

PATRICIA BRANDELLI GALLINA

A UTILIZAÇÃO DE *DATA MINING* EM EMPRESAS DE *E-COMMERCE*  
OBJETIVANDO A FIDELIZAÇÃO DO CLIENTE

São Paulo

2012

PATRICIA BRANDELLI GALLINA

A UTILIZAÇÃO DE *DATA MINING* EM EMPRESAS DE *E-COMMERCE*  
OBJETIVANDO A FIDELIZAÇÃO DO CLIENTE

Monografia apresentada à Escola Politécnica  
da Universidade de São Paulo, como parte  
dos requisitos para obtenção do título de  
MBA em Tecnologia da Informação.

São Paulo

2012

PATRICIA BRANDELLI GALLINA

A UTILIZAÇÃO DE *DATA MINING* EM EMPRESAS DE *E-COMMERCE*  
OBJETIVANDO A FIDELIZAÇÃO DO CLIENTE

Monografia apresentada à Escola Politécnica da Universidade de São Paulo, como parte dos requisitos para obtenção do título de MBA em Tecnologia da Informação.

Área de Concentração: Tecnologia da Informação

Orientação: Prof. Dr. Alexandre Nardi

São Paulo

2012

MBA/TI

2012

G-138 u

DEDALUS - Acervo - EPEL



31500022091

## FICHA CATALOGRÁFICA

m2012 AS x

**Gallina, Patricia Brandelli**

**A utilização de data mining em empresas de e-commerce  
objetivando a fidelização do cliente / P.B. Gallina – São Paulo,  
2012.**

**47 p.**

**Monografia (MBA em Tecnologia da Informação) – Escola  
Politécnica da Universidade de São Paulo. Programa de Educação  
Continuada em Engenharia.**

**1.Mineração de dados 2.Tecnologia da informação 3.Banco de  
dados I. Universidade de São Paulo. Escola Politécnica. Programa  
de Educação Continuada em Engenharia II. t.**

## AGRADECIMENTOS

Ao meu orientador Alexandre Nardi pelo apoio, conhecimento e direcionamento, essenciais para a construção deste trabalho.

Aos meus professores do curso de MBA, pelos valiosos ensinamentos.

Aos meus colegas do curso de MBA, pela alegria, companheirismo, amizade e persistência nos momentos em que o cansaço parecia maior do que tudo, principalmente à Elaine, minha fiel companheira de mesa nos últimos 2 anos.

Ao meu marido Angelo, pelo amor e paciência.

A minha mãe Eda, e meus irmãos Cristiane e Daniel, por estarem sempre ao meu lado.

## RESUMO

Com o crescimento da Internet e do comércio eletrônico está surgindo uma nova maneira das empresas conhecerem e se relacionarem com seus clientes. O marketing direto, ou marketing de relacionamento, está tendo que se adaptar aos novos tempos de vendas pela *Web*. Como as empresas podem conhecer as pessoas que visitam os seus *sites*? Como descobrir suas preferências, suas necessidades, seus desejos de consumo? O *Data Mining* está mostrando ser uma poderosa ferramenta para auxiliar as empresas na busca deste conhecimento. Este trabalho tem por objetivo demonstrar como as empresas estão utilizando o *Data Mining* e o *Web Mining* para melhorarem o conhecimento sobre seus clientes, e para conseguirem aumentar as suas vendas através de uma propaganda mais direcionada, personalizada e eficiente. Inicialmente são apresentados alguns conceitos explicativos sobre *e-commerce*, fidelização de clientes, como estão estruturadas as informações que trafegam na *Web*, e conceitos sobre *Data Warehouse*, *Knowledge Discovery in Database (KDD)*, *Data Mining* e *Web Mining*. E por fim são apresentados estudos de casos de empresas reais que utilizaram o *Data Mining* em seus processos e obtiveram resultados muito satisfatórios.

**Palavras-chave:** *Data Mining*, *Web Mining*, Fidelização de Clientes, *e-commerce*.

## ABSTRACT

Considering the constant growth of the Internet and e-commerce, the companies have found an arising way to introduce themselves and engage with their customers. Direct marketing, also known as customer relationship marketing, is adapting itself to the new age of online sales. How can companies know who the people that visit their websites are? How is it possible to find out their preferences, needs, consumption desires? Data Mining is proving to be a powerful tool to help companies that seek this kind of information. This study aims to show how companies are using the Data Mining and Web Mining to improve their knowledge about their customers and also to increase their sales by using a more focused, personalized and efficient advertising. Initially will be showed some explanatory concepts about e-commerce, customer loyalty, how is structured the information that travels on the Web, and concepts about Data Warehouse, Knowledge Discovery in Database (KDD), Data Mining and Web Mining. At last will be demonstrated case studies about companies that used Data Mining in their processes and got very satisfactory results.

