foram testadas várias configurações de rede, e a que apresentou melhores resultados foi uma rede neural de regressão geral (GRNN) - e não uma rede retropropagada.	O modelo híbrido resultou em MAPE de 3.8%, contra MAPE de 6.1% da modelagem qualitativa anteriormente utilizada pela empresa. Analisando os modelos isoladamente, os resultados foram, em termos de MAPE: 5% para o ES, 4.2% para a regressão, e 4.6% para a rede neural. Isso demonstra que o modelo híbrido tem melhor performance que os demais isoladamente, porém a rede neural, isoladamente, não obteve melhora significativa.
CHURCH & CURRAM, 1996	Compararam quatro modelos econométricos entre si, e então utilizam os mesmos dados na geração de modelos baseados em redes neurais. Foi gerada uma rede neural para cada modelo econométrico, pois estes utilizavam variáveis independentes distintas.	Os resultados sugerem pouca ou nenhuma melhoria nas previsões quando da utilização de redes neurais. Todos os modelos baseados em redes neurais gerados, apesar de utilizarem as mesmas variáveis e expostos aos mesmos conjuntos de dados de treinamento, geraram resultados semelhantes aos modelos econométricos.
ALMEIDA & SIQUEIRA, 1996	Compararam as técnicas de regressão logística e redes neurais na previsão de falência de bancos brasileiros.	Os resultados mostraram que as redes neurais podem ser aplicadas com sucesso na previsão de falência, apesar de não superarem significativamente outras técnicas estatísticas.
Outros estudos metodológicos		

CHAKRABORTY et al., 1992	Aplicação de redes neurais na previsão de séries temporais multivariadas. A série utilizada era composta por três variáveis (ou, de fato, três séries distintas), e os autores compararam a performance de redes neurais 1) que combinassem as três variáveis e 2) construídas isoladamente, uma para cada série; na previsão das saídas futuras das séries.	A conclusão principal é que a rede combinada - que toma as três variáveis como entradas - possui desempenho significativamente superior às três redes isoladas.
ZHANG, 1994	Testou várias topologias de redes neurais, a fim de determinar se o aumento no número de camadas ocultas melhora a performance das redes. As redes foram construídas para previsão de séries temporais.	A conclusão é que a melhor topologia é aquela com duas camadas ocultas. Esta topologia mostrou-se melhor que apenas uma camada, porém o aumento no número de camadas ocultas (mais de duas) não trouxe benefícios. Comprova que o melhor é usar até duas camadas ocultas, não mais que isso.
NELSON et al., 1994	Testaram a capacidade de detecção de sazonalidade das redes neurais. Diversas redes com dados desazonalizados e também com forte sazonalidade foram analisadas, utilizando dados mensais com 68 pontos (mais de 5 anos).	Os resultados indicaram que a desazonalização prévia dos dados é benéfica para a capacidade preditiva das redes. Ou seja, elas não detectam a sazonalidade automaticamente e com precisão.

Tabela 3 – Comparação de desempenho da técnica de redes neurais artificiais frente a outras técnicas estatísticas

CAPÍTULO III – METODOLOGIA DA PESQUISA

1. O Método de Pesquisa

Este trabalho está fundamentalmente calcado em uma abordagem quantitativa de pesquisa, a fim de atender ao objetivo básico de criação de uma metodologia capaz de identificar padrões de relacionamento úteis na previsão de vendas individual no varejo, com o uso do processo de mineração de dados.

Os estudos quantitativos são aqueles onde o pesquisador estabelece hipóteses e as testa por meio da mensuração de variáveis operacionalmente definidas, quantificando o resultado com o uso de variados métodos quantitativos e estatísticos. É um tipo de pesquisa tradicionalmente aceita como tendo confiabilidade (fidedignidade) e validade, desde que realizada seguindo uma metodologia rigorosa. Normalmente, utilizam-se testes estatísticos e análise de variância para verificação das hipóteses estatísticas. A validação dá-se pelo nível de significância adotado nos testes estatísticos (KIRK & MILLER, 1988).

As pesquisas quantitativas podem ser orientadas por delineamentos experimentais ou quase-experimentais, ou serem estudos descritivos — exploratórios — isto é, buscarem a descrição da população ou fenômeno e o estabelecimento de relações entre variáveis e fatos.

Esta pesquisa baseou-se numa metodologia orientada por um estudo exploratório dos dados disponíveis, isto é, buscou a descrição do fenômeno — demanda no varejo — e o estabelecimento de relações entre as variáveis explicativas e as vendas. Apesar de tratar-se de uma pesquisa eminentemente quantitativa, não foram estabelecidas hipóteses *a priori*, mas sim questões de pesquisa (veja seção Questões de Pesquisa, pág. 7), que foram respondidas por meio de um processo efetivo de mineração dos dados, onde o pesquisador realiza levantamentos e análises iterativas buscando informações “escondidas” nos dados.

Construiu-se, através do uso da técnica de redes neurais artificiais, um modelo de previsão de vendas a fim de prever a demanda individualizada de produtos em uma loja de varejo, levando em consideração fatores explicativos da demanda (modelagem causal) e também o relacionamento entre diversos produtos.

2. Desenvolvimento da Pesquisa

A metodologia de pesquisa seguiu os seguintes passos para a elaboração do modelo de previsão de vendas:

1. Obtenção dos dados de vendas atomizadas no varejo;
2. Seleção e limpeza dos dados;
3. Análise do cesto de compras para seleção de um grupo de produtos relacionados entre si. A determinação de um grupo de produtos estreitamente relacionados também serviu para responder a uma das questões de pesquisa levantadas: esta informação pode auxiliar na capacidade preditiva dos modelos de previsão baseados em redes neurais?;
4. Tratamento e preparação dos dados para modelagem através de redes neurais;
5. Geração dos modelos baseados em redes neurais, testando-se diferentes combinações de topologia e variáveis de entrada. É selecionada a rede neural com maior acurácia (menor erro nas previsões);
6. Comparação do desempenho medindo-se a acurácia alcançada dentro e fora dos dados de treinamento. Para comparação relativa dos seus resultados, a rede com maior acurácia foi confrontada com as modelagens *naive* e de regressão linear. Tais comparações foram escolhidas por serem técnicas bastante utilizadas para determinação da capacidade preditiva mínima de modelos de previsão (MAKRIDAKIS et al., 1983), e amplamente utilizadas na prática pelas empresas (HUSSEY & HOOLEY, 1995).

2.1. Modelo de Pesquisa

A figura a seguir resume o processo de análise de dados executado.

Figura 17 – Modelo da Pesquisa

Conforme mostrado no modelo acima, a pesquisa foi conduzida por meio da geração de redes neurais artificiais que pudessem prever o comportamento das vendas no nível individualizado, tomando como entrada variáveis explicativas da demanda — a própria demanda histórica, e as variáveis do composto de marketing — agrupadas de acordo com um conjunto de produtos inter-relacionados — pertencentes a um mesmo centro de interesse — selecionados através da técnica de análise do cesto de compras.

O processo completo pode ser observado na figura a seguir, que detalha os passos 1 a 6 descritos anteriormente e resumidos na Figura 17.

Figura 18 – Metodologia conduzida para a Pesquisa

CAPÍTULO IV – PESQUISA EMPÍRICA

1. Contextualização da Pesquisa

A pesquisa realizada foi desenvolvida com base nas vendas de uma grande empresa varejista brasileira, doravante denominada simplesmente EMPRESA. Sendo uma rede varejista de grande porte, os produtos comercializados pela EMPRESA são bastante variados, passando por eletro-eletrônicos, utilidades domésticas, móveis, vestuário, brinquedos, entre outros. A rede de pontos de venda é distribuída em todo o território nacional, totalizando mais de 88.000 m² de área de venda distribuídas em mais de 100 pontos de venda. Além da venda nos pontos de venda, a EMPRESA realiza também a venda através da Internet e por pedidos.

1.1. *Metodologia de Previsão Utilizada pela EMPRESA*

A fim de obter uma visão geral do processo de previsão e da capacidade preditiva atual da EMPRESA, foi aplicado um questionário aos diretores geral, de vendas e de logística, e aos responsáveis pelas áreas de compras e vendas. O questionário — descrito no ANEXO A — visou essencialmente descobrir como é executada a função de previsão dentro da EMPRESA, e se esta função é vista com uniformidade dentro da empresa.

De acordo com as respostas dadas, a previsão é realizada pela EMPRESA de duas formas distintas, e com dois objetivos, por departamentos diferentes. Existe uma previsão de vendas agregada, formulada pela diretoria da empresa anualmente, com base nos resultados anteriores e expectativas para os próximos meses, especialmente macroeconômicas e de desempenho da empresa. Essa previsão serve de base para tomada de decisões estratégicas, como a abertura ou fechamentos de lojas, ou incremento em linhas de produtos. É aferida e recalculada mensalmente. Existe, entretanto, outra previsão de vendas, num nível operacional, que é utilizada basicamente no planejamento de compras, e é de responsabilidade dos gerentes de produto, que são alimentados por dados dos compradores. Os compradores respondem pela previsão (e conseqüentemente pelos pedidos de compra), sendo subordinados pela área de Logística.

A previsão agregada é feita anualmente de forma totalmente qualitativa, tomando-se como base números do ano anterior e a percepção dos gestores. Já a previsão detalhada é realizada por produto, utilizando-se a técnica *naive* simples com a média das vendas dos

últimos 7 dias. Em outras palavras, a previsão de vendas operacional é exatamente igual às vendas dos últimos 7 dias. Como era de se esperar, este modelo gera graves distorções, principalmente porque a EMPRESA externou que suas vendas são caracterizadas por picos acentuados, seja devido a datas comemorativas ou devido a campanhas de marketing muito agressivas e com duração muito curta. A previsão *naive* é então modificada qualitativamente pelos gestores de compras da EMPRESA, tendo em vista seu sentimento com relação ao mercado, *feedbacks* dados pelos fornecedores, pelos vendedores e gerentes, e principalmente tendo em vista desempenhos em anos anteriores.

As previsões não são aferidas, ou seja, a EMPRESA não sabe exatamente quanto está errando nas previsões. Há o sentimento geral de que o erro é grande, principalmente evidenciado pelo excesso de alguns produtos em estoque, enquanto os produtos mais vendidos freqüentemente faltam nas prateleiras. Em resumo, a EMPRESA vê a função de previsão como sendo uma questão muito relevante, mas não vê possibilidade de melhoria imediata.

2. Obtenção dos Dados

A criação de um modelo causal de previsão de vendas quantitativo dá-se a partir da coleta de seqüências temporais de dados sobre as vendas efetuadas, conjugados com informações que possam explicar as vendas. Ou seja, dado um determinado período de tempo, a quantidade de produtos vendidos é sumarizada em totais por período — semanal, por exemplo — e tabulada junto com as demais variáveis explicativas, como o preço praticado no período, a divulgação (esforço de marketing) realizada, dados macroeconômicos, entre outros.

Neste trabalho, o objetivo visado foi a criação de modelos de previsão não-agregados, e, portanto, os dados utilizados foram dispostos no nível de produto. Ou seja, as vendas foram tabuladas por SKU (*Stock Keeping Unit*), que é o identificador único de um produto. Adicionalmente, buscou-se por dados obtidos no varejo, onde o gestor tem maior área de manobra sobre as variáveis independentes e a reação às mudanças sobre o composto de marketing é mais rápida, já que se lida diretamente com o cliente final.

Foram coletados dados relativos às vendas da EMPRESA em nível atomizado — venda a venda — no período de Novembro de 1998 a Junho de 2002, totalizando 43 meses, sendo coletados dados referentes a todos os produtos vendidos em cinco pontos de venda selecionados. O dicionário dos dados coletados encontra-se no ANEXO B.

Os dados coletados estavam distribuídos em múltiplos bancos de dados, incluindo diferentes tabelas constando: dados sobre os clientes, dados relativos às vendas (transação por transação, numa tabela de volume expressivo), dados sobre as transações a prazo, dados sobre as lojas, tabelas de produtos, tabelas de fornecedores, dados dispersos sobre as promoções realizadas, entre outros. Todos estes bancos de dados foram unificados e convertidos. Por exemplo, tabelas existentes em bases de dados no formato IBM DB2 e no formato Excel foram todas convertidas para o formato Microsoft SQL Server, num único banco de dados consolidado.

3. Seleção e Limpeza dos Dados

Buscando detectar comportamentos de venda regionais, optou-se por realizar a análise de todos os produtos comercializados em apenas um ponto de venda. A loja escolhida foi aquela com maior número de registros, e seus dados foram separados dos demais. Os dados foram, num primeiro momento, mantidos com suas escalas originais, já que a primeira análise realizada foi a análise do cesto de compras, que visa apenas o estudo das frequências dos registros para detectar os produtos inter-relacionados, sinalizando os centros de interesse do consumidor.

Obteve-se primeiramente uma visão geral dos dados, incluindo a avaliação de possíveis *outliers* por meio de estudos de frequência e plotagem de gráficos temporais e *boxplot*³. A primeira observação relevante diz respeito à forte curva ABC apresentada, ou seja, apesar de existirem muitos produtos distintos — são exatos 48.287 SKU's em 430 linhas de produtos — apenas uma fração deles possui movimentação frequente. Outra característica importante são as vendas realizadas a prazo, responsáveis por 61,8% do total. As vendas à prazo são especialmente interessantes por tratar-se de transações identificadas — onde há informações precisas sobre o consumidor — que permitem análises de perfis e padrões de consumo.

Em muitos registros foram encontradas inconsistências que forçaram seu descarte do banco de dados, como inconsistências de datas entre a transação de venda a prazo e seu

³ Todas as análises estatísticas e plotagem de gráficos foram realizadas utilizando-se o pacote estatístico SPSS 11.0.

registro na tabela correspondente, juros negativos, estoque sem registro ou negativo, ausência de classificação de produto, entre outros.

Ao final desta etapa, obteve-se um banco de dados unificado e livre de inconsistências, contendo informações sobre todas as vendas realizadas em apenas uma loja, as quais totalizaram 279.149 transações, constituídas por vendas de 8.106 produtos de 558 marcas distintas.

4. Análise do Cesto de Compras

A análise do cesto de compras é uma técnica de mineração de dados utilizada para encontrar grupos de itens que ocorrem simultaneamente (e em que ordem) em transações, conforme mostrado na seção “Indução de Regras / Análise do Cesto de Compras” (pág. 67). Com ela são geradas regras do tipo “se o consumidor compra o produto A , há uma probabilidade P de que ele também compre o produto B ”. A análise do cesto de compras foi utilizada objetivando formar os centros de interesse de compras, ou seja, para encontrar grupos de produtos inter-relacionados. A utilização da técnica de análise do cesto de compras teve como propósitos:

(1) separar um grupo coeso de produtos, minimizando o total de produtos a serem previstos pelo modelo, uma vez que o objetivo da pesquisa foi testar a aplicabilidade das redes neurais e da metodologia de modelagem empregada, e não construir um modelo de previsão para todos os produtos da EMPRESA;

(2) determinar quais produtos fazem parte de um mesmo centro de interesse, e assim investigar se essa informação traz alguma melhoria na capacidade preditiva de modelos de previsão baseados em redes.

Duas estratégias de abordagem dos dados foram tomadas: (1) análise através das transações, buscando quais itens são vendidos em conjunto numa mesma transação; e (2) análise horizontal por cliente, procurando quais produtos um mesmo cliente compra ao longo do tempo. As análises realizadas e os resultados obtidos seguem abaixo⁴.

⁴ Todas as análises foram realizadas com o uso do sistema de mineração de dados *IBM Intelligent Miner 6.1*.