**Keywords:** Data Mining, Web Mining, Customer Loyalty, e-commerce.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Comparativo de preços de lojas de <i>e-commerce</i> .....	10
Figura 2 – Faturamento do <i>e-commerce</i> no Brasil – em Bilhões de Reais .....	18
Figura 3 – Acesso à Internet no Brasil – Indicadores Gerais .....	18
Figura 4 – Comparação entre bancos de dados operacionais e <i>Data Warehouse</i> ...	22
Figura 5 – Processo de <i>KDD</i> .....	23
Figura 6 – Utilização de cookies para identificar o usuário .....	30
Figura 7 – Recomendações de produtos .....	30
Figura 8 – Itens sugeridos pelo <i>site</i> .....	32
Figura 9 – Sessão “Seu Histórico Recente” .....	33

## SUMÁRIO

1. CAPÍTULO 1 – CONSIDERAÇÕES INICIAIS.....	10
1.1 INTRODUÇÃO .....	10
1.2 OBJETIVOS .....	11
1.3 JUSTIFICATIVA .....	12
1.4 ABRANGÊNCIA .....	13
1.5 METODOLOGIA.....	14
1.6 ORGANIZAÇÃO DO TEXTO.....	14
2. CAPÍTULO 2 – CONCEITUAÇÃO TEÓRICA.....	16
2.1 MARKETING DIRETO E <i>e-COMMERCE</i> .....	16
2.2 FIDELIZAÇÃO DE CLIENTES.....	19
2.3 A ESTRUTURAÇÃO DOS DADOS TRAFEGADOS NA <i>WEB</i> .....	20
2.4 <i>DATA WAREHOUSE</i> .....	21
2.5 <i>KNOWLEDGE DISCOVERY IN DATABASE</i> .....	22
2.5.1 Etapas do processo de <i>KDD</i> .....	23
2.6 <i>DATA MINING</i> .....	24
2.7 <i>WEB MINING</i> .....	26
3. CAPÍTULO 3 – <i>DATA MINING</i> EM EMPRESAS DE <i>e-COMMERCE</i> .....	28
3.1 ESTUDO DE CASO: B2W Companhia Global do varejo .....	28
3.2 ESTUDO DE CASO: Amazon.com.....	31
3.3 ESTUDO DE CASO: WalMart .....	34
3.4 OUTROS EXEMPLOS DE APLICAÇÃO .....	36
4. CAPÍTULO 4 – CONSIDERAÇÕES FINAIS.....	42
4.1 CONCLUSÕES .....	42
4.2 TRABALHOS FUTUROS .....	44
REFERÊNCIAS BIBLIOGRÁFICAS.....	45

# 1. CAPÍTULO 1 – CONSIDERAÇÕES INICIAIS

## 1.1 INTRODUÇÃO

Com o número crescente de empresas de *e-commerce* no mercado, está se tornando cada vez mais difícil conquistar e manter clientes. As vantagens do *e-commerce* para as empresas são inúmeras: a loja (ou o *site*) está “aberta” 24 horas por dia, 7 dias por semana, não há custos de manter um local físico e vendedores, é possível “visitar” toda a loja e ver seus produtos de maneira rápida e com comodidade, apenas com alguns cliques. Porém, ao analisar algumas das maiores e já consolidadas empresas que atuam neste segmento, como Submarino, Americanas, Fast Shop, Ponto Frio, Carrefour, percebe-se que não existem muitas diferenças significativas entre elas. Levando-se em conta seu ramo de atuação, todas possuem praticamente os mesmos produtos e pouca variação em relação a preços, valores de frete e condições de pagamento, o que pode ser exemplificado através da Figura 1. As informações nela demonstradas foram obtidas nos *sites* das respectivas lojas na Internet, no dia 12 de novembro de 2011. A coluna com valores de frete foi suprimida pois na data em questão todos os *sites* pesquisados estavam com valor de frete “Grátis”.

Lojas	Câmera Fotográfica Sony Cyber Shot W510		Celular Samsung Galaxy Ace S5830		Livro "Steve Jobs: A Biografia" - Walter Isaacson		Refrigerador Brastemp BRW50NB Inox	
	Preço	Cond.Pag.	Preço	Cond.Pag.	Preço	Cond.Pag.	Preço	Cond.Pag.
Submarino	399,00	12x s/juros	696,00	12x s/juros	37,50	7x s/juros	3499,00	12x s/juros
Americanas	399,00	12x s/juros	696,00	12x s/juros	33,75	-	-	-
Fast Shop	403,53	10x s/juros	718,08	12x s/juros	-	-	3499,00	12x s/juros
Ponto Frio	449,00	12x s/juros	759,75	10x s/juros	34,40	3x s/juros	3199,00	10x s/juros
Carrefour	399,00	9x s/juros	799,00	12x s/juros	-	-	3099,00	12x s/juros
Wallmart	449,00	12x s/juros	799,00	12x s/juros	37,40	-	3299,00	12x s/juros

Figura 1 – Comparativo de preços de lojas de *e-commerce*

Fonte: Sites do Submarino, Americanas, Fast Shop, Ponto Frio, Carrefour e Wallmart, 2012

Somando-se a isso temos a Internet, que produz diariamente um volume de dados potencialmente muito rico no que se refere à informações sobre negócios, mercados, clientes, preferências e novas oportunidades. No geral estas informações encontram-se de maneira desestruturada, com diferentes tipos de dados, tais como números, textos, imagens, sons, vídeos, *logs* de acesso a servidores, tornando-se essencial conseguir identificar que informações são relevantes para se extrair delas um conhecimento útil.

Sem dúvida, um dos maiores desafios para as empresas de *e-commerce* é justamente captar, armazenar, processar e analisar estes dados de forma que possam ser revertidos em reais benefícios ao negócio, proporcionando uma estratégia de marketing mais adequada ao perfil das pessoas que estão acessando o *site*, moldando o conteúdo das páginas de forma a agradar cada usuário, transformando-o de simples visitante em comprador potencial. Se as empresas conhecerem melhor seus clientes, seus desejos e necessidades, a comunicação entre eles se tornará mais assertiva e mais eficiente.

## 1.2 OBJETIVOS

*Data Mining*, ou Mineração de Dados, pode ser entendido como o processo de extração de informações, sem conhecimento prévio, de um grande banco de dados e seu uso para tomada de decisões. É uma metodologia aplicada em diversas áreas que usam o conhecimento, como empresas, indústrias e instituições de pesquisa. Segundo Sferra e Corrêa (2003), *Data Mining* define um processo automatizado de captura e análise de grandes conjuntos de dados para extrair um significado, sendo usado tanto para descrever características do passado como para prever tendências para o futuro.

A Mineração de Dados é uma das etapas do processo de *KDD* (*Knowledge Discovery in Database*), e possui como resultado final a extração do conhecimento. Sua principal característica é a extração de informações não triviais que estão implicitamente contidas nas bases de dados, o que é difícil de se fazer com métodos tradicionais de análise. O processo é iterativo e composto de uma série de etapas

sequenciais, podendo inclusive retornar às etapas anteriores. Encontrar padrões exige que haja uma simplificação e organização nos dados, desconsiderando aquilo que é específico e privilegiando o que é genérico, ou seja, o objetivo maior é poder generalizar o conhecimento adquirido para novas ocorrências do fenômeno ou para outros contextos ou situações parecidas com a utilizada, na construção do modelo computacional. Na prática, ao conseguir identificar grupos de consumidores com características ou comportamentos semelhantes, as empresas podem direcionar uma oferta de algum produto ou serviço com mais precisão, pois será mais adequada às necessidades daquele grupo.

Através do *Data Mining* as grandes empresas podem conhecer melhor seus clientes, seus padrões de consumo e necessidades. O *DM* resgata em organizações grandes o papel do dono da loja atendendo no balcão e conhecendo seus clientes. Como resultado prático disso tem-se um melhor direcionamento das estratégias de marketing para cada tipo de usuário. Pode-se oferecer aos clientes exatamente o que eles querem ou necessitam, ou melhor, pode-se antecipar os seus desejos de consumo.

O objetivo deste trabalho é demonstrar como as técnicas de *Data Mining* podem ser utilizadas para auxiliar empresas de *e-commerce* a capturar, processar e entender os dados obtidos através da navegação em seus *sites*, definir padrões de comportamento, tendências e associações comuns, objetivando a fidelização do cliente.

### 1.3 JUSTIFICATIVA

Todos os dias existem milhões de usuários conectados à Internet de forma cada vez mais intensa e inerente ao seu dia a dia. A *Web* é utilizada para a troca de informações, fazer compras, pagar contas, conversar através de voz e mensagens de texto, publicar textos, imagens e vídeos, manter relacionamentos pessoais e profissionais. Somando-se a isso um ecossistema cada vez mais sofisticado e moderno, tem-se uma verdadeira explosão de dados sendo gerados e armazenados diariamente.

Segundo Póvoa (2012), a Internet produziu mais dados em 2010 do que em toda a sua história. Somente o *site Youtube* tem 48 horas de *upload* de vídeos por minuto. O *Facebook* já tem 100 bilhões de fotos em seus servidores, e uma média de 6 bilhões de imagens adicionadas mensalmente. Ao terminar de ler esta linha, cerca de 34 mil buscas foram executadas no *Google*, em apenas 1 segundo.