4.1. **Análise do Cesto de Compras por Transação**

A análise do cesto de compras por transação foi executada tomando-se os campos de “série da nota fiscal” e “nota fiscal” concatenados, a fim de se obter um identificador único para cada transação, um identificador do *ticket* de venda. Uma mesma transação — que tem conseqüentemente a mesma nota fiscal — pode conter mais de um produto; aqueles produtos que são vendidos juntos com maior frequência devem, naturalmente, ter maior relação entre si.

As seguintes dimensões foram tomadas na análise do cesto de compras por transação:

- Análise por produto, ou seja, atomizada ao nível de SKU's;
- Análise por linha de produto — que representa os produtos substitutos — de acordo com a classificação adotada pela EMPRESA;
- Análise por setor da loja, de acordo com a classificação adotada pela EMPRESA.

Como mostrado anteriormente, duas medidas são essenciais na análise do cesto de compras: o suporte e a confiança. O suporte representa a porcentagem de vezes em que a regra aparece, ou seja, sua frequência. A confiança representa, numa regra do tipo “se A então B ”, a porcentagem P de transações onde, dado que ocorre A , também ocorre B — a grosso modo, a probabilidade de que quem compra o produto A também irá comprar o produto B . O pacote de mineração de dados utilizado infere todas as regras existentes automaticamente, tomando sempre como entrada o campo que identifica a transação, e quais os parâmetros de suporte e confiança mínimos desejados. Se estes parâmetros forem muito altos, o número de regras gerado é pequeno, não trazendo resultados interessantes; e por outro lado, se o suporte e/ou confiança exigidos forem muito baixos, são geradas regras em demasia, dificultado sua interpretação e análise. Por isso, em todas as análises efetuadas os parâmetros de suporte mínimo e confiança mínima foram selecionados iterativamente, de acordo com o número de regras gerado.

Os resultados foram sempre ordenados segundo o multiplicador *Support * Confidence*, que balanceia modelos com maior suporte e confiança. A melhoria encontrada com cada regra — *lift* — também foi medida, representando o quanto a frequência medida é maior que o esperado. Na modelagem buscou-se por regras com associação positiva — onde a compra do item A influencia positivamente a compra do item B — ou negativa — onde quem compra o item A não compra o item B .

Análise por Produto

Support(%)	Confidence(%)	Rule Support* Confidence	Lift	Type	Rule Body	Rule Head
0.4613	60.0000	0.0028	11.4000	+	[VENT ARNO30]	==> [PANELA PR CLOCK]
0.1538	59.0900	0.0009	11.2300	+	[VENT ARNO30(0,08KW/H)]	==> [PANELA PR CLOCK]
0.0828	70.0000	0.0006	538.0500	+	[PANELA AV NIGRO]	==> [PANELA AV NIGRO]
0.0946	66.6700	0.0006	12.6700	+	[VENT MALLORY 110]	==> [PANELA PR CLOCK]

Tabela 4 – Resultados da análise do cesto de compras por produto, abordados por transação

Como se pode observar na tabela acima, aparentemente há forte relacionamento entre os produtos “ventiladores” e “painéis”. Tomando como exemplo a primeira linha da tabela, vemos que a regra “VENT ARNO30” → “PANELA CLOCK” é positiva, e ocorreu em 0,46% das transações. De cada 100 compradores do ventilador, 60 também compraram a panela, já que a confiança é de 60%. De acordo com a melhoria (*lift*) alcançada, isso representa, uma ocorrência 11,4 vezes maior do que a frequência esperada estatisticamente, dadas as frequências isoladas dos produtos.

Os resultados mostrados na Tabela 4 foram obtidos adotando-se suporte mínimo de 0,05% — um limite inferior consideravelmente baixo. Mesmo assim, o número de regras obtidas foi pequeno — apenas 4⁵ — fato este que deve ser atribuído à grande quantidade de produtos presentes no universo de dados, e conseqüentemente a frequência de cada produto, independente de estar associado ou não, é sempre baixa. Mesmo assim, nas regras obtidas há boa confiança e grande melhoria.

Análise por Linha de Produto

Support(%)	Confidence(%)	Rule Support*Confidence	Lift	Type	Rule Body	Rule Head
0.7688	26.3200	0.0020	3.9300	+	Ventiladores	==> Painéis
0.0710	42.8600	0.0003	47.6800	+	Acessórios para videogames	==> Videogames
0.0710	37.5000	0.0003	25.3700	+	Bolas de jogar	==> Brinquedos

Tabela 5 - Resultados da análise do cesto de compras por linha de produto, abordados por transação

⁵ Outras regras foram geradas pelo software utilizado, mas como seu tipo é neutro e a melhoria diferente de 1, as regras não podem ser consideradas válidas e foram excluídas. O mesmo procedimento foi efetuado em análises semelhantes.

As linhas de produtos são agrupamentos de produtos definidos pela EMPRESA, e foram utilizadas como forma de aglutinar os produtos semelhantes. A agregação de produtos é um artifício usual na análise do cesto de compras, e normalmente faz com que o número de regras distintas encontradas aumente, já que as ocorrências de transações de suporte às regras aumenta com o uso de registros semelhantes agregados.

Os resultados obtidos na análise por linha de produto confirmam as análises por produto, uma vez que de cada 100 compradores de ventiladores (de qualquer marca e modelo), cerca de 26 também compraram painéis. Os resultados também mostram relacionamentos importantes entre acessórios para videogames e videogames; e entre bolas de jogar e brinquedos educativos.

Análise por Setor

Na análise por setor da loja, o número de regras encontradas foi muito baixo, forçando a diminuição dos limites mínimos de suporte e confiança. Adotando-se suporte mínimo de 0,01% e confiança mínima de 5%, obteve-se:

<i>Support(%)</i>	<i>Confidence (%)</i>	<i>Rule Support* Confidence</i>	<i>Lift</i>	<i>Type</i>	<i>Rule Body</i>	<i>Rule Head</i>
0.0118	50.0000	0.0001	100.6500	+	Telefonia + Acessórios de lazer diversos	==> Produtos para casa e outros
0.0118	50.0000	0.0001	3.6000	+	Decoração, cozinha e outros + Cama, mesa, banho e malas	==> Roupas e acessórios para crianças
0.0118	20.0000	0.0000	0.9000	+	Utilidades domésticas + Produtos de lazer diversos	==> Decoração, cozinha e outros
0.0355	8.3300	0.0000	10.5200	+	Aparelhos de som	==> Eletro-eletrônicos e acessórios

Tabela 6 - Resultados da análise do cesto de compras por setor, abordados por transação

Os resultados indicam que a análise por setor não parece ser ideal, dado o grande número de produtos (aparentemente não relacionados) dentro de um mesmo setor, fazendo com que as regras obtidas não sejam fortes. Possivelmente, uma reorganização dos setores, visando a colocação de produtos relacionados dentro de um mesmo setor, traria resultados consistentes com as análises por produto e por linha de produto.

Análises por Taxonomias

Uma taxonomia de produtos representa a divisão hierárquica subentendida entre eles; ou, em outras palavras, os níveis de agrupamentos existentes entre os produtos. De acordo com a taxonomia adotada pela EMPRESA, os produtos são separados por setor, e cada setor é então dividido em várias linhas de produtos. Cada linha de produto é constituída por produtos substitutivos; cada setor, idealmente, deve ser constituído por vários grupos de produtos substitutos, que são complementares entre si, conforme mostrado na seção “Elasticidades cruzadas”, pág. 55.

O software de mineração de dados utilizado permite que seja informada a taxonomia existente entre os produtos, para que ela seja considerada na análise do cesto de compras. Desta forma, as frequências de ocorrência são pesquisadas não só num nível hierárquico específico, mas também de maneira cruzada, buscando relacionamentos entre produtos e linhas de produtos; entre setores e linhas de produtos, e assim por diante.

Além das análises mostradas, outras análises foram efetuadas levando-se em consideração a taxonomia existente entre produtos, linhas de produto e setores. Foram conduzidas análises considerando as taxonomias: (1) Linhas de Produto → Produto; (2) Setor → Linha de Produto; e finalmente (3) Setor → Linha de Produto → Produto. Todos os resultados confirmaram e reforçaram as análises anteriores (por produto, linha de produto e setor, sem considerar a taxonomia entre eles), conforme pode ser observado em algumas das regras evidenciadas, na tabela abaixo.

<i>Support(%)</i>	<i>Confidence (%)</i>	<i>Type</i>	<i>Rule Support* Confidence</i>	<i>Lift</i>	<i>Rule Body</i>	<i>Rule Head</i>
0.4613	60.0000	+	0.0028	11.4000	[VENT ARNO30]	==> [PANELA PR CLOCK]
0.7688	26.3200	+	0.0020	5.0000	Ventiladores em geral	==> [PANELA PR CLOCK]
0.1538	59.0900	+	0.0009	11.2300	[VENT ARNO30(0,08KW /H]	==> [PANELA PR CLOCK]
0.0946	66.6700	+	0.0006	12.6700	[VENT MALLORY 110]	==> [PANELA PR CLOCK]
0.0710	42.8600	+	0.0003	47.6800	Acessórios para videogames	==> Videogames

Tabela 7 – Exemplos de regras da análise do cesto de compras abordado por transação, com uso de taxonomias entre os produtos

4.1.1. Conclusões da Análise do Cesto de Compras por Transação

A partir de todas as análises efetuadas por transação — onde as regras são geradas a partir de uma compra casada de produtos — os seguintes grupos de produtos puderam ser considerados inter-relacionados:

- Ventiladores X Painelas. Aparentemente o relacionamento mais forte, já que pôde ser observado até no nível individual de produtos (Tabela 4), e repetiu-se na análise por linha de produtos com confiança de 26% (Tabela 5), e também na análise com uso das taxonomias, em diversas regras.
- Acessórios para videogames X Videogames. Observada inicialmente na análise por linha de produto com confiança de quase 43% e uma melhoria de mais de 47 vezes (Tabela 5), repetiu-se na análise com uso das taxonomias, como pode ser visto na Tabela 7.
- Bolas de jogar X Brinquedos educativos. Também com forte confiança (37,5%) e melhoria (25,4), segundo a análise por linha de produto (Tabela 5).

4.2. Análise do Cesto de Compras por Cliente

A análise do cesto de compras por cliente visa estabelecer quais produtos um mesmo cliente compra ao longo do tempo. Como muitas das transações da EMPRESA são identificadas (sabe-se quem é o comprador), a análise do cesto de compras pode ser executada horizontalmente, estabelecendo freqüências de compra ao longo do tempo. Ou seja, ao invés de registrar as regras observadas através de compras casadas (numa mesma transação), a análise por cliente busca por regras observadas ao longo do tempo; quando um mesmo cliente, após adquirir o produto *A*, retorna à loja para adquirir o produto *B*.

Para executar esta análise foram levantados os identificadores de clientes: o documento apresentado (RG) ou o número do cartão do cliente na rede da EMPRESA. Uma vez que nem todos os registros dispunham de número do cartão do cliente (apenas os clientes que já realizaram alguma compra à prazo na EMPRESA), os registros foram cruzados e comparados, obtendo-se um identificador único para cada cliente. Mesmo assim, muitos registros (cerca de 40%) não puderam ser identificados: tratam-se de transações anônimas, e que, portanto, não puderam ser analisadas. Por este motivo, e também porque o comportamento de compra dos consumidores pode ser muito diferente quando realiza

compras casadas e quando as faz ao longo do tempo, já era esperado que a análise do cesto de compras por clientes não trouxesse os mesmos resultados que a análise por transações.

Com a criação dos identificadores únicos dos clientes, as compras realizadas por um mesmo cliente ao longo do tempo foram computadas pelo sistema de mineração de dados como pertencentes a uma mesma transação, e assim as análises do cesto de compras por clientes puderam ser conduzidas, levando-se em consideração as mesmas dimensões das análises anteriores.

Análise por Produto

<i>Support (%)</i>	<i>Confidence (%)</i>	<i>Type</i>	<i>Rule Support * Confidence</i>	<i>Lift</i>	<i>Rule Body</i>	<i>Rule Head</i>
0.3194	51.5200	+	0.0016	60.9300	[CANECA OXFORD]	==> [CANECA OXFORD]
0.2067	25.5800	+	0.0005	2.2100	[TEL PADRAO CHAVE IBR]	==> [TEL PADRAO CHAVE IBR]
0.1315	35.0000	+	0.0005	3.0300	[MOT BAND TT GATINHA]	==> [TEL PADRAO CHAVE IBR]
0.0752	57.1400	+	0.0004	4.9400	[POP MAG TOY MC LAREN]	==> [TEL PADRAO CHAVE IBR]
0.0939	45.4500	+	0.0004	3.9300	[TAPETE J SERRANO]	==> [TEL PADRAO CHAVE IBR]
0.0564	75.0000	+	0.0004	665.2500	[SUPORTE GALZERANO]	==> [BANH BABY PLASUTIL]
0.0564	50.0000	+	0.0003	665.2500	[BANH BABY PLASUTIL]	==> [SUPORTE GALZERANO]
0.0564	50.0000	+	0.0003	102.3500	[BOLA VIN PIU PIU]	==> [BOLA VIN VOLLEY]
0.0939	26.3200	+	0.0002	4.5500	[BIC ERG VITALLI]	==> [LIQ WAL 1(0,250KWH)]
0.0564	42.8600	+	0.0002	10.2300	[VIDEO GAME DYNACOM]	==> [FERRO B&D V1(1,20KWH)]
0.0564	42.8600	+	0.0002	3.7100	[PT RS SCHMIDT]	==> [TEL PADRAO CHAVE IBR]
0.0752	25.0000	+	0.0002	57.8500	[CJ ARROZ ATHENAS TRA]	==> [CJ MASSA ATHENAS TRA]
0.0564	30.0000	+	0.0002	114.0400	[CJ COPOS CISPER]	==> [JARRA CISPER]
0.0564	27.2700	+	0.0002	6.5100	[POP LON JUN PEAOGUGU]	==> [FERRO B&D V1(1,20KWH)]
0.0564	27.2700	+	0.0002	2.3600	[POP ROS CAM CEBOLINH]	==> [TEL PADRAO CHAVE IBR]
0.0564	25.0000	+	0.0001	51.1700	[EDUC EL PAU CARIMBOS]	==> [EDUC EL PAU CARIMBOS]
0.0564	25.0000	+	0.0001	27.7200	[PROCES WALITA 110]	==> [LIQ WAL 1(0,350KWH)]

Tabela 8 - Resultados da análise do cesto de compras por produto, abordados por cliente

De acordo com o conjunto de regras formadas, percebe-se que de maneira geral o mesmo cliente retorna à loja para comprar o mesmo produto, ou um produto muito semelhante, ou um produto complementar.

Os relacionamentos encontrados na análise por transação a princípio não se repetiram, conforme pode ser observado na Tabela 8. Tomando por exemplo a última regra da tabela, vemos que um mesmo cliente compra o produto “processador Walita”, e também compra o produto “liquidificador Walita”, sendo que esta regra tem suporte de 0,0564% — total de observações da regra — e confiança de 25% — de cada 4 clientes que compraram o processador, 1 também comprou o liquidificador. A melhoria foi de 27,7, o que significa que a regra ocorreu cerca de 27 vezes mais do que o esperado estatisticamente, de acordo com a frequência de observações dos produtos.