Toda essa imensa massa de dados cresce rapidamente, e carrega consigo uma diversidade de informações a respeito de seus usuários, como se comportam, com quem conversam, o que compram, o que publicam, quais são seus interesses. As empresas que souberem identificar esses consumidores, entendê-los, e prever ou sugerir suas tendências de compras, terão uma mina de ouro nas mãos. Como disse Póvoa (2012): "Informação existe à vontade, o desafio é o que fazer com ela".

Este trabalho se justifica pela participação cada vez maior da internet e do *e-commerce* em nossas vidas, e pela importância em se aproveitar a grande quantidade e diversidade de informações que podem ser obtidas através de *logs* de navegação *Web*, com o objetivo de tornar o marketing direto mais assertivo.

#### 1.4 ABRANGÊNCIA

De acordo com os objetivos propostos, inicialmente será feita uma análise conceitual sobre Marketing Direto, Fidelização de Clientes e *e-commerce* e, considerando-se o tema proposto para este trabalho, será dada maior ênfase ao tipo de comércio eletrônico *Business to Consumer* (B2C), pois o escopo desta análise é bem menos abrangente do que o *e-commerce* na sua totalidade.

Também serão analisados brevemente como estão estruturados os dados de *logs* de servidores *Web* e como devem ser armazenados em *Data Warehouses*, para que facilitem a extração de detalhes valiosos sobre cada gesto efetuado pelo usuário em um *site*.

E finalmente, será feito um detalhamento conceitual sobre o processo de *KDD* (*Knowledge Discovery in Database*), enfatizando-se a etapa de *Data Mining*, ou *Web Mining*, que é o *Data Mining* aplicado à dados da *Web*.

Após esta etapa serão apresentados casos práticos do uso de *Data Mining* em empresas existentes no mercado brasileiro, e também empresas estrangeiras.

É importante salientar que as informações sobre as empresas citadas foram obtidas em artigos de domínio público e através da Internet, não sendo, portanto, informações oficiais fornecidas pelas empresas em questão.

## 1.5 METODOLOGIA

Para o desenvolvimento da parte teórica deste trabalho foram efetuadas pesquisas em diversas fontes de informações que continham tópicos relacionados ao *Data Mining* e *KDD*, *Data Warehouse*, *e-commerce* e Marketing direto, tais como livros, artigos científicos, monografias, revistas, etc.

No entanto, houve um pouco de dificuldade em se encontrar informações de fontes confiáveis sobre como as empresas estão usando o *Data Mining*, por se tratar de um assunto confidencial e estratégico para o negócio. Em função disso foram necessárias muitas pesquisas na Internet, em *sites*, revistas e artigos especializados.

## 1.6 ORGANIZAÇÃO DO TEXTO

Este trabalho está organizado da seguinte forma:

No Capítulo 1 é feita uma apresentação do assunto que será tratado, do contexto no qual ele está sendo analisado, os objetivos deste estudo e sua justificativa, e a metodologia que foi utilizada.

No Capítulo 2 será apresentada uma visão geral da parte teórica dos assuntos abordados neste trabalho, com conceitos de: Marketing Direto e *e-commerce*, Fidelização de Clientes, como os dados na *Web* estão estruturados, *Data*

*Warehouse*, o *KDD (Knowledge Discovery in Database)* e suas etapas, *Data Mining* e *Web Mining*.

No Capítulo 3 serão analisados alguns exemplos práticos de empresas que utilizam o *Data Mining* como ferramenta de apoio ao marketing mais direcionado e à fidelização de clientes.

No Capítulo 4 serão apresentadas as conclusões finais do trabalho.

## 2. CAPÍTULO 2 – CONCEITUAÇÃO TEÓRICA

### 2.1 MARKETING DIRETO E e-COMMERCE

Nos dias de hoje as empresas estão tendo acesso à seus clientes de forma mais direta, sem intermediários, e o surgimento da Internet contribuiu com isso significativamente. Através dos meios eletrônicos as empresas podem fazer ofertas diariamente a seus clientes, e medir os resultados com mais precisão. Segundo Kotler (2000) marketing direto é um sistema interativo que usa um ou mais meios de propaganda para obter uma resposta mensurável e/ou uma transação em qualquer local. Quando fala-se em “resposta mensurável” quer-se dizer, em geral, um pedido de cliente. No entanto, alguns profissionais de marketing referem-se ao marketing direto de maneira mais abrangente, como tendo a função de construir e manter um relacionamento duradouro com o cliente (marketing direto de relacionamento). Em função da Internet a relação entre empresa e cliente está se tornando mais estreita e personalizada.