Análise por Linha de Produto

Os resultados da análise horizontal por linha de produtos também diferiram bastante daqueles obtidos na análise por transação, como mostrado abaixo:

Support (%)	Confidence (%)	Type	Rule Support* Confidence	Lift	Rule Body	Rule Head
0.0939	83.3300	+	0.0008	5.9200	Ferros de passar	==> Telefonia
0.0752	100.0000	+	0.0008	32.2500	Liquidificadores + Batedeiras	==> Liquidificadores
0.0752	100.0000	+	0.0008	6.6400	Processadores + Batedeiras	==> Processadores
0.1691	39.1300	+	0.0007	2.6000	Videogames	==> Ferros de passar
0.0752	66.6700	+	0.0005	4.7400	Panelas + Brinquedos	==> Telefonia
0.0939	50.0000	+	0.0005	3.3200	Telefonia	==> Liquidificadores
0.0564	75.0000	+	0.0004	221.7500	Criado mudo	==> Mesas
0.0564	75.0000	+	0.0004	4.9800	Liquidificadores	==> Ferros de passar
0.0564	60.0000	+	0.0003	4.2600	Ferros de passar + Brinquedos	==> Telefonia
0.0939	35.7100	+	0.0003	6.7900	Brinquedos	==> Brinquedos
0.0564	50.0000	+	0.0003	5.0900	Ferros de passar	==> Brinquedos
0.0752	28.5700	+	0.0002	5.4100	Telefonia + Processadores	==> Ferros de passar
0.0564	33.3300	+	0.0002	253.4300	Acessórios para banheiro	==> Banheiras
0.0564	33.3300	+	0.0002	3.0200	Cadeira de bebê	==> Brinquedos

Tabela 9 - Resultados da análise do cesto de compras por linha de produto, abordados por cliente

Análise por Setores

Repetindo os resultados na análise por transações, esta análise não trouxe regras válidas ou significativas, comprovando que a divisão de setores adotada pela EMPRESA parece não corresponder aos relacionamentos existentes entre os produtos, do ponto de vista do comportamento de compra dos consumidores.

Análises por Taxonomias

Repetindo o procedimento adotado na análise por transação, foram conduzidas análises considerando as taxonomias Linhas de Produto → Produto; Setor → Linha de Produto; e finalmente Setor → Linha de Produto → Produto. Os resultados das análises por produto e por linha de produto foram confirmados.

4.2.1. Conclusões da Análise do Cesto de Compras por Cliente

Durante as análises por cliente efetuadas, as regras obtidas nas análises por transação não foram confirmados, indicando que o comportamento de compra do consumidor é diferente quando ele compra ao longo do tempo e quando compra de imediato (compra casada ou compra por impulso). Tendo em vista o objetivo final de previsão de vendas com dados atomizados de produtos relacionados, os resultados da análise de cesto de compras indicaram que as seguintes linhas de produtos merecem ser estudados em conjunto:

<i>Linhas de Produtos</i>
Ventiladores
Processadores
Liquidificadores
Ferros de passar
Batedeiras
Videogames
Acessórios para videogames
Tapetes
Mesa
Criado mudo
Telefonia
Pratos
Panelas
Bicicleta ergométrica
Banheira
Cadeira bebê
Acessórios para banheiro
Bolas
Brinquedos diversos

Tabela 10 – Lista de Linhas de Produtos relacionadas – resultados preliminares

4.3. **Confirmação da Análise do Cesto de Compras**

Uma nova fase de análise foi conduzida buscando confirmar e refinar os resultados. Para esta nova fase de testes, apenas os produtos pertencentes às linhas de produtos aparentemente relacionadas — listadas na Tabela 10 — foram considerados. Todos os registros dos demais produtos foram descartados.

A idéia central por trás desta confirmação de resultados é que separando-se os produtos com alguma afinidade previamente selecionados, e conduzindo-se uma análise do cesto de compras apenas com tais produtos, a frequência de observação das regras deve ser aumentada, já que os demais produtos foram descartados do universo analisado. Assim, aumenta-se o suporte das regras, e é possível notar com mais atenção os relacionamentos mais recorrentes.

Todas as análises anteriores foram repetidas: análises orientadas por transação e por cliente, com base nos produtos, linhas de produtos e setores, e nas taxonomias correspondentes. Os resultados confirmaram a maior parte das regras levantadas anteriormente, sendo que as seguintes linhas de produtos apresentam os relacionamentos mais fortes:

<i>Linhas de Produtos</i>
Ventiladores
Processadores
Liquidificadores
Ferros de passar
Batedeiras
Videogames
Acessórios para videogames
Tapetes
Mesa
Criado mudo
Telefonia
Pratos
Panelas
Brinquedos diversos

Tabela 11 - Lista de Linhas de Produtos relacionadas

A Figura 19 mostra um exemplo de gráfico esquemático dos produtos e relacionamentos encontrados durante as análises do cesto de compras, a título de ilustração dos relacionamentos obtidos. Ela é gerada pelo software de mineração de dados, e mostra através da espessura das setas a força dos relacionamentos obtidos, e através de cores as diferentes análises processadas. Como os resultados incluem análises processadas com uso das taxonomias de produtos, além da sinalização dos produtos, há também em alguns casos a

indicação da linha de produto ou setor da loja. Para a obtenção da figura, segue-se um processo interativo, onde os relacionamentos são plotados pelo sistema de mineração de dados, e são manipulados através de arraste na tela pelo usuário, separando e agrupando os relacionamentos mais importantes.

Figura 19 – Relacionamentos entre produtos detectados nas análises do cesto de compras

Tendo em vista o objetivo final de previsão de vendas com dados atomizados de produtos relacionados, e buscando reduzir o número de produtos a serem considerados nas

previsões, foi realizada uma inspeção visual na Figura 19, buscando por produtos que, apesar de relacionados entre si, não formam um grupo coeso com a maioria dos produtos relacionados.

A inspeção visual mostrou que as seguintes linhas de produtos estão conectadas, formando um grupo coeso — e conseqüentemente um centro de interesse aparente:

<i>Linhas de Produtos</i>	<i>Código</i>
Ventiladores	204
Processadores	212
Liquidificadores	217
Ferros de passar	219
Batedeiras	237
Telefonia	1300
Panelas	1443
Brinquedos diversos	1805, 1810, 1817, 1826, 1838

Tabela 12 – Centro de interesse detectado

Os demais produtos, como por exemplo videogames e acessórios para videogames, apesar de possuírem forte relação entre si, não formam um grupo coeso com os produtos acima, fazendo parte portanto de outro centro de interesse. A fim de minimizar o número de produtos a ser modelado para previsão de vendas, e analisar apenas produtos que estejam relacionados entre si, apenas um centro de interesse foi considerado. Todos os registros dos demais produtos foram descartados.

Dessa forma, concluiu-se a análise do cesto de compras, que teve como consequência direta a diminuição no volume de dados a serem previstos — apenas os produtos com grande afinidade, listados na Tabela 12, foram preservados para as fases de tratamento e previsão seguintes. Além da diminuição na quantidade de produtos analisados, a separação do grupo de produtos pertencentes a um centro de interesse foi utilizada posteriormente como uma informação adicional, na tentativa de melhorar a capacidade preditiva dos modelos de previsão baseados em redes neurais.

5. Tratamento dos Dados para Previsão

O primeiro passo na preparação de dados foi agregar as vendas atomizadas semanalmente, para dessa forma trabalhar com a variação semanal das variáveis independentes. Algumas variáveis tiveram de ser transformadas para representar as peculiaridades do composto de marketing e demais características das vendas ao longo da

semana considerada. Por exemplo, foram tomadas as médias dos preços praticados por produto ao longo da semana, os juros médios por produto foram calculados de acordo com as prestações da venda, e assim por diante. Um novo banco de dados foi então formado, e seu dicionário consta do ANEXO C.

Em resumo, as variáveis independentes disponíveis pela EMPRESA eram:

<i>Variável</i>	<i>Utilização</i>
Semana	Variável de tempo, contabilizando as semanas sequencialmente. A contagem inicia-se em 01/11/1998, considerada como semana 1. A última semana, que se inicia no dia 26/05/2002, é a semana 187.
PrecoMedioSemanal	Média de preço praticado por produto, durante a semana correspondente.
JurosMedioSemanal	Taxa de juros média praticada por produto, durante a semana correspondente.
Pprazo	Porcentagem de vendas realizadas à prazo na semana anterior.
Id_produto, Produto, Linha, Setor, Id_marca	Variáveis de identificação do produto.
Classe, Classe_distribuicao	A EMPRESA adota duas classificações distintas para um mesmo produto: uma baseada no giro do produto, e outra baseada no tipo de compra/distribuição a que ele está submetido. Ambas, entretanto, identificam parâmetros de alta, média ou baixa venda, venda sob encomenda, venda sem mostruário, venda direta do fornecedor, e produto fora de linha.
Comissao	Indica se o produto entra na comissão dos vendedores na semana correspondente.
Estoque, Estoque_dias, Reposição_dias, Mostruario, Grade Campanha, Camapanha_agressiva, Veiculações	Representam a disponibilidade do produto, seja em estoque, em mostruário, ou em prazo de entrega/reposição. Variáveis que medem o esforço promocional efetuado na semana. Para esforços corporativos, todos os produtos receberam o mesmo valor para a variável “Campanha” no período correspondente; as campanhas agressivas focaram apenas alguns grupos de produtos; enquanto “Veiculacoes” mede a exposição das campanhas na mídia.
Variáveis com lag de tempo	Foram testadas variáveis com <i>lag</i> de tempo para aquelas onde se detectou correlações horizontais.

Tabela 13 – Principais variáveis independentes usadas nos modelos de previsão semanal

Com a sumarização dos produtos pertencentes ao centro de interesse evidenciado, e agregação dos dados em semanas, restaram 12.560 registros representando as vendas semanais de 881 produtos distintos.

Foram realizadas investigações exploratórias nos dados, onde se detectou que as linhas de produtos estão relativamente equilibradas quanto ao número de registros, com certa predominância de transações nas linhas de brinquedos. Foram traçados *boxplots* das vendas por produto, linha de produto e setor. Todos mostraram muitos *outliers*, inviabilizando a retirada manual dos *outliers*. Todos os desvios-padrão mostraram-se altos, o que indica potencialmente uma baixa previsibilidade nos dados.

Figura 20 – *Boxplot* do total de vendas semanal por setor, indicando a presença de muitos *outliers*

A Figura 20, por exemplo, mostra o total de produtos vendidos numa mesma semana, separados por setor. No eixo horizontal, vê-se os quatro setores selecionados (setor 2, setor 13, setor 14 e setor 18), com o total de registros em cada setor. Como pode ser observado na figura, não são evidenciados os quartis usuais do gráfico *boxplot*; ao invés disso, o que se vê são muitos *outliers* (mostrados como números correspondentes às semanas em que ocorreram), e uma média de vendas semanal por produto próxima de zero.

Também foram traçados gráficos lineares de vendas semanais e mensais por linha e setor. Os gráficos mostraram que há forte sazonalidade, e os picos de vendas são muito expressivos, conforme pode ser visto na Figura 21. A figura mostra as vendas médias semanais por setor da EMPRESA, cada setor representado por uma cor diferente. Como pode ser visto em algumas semanas (como no fim do ano, nas semanas 49 a 52) ocorrem picos de vendas acentuados em todos os setores, porém os picos diferem de um setor para o outro, o mesmo se repetindo para as linhas de produtos analisadas.

Figura 21 – Médias de vendas semanais por setor, indicando forte sazonalidade e picos de vendas

5.1. Transformações Realizadas

Vários autores mostraram que os padrões nos dados cuja distribuição difere muito da normal são de difícil assimilação por modelos quantitativos, incluindo-se aqueles gerados com uso de técnicas de mineração de dados (TUKEY et al., 1983; BERRY & LINOFF, 1997; HAND, 1998). A fim de contornar esta limitação, é comum a aplicação de transformações nos dados, para que estes tenham uma distribuição mais próxima da normal, e assim possam gerar modelos mais precisos (mais aderentes à realidade).

A fim de avaliar a necessidade de transformações nos dados, todas as variáveis numéricas foram apreciadas, sendo realizadas as seguintes operações:

- Plotagem no tempo;
- Construção de *boxplot* total, separado por setor e por linha;
- Construção de histogramas total, separado por setor e por linha.

Todas as variáveis, com exceção de “JurosMedioSemanal”, apresentaram comportamentos não-desejados, com distribuições muito distantes da normal. Partiu-se então para uma transformação nas variáveis. As transformações necessárias foram obtidas através do gráfico dispersão versus nível (LogM X LogD_f), de acordo com o procedimento descrito por TUKEY et al. (1983):

$$\text{Log } D_f = k + b \cdot \text{Log } M \text{ (Equação XV)}$$

onde M é a mediana da variável tomada — neste caso, por linha de produto — e D_f é a medida do inter-quartil da mesma variável. Uma vez traçada a melhor reta para a Equação XV, é aproximado o valor de *b*, que indica a melhor transformação nos dados:

$$\text{Transformação: } Z = x^{1-b} \text{ (Equação XVI)}$$

A descrição completa das transformações testadas e resultados obtidos encontra-se no ANEXO D. A tabela a seguir sintetiza as transformações que melhor ajustaram os dados:

<i>Variável</i>	<i>Transformação</i>	<i>Nova variável</i>	<i>Testar original?</i>
Estoque	$Z = \log x$	Estoque_log	N
Estoque_dias	$Z = \frac{1}{x^2}$	Estoque_dias-2	S
Reposicao_dias	Nenhuma		
Mostruario	$Z = \log x$	Mostruario_log	S
Grade	$Z = \log x$	Grade_log	n
Total7	$Z = \log x$	Vendas_log	n
PrecoMedioSemanal	$Z = x^{0,2}$	Preco_pot	n
Pprazo	Nenhuma		
JurosMedioSemanal	Nenhuma		

Tabela 14 – Transformações aplicadas aos dados

5.2. Testes de Correlações

Foram realizados testes de correlação entre todas as variáveis independentes a serem utilizadas, a fim de visualizar aquelas com maior impacto para a modelagem e estabelecer uma linha mestra para a seleção das variáveis a serem testadas nos modelos baseados em redes neurais.

A matriz de correlação entre as variáveis independentes mostrou forte relacionamento entre:

Correlações entre variáveis independentes		Correlação Pearson	Nível significância
Estoque_log	Grade_log	0,590	0,01
Mostruario_log	Preco_pot	-0,459	0,01
Classe_distribuicao	Classe	0,556	0,01

Tabela 15 – Correlações significativas detectadas entre as variáveis independentes

Como as correlações *Pearson* são altas (maiores que 0,4 em módulo), nos modelos criados foi tentada a retirada de uma das variáveis relacionadas. Isso torna a modelagem mais simples e com maior capacidade de extrapolação, e supõe-se que não ocasione perda de capacidade preditiva e (GROSS & PETERSON, 1983).

Já as correlações entre as variáveis independentes e a variável dependente (total de vendas na semana) resultaram em apenas duas correlações estatisticamente significativas:

Variável independente	Correlação Pearson com Vendas	Nível significância
Estoque_log	0,439	0,01
Pprazo	-0,648	0,01

Tabela 16 - Correlações significativas entre as variáveis independentes e a variável dependente

Visando confirmar visualmente os resultados das análises de correlação, foram desenhados gráficos do tipo *scatterplot* para todas as variáveis correlacionadas. Os gráficos mostraram claramente que as correlações são verdadeiras, com exceção possivelmente da correlação Classe_distribuição X Classe, onde não é possível identificar padrões no gráfico. Os gráficos *scatterplot* das correlações estão mostrados no ANEXO E.

5.3. Testes de Autocorrelação e Correlações com Lag de Tempo

Quanto maior o coeficiente de correlação (em módulo), mais forte a associação entre duas variáveis. Entretanto, não basta analisar apenas as correlações das variáveis num dado instante — numa determinada semana, como é o caso desta pesquisa. Uma variável pode ter impacto em outra apenas depois de determinado período de tempo, ou seja, sua mudança ou ocorrência pode afetar as vendas após determinado período (*lag*) de tempo.

A fim de detectar possíveis associações com *lag* de tempo, faz-se necessário o estudo das autocorrelações — correlação da variável dependente com ela mesma — e das correlações com lag de tempo — correlações das variáveis independentes com as vendas, sendo as primeiras tomadas com *lag* de tempo.

As correlações parciais com o total de vendas na semana foram computadas para todas as variáveis, sendo obtidos os seguintes resultados estatisticamente significativos:

<i>Variável</i>	<i>Semanas anteriores correlacionados com vendas</i>
Grade	1,4
Estoque	1
PrecoMedio	2,3
JurosMedio	4
Vendas	1, 2, 10, 42, 52

Tabela 17 – Autocorrelações e correlações parciais verificadas

Os resultados indicaram a necessidade de modelagem com a utilização das variáveis acima introduzidas com os *lags* de tempo respectivos, uma vez que elas podem ampliar a capacidade preditiva dos modelos. Dessa forma, serão testadas, por exemplo, modelagens onde a variável “Grade” é introduzida com *lags* de tempo de 1 e 4 semanas, indicando que, possivelmente, uma alteração no seu valor afeta as vendas após 1 semana, e também após 4 semanas. Somente através de um processo de tentativa e erro — introduzindo as variáveis acima com os *lags* respectivos nos modelos de previsão — é que se poderá verificar se há melhoria na capacidade preditiva. As autocorrelações são apenas um indicativo dos *lags* a serem testados.

6. Modelagem com Uso de Redes Neurais

Uma vez preparados os dados e selecionadas as variáveis a serem utilizadas, deu-se início ao processo de criação dos modelos de previsão baseados em redes neurais artificiais. A criação da rede neural propriamente dita é feita através de um processo totalmente automático nos pacotes de mineração de dados, como foi o caso do software utilizado⁶. Uma vez fornecidas as entradas (normalmente importadas através de Bancos de Dados ou mesmo arquivos texto), e saídas a serem previstas (no caso, as vendas de cada produto, semana a semana), o software de geração de redes neurais inicia um processo de treinamento automático, onde são testadas inúmeras topologias de rede e parâmetros de treinamento.

Após o treinamento (que para o conjunto de dados utilizado levou de 3 a 12 horas, por rede neural) o sistema fornece qual a melhor rede obtida — número de camadas intermediárias, número de nós nestas camadas, quais parâmetros de treinamento foram utilizados — e quais os resultados que ela gera para o conjunto de dados de treinamento: coeficiente de ajuste (R^2), erro quadrado médio, desvio-padrão dos erros, entre outros. Com a

⁶ Todos os modelos foram gerados com o uso do software de criação de redes neurais artificiais *Megaputer PolyAnalyst 4.5*, que gera redes *Perceptron* multicamadas.

rede já treinada, pode-se utilizá-la para prever outros conjuntos de dados, verificando sua real performance.

A primeira modelagem realizada foi dirigida somente à avaliação das transformações utilizadas. Duas redes foram construídas, tomando como entrada as mesmas variáveis, porém, na primeira rede as entradas foram tomadas como variáveis literais (sem transformação), enquanto na segunda, elas foram transformadas como mostrado na Tabela 14. As redes foram construídas utilizando-se todos os dados disponíveis — os softwares modernos de redes neurais já separam os dados em grupos de treinamento e validação — para previsão com prazo de uma semana.

Os resultados mostraram claramente os benefícios advindos das transformações aplicadas aos dados, como mostra a Tabela 18. A rede 1 — sem transformações — obteve Erro Quadrado Médio (MSE) superior, e ajuste aos dados de treinamento sensivelmente inferior à rede 2 — com todas as entradas numéricas transformadas. Todos os modelos gerados a partir de então utilizaram como entradas as variáveis transformadas.

<i>Rede</i>	<i>Variáveis de entrada</i>	<i>Variável Prevista</i>	<i>Nós</i>	<i>Camadas</i>	<i>R2</i>	<i>MSE⁷</i>
1	Id_produto, Comissao, Setor, Linha, Id_marca, Classe, Classe_distribuicao, Estoque, Estoque_dias, Reposicao_dias, Mostruario, Grade, Semana, PrecioMedioSemanal, JurosMedioSemanal, Pprazo	Venda7	3	3	0,2194	17,5000
2	Id_produto, Comissao, Setor, Linha, Id_marca, Classe, Classe_distribuicao, Estoque_log, Estoque_dias-2, Reposicao_dias, Mostruario_log, Grade_log, Semana, Precio_pot, JurosMedioSemanal, Pprazo	Vendas_log	14	3	0,6052	1,9306

Tabela 18 – Resultados das redes 1 e 2, comparando os efeitos das transformações nos dados

6.1. Seleção de Variáveis de Entrada

Para as variáveis cuja transformação não se mostrou suficiente para aproximar os dados da curva Normal — “Estoque_dias” e “Mostruario” — foram gerados modelos com todas as variáveis transformadas, com exceção delas, uma a uma e em conjunto:

- Rede Neural 3 – variável “Estoque_dias” sem transformações;

⁷ O erro MSE teve foi ajustado à variável original, para fins de comparação na mesma unidade de medida.

- Rede Neural 4 – variável “Mostruario” sem transformações.

Os resultados provaram que, mesmo quando a transformação parece não ser adequada, há melhoria no desempenho dos modelos de previsão baseados em redes neurais, uma vez que todos os modelos tiveram desempenho inferior àquele obtido com a rede 2 — com todas as entradas transformadas.

Também foi testada a retirada das variáveis com forte correlação entre si — “Estoque” X “Grade” e “Mostruario” X “Preco” (veja Tabela 15). Os resultados mostraram melhorias quando se retira uma das variáveis correlacionadas. As melhorias foram mais profundas na retirada isolada da variável “Grade”, o que de certa forma era esperado uma vez que a quantidade em estoque tem forte correlação com as vendas (veja Tabela 16). Os novos modelos construídos foram:

- Rede Neural 5 – variável “Estoque_log” retirada;
- Rede Neural 6 – variável “Grade_log” retirada;
- Rede Neural 7 – variável “Mostruario_log” retirada;
- Rede Neural 8 – variável “Preco_pot” retirada;
- Rede Neural 9 – variáveis “Mostruario_log” e “Grade_log” retiradas (provocaram melhorias mais sensíveis dentre as anteriores).

Rede	Variáveis de entrada	Variável Prevista	Nós	Camadas	R2	MSE
6	Id_produto, Comissao, Setor, Linha, Id_marca, Classe, Classe_distribuicao, Estoque_log, Estoque_dias-2, Reposicao_dias, Mostruario_log, Semana, Preco_pot, JurosMedioSemanal, Pprazo	Vendas_log	3	3	0,6127	1,9187

Tabela 19 - Resultados da rede neural 6, sem a presença da variável “Grade”

Como dito anteriormente, os softwares de geração de redes neurais modernos têm a capacidade de testar inúmeras topologias de rede e configurações de treinamento automaticamente, fornecendo aquela com melhor ajuste (R^2) possível. Como se pode observar nos resultados da Rede 6 (Tabela 19), a melhor topologia obtida foi aquela com 3 camadas — 1 camada intermediária — sendo que a camada intermediária foi composta de apenas 3 nós. O erro nas previsões é dado pelo valor do MSE (Erro Quadrado Médio). Nos modelos criados a partir de variáveis transformadas o erro não pode ser lido literalmente (em termos de unidades

de produtos), já que a diferença entre o previsto e o real foi tomada numa variável transformada (logarítmica, no caso).

Além das transformações visando à melhoria no comportamento das variáveis de entrada, alguns autores sugerem também o procedimento de padronização das entradas (WRAY et al., 1994), como forma de deixar todas as variáveis numa mesma ordem de grandeza, o que, dadas as características inerentes à forma de treinamento e operação das redes neurais, tende a melhorar seu desempenho. Outros dois modelos foram então gerados, para:

- Entradas numéricas padronizadas entre 0 e 1 (rede 10);
- Todas as entradas padronizadas entre 0 e 1, inclusive as categóricas (rede 11).

Os resultados mostraram que os procedimentos de normalização e padronização são benéficos para o desempenho das redes neurais. Entretanto, os modelos não conseguiram assimilar as variáveis categóricas como numéricas padronizadas, piorando seu desempenho. Isso indica um bom trabalho dos softwares em lidar com as variáveis categóricas, mesmo que estas estejam representadas por números com ordem de grandeza muito superior às demais variáveis. A rede neural com melhor desempenho foi aquela com todas as variáveis numéricas padronizadas entre 0 e 1:

<i>Rede</i>	<i>Variáveis de entrada</i>	<i>Variável Prevista</i>	<i>Nós</i>	<i>Camadas</i>	<i>R2</i>	<i>MSE</i>
10	Id_produto, Comissao, Setor, Linha, Id_marca, Classe, Classe_distribuicao, Estoque_logP, Estoque_dias-2P, Reposicao_diasP, Semana, Preco_potP, JurosMedioSemanalP, PprazoP, Mostruario_logP	Vendas_log	26	4	0,6176	1,9108

Tabela 20 - Resultados da rede neural 10, com as entradas numéricas normalizadas e padronizadas

Finalmente, buscou-se o tratamento dos registros contendo valores faltantes (*missing*) de todas as variáveis, com base nos demais registros existentes. Vários métodos de preenchimento de valores faltantes, como tendência linear e média foram tentados, mas em nenhum deles obteve-se modelos com capacidade preditiva superior àquela da rede 10.

6.2. Incorporação de Variáveis de Entrada

Após a determinação do conjunto mínimo de variáveis independentes e da melhor forma de tratamento das entradas, deu-se início ao processo de incorporação de novas

variáveis que pudessem explicar as vendas, na tentativa de melhorar a capacidade preditiva dos modelos através do fornecimento de informações adicionais, aumentando a complexidade dos modelos.

Foram geradas redes neurais incorporando-se as informações complementares a seguir.

1. Dados sobre a sazonalidade e passagem do tempo:

- “Semana_ano”, representando a semana do ano, variando de 1 a 52;
- “Mes_ano”, representando o mês do ano, variando de 1 a 12;
- 12 entradas binárias (0 ou 1), representando os meses do ano;
- 52 entradas binárias (0 ou 1), representando as semanas do ano;

➤ Não houve melhorias com o acréscimo destas variáveis. A rede com melhor resultado foi aquela com a medida original e isolada da variável “Semana”, mostrando que o modelo original foi capaz de assimilar a sazonalidade.

2. Análise de *lags* de tempo das vendas (seguindo resultados da Tabela 17):

- Vendas da semana anterior;
- Vendas de 2ª semana anterior;
- Vendas de 10ª semana anterior;
- Vendas da 42ª semana anterior;
- Vendas da 52ª semana anterior;
- Vendas da 52ª, 42ª, 10ª, 2ª e semana imediatamente anterior;
- Todos os casos anteriores conjugados;
- Todos os casos anteriores, preenchendo-se valores *missing*;

➤ Os resultados mostraram que houve melhoria com a adoção das variáveis de vendas com *lag* de tempo, desde que os *missing values* fossem preenchidos. Neste caso o valor adotado nos registros faltantes foi igual ao último (próximo no tempo) valor conhecido, ou igual a 0 se não houver nenhum outro valor disponível. Ressalta-se que a melhoria de desempenho foi visível apenas nos lags de 1 e 2 semanas, sendo muito mais pronunciado com o uso apenas do *lag* de 1 semana (Rede 26), e sendo indiferente a adoção de ambos.

3. Análise de lags de tempo das variáveis independentes (conforme Tabela 17):

- Grade do produto (quantidade especificada a ser mantida por loja) da semana anterior;
 - Grade do produto na 4ª semana anterior;
 - Estoque na semana anterior;
 - Preço na 2ª semana anterior;
 - Preço na 3ª semana anterior;
 - Juros médio na 4ª semana anterior;
 - Todos os casos anteriores, preenchendo-se valores *missing*;
- Não houve resultados positivos nos modelos gerados com o uso dos lags cruzados. Eles foram então descartados.
4. Incorporação de dados de promoção (esforço de marketing):
- Dados de campanhas promocionais institucionais, atribuídos com valores 0 ou 1 para todos os produtos;
 - Dados de campanhas de marketing agressivas, direcionadas somente a alguns produtos;
 - Número de veiculações na mídia;
 - Todos os casos anteriores conjugados;
- As variáveis promocionais não trouxeram melhora na capacidade preditiva das redes. A explicação mais provável deve-se à dificuldade em atribuir a um produto a campanha realizada, já que os registros de campanhas realizadas não eram completos nem precisos. Isso obrigou a atribuir a campanha a toda uma linha ou mesmo setor de produtos na maior parte dos casos, “confundindo” a rede.

No total, foram geradas 66 redes neurais, com a incorporação das variáveis acima citadas. Todas utilizaram o mesmo conjunto de dados, sempre prevendo para uma semana à frente, e utilizando entradas numéricas normalizadas e padronizadas.

Em resumo, de todas as variáveis incorporadas ao modelo visando ampliar as informações disponíveis sobre o composto de marketing e sobre a sazonalidade, apenas a informação sobre o total do produto vendido na semana anterior interferiu significativamente na capacidade preditiva da rede neural. Esta variável passou a ser incorporada em todas as análises posteriores.

<i>Rede</i>	<i>Variáveis de entrada</i>	<i>Variável Prevista</i>	<i>Nós</i>	<i>Camadas</i>	<i>R2</i>	<i>MSE</i>
26	Id_produto, Comissao, Setor, Linha, Id_marca, Classe, Classe_distribuicao, Estoque_logP, Estoque_dias-2P, Reposicao_diasP, Semana, Preco_potP, JurosMedioSemanalP, PprazoP, Mostuario_logP, Lag1	Vendas_log	7	3	0,6585	1,8441

Tabela 21 - Resultados da rede neural 26, com incorporação das vendas na semana anterior e *missing values* preenchidos

6.3. Incorporação de Informações do Centro de Interesse

Um dos objetivos da presente pesquisa é avaliar se um modelo de previsão de vendas baseado em redes neurais artificiais é capaz de detectar os relacionamentos presentes entre os produtos que fazem parte de um centro de interesse do consumidor — grupo de produtos inter-relacionados, como aqueles selecionados durante o processo de análise do cesto de compras (veja “Análise do Cesto de Compras”, pág. 87).

Apesar dos modelos anteriormente construídos levarem em consideração apenas os produtos do centro de interesse detectado (Tabela 12), nenhuma informação foi apreciada para incorporar as informações sobre os possíveis relacionamentos existentes entre os produtos, uma vez que as redes foram treinadas com registros individuais dos produtos. Para que o conhecimento acerca do centro de interesse seja considerado, faz-se necessária a utilização de novas entradas, que representem as informações agregadas sobre o grupo de produtos considerado.

Os seguintes dados foram aglutinados na tentativa de representar o conhecimento sobre as elasticidades cruzadas entre os produtos:

- Vendas totais da linha de produtos correspondente (rede 79);
- Vendas totais do setor da loja correspondente (rede 80);
- Vendas totais de todo o centro de interesse (rede 81);
- Preço médio da linha de produtos correspondente (rede 82);
- Preço médio do setor da loja correspondente (rede 83);
- Preço médio de todo o centro de interesse (rede 84);
- Diferença de preço entre o produto e a média da linha de produto (rede 85);
- Diferença de preço entre o produto e a média do setor da loja (rede 86);
- Diferença de preço entre o produto e a média de todo o centro de interesse (rede 87).

- Idem, para entradas padronizadas (redes 88 a 105);
- Idem, com preenchimento de *missing values* assumindo o último valor existente (redes 106 a 141).