Conforme Martin apud Magalhães (2007), o comércio eletrônico consiste na compra e venda de produtos, informações e serviços pela Internet. O comércio *online* mantém as mesmas características de troca entre compradores e vendedores, com a diferença de que ampara-se totalmente na Internet para existir.

Segundo Potter, Turban e Rainer (2005), existem vários tipos de comércio eletrônico, sendo os mais comuns:

**B2B – Business to Business:** É a negociação eletrônica entre as empresas. Sendo bastante comum, é a modalidade que mais movimentava importâncias monetárias.

**B2C – Business to Consumers:** É a negociação eletrônica entre empresas e consumidores. Representa a virtualização da compra e venda, as pessoas escolhem e pagam pelos produtos pela Internet. Como exemplo tem-se os sites [www.submarino.com](http://www.submarino.com), [www.voegol.com.br](http://www.voegol.com.br), [www.amazon.com](http://www.amazon.com), entre muitos outros.

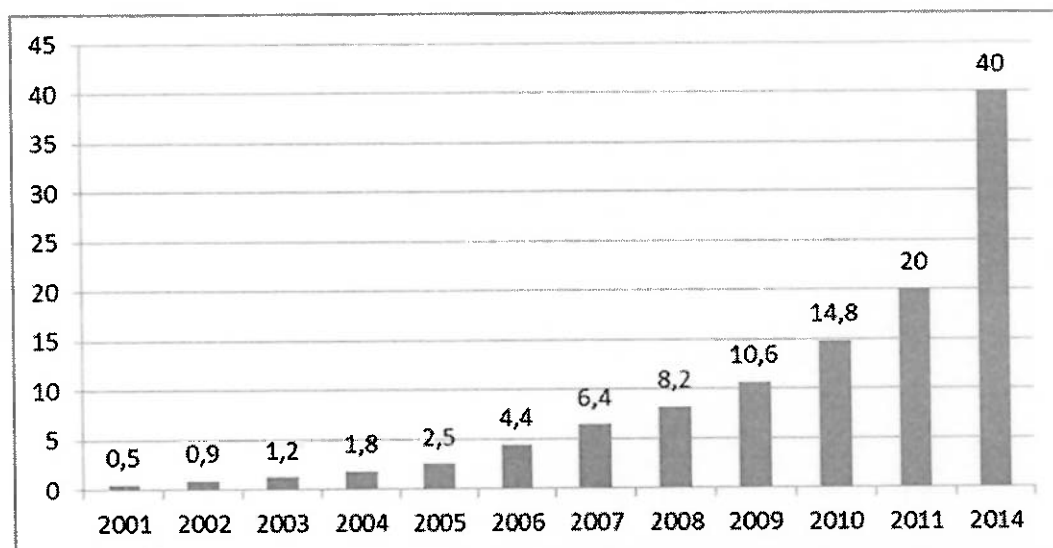
**C2B – Consumers to Business:** Consiste na negociação eletrônica entre consumidores e empresas. É o reverso do B2C, também chamado de “Leilão reverso”. Acontece quando consumidores vendem para empresas.

**C2C – Consumers to Consumers:** É a negociação eletrônica entre consumidores. É uma modalidade muito comum e efetua muitas negociações, porém de valores pequenos. O exemplo mais conhecido no Brasil é o site [www.mercadolivre.com.br](http://www.mercadolivre.com.br).

Entre os fatores que mais estão contribuindo para o crescimento do comércio eletrônico podemos citar a popularização da Internet, o aumento no consumo de computadores pelas classes sociais mais baixas, a comodidade de se fazer compras em casa, sem enfrentar filas, sem custo de estacionamento nos grandes centros de compras, sem congestionamentos, de maneira rápida e individualizada, e ainda permitindo que o consumidor faça pesquisa de preços em lojas similares, sem precisar sair de casa.

Para as empresas o comércio eletrônico também traz vantagens. O direcionamento das ofertas de produtos pode ser feito especificamente para um grupo de potenciais clientes previamente identificado. As empresas selecionam grupos de pessoas que lhes interessam, como mulheres grávidas, pessoas com obesidade, clientes com maior poder aquisitivo, etc, e criam campanhas customizadas para o grupo que se quer atingir. Isso permite que se tenha um grau de acerto maior na oferta que foi feita, além de que é um tipo de “contato” que não é tão visível para seus concorrentes, o que pode significar uma vantagem competitiva.

Segundo dados da empresa de consultoria e-bit citados em Facchini (2011), houve um crescimento de 40% no faturamento de comércio eletrônico no Brasil em 2010, totalizando R\$ 14,8 bilhões. Estima-se que em 2011 o volume de operações gire em torno de R\$ 20 bilhões, e R\$ 40 bilhões em 2014, conforme está demonstrado na Figura 2.



**Figura 2 – Faturamento do e-commerce no Brasil – em Bilhões de Reais**

Fonte: Site <http://www.e-commerce.org.br/stats.php>, 2012.

Conforme apontado na 24<sup>a</sup> edição do Relatório *WebShoppers* (2011), no que depender dos consumidores, o comércio eletrônico não deve parar de evoluir. Apesar de existirem problemas de logística que interferem no bom atendimento das entregas por parte das empresas, levantou-se que 86% dos consumidores brasileiros ficaram satisfeitos com o comércio virtual no primeiro semestre de 2011.

Na figura 3 é apresentado um detalhamento um pouco maior sobre como os brasileiros navegam na Internet, qual é o tempo de navegação no mês (em horas) e o tempo médio gasto em cada página visualizada (em segundos), conforme informações obtidas do site [www.e-commerce.org.br](http://www.e-commerce.org.br). Estão sendo chamados de “Usuários Ativos” os usuários de Internet que tiveram pelo menos um acesso no mês anterior.

Setembro - 2011 Brasil	
Pessoas >16 anos c/ acesso a Internet - (milhões)	78 mi
Usuários Ativos (milhões)	46,3 mi
Tempo de navegação no mês (hs)	48:26 hs
Tempo médio gasto em cada página visualizada (seg)	00:47 seg

**Figura 3 – Acesso à Internet no Brasil – Indicadores Gerais**

Fonte: Site <http://www.e-commerce.org.br/stats.php>, 2012.

Ainda com base em dados obtidos da empresa de consultoria e-bit (2011), a categoria de produtos mais vendida pela internet nos 6 primeiros meses de 2011 é a de "Eletrodomésticos", com 13% do volume total de pedidos, seguida pela categoria "Informática", com 12%, depois "Saúde, Beleza e Medicamentos", com 11%, logo após "Livros e assinaturas de jornais e revistas", com 8%, e por último a categoria "Eletrônicos", com 6%.

## 2.2 FIDELIZAÇÃO DE CLIENTES

No cenário atual de globalização que se observa, com uma diversidade cada vez maior de produtos, serviços, marcas, preços e fornecedores, a concorrência entre as empresas está cada vez mais acirrada, e conquistar a fidelidade dos clientes tornou-se uma questão de sobrevivência.

Para Kotler (2000) a probabilidade de satisfação e repetição da compra por parte de um cliente depende de a oferta atender ou não suas expectativas de valor. Os clientes sempre procuram maximizar o valor, dentro dos limites de custos envolvidos na procura pelo produto, e limitações de conhecimento, mobilidade e receita. Na prática isso quer dizer que, se o desempenho não alcançar as expectativas, o cliente ficará insatisfeito. Por outro lado, se o desempenho for além das expectativas, o cliente ficará altamente satisfeito e encantado.

Seguindo a idéia de satisfazer e encantar seus clientes, se uma empresa consegue surpreendê-los e se antecipar às suas necessidades, maiores serão as suas chances de fidelização. Para isso as empresas precisam acima de tudo conhecer seus clientes potenciais, entender seu comportamento, suas condições de vida, seus familiares, suas motivações e desejos.

Conforme Andreassen e Lindestad apud Baptista (2005), a relação entre a taxa de retenção de clientes e a rentabilidade de longo prazo não é linear, e sim exponencial, porque mesmo uma pequena alteração na taxa de retenção tem efeitos significativos na rentabilidade futura de longo prazo.

### 2.3 A ESTRUTURAÇÃO DOS DADOS TRAFEGADOS NA WEB

O surgimento e a popularização da Internet mudaram definitivamente a forma como se lida e utiliza a informação. Antes dela, para se fazer uma pesquisa por exemplo, eram necessários livros, artigos, enciclopédias, tudo isso sendo consultado provavelmente em uma biblioteca. Com a *Web* é possível ler, pesquisar, acessar e compartilhar uma quantidade de informações que provavelmente não se conseguiria com os métodos anteriormente citados.

As informações que trafegam na *Web* apresentam-se em formatos diversos: são textos, tabelas de dados, fotos, arquivos de áudio, arquivos de vídeo, etc. Em uma única página da Internet tem-se vários “pedaços” de informações distintas, como *links* de navegação, propagandas, políticas de privacidade, além do conteúdo da página propriamente dito.

Outro tipo de informação muito valiosa, porém transparente aos usuários, são os *logs* de servidores *Web*. Os indivíduos interagem com os *sites* da *Web* através de um programa que é chamado de Navegador. Todas as informações da “comunicação” que foi estabelecida entre um usuário e o servidor onde a página *Web* está armazenada ficam registradas, é o que se chama *log*.