Os resultados mostraram uma grande melhora das redes ao se utilizar a informação de total de vendas dos produtos substitutos (pertencentes a uma mesma linha de produtos, e, portanto, produtos concorrentes diretos), ao mesmo tempo em que não houve melhoria quando da utilização dos preços médios nem das diferenças de preços. A rede neural 108, cujos resultados são mostrados a seguir, comprova que a incorporação de informações a respeito de produtos inter-relacionados pode trazer benefícios na previsão de vendas com uso de redes neurais artificiais.

Rede	Variáveis de entrada	Variável Prevista	Nós	Camadas	R2	MSE
108	Id_produto, Comissao, Setor, Linha, Id_marca, Classe, Classe_distribuicao, Estoque_logP, Estoque_dias-2P, Reposicao_diasP, Semana, Preco_potP, JurosMedioSemanalP, PprazoP, Mostuario_logP, Lag1, Vendas_linhaP	Vendas_log	7	3	0,7164	1,7457

Tabela 22 - Resultados da rede neural 108, com incorporação do total de vendas de produtos substitutos (*missing values* preenchidos)

6.4. Conclusões da Modelagem Baseada em Redes Neurais

Após o processo de modelagem, foi observado que o melhor desempenho das redes neurais artificiais até o momento (rede 108), no conjunto de dados disponível, foi aquele obtido com:

- Dados transformados, normalizados e padronizados entre 0 e 1;
- *Lag* de tempo das vendas, com 1 semana de diferença;
- Média de vendas da linha de produtos (produtos substitutos);

A tabela a seguir resume o processo de geração de redes neurais para previsão executado:

<i>Redes</i>	<i>Testes realizados</i>	<i>Melhor rede</i>	<i>MSE da melhor rede</i>
1 ~ 4	Comparação entre entradas transformadas e não transformadas	2 – Todas as entradas transformadas	1,9306
5 ~ 9	Teste de retirada de variáveis correlacionadas	6 – Retirada isolada da variável “Grade”	1,9187
10 ~ 11	Comparação entre entradas padronizadas e não padronizadas	10 – Entradas numéricas padronizadas entre 0 e 1	1,9108
12 ~ 78	Incorporação de novas variáveis para representar: sazonalidade, autocorrelações das vendas, correlações com <i>lags</i> de tempo, dados promocionais	26 – Incorporação das vendas na semana anterior (autocorrelação com <i>lag</i> de 1 semana)	1,8441
79 ~ 141	Incorporação de informações agregadas sobre o centro de interesse	108 – Incorporação do total de vendas de produtos substitutos (<i>missing values</i> preenchidos)	1,7457

Tabela 23 – Resumo das redes neurais geradas e performance obtidas

O Modelo 108 foi tomado então como referência, para medição de capacidade preditiva frente a dados desconhecidos e frente a outras modelagens. A fim de se avaliar a capacidade preditiva do modelo baseado em redes neurais, o conjunto de dados foi dividido em dois grupos:

1. Grupo de criação do modelo, constituído internamente por registros de treinamento e validação (automaticamente separados pelo software de mineração de dados). Para criação do modelo, foram utilizados os registros de Novembro de 1998 a Maio de 2001, num total de 135 semanas consecutivas, sempre prevendo para uma semana à frente.
2. Grupo de avaliação externo, simulando a utilização real do modelo de previsão em dados “desconhecidos”, nunca apresentados à rede neural. Os registros subsequentes à semana 135 (Maio de 2001 a Junho de 2002), totalizando 3140 casos, foram usados para pôr à prova a capacidade preditiva do modelo gerado.

Após a criação da rede com o conjunto de dados reduzidos, a mesma foi testada frente aos registros desconhecidos, e os resultados obtidos foram transformados de volta à unidade de medida original, com a retirada da padronização, normalização e função logarítma, alcançando assim a previsão de quantos produtos serão vendidos, semana a semana, por SKU. Os resultados geraram uma nova medida de erro, uma vez que agora estão tomados por unidades de produtos vendidos.

<i>Modelo</i>	<i>Dados</i>	<i>MSE</i>	<i>RMSE</i>	<i>MAPE</i>
Rede 108	Dados de criação (até semana 135)	55,84	10,33	31,61%
Rede 108	Dados de validação (a partir da semana 136)	108,12	10,40	41,21%

Tabela 24 – Capacidade preditiva do Modelo Baseado em Redes Neurais (rede 108)

De acordo com a Tabela 24, o modelo baseado em redes neurais erra em sua previsão de vendas, em média, um total de 10,4 unidades de produto para previsão de uma semana à frente, o que corresponde a um erro médio percentual de 41,21%.

Se a modelagem for eficiente (capaz de “aprender” com os erros do passado), não deve haver correlação em série entre os erros de um período para outro (PUTSIS, 1998). Os testes de autocorrelação e autocorrelação parcial não mostraram autocorrelação nos erros, indicando que o modelo é eficiente. Se a modelagem for não-enviesada, haverá um padrão consistente de erros para mais e para menos, e conseqüentemente a distribuição dos erros deve ter uma forma próxima da normal (LAWRENCE et al., 2000). O modelo gerado parece também respeitar esta premissa, como pode ser visto na Figura 22.

A Figura 23 mostra a performance da Rede 108 ao tentar prever as vendas das semanas posteriores à 135, que não foram utilizadas para treinamento da rede.

Figura 22 – Histograma da dispersão dos erros, na escala da variável prevista

Figura 23 – Gráfico de vendas previstas X vendas reais

7. Comparação com Outras Modelagens

Com o intuito de estabelecer uma referência para a capacidade preditiva do modelo baseado em redes neurais, duas comparações foram realizadas: com a modelagem *naïve* não ajustada e com a regressão linear usando como entradas as mesmas variáveis da rede neural 108.

7.1. Comparação com a Modelagem Naïve

Um modelo de previsão *naïve* não ajustado é aquele onde se toma o último registro conhecido como sendo a previsão para o próximo período, sem considerar a sazonalidade da variável dependente. A comparação com a previsão *naïve* é uma referência comum encontrada na literatura a fim de medir o ganho de performance ao se adotar modelagens mais sofisticadas (LAWRENCE et al., 2000). É também importante já que é o método de previsão atualmente utilizado na EMPRESA onde os dados foram tomados.

Observou-se que a modelagem baseada em redes neurais artificiais obteve desempenho cerca de 54% superior à previsão *naïve*, tomando-se por base a raiz do erro quadrado médio (RMSE), e 64% superior se for considerado o erro médio percentual (MAPE), ponderando os resultados para os dados de validação.

<i>Modelo</i>	<i>Dados</i>	<i>MSE</i>	<i>RMSE</i>	<i>MAPE</i>
<i>Naive</i>	Dados de criação (até semana 135)	133,98	11,58	126,24%
<i>Naive</i>	Dados de validação (a partir da semana 136)	506.44	22,50	116.09%

Tabela 25 – Capacidade preditiva da previsão *naive* não ajustada

7.2. Comparação com a Regressão Linear

A regressão linear talvez seja a técnica mais utilizada para previsão quantitativa, por isso também é uma boa referência comparativa para avaliação do desempenho das redes neurais na previsão de vendas atomizada. Para esta regressão, foram utilizadas as mesmas variáveis do Modelo 108, que obteve a melhor capacidade preditiva dentre aqueles testados. O procedimento de tratamento e separação também seguiu aquele do Modelo 108, com a utilização de dados normalizados e padronizados, e separação em dois grupos de dados: dados de criação até a semana 135 e avaliação do desempenho com dados da semana 136 à semana 187.

O software utilizado para a regressão⁸ gera, a partir da matriz de correlações, várias equações de regressão, utilizando como variáveis independentes apenas aquelas com maior impacto no modelo, selecionadas através do método *stepwise*. Foi gerado um total de 5 equações de regressão válidas, mostradas abaixo:

<i>Equação</i>	<i>Variáveis Independentes</i>
1	Vendas da linha de produtos
2	Vendas da linha de produtos, porcentagem de vendas a prazo
3	Vendas da linha de produtos, porcentagem de vendas a prazo, Vendas no período anterior
4	Vendas da linha de produtos, porcentagem de vendas a prazo, Vendas no período anterior, Estoque
5	Vendas da linha de produtos, porcentagem de vendas a prazo, Vendas no período anterior, Estoque, Prazo de reposição

Tabela 26 – Equações de regressão linear geradas

Levando em consideração o grupo de dados de validação, todos eles desempenharam pior que o modelo baseado em redes neurais, como mostrado na tabela abaixo:

<i>Equação</i>	<i>MSE</i>	<i>RMSE</i>	<i>MAPE</i>
1	283,65	16,84	79,52%
2	271,26	16,47	58,05%
3	264,58	16,27	51,97%
4	248,29	15,76	50,15%
5	189,03	13,75	49,10%

Tabela 27 – Capacidade preditiva das equações de regressão linear

⁸ As análises de regressão foram realizadas utilizando-se o pacote estatístico SPSS 11.0.

Avaliando-se a equação de regressão com maior capacidade preditiva — Equação 5 — observa-se que ela possui erro de ordem 32% maior que o modelo baseado em redes neurais, comparando-se o erro quadrado médio, e 19,1% maior em termos percentuais.

7.3. *Resumo do Desempenho das Redes Neurais*

A tabela a seguir resume o desempenho da técnica de redes neurais artificiais frente às técnicas de previsão *naive* e regressão linear, mostrando que as redes neurais obtiveram erros menores que as demais modelagens, seja qual for a medida de erro adotada.

<i>Técnica de Previsão</i>	<i>MSE</i>	<i>RMSE</i>	<i>MAPE</i>
<i>Naive</i> não ajustada	506,44	22,50	116,09 %
Regressão linear – equação 5	189,03	13,75	49,10 %
Redes neurais – rede 108	108,12	10,40	41,21 %

Tabela 28 – Performance comparativa das redes neurais

CAPÍTULO V - CONCLUSÕES

1. Resultados Alcançados

Este trabalho investigou acerca da possibilidade de utilização de redes neurais artificiais para previsão de vendas de produtos no nível atomizado. O objetivo geral foi o estabelecimento de uma metodologia capaz de identificar padrões de relacionamento úteis na previsão de vendas individual no varejo. Tais relacionamentos, uma vez assimilados por um modelo de previsão causal, permitem ao gestor a identificação do composto de marketing adequado (preço, produtos e promoção), através da visualização do impacto nas vendas a partir de simulações de mudanças no composto de marketing.

Durante o trabalho de revisão bibliográfica, observou-se que muitos autores já estudaram a previsão de demanda agregada — para toda corporação — utilizando diversas técnicas de previsão e modelagem de dados, e prevendo em diferentes horizontes de tempo (TANG et al., 1991; LUXHOJ et al., 1996; ANSUI et al., 1996; KUO & XUE, 1998; SPEDDING & CHAN, 2000, entre outros). Tais estudos mostraram que a previsão quantitativa pode ser bem sucedida tanto numa abordagem de séries temporais — quando a única entrada é a própria informação da venda, em diferentes *lags* de tempo — quanto numa abordagem causal — quando variáveis explicativas da demanda são usadas como entrada. As previsões são usualmente consideradas como bem sucedidas quando oferecem melhoria considerável se comparadas com a previsão *naive*, que considera que as vendas no próximo período serão iguais àsquelas verificadas no período anterior.

Alguns autores também já investigaram a previsão de demanda no nível individual de produtos, porém tais estudos voltaram-se à modelagem das vendas de apenas um produto, seja numa abordagem de séries de tempo (BARASH & MITCHELL, 1998; GORDON, 1998; ALIBAIG & LILLY, 1999) ou causal (MENTZER & KENT, 1999; SUBRAHMANYAN, 2000). Sendo assim, tais modelos são incapazes de detectar os relacionamentos entre os produtos, os quais mostraram-se importantes nos estudos (não voltados à previsão) de BEN-AKIVA & GERSHENFELD (1998) e GUADAGNI & LITTLE (1998).

Os trabalhos citados ajudaram no delineamento da presente pesquisa, mostrando que havia uma lacuna para o desenvolvimento de uma previsão de vendas no nível individual que levasse em conta o relacionamento entre diversos produtos. Duas técnicas foram selecionadas

para o desenvolvimento dos modelos: a análise de cestos de compra, capaz de identificar quais produtos estão inter-relacionados, e a geração de redes neurais artificiais para previsão da demanda, dadas suas qualidades de autoadaptação a dados não-lineares e boa capacidade de generalização.

Durante a realização das análises, obteve-se uma metodologia que pode ser replicada para outro conjunto de dados sem grandes alterações. A metodologia consiste basicamente em:

1. Selecionar o conjunto de dados em registros agrupados com certa frequência, por exemplo semanal ou mensal, aglutinando o maior número de variáveis explicativas possível;
2. Executar a análise do cesto de compras, usando as abordagens por transação (produtos com venda casada) e por cliente (análise horizontal no tempo). A inspeção visual dos relacionamentos ajuda a separar apenas produtos que façam parte de um mesmo centro de interesse.
3. Tratar os dados de entrada, estabelecendo a necessidade de transformações nos dados. As entradas numéricas, quando normalizadas e padronizadas, tendem a ocasionar uma melhor performance das redes neurais.
4. Gerar redes neurais testando a retirada de variáveis correlacionadas.
5. Testar a incorporação do maior número possível de variáveis explicativas e também de variáveis que possam conter informações sobre o grupo de produtos considerado (centro de interesse).

De acordo com os resultados alcançados nesta pesquisa, a metodologia aplicada parece ser bem sucedida na previsão de vendas de produtos no nível individual, superando as modelagens *naive* e de regressão linear. A topologia de rede que obteve o melhor desempenho quando avaliada em dados externos aos de criação, na previsão com horizonte de 1 semana à frente, foi composta por 1 camada intermediária contendo 7 nós. Sua performance resultou em boa aderência aos dados de criação (R^2 de 0,72) e erro médio RMSE de 10,4 — erro de cerca de 10 unidades de produto para mais ou para menos, equivalentes a 41,2% do total de vendas.

Também de acordo com os resultados obtidos, algumas informações a respeito do centro de interesse podem ser utilizadas com sucesso na previsão de vendas, uma vez que, no

conjunto de dados usado, a informação sobre as vendas de todos os produtos substitutos colaborou para o desempenho da rede neural.

2. Conclusões da Previsão Baseada em Redes Neurais

A modelagem através de redes neurais artificiais pode ser considerada adequada para a previsão de demanda de produtos no nível individual. Com o uso de dados separados por produto, pode-se treinar uma rede neural para simular entradas nas variáveis explicativas, e assim planejar ações e antecipar seus resultados com alguma precisão. Tomando como exemplo os dados disponíveis nesta pesquisa, o gestor pode simular mudanças no preço, no nível de estoque, na exposição dos produtos na loja (mostruário), na frequência de reposição, na taxa de juros e incentivo ao pagamento parcelado, no comissionamento dos vendedores, entre outras, e prever as alterações na demanda por produto.

Os resultados sugerem que as redes neurais podem prever as vendas de produtos atomizados no curto prazo com maior precisão do que as técnicas *naïve* não-ajustada e de regressão linear. A equação de regressão que mais se aproximou das redes neurais obteve desvios da ordem de 19% maiores, o que pode impactar de maneira significativa o planejamento logístico, de suprimentos e finalmente o faturamento e lucro de uma corporação.

Outras conclusões importantes das análises efetuadas foram que as redes neurais podem tratar facilmente os dados de diferentes produtos, e que informações agregadas sobre um centro de interesse podem auxiliar na previsão, comprovando a existência de relacionamentos entre os produtos, e que tais relacionamentos podem ser captados pelas redes neurais.

Entretanto, os erros verificados ainda são altos: a rede neural errou, em termos percentuais, cerca de 41% nas previsões de uma semana à frente — certamente os desvios seriam ainda maiores para horizontes de tempo mais amplos. Isso pode significar que ainda há espaço para melhorias significativas na metodologia e técnica de previsão.

Vários são os fatores que podem explicar o (ainda) alto erro nas previsões obtidas com a metodologia adotada. O primeiro ponto a ser levantado é a qualidade dos dados de entrada disponíveis. O conjunto de dados utilizado dispunha de grande quantidade de registros com valores faltantes (*missing*) — algumas variáveis chegaram a apresentar até 81% de valores

missing — e um número grande de inconsistências também foi detectado, indicando uma baixa confiabilidade nos dados. A presença de dados confiáveis é um dos fatores de maior ganho em acurácia (BROWN, 1963; CHURCH & CURRAM, 1996).

Outro ponto importante foi o (relativamente) baixo número de registros por produto. O tempo de ciclo de vida dos produtos — normalmente menos de 1 ano, é baixo para que se tenha registros suficientes para um treinamento adequado das redes neurais. Esta característica negativa foi ainda mais salientada na seleção dos dados, onde se optou pela utilização exclusiva dos registros de apenas uma loja; talvez a inclusão dos registros de todas as lojas aumentasse o corpo de dados e minimizasse o problema.

As informações sobre promoções realizadas também não puderam ser aproveitadas nos modelos, principalmente pelo fato de que os registros disponíveis das promoções não estavam vinculados precisamente aos produtos anunciados. O histórico de vendas da EMPRESA mostra a presença de picos acentuados nas vendas, picos estes que segundo a própria EMPRESA devem-se reconhecidamente a promoções agressivas, direcionadas a produtos específicos. De acordo com GEURT & WHITLARK (1996), a presença de *outliers* é de difícil assimilação pelos modelos de previsão, e conseqüentemente os picos de vendas devido a tais promoções geraram picos de erros nas previsões, que forçaram a um aumento no erro médio.

Outros dados que sabidamente têm papel importante no desempenho do varejo são aqueles referentes às movimentações macroeconômicas. Sua incorporação talvez pudesse melhorar a performance das redes neurais.

Dados sobre a concorrência, como preço médio, promoções realizadas, entre outros, também poderiam melhorar a capacidade preditiva das redes neurais, uma vez que o varejo é altamente sensível às ações da concorrência. Tais dados não foram coletados sistematicamente pela EMPRESA, o que impediu que eles fossem considerados.

3. Sugestões para Trabalhos Futuros

Neste trabalho ficou claro que o processo de mineração de dados, através do uso de técnicas de análise de cesto de compras e previsão com uso de redes neurais artificiais, pode ser usado com sucesso com objetivo de previsão de demanda no nível individual. Esta pesquisa, por ser de certa forma inédita, requer confirmação com outro grupo de dados. Há

indícios de que uma maior quantidade de dados (quantidade de registros de um mesmo produto), presença de informações mais ricas (maior número de variáveis explicativas, como dados sobre a concorrência), e dados mais confiáveis (com menor quantidade de dados faltantes e/ou inconsistentes) podem trazer grandes benefícios na capacidade preditiva dos modelos; tais possibilidades merecem estudo mais aprofundado.

O benefício trazido pela análise conjunta de produtos relacionados também pode ser investigado com maior profundidade. Uma possibilidade é a inclusão de entradas individuais sobre os demais produtos de um mesmo centro de interesse, tentando captar, por exemplo, o preço, a demanda e demais características dos produtos relacionados, em dados individualizados (e não agregados como a metodologia aqui proposta). A análise de *lags* de tempo neste caso também pode ser benéfica, uma vez que as alterações em um produto podem trazer efeitos nos produtos complementares e substitutos apenas depois de determinado período de tempo.

A análise de *lags* de tempo também pode ser vastamente investigada na incorporação de dados macroeconômicos, já que estes podem ter efeito retardado sobre a demanda no varejo.

Finalmente, a abordagem tradicional de modelagem individual de produtos — gerando um modelo ou rede neural para cada produto — também pode ser averiguada quanto à adição de informações sobre o centro de interesses.

Referências Bibliográficas

- ACZEL, Amir D., 1993. **Complete Business Statistics**. Irwin Homewood, IL, 2a. ed.
- ALIBAIG, Shahbaz; LILLY, Bryan, 1999. **Updating your forecasting system: Wisconsin tissue's experience**. The Journal of Business Forecasting Methods & Systems, Fall, v. 18, n. 3, págs. 13-18.
- ALMEIDA, Fernando Carvalho de, 1993. **L'Evaluation des risques de défaillance des entreprises à partir des réseaux de neurones insérés dans les systèmes d'aide à la décision**. Tese de doutorado apresentada à Université Pierre Mendès France de Grenoble, Ecole Supérieure des Affaires.
- ALMEIDA, Fernando Carvalho de, 1995. **Desvendando o uso de redes neurais em problemas de Administração de Empresas**. Revista Administração de Empresas, v.35, n.1, Jan-Fev, págs 40-55, São Paulo.
- ALMEIDA, Fernando Carvalho de; SIQUEIRA, José de Oliveira, 1996. **Comparação entre regressão logística e redes neurais na previsão de falência de bancos brasileiros**. Jornadas de Classificação e Análise de Dados, n. 4, págs. 1-10, Lisboa.
- ANSUJ, A. P.; CAMARGO, M.E.; RADHARAMANAN, R.; PETRY, D.G., 1996. **Sales forecasting using time series and neural networks**. Computers and Industrial Engineering, Elsevier Science Ltda., v. 31, n. 1/2, págs. 421-425.
- BAETS, Walter R.J.; VENUGOPAL, V., 1994. **Neural Networks and Statistical Techniques in Marketing Research: A Conceptual Comparison**. Marketing Intelligence & Planning, MCB University Press, v. 12, n. 7, págs. 30-38.
- BARASH, Mark; MITCHELL, Donald H., 1998. **Account based forecasting at Nabisco biscuit company**. Journal of Business Forecasting, n. 31, Summer.
- BASS, Frank M., 1969. **A New Product Growth Model for Consumer Durables**. Management Science, n. 15, Janeiro, págs. 215-227.
- BEN-AKIVA, M.; GERSHENFELD, Shari, 1998. **Multi-featured Products and Services: Analysing Pricing and Bundling Strategies**. Journal of Forecasting, John Wiley & Sons, n. 17, págs. 175-196.
- BERRY, Michael J. A.; LINOFF, Gordon, 1997. **Data Mining Techniques for Marketing, Sales and Customer Support**. John Wiley & Sons, Inc., 1a. ed.
- BIDARKOTA, P. V., 1998. **The comparative forecast performance of univariate and multivariate models: an application to real interest rate forecasting**. International Journal of Forecasting, Elsevier Science B.V., n. 14, págs. 457-468.
- BOX, G. E. P.; JENKINS, G. M. **Time-Series Analysis, Forecasting and Control**. Ed. San Francisco: Holden-Day, 1976.
- BRENTNALL, G. J.; PICKSLEY, J. D., 1999. **Bayesian modeling enhances supply chain forecasting and planning**. The Journal of Business Forecasting Methods & Systems, Fall, v. 18, n. 3, págs. 19-23.
- BROWN, R. Goodell, 1959. **Statistical Forecasting for Inventory Control**. The Maple Press Company, New York, 1a. ed.
- BROWN, R. Goodell, 1963. **Smoothing, forecasting and prediction of discrete time series**. Prentice-Hall, Englewood Cliffs, N.J.

- CERULLO, M.J.; AVILA, A., 1975. **Sales Forecasting Practices: a survey**. Managerial Planning, n. 24, págs. 33-39.
- CHAKRABORTY, K.; MEHROTRA, K; MOHAN, C. K, 1992. **Forecasting the behaviour of multivariate time series using neural networks**. Neural Networks, v. 5, n. 6, págs. 961-970.
- CHANG, P.T., 1997. **Fuzzy seasonality forecasting**. Fuzzy Sets and Systems, Elsevier Science B.V., n. 90, págs. 1-10.
- CHASE, Charles W. Jr., 1996. **What you need to know when building a sales forecasting system**. The Journal of Business Forecasting Methods & Systems, Fall, v. 15, n. 3, págs. 2-23.
- CHASE, Charles W. Jr., 1999. **Sales forecasting at the dawn of the new millennium**. The Journal of Business Forecasting Methods & Systems, Fall, v. 18, n. 3, págs. 2-5.
- CHURCH, Keith B.; CURRAM, Stephen, 1996. **Forecasting consumer's expenditure: A comparison between econometric and neural network models**. International Journal of Forecasting, Elsevier Science B. V. n. 12, págs. 255-267.
- CURLEY, M. Louise; SILK, Leonard S., 1970. **A primer on business forecasting, with a guide to sources of business data**. Random House, New York, 1a. ed.
- DAGANZO Carlos, 1979. **Multinomial probit: The Theory and its Application to Demand Forecasting**. Academic Press, New York, 1a. ed.
- DE VILLE, Barry, 1996. **Predictive Models in Market Research**. Marketing Research, v. 8, n. 2, págs. 43-45.
- DIAS, George P. P., 2002. **Proposta De Processo De Previsão De Vendas Para Bens De Consumo**. Dissertação de Mestrado apresentada à Escola Politécnica da Universidade de São Paulo.
- DOYLE, A. Conan, 2002. **The Adventure of the Dancing Men**. In: The Return of Sherlock Holmes, ed. IndyPublish.com. Disponível em <<http://www.bakerstreet221b.de/canon/danc.htm>>. Publicado pela primeira vez na *Collier's Weekly*, Dez. 1903.
- DRURY, D. H., 1990. **Issues in forecasting management**. Management International Review, v. 30, págs. 317-329.
- DUTTA, Siddhartha; AMARAVADI, Chandra S; SAMADDAR, Subhashish, 1995. **Intelligent marketing information systems: computerized intelligence for marketing decision making**. Marketing Intelligence & Planning, University Press Limited; v. 13, n. 2, págs. 4-13.
- ENGEL, J. F.; BLACKWELL, R. D.; MINIARD, P. W., 1993. **Consumer behavior**. Ed. Dryden Press, Hinsdale.
- FEELDERS, A.; DANIELS, H.; HOLSHEIMER, M., 2000. **Methodological and practical aspects of data mining**. Information and Management, Elsevier Science B. V., n. 37, págs. 271-281.
- FILDES, Robert; BOTTOMLEY, Paul A., 1998. **The Role of Prices in Models of Innovation Diffusion**. Journal of Forecasting, John Wiley & Sons, n. 17, págs. 539-555.
- FILDES, Roberts; Hastings, R., 1994. **The organization and improvement of market forecasting**. Journal of Operational Research Society, v. 45, págs. 1-16.

- FOURT, L. A.; WOODLOCK, J. W., 1960. **Early prediction of market sucess of grocery products**. Journal of Marketing, n. 25, Outubro, págs. 31-38.
- GARGANO, Michael L.; RAGGAD, Bel G., 1999. **Data mining - a powerful information creating tool**. OCLC Systems and Services, MCB University Press, v. 15, n. 2, págs. 81-90.
- GAYLE, Sanfort, 2003. **The Marriage of Market Basket Analysis to Predictive Modeling**. SAS Institute White Paper, SAS Institute Inc. Disponível em <<http://www.sas.com.br>>. Acessado em Janeiro de 2003.
- GEURTS, Michael D.; WHITLARK, David, 1996. **Improving sales forecasts by improving the input data**. The Journal of Business Forecasting Methods & Systems, Fall, v. 15, n. 3. 15-18.
- GORDON, Rich, 1998. **A role for the forecasting function**. The Journal of Business Forecasting Methods & Systems, Winter, v. 16, n. 4, págs. 3-7.
- GRAY, Paul, 1997. **Data Warehouses, OLAP, Data Mining, and the New DSS**. Claremont Graduate University Program in information Science, ISDSS'97, págs. 7-24.
- GROSS, Charles W.; PETERSON, Robin T., 1983. **Business forecasting**. Houghton Mifflin Co., Boston, 2a. ed.
- GUADAGNI, P. M.; LITTLE, J. D. C., 1998. **When and What to Buy: a Nested Logit Model of Coffee Purchase**. Journal of Forecasting, John Wiley & Sons, v. 17.
- HAND, D. J., 1998. **Data Mining: Statistics and More?** The American Statistician, Maio, v.52, n.2, págs. 112-118.
- HANKE, John E.; REITSCH, Arthur G, 1995. **Business Forecasting**. Ed. Prentice Hall, Englewood Cliffs, NJ, 5a. ed.
- HANSSENS, D. M.; PARSONS, J. L.; SCHULTZ, R. L., 1990. **Market Response Models: Econometric and Time Series Analysis**. Ed. Kluwer, Boston.
- HARDIE, Bruce G. S.; FADER, Peter S.; WISNIEWSKI, Michael, 1998. **An Empirical Comparison of New Product Trial Forecasting Methods**. Journal of Forecasting, John Wiley & Sons, London, n. 17, págs. 209-229.
- HEELER, R. M.; HUSTAD, T. P., 1980. **Problems in predicting new products growth for consumer durables**. Management Science, n. 26, Outubro, págs. 1007-1020.
- HIBBERT, Brynn; WILKINSON, Ian F., 1994. **Chaos Theory and the Dynamics of Marketing Systems**. Journal of Academy of Marketing Science, Academy of Marketing Science, v. 22, n. 3, págs. 218-233.
- HUSSEY, Michael; HOOLEY, Graham, 1995. **The diffusion of quantitative methods into marketing management**. Journal of Marketing Practice: Applied Marketing Science, MCB University Press, v. 1, n. 4, págs. 13-31.
- IMMON, W. H., 1997. **Como construir o Data Warehouse**. Campinas.
- KIRK, Jerome; MILLER, Marc L, 1988. **Reliability and Validity in Qualitative Research**. Sage University Paper Series on Qualitative Research Methods, v. 1, Sage Publications, California.

- KLEIN, Lawrence R.; YOUNG, Richard M., 1980. **An introduction to econometric forecasting and forecasting models**. The Wharton Econometric Studies Series, Lexington Books, Lexington, MA, 2a. ed.
- KOTLER, Philip, 1999. **Marketing Management: analysis, planning, implementation and control**. Prentice-Hall International, London, UK, Millennium ed.
- KUO, R. J.; XUE, K. C., 1998. **A decision support system for sales forecasting through fuzzy neural networks with asymmetric fuzzy weights**. Decision Support Systems, Elsevier Science B. V., n. 24, págs. 105-126.
- KUO, R. J.; XUE, K. C., 1999. **Fuzzy neural networks with application to sales forecasting**. Fuzzy Sets And Systems, Elsevier Science B. V., n. 108, págs. 123-143.
- LAPEDES, A.; FARBER, R., 1987. **Nonlinear signal processing using neural networks prediction and system modeling**. Technical Report LA-UR-87-2662, Los Alamos national Library, Los Alamos, NM.
- LAWRENCE, K. D.; LAWTON, W. H., 1981. **Application of diffusion models: some empirical results**. In Wind, Y.; Mahajan, V.; Cardozo, R. N., New Product Forecasting. Lexington Books, Lexington, MA, págs. 529-541.
- LAWRENCE, Michael; EDMUNDSON, Bob; O'CONNOR, Marcus, 2000. **A field study of sales forecasting accuracy and processes**. Amsterdam European Journal of Operational Research, Elsevier Science, B. V., Apr 1, Vol. 122, n. 1, págs. 151-160.
- LEWISON, D. M., 1997. **Retailing**. Ed. Prentice-Hall, Englewood Cliffs.
- LINDBLAD, Thomas; JOHANSSON, Robert; LINDSEY, Clark S.; EIDE, Age, 1997. **Data mining and neural networks for knowledge discovery**. Nuclear Instruments & Methods in Physics Research, Elsevier Science B. V., n. 389, págs. 251-254.
- LUXHOJ, J. T.; RIIS, J. O.; STENSALLE, B., 1996. **A hybrid econometric-neural network modeling approach for sales forecasting**. Amsterdam International Journal of Production Economics, Elsevier Science B. V., n. 43, págs. 175-192.
- MAHAJAN, V.; BASS, Frank M.; MULLER, E., 1990. **New product diffusion models in marketing: a review and directions for research**. Journal of Marketing, v. 54, Janeiro, págs. 1-26.
- MAKRIDAKIS, S.; WHEELWRIGHT, Steven C.; MCGEE, Victor E., 1983. **Forecasting: Methods and Application**. John Wiley & Sons, New York, 2a. ed.
- MAKRIDAKIS, S.; WHEELWRIGHT, Steven C., 1987. **The handbook of forecasting: a manager's guide**. Ed. Wiley, New York, 2a. Edição.
- MANSFIELD, E., 1961. **Technological change and the rate of imitation**. Econometrica, v. 29, n. 10, págs. 741-766.
- MASON, J. B.; MAYER, M. L.; WILKINSON, J. B., 1993. **Modern retailing: theory and practice**. Ed. Irwin, Chicago, 6a. ed.
- MATIAS, Alberto B.; SIQUEIRA, José de Oliveira, 1996. **Risco bancário: modelo de previsão de insolvência de bancos no Brasil**. Revista de Administração da USP, v. 31, n. 2, Abril/Junho, págs. 19-28, São Paulo.
- MCHUGH, A. K.; SPARKES, J. R., 1983. **The forecasting dilemma**. Management Accounting, n. 61, págs. 30-34.