Os *logs* registram qual página o usuário solicitou, e toda a sequência cronológica de cliques que ele efetuou. A partir dessas informações pode-se analisar, por exemplo, se a navegação resultou em uma compra, quanto tempo ele levou para localizar seu produto no *site*, se o usuário saiu da página antes de todo o conteúdo ser carregado, etc. Conforme definido por Chiara (2003), é como se fosse possível olhar por sobre o ombro do cliente enquanto ele faz suas compras. E a melhor parte é que pode-se “modificar” a loja enquanto o cliente “anda” por ela. Pode-se reorganizar as “gôndolas virtuais” para mostrar itens que provavelmente ele compraria.

Os dados que trafegam na *Web* classificam-se como estruturados, semi-estruturados e não-estruturados. Os dados não estruturados são os “documentos” da *Web*, os textos. Os dados estruturados são armazenados em estruturas bem definidas, os bancos de dados. E os dados semi-estruturados são um ponto de

convergência entre os dois citados anteriormente, por exemplo os documentos XML, e naturalmente encorajam a adaptação ou extensão, e posterior aplicação, de técnicas conhecidas de mineração de dados para dados estruturados, a fim de se descobrir padrões interessantes em dados semi-estruturados.

## 2.4 DATA WAREHOUSE

Um *Data Warehouse* é um grande aglomerado de dados provenientes das mais diversas bases de dados existentes e mantidas por uma empresa, passando antes por processos de extração, crítica, transformação e padronização dos dados. Segundo Addrians e Zantinge (1996) apud Sferra e Corrêa (2003), *Data Warehouse* é um depósito central de dados, extraídos de dados operacionais, em que a informação é orientada a assuntos, não volátil e de natureza histórica. Devido a essas características, os *Data Warehouses* são repositórios de dados extremamente organizados, o que facilita a aplicação do *Data Mining*.

Os bancos de dados operacionais armazenam os dados necessários para as operações do dia-a-dia da empresa, são utilizados pelos diferentes sistemas de aplicação que a empresa utiliza, podendo assim sofrer constantes atualizações nos seus dados. Já o *Data Warehouse* armazena dados analíticos, tanto detalhados como resumidos, e normalmente são destinados às necessidades da gerência nos processos de tomada de decisões. Na Figura 4 pode-se ver um resumo das características dos dois tipos de bancos de dados.

Características	Bancos de dados Operacionais	Data Warehouse
Objetivo	Operações diárias do negócio	Analisar o negócio
Uso	Operacional	Informativo
Tipo de processamento	OLTP	OLAP
Unidade de trabalho	Inclusão, alteração, exclusão	Carga e consulta
Número de usuários	Milhares	Centenas
Tipo de usuário	Operadores	Comunidade gerencial
Interação do usuário	Somente pré-definida	Pré-definida e <i>ad-hoc</i>
Condições dos dados	Dados operacionais	Dados analíticos
Volume	Megabytes - gigabytes	Gigabytes - terabytes
Granularidade	Detalhados	Detalhados e resumidos

Redundância	Não ocorre	Ocorre
Estrutura	Estática	Variável
Acesso a registros	Dezenas	Milhares
Atualização	Contínua (tempo real)	Periódica (em <i>batch</i> )
Número de índices	Poucos/Simples	Muitos/Complexos
Intenção dos índices	Localizar um registro	Aperfeiçoar consultas

**Figura 4 – Comparação entre bancos de dados operacionais e *Data Warehouse***

Fonte: Thomé.

Com base nesses conceitos observa-se que o *Data Warehouse* não é um fim, mas sim um meio para que as empresas consigam dispor e analisar informações históricas objetivando a melhoria dos seus processos e negócios. É neste contexto que a utilização do *KDD* torna-se essencial, proporcionando a geração de um novo conhecimento para as empresas a partir da busca de padrões e correlações que não eram possíveis de serem identificadas com os bancos de dados operacionais.

## 2.5 KNOWLEDGE DISCOVERY IN DATABASE

*Data Mining* é uma das etapas de um processo maior chamado *KDD* (*Knowledge Discovery in Database*), em português “Descoberta do Conhecimento em Bases de Dados”. Segundo Addrians e Zantinge (1996) apud Sferra e Corrêa (2003), o *KDD* permite a extração não trivial de conhecimento previamente desconhecido e potencialmente útil de um banco de dados. Na prática isso significa extrair conhecimento útil de grandes conjuntos de dados, descobrir padrões de comportamento de acordo com regras previamente estabelecidas. Este conceito é enfatizado por Chiara (2003), que diz que o *KDD* é um processo de descoberta de novas correlações através da mineração de grandes quantidades de dados armazenados.

O processo de *KDD* é dividido em algumas etapas, que serão detalhadas em seguida, e que estão sendo demonstradas na Figura 5.

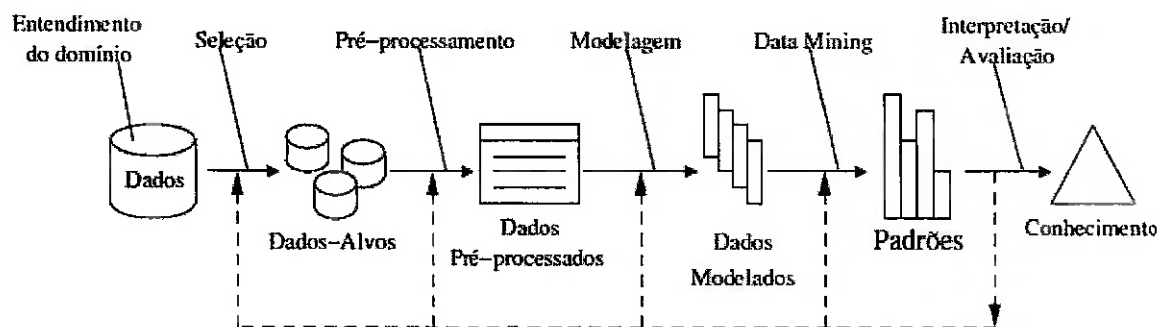


Figura 5 – Processo de *KDD*

Fonte: Chiara, 2003.

### 2.5.1 Etapas do processo de *KDD*

As etapas do processo de *KDD* são executadas de maneira iterativa e interativa. Conforme citado por Brachnad e Anand (1996) apud Sferra e Corrêa (2003), as etapas são interativas pois envolvem a cooperação da pessoa responsável pela análise dos dados, cujo conhecimento sobre o domínio orientará a execução do processo. Já a iteração existe porque este não é um processo que ocorre de forma sequencial e, às vezes, necessita que algumas etapas sejam repetidas, com novas seleções de parâmetros, novos conjuntos de dados, aplicações de diferentes técnicas de *Data Mining* e posterior análise dos resultados.

**Etapa 1 - Entendimento do domínio:** Consiste em analisar e entender qual o domínio dos dados que serão tratados e quais os objetivos do processo de análise que estará sendo iniciado.

**Etapa 2 – Seleção dos dados:** é a escolha dos dados que serão analisados, dentro da base de dados existente.

**Etapa 3 – Pré-processamento dos dados:** nesta etapa é feita a limpeza dos dados, que consiste na remoção de informações interferentes (inconsistências e duplicações), e na decisão de como manipular as informações que estão faltando. A qualidade dos dados analisados é essencial para se obter resultados confiáveis. Segundo Mannila (1996) apud Sferra e Corrêa (2003), esta etapa pode tomar até

80% do tempo necessário para todo o processo, devido às dificuldades de integração de bases de dados heterogêneas.

**Etapa 4 – Modelagem dos dados:** Também chamada de Transformação dos dados, nesta etapa os dados devem passar por outra transformação, que os armazena adequadamente visando facilitar o uso das técnicas de *Data Mining*. Conforme citado por Sferra e Corrêa (2003), nesta fase é possível que haja uma considerável expansão do *Data Warehouse*, já que as informações são alocadas de maneira mais eficiente.

**Etapa 5 – Data Mining:** consiste na aplicação de técnicas e algoritmos computacionais específicos, que devem ser escolhidos de acordo com o problema que se está tentando resolver. Para cada problema particular existe uma técnica particular. Nesta etapa é muito importante a experiência e intuição do analista.

**Etapa 6 – Interpretação dos dados:** é a análise e interpretação dos padrões resultantes da etapa de *Data Mining*. Se estes resultados não estiverem satisfatórios ou não fizerem sentido, deve-se retornar a qualquer uma das etapas anteriores. Lembrando do próprio conceito do *KDD*, o resultado deve ser um conhecimento útil para aplicação em um determinado problema.

## 2.6 DATA MINING

Segundo Han & Kamber (2006), *Data Mining*, ou Mineração de Dados, consiste na extração ou "mineração" de conhecimento de grandes conjuntos de dados. É uma das etapas do processo de *KDD*, onde algoritmos computacionais são aplicados aos dados para a obtenção de padrões.

Para que a descoberta do conhecimento seja relevante é importante utilizar o método mais apropriado ao resultado que se deseja obter. Em seguida serão detalhados alguns métodos de *Data Mining* conforme definição e citação feita por Fayyad et al. (1996) apud Sferra e Corrêa (2003):

**Classificação:** associa ou classifica um item a uma ou várias classes categóricas pré-definidas. A construção