- MENTZER, J. T.; COX Jr., J. E., 1984. **Familiarity, application and performance of Sales forecasting techniques**. Journal of Forecasting, n. 3, págs. 27-36.
- MENTZER, John T., 1999. **The impact of forecasting on return on shareholders's value**. The Journal of Business Forecasting Methods & Systems, Fall, v. 18, n. 3, págs. 8-12.
- MENTZER, John T.; KENT, John L., 1999. **Forecasting demand in the Longaberger Company**. Marketing Management, Summer, v. 8, n. 2, págs. 46-50, Chicago.
- MOON, Mark A.; MENTZER, John T., 1999. **Improving salesforce forecasting**. The Journal of Business Forecasting Methods & Systems, Summer, v. 18, n. 2, págs. 7-12.
- NAYLOR, T.H., 1981. **Experience with corporate econometric models: a survey**. Business Economics, n. 16, págs. 79-83.
- NELSON, M.; O'CONNOR, M.; HILL, T.; REMUS, B., 1994. **Can neural networks be applied to time series forecasting and learn seasonal patterns: An empirical investigation**. Proceedings of the Twenty seventh Annual Hawaii International Conference on System Sciences, págs. 649-655.
- NEWBOLD, Paul, 1995. **Statistics for Business & Economics**. Prentice-Hall, Inc. Englewood Cliffs, N.J., 4a. ed.
- PAN, J.; NICHOLS, D. R.; JOY, O., 1977. **Sales Forecasting Practices of Large U.S. industrial firms**. Financial Management, n. 6, págs. 72-77.
- PARENTE, Juracy, 1997. **Gerenciamento de categorias**. Revista SuperHiper, págs. 230-234, Abril.
- PARENTE, Juracy, 2000. **Varejo no Brasil: Gestão e Estratégia**. Ed. Atlas, São Paulo.
- PARK, J.; SANDBERG, I. W., 1991. **Universal approximation using radial basis function networks**. Neural Computation, v. 3, págs. 246-257.
- PEACOCK, Peter R., 1998a. **Data Mining in Marketing: Part 1 – The revolution is upon us, so choose your weapons carefully**. Marketing Management, Winter, págs. 9-18.
- PEACOCK, Peter R., 1998b. **Data Mining in Marketing: Part 2 – Dig deep to unearth knowledge inherent in databases**. Marketing Management, Spring, págs. 15-25.
- PETERSON, R.T., 1993. **Forecasting practices in retail industry**. Journal of Business Forecasting, n. 12, págs. 11-14.
- PETERSON, Robin T.; JUN, Minjoon, 1999. **Forecasting sales in wholesale industry**. The Journal of Business Forecasting Methods & Systems, Summer, v. 18, n. 2, págs. 15-18.
- PETRON, E., 1999. **Stuttgart Neural Network Simulator**. Linux Journal, Julho, págs. 44-48.
- PIDD, Michael, 1998. **Modelagem Empresarial: ferramentas para Tomada de Decisão**. Ed. Bookman, Porto Alegre, RS.
- POWERS, Anthony, 1988. **Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event**. The Journal of Consumer Research, v. 10, Dezembro, págs. 6-9.
- PUTSIS Jr., Willian P., 1998. **Parameter Variation and New Product Diffusion**. Journal of Forecasting, John Wiley & Sons, n. 17, págs. 231-257.
- ROBERTS, John, 1998. **Marketing Approaches to Forecasting Problems**. Journal of Forecasting, John Wiley & Sons, n. 17, págs. 169-174.

- ROSATI, Anna Maria, 1996. **Forecasting at Segix, Italia: A pharmaceutical company.** The Journal of Business Forecasting Methods & Systems, Fall, v. 15, n. 3, págs. 7-9.
- SAKAI, Hidekata; NAKAJIMA, Hideki; YASUDA, Masashi; HIGASHIHARA, Minoru; OOSUMI, Masato, 1999. **Development of a fuzzy sales forecasting system for vending machines.** Computers & Industrial Engineering, Elsevier Science Ltda., n. 36, págs. 427-449.
- SANDERS, N. R., 1992. **Corporate forecasting practices in the manufacturing industry.** Production and Inventory Management, n. 33, págs. 54-57.
- SANDERS, N. R.; MANRODT, K., 1994. **Forecasting practices in US corporations: survey results.** Interfaces, v. 24, págs. 92-100.
- SAS, 1996. **Data Mining with the SAS System: From Data to Business Advantage.** SAS Institute White Paper, SAS Institute Inc. Disponível em <<http://www.sas.com.br>>. Acessado em Janeiro de 2001.
- SAS, 1998. **Finfing the Solution to Data Mining.** SAS Institute White Paper, SAS Institute Inc. Disponível em <<http://www.sas.com.br>>. Acessado em Janeiro de 2001.
- SPEEDING, T. A.; CHAN, K. K., 2000. **Forecasting demand and inventory management using Bayesian time series.** Integrated Manufacturing Systems, MCB University Press, v. 11, n. 5, págs. 331-339.
- SPSS, 1999. **Data Mining with Confidence.** Disponível em <<http://www.spss.com>>. Acessado em Fevereiro de 2001.
- SUBRAHMANYAN, Saroja, 1998 **An exploratory empirical study on how pricing and inventory decisions are made in a simulated environment.** Research paper Series, Faculty of Business Administration, National University of Singapore, Maio, RPS, págs. 98-123.
- SUBRAHMANYAN, Saroja, 2000. **Using quantitative models for setting retail prices.** Journal of Product and Brand Management, MCB University Press, v. 9, n. 5, págs. 304-320.
- TAMASCHKE, Rick; BUTTERY, Alan, 1995. **Marketing decision support systems in a small trading nation: an Australian case study.** Marketing Intelligence & Planning, MCB University Press Limited, v. 13, n. 2, págs. 14-28.
- TANG, Z.; Almeida, C.; FISHWICK, P. A, 1991. **Time series forecasting using neural networks vs. Box-Jenkins methodology.** Simulation Council, v. 57, n. 5, págs. 303-310.
- THALL, N., 1992. **Neural Forecasts: A Retail Sales Booster.** Discount Merchandiser, v. 32, n. 10, págs. 41-42.
- TUKEY, J. W.; HOAGLIN, D. C.; MOSTELLER, F., 1983. **Análise exploratória de dados: técnicas robustas.** Ed. Salamandra, Lisboa.
- VOLKERT, J.; ESPAHBODI, P.; ESPAHBODI, H., 1995. **Early Identification of Firms Requiring Debt Restructuring.** The Journal of Business Forecasting, Fall, v. 14, n.3.
- WHEELRIGHT, Steven C.; CLARKE, D. G., 1976. **Corporate forecasting: promise and reality.** Harvard Business Review, v. 54, págs. 40-64.
- WHITE, H. R., 1986. **Sales Forecasting: timesaving and profitmaking strategies that work.** Scott, Foresman and Company, London.
- WILSON, M.; DAUBECK, H. G., 1989. **Marketing managers evaluate forecasting models.** Journal of Business Forecasting, n. 8, págs. 19-22.

WINKLHOFER, Heidi; WITT, Stephen F.; DIAMANTOPOULOS, Adamantios, 1996. **Forecasting Practice: a review of the empirical literature and an agenda for future research**. International Journal of Forecasting, Elsevier Science B. V., n. 12, págs. 193-221.

WRAY, Barry; BEJOU, David; PALMER, Adrian, 1994. **Using Neural Network Analysis to Evaluate Buyer-Seller Relationships**. European Journal of Marketing, v. 28, n. 10, págs. 32-48.

XU, X. Mark, 1999. **The strategic orientation of marketing information systems - an empirical study**. Marketing Intelligence & Planning, MCB University Press, v. 17, n. 6, págs. 262-272.

ZHANG, Guoqiang; PATUWO, B. Eddy; HU, Michael Y., 1998. **Forecasting with artificial neural networks: The state of the art**. International Journal of Forecasting, Elsevier Science B. V., n. 14, págs. 35-62.

ZHANG, X., 1994. **Time series analysis and prediction by neural networks**. Optimization Methods and Software, v. 4, págs. 151-170.

ZHOU, Wei, 1999. **Integration of different forecasting models**. The Journal of Business Forecasting Methods & Systems, Fall, v. 18, n. 3, págs. 26-29.

ANEXO A — Questionário aplicado à EMPRESA

1. O senhor poderia dar algum *background* sobre sua empresa: produtos, número de funcionários, volume de vendas, faturamento, lojas etc.
2. Onde reside a função de previsão — quem são os responsáveis?
3. Quantas pessoas trabalham na área?
4. Qual a porcentagem de tempo dedicada à função?
5. Quantos produtos entram nas previsões?
6. Como são feitas as previsões? Top-down ou botton-up? Nível de agregação?
7. Qual o horizonte de tempo previsto?
8. O Sr. pode explicar o processo de previsão: quem prepara, quem participa, se é permitida alteração baseada em julgamento etc.
9. Qual o volume de dados utilizados?
10. O Sr. utiliza dados externos e sobre a concorrência, e como eles são obtidos?
11. E dados macroeconômicos?
12. O Sr. sabe dizer quais são os erros observados no momento, a nível agregado, por categoria e por SKU?
13. Quais itens são mais fáceis de serem previstos e por que?
14. Na sua opinião, as previsões são bem utilizadas?
15. O processo prevê demanda, entregas ou compras? Por que?
16. Qual o papel da força de vendas nas previsões?
17. Sabe dizer se existe sazonalidade bem característica nas vendas?
18. A alta administração considera as previsões como uma ferramenta de gestão importante? Utilizam-nas na tomada de decisões?
19. O Sr. experimentou algum conflito de interesses entre as áreas funcionais com relação às previsões?
20. O Sr. está satisfeito com o processo de previsão? Se não, como o Sr. pretende melhorá-lo?
21. Qual o tempo do ciclo de vida dos produtos?
22. Qual o nível de escolaridade dos envolvidos nas previsões? A empresa provê treinamento para os mesmos?
23. Saberá listar as variáveis que mais influem nas vendas?

ANEXO B — Dicionário de Dados Originais

Cadcar

Descreve as compras realizadas a prazo pelos clientes.

<i>Nome</i>	<i>Tipo</i>	<i>Descrição</i>
FILIAL_VENDA	SMALLINT	Identificador da filial que registrou a venda.
CARTAO	INTEGER	Número do cartão do cliente que realizou a compra.
COMPRA_DATA	DATE	Data da compra.
ENTRADA_VALOR	DECIMAL	Valor dado como entrada (à vista) para a compra.
NF_VALOR_TOTAL	DECIMAL	Valor total da compra, em reais.
PRESTACAO_QTD	SMALLINT	Quantidade de prestações da compra.
PRESTACAO_VALOR	DECIMAL	Valor de cada prestação.
SLIP	INTEGER	Identificador do crediário realizado, serve para identificar a compra.
NF_NUMERO	INTEGER	Número da nota fiscal de venda.
NF_SERIE	CHARACTER	Série da nota fiscal de venda.

Estoque

Contém *snapshots* das posições de estoque das filiais. Serve para saber quando um produto não estava disponível para venda numa data específica por falta de estoque.

<i>Nome</i>	<i>Tipo</i>	<i>Descrição</i>
FILIAL	SMALLINT	Identificador da filial que registrou a venda.
SETOR	SMALLINT	Identificador do setor relacionado ao produto.
LINHA	SMALLINT	Identificador da linha relacionada ao produto.
PRODUTO	INTEGER	Identificador do produto.
QUANTIDADE	INTEGER	Quantidade em estoque na filial.
DATA_BASE	DATE	Data da posição de estoque.
MOSTRUARIO	SMALLINT	Quantidade de itens do produto em mostruário.
GRADE	SMALLINT	Quantidade de itens do produto especificada para ser mantida na filial.
VENDA_FIS_ULT_07	SMALLINT	Unidades vendidas do produto nos últimos 7 dias.
VENDA_FIS_ULT_30	SMALLINT	Unidades vendidas do produto nos últimos 30 dias.

Linhas

Cadastra as linhas de produtos comercializadas por toda a rede.

<i>Nome</i>	<i>Tipo</i>	<i>Descrição</i>
SETOR	SMALLINT	Identificador do setor relacionado à linha de produtos.
LINHA	INTEGER	Identificador da linha de produtos.
DESCRICAO	CHARACTER	Descrição da linha de produtos.
CONTAGEM	INTEGER	Número de produtos comercializados dentro da linha específica, em junho de 2002.

LinhasXLoja

Cadastra as linhas de produtos comercializadas em cada loja.

<i>Nome</i>	<i>Tipo</i>	<i>Descrição</i>
FILIAL	SMALLINT	Identificador da filial.
LINHA	INTEGER	Identificador da linha de produto que a loja comercializa.

Lojas

Cadastro das lojas ativas em Junho de 2002.

<i>Nome</i>	<i>Tipo</i>	<i>Descrição</i>
FILIAL	SMALLINT	Identificador da filial.
REGIÃO	SMALLINT	Identificador do região relacionada à filial.
DESCRIÇÃO	CHARACTER	Nome e identificação da filial.
AREA	INTEGER	Área ocupada pela filial, em metros quadrados.
TIPO	CHARACTER	Tipo de filial. Pode assumir os seguintes valores: “C” – loja convencional, e “E” – loja eletrônica.
DEPOSITO	SMALLINT	Identificador do depósito que atende a loja.
CL	CHARACTER	Classe de loja. Pode assumir os seguintes valores: “C” – loja convencional, “E” – loja eletrônica, e “S” – loja de shopping.
VEND	SMALLINT	Número de vendedores da filial.

Prod_mov

Contém todas as vendas efetuadas nas filiais selecionadas.

<i>Nome</i>	<i>Tipo</i>	<i>Descrição</i>
PRODUTO	INTEGER	Identificador do produto vendido.
FILIAL	SMALLINT	Identificador da loja onde foi efetuada a venda.
MOVIMENTO_DATA	DATE	Data da venda.
DOCUMENTO	INTEGER	Número do documento apresentado.
NOTA_FISCAL_SERIE	SMALLINT	Série da nota fiscal de venda.
MOVIMENTO_QTD	SMALLINT	Unidades vendidas do produto.
DOCUMENTO_TIPO	SMALLINT	Mostra se foi venda à vista ou a prazo. Pode assumir os seguintes valores: “16” – venda a vista; e “12” – venda a prazo.
PRECO	DECIMAL	Preço unitário do produto vendido.
CUSTO_TECNICO	DECIMAL	Custo técnico considerado para o produto.
CUSTO_GERENCIAL	DECIMAL	Custo gerencial considerado para o produto.
VENDEDOR_NUMERO	INTEGER	Identificador do vendedor que realizou a venda.
CARTAO	INTEGER	Número do cartão que identifica o cliente.
NOTA_FISCAL	INTEGER	Número da nota fiscal.
ICMS	DECIMAL	Valor do ICMS cobrado.
ENTRA_MARGEM	CHARACTER	Identifica se o produto entra nas comissões do vendedor. Pode assumir os seguintes valores: “S” – SIM; “N” – NÃO; e “B” – brinde/bonificação (indireta).
PRODUTO_VALIDADE	DATE	Data de validade do produto.

SETOR	SMALLINT	Setor relacionado ao produto.
LINHA	SMALLINT	Linha relacionada ao produto.
SLIP	INTEGER	Identificador do crediário para vendas a prazo.
SLIP_SERIE	CHARACTER	Série do crediário para identificação das vendas a prazo.

Produtos

Cadastro de todos os produtos comercializados.

<i>Nome</i>	<i>Tipo</i>	<i>Descrição</i>
PRODUTO	INTEGER	Identificador do produto.
FORNECEDOR	INTEGER	Identificador do fornecedor do produto.
DESCRICAO	CHARACTER	Descrição abreviada do produto.
CLASSE	CHARACTER	Classificação do produto. Pode assumir os seguintes valores: 0, A, B, D, E, F, J, L, M, S, W.
ESTOQUE	INTEGER	Estoque planejado (ideal) do produto, em dias.
REPOSICAO	INTEGER	Quantidade de reposição planejada (ideal) do produto (mínimo de compra), em dias.
MOSTRUARIO	INTEGER	Número total de produtos em mostruário na rede.
FANTASIA	CHARACTER	Nome fantasia do fornecedor do produto.
CLAS_DISTRIBUICAO	CHARACTER	Classificação do produto. Pode assumir os seguintes valores: ?, A, B, C, E, F, G, M, S, W.
SETOR	INTEGER	Identificador do setor relacionado ao produto
LINHA	INTEGER	Identificador da linha relacionada ao produto

ANEXO C — Dicionário de Dados – Vendas Semanais

<i>Nome</i>	<i>Tipo</i>	<i>Descrição</i>
Id_produto	INT	Identificador do produto vendido.
Produto	NVARCHAR (50)	Descrição do produto vendido.
Comissao	SMALLINT	Identifica se o produto entra nas comissões do vendedor. Pode assumir os seguintes valores: “0” – vendedor não recebe comissão sobre o produto; “1” – vendedor recebe comissão sobre a venda; e “2” – vendedor recebe brinde/bonificação (comissão indireta).
Setor	SMALLINT	Identificador do setor relacionado ao produto.
Linha	SMALLINT	Identificador da linha relacionada ao produto.
Id_marca		Identificador do fornecedor do produto.
Classe	SMALLINT	Classificação do produto. Pode assumir os seguintes valores: 0 (0 – missing?), 1 (A), 2 (B), 3 (D), 4 (E), 5 (F), 6 (J), 7 (L), 8 (M), 9 (S), 10 (W).
Classe_distribuicao	SMALLINT	Classificação do produto segundo o critério de distribuição. Pode assumir os seguintes valores: 0 (missing), 1 (A), 2 (B), 3 (C), 4 (E), 5 (F), 6 (G), 7 (S), 8 (W).
Estoque	INTEGER	Quantidade em estoque na loja, tomada a data do início do mês.
Estoque_dias	INTEGER	Estoque planejado (ideal) do produto, em dias.
Reposicao_dias	INTEGER	Quantidade de reposição planejada (ideal) do produto (mínimo de compra), em dias.
Mostruario	INTEGER	Quantidade de itens do produto em mostruário.
Grade	INTEGER	Quantidade de itens do produto especificada para ser mantida na filial.
Total7	INTEGER	Unidades vendidas do produto na semana.
Semana	INTEGER	Identificação da semana correspondente. Assume valores de 1 (semana de 01/11/1998) a 187 (semana de 27/05/2002).
PrecoMedioSemanal	REAL	Média de preço praticado nas vendas do produto durante a semana correspondente.
JurosMedioSemanal	REAL	Média de taxa de juros composto praticado nas vendas a prazo do produto na semana correspondente.
Pprazo	REAL	Porcentagem das vendas que foram realizadas a prazo na semana anterior.

ANEXO D — Transformações nos Dados

Estoque

A distribuição da variável original (sem transformações) foge muito da distribuição normal, como pode ser observado pelo histograma abaixo:

Figura 24 – Histograma da variável “Estoque” sem transformações

Tal distribuição não é adequada para modelagem com uso de técnicas quantitativas, e em especial das técnicas de mineração de dados. Ela provoca um comportamento excessivamente patológico da variável, como demonstrado pela seu gráfico *boxplot*, mostrado na Figura 25. Este tipo de comportamento é de difícil assimilação pelas técnicas de mineração de dados — dentre elas a técnica de redes neurais artificiais — e acarreta na geração de modelos pouco prováveis de serem aderentes aos dados, e conseqüentemente com baixa performance, seja qual for sua finalidade.

Figura 25 – *Boxplot* da variável “Estoque” sem transformações

A fim de melhorar o comportamento da variável, são aplicadas transformações nos dados. A transformação sugerida é aquela dada pelo resultado do gráfico dispersão versus nível:

Equação linear ajustada	$y = 0,5398x + 0,5203$
Valor de b	0,5
Transformação sugerida	$Z = \sqrt{x}$

Apesar do gráfico dispersão versus nível mostrar a transformação raiz como ideal, outras transformações foram tentadas e a que melhor ajustou os dados foi a transformação log x.

Transformação	$Z = \log x$
Nova variável	Estoque_log

A transformação aproxima a distribuição da variável de uma distribuição normal, melhorando o comportamento dos dados, como pode ser visto nas figuras a seguir.

Estoque_dias

Resultado do gráfico dispersão versus nível:

Equação linear ajustada	$y = 2,7411x - 2,7196$
Valor de b	3
Transformação sugerida	$Z = \frac{1}{x^2}$

Neste caso a transformação resultou em melhoria aparente dos dados; porém esta melhoria não foi profunda. Outras transformações foram tentadas, sem sucesso. Uma nova variável foi criada com a transformação sugerida, porém ambas foram testadas nos modelos criados.

Transformação	$Z = \frac{1}{x^2}$
Nova variável	Estoque_dias-2

Reposicao_dias

Resultado do gráfico dispersão versus nível:

Equação linear ajustada	$y = -0,0845x + 0,0845$
Valor de b	0
Transformação sugerida	nenhuma

Como o gráfico dispersão versus nível não trouxe resultados satisfatórios, várias transformações foram tentadas, sem sucesso. A variável permaneceu sem transformações.

Mostruario

Resultado do gráfico dispersão versus nível:

Equação linear ajustada	$y = 0,7087x + 0,3868$
Valor de b	0,7
Transformação sugerida	$Z = x^{0,3}$

Neste caso a transformação resultou em melhoria aparente dos dados; porém esta melhoria não foi profunda. Outras transformações foram tentadas, sendo a que melhor adaptou os dados a transformação logarítma.

Transformação	$Z = \log x$
Nova variável	Mostruario_log

Grade

Resultado do gráfico dispersão versus nível:

Equação linear ajustada	$y = 0,9232x + 0,4442$
Valor de b	1
Transformação sugerida	$Z = \log x$

A transformação sugerida mostrou ser a que melhor ajusta os dados.

Transformação	$Z = \log x$
Nova variável	Grade_log

Total7

Resultado do gráfico dispersão versus nível:

Equação linear ajustada	$y = 1,1243x + 0,0695$
Valor de b	1
Transformação sugerida	$Z = \log x$

A transformação sugerida mostrou ser a que melhor ajusta os dados.

Transformação	$Z = \log x$
Nova variável	Vendas_log

PrecoMedioSemanal

Resultado do gráfico dispersão versus nível:

Equação linear ajustada	$y = 0,794x + 0,1608$
Valor de b	0,8
Transformação sugerida	$Z = x^{0,2}$

A transformação sugerida mostrou ser a que melhor ajusta os dados, mesmo tentando-se outras transformações, como a logarítma.

Transformação	$Z = x^{0,2}$
Nova variável	Preco_pot

Pprazo

Resultado do gráfico dispersão versus nível:

Equação linear ajustada	$y = -0,0646x + 1,7929$
Valor de b	0
Transformação sugerida	nenhuma

De fato, mesmo tentando várias transformações, nenhuma mostrou ajustar os dados.

ANEXO E — Correlações e Autocorrelações Verificadas

Matriz de Correlações

Correlations															
		COMISSAO	ID_MARCA	CLASSE	Classe_ distribuicao	Reposicao_ dias	SEMANA	JurosMedio Semanal	PPRAZO	Estoque_log	Estoque_ dias-2	Mostruario_ log	Grade_log	Vendas_log	Preco_ pot
COMISSAO	Pearson	1	,033	-,001	-,030	,006	-,034	-,002	-,040	,059	,010	-,002	-,011	,033	,033
	Correlation														
	Sig. (2-tailed)	.	,000	,916	,001	,510	,000	,808	,000	,000	,262	,937	,690	,000	,000
ID_MARCA	N	12559	12559	12559	12559	12559	12559	12559	12559	4033	12559	2352	1249	12559	12559
	Pearson	,033	1	,120	,074	-,118	-,067	-,002	-,012	,072	-,046	-,103	,080	,051	,163
	Correlation														
CLASSE	Sig. (2-tailed)	,000	.	,000	,000	,000	,000	,859	,177	,000	,000	,000	,005	,000	,000
	N	12559	12559	12559	12559	12559	12559	12559	12559	4033	12559	2352	1249	12559	12559
	Pearson	-,001	,120	1	,556	-,143	-,265	-,012	,145	-,041	,072	-,222	,265	-,190	,040
Classe_distribuiçao	Correlation														
	Sig. (2-tailed)	,916	,000	.	,000	,000	,000	,195	,000	,010	,000	,000	,000	,000	,000
	N	12559	12559	12559	12559	12559	12559	12559	12559	4033	12559	2352	1249	12559	12559
Reposicao_dias	Pearson	-,030	,074	,556	1	-,116	-,304	,024	,133	-,119	,047	,177	-,193	-,197	-,168
	Correlation														
	Sig. (2-tailed)	,001	,000	,000	.	,000	,000	,007	,000	,000	,000	,000	,000	,000	,000
SEMANA	N	12559	12559	12559	12559	12559	12559	12559	12559	4033	12559	2352	1249	12559	12559
	Pearson	,006	-,118	-,143	-,116	1	,018	-,002	-,118	,213	-,011	,236	,033	,223	-,047
	Correlation														
JurosMedioSemanal	Sig. (2-tailed)	,510	,000	,000	,000	.	,038	,838	,000	,000	,202	,000	,245	,000	,000
	N	12559	12559	12559	12559	12559	12559	12559	12559	4033	12559	2352	1249	12559	12559
	Pearson	-,034	-,067	-,265	-,304	,018	1	-,050	,042	-,019	-,005	-,062	-,008	-,020	,073
PPRAZO	Correlation														
	Sig. (2-tailed)	,000	,000	,000	,000	,038	.	,000	,000	,216	,553	,003	,773	,026	,000
	N	12559	12559	12559	12559	12559	12559	12559	12559	4033	12559	2352	1249	12559	12559
	Pearson	-,002	-,002	-,012	,024	-,002	-,050	1	-,005	,019	,004	,022	,098	-,029	-,119
	Correlation														
	Sig. (2-tailed)	,808	,859	,195	,007	,838	,000	.	,562	,222	,670	,295	,001	,001	,000
	N	12559	12559	12559	12559	12559	12559	12559	12559	4033	12559	2352	1249	12559	12559
	Pearson	-,040	-,012	,145	,133	-,118	,042	-,005	1	-,279	,032	-,148	-,059	-,648	,122
	Correlation														

	Sig. (2-tailed)	,000	,177	,000	,000	,000	,000	,562	.	,000	,000	,000	,037	,000	,000
Estoque_log	N	12559	12559	12559	12559	12559	12559	12559	12559	4033	12559	2352	1249	12559	12559
	Pearson	,059	,072	-,041	-,119	,213	-,019	,019	-,279	1	-,036	,076	,590	,439	-,058
	Correlation														
	Sig. (2-tailed)	,000	,000	,010	,000	,000	,216	,222	,000	.	,022	,000	,000	,000	,000
Estoque_dia	N	4033	4033	4033	4033	4033	4033	4033	4033	4033	4033	2143	1218	4033	4033
s-2	Pearson	,010	-,046	,072	,047	-,011	-,005	,004	,032	-,036	1	,044	,125	-,048	-,060
	Correlation														
	Sig. (2-tailed)	,262	,000	,000	,000	,202	,553	,670	,000	,022	.	,034	,000	,000	,000
Mostruario_l	N	12559	12559	12559	12559	12559	12559	12559	12559	4033	12559	2352	1249	12559	12559
og	Pearson	-,002	-,103	-,222	,177	,236	-,062	,022	-,148	,076	,044	1	-,032	,143	-,459
	Correlation														
	Sig. (2-tailed)	,937	,000	,000	,000	,000	,003	,295	,000	,000	,034	.	,272	,000	,000
Grade_log	N	2352	2352	2352	2352	2352	2352	2352	2352	2143	2352	2352	1149	2352	2352
	Pearson	-,011	,080	,265	-,193	,033	-,008	,098	-,059	,590	,125	-,032	1	,219	-,271
	Correlation														
	Sig. (2-tailed)	,690	,005	,000	,000	,245	,773	,001	,037	,000	,000	,272	.	,000	,000
Vendas_log	N	1249	1249	1249	1249	1249	1249	1249	1249	1218	1249	1149	1249	1249	1249
	Pearson	,033	,051	-,190	-,197	,223	-,020	-,029	-,648	,439	-,048	,143	,219	1	-,033
	Correlation														
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,026	,001	,000	,000	,000	,000	,000	.	,000
Preco_pot	N	12559	12559	12559	12559	12559	12559	12559	12559	4033	12559	2352	1249	12559	12559
	Pearson	,033	,163	,040	-,168	-,047	,073	-,119	,122	-,058	-,060	-,459	-,271	-,033	1
	Correlation														
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	.
	N	12559	12559	12559	12559	12559	12559	12559	12559	4033	12559	2352	1249	12559	12559

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Correlação Estoque_log X Grade_log

Figura 28 – Gráfico *scatterplot* entre “Estoque_log” e “Grade_log”

Correlação Mostruario_log X Preço_pot

Figura 29 – Gráfico *scatterplot* entre “Mostruario_log” e “Preço_pot”

Correlação Classe_distribuicao X Classe

Figura 30 – Gráfico *scatterplot* entre “Classe_distribuicao” e “Classe”

Correlação Estoque_log X Vendas_log

Figura 31 – Gráfico *scatterplot* entre “Estoque_log” e “Vendas_log”

Correlação Pprazo X Vendas

Figura 32 – Gráfico *scatterplot* entre “Pprazo” e “Vendas_log”

Autocorrelação da variável “Vendas_log”

Figura 33 – Gráfico de autocorrelação da variável “Vendas_log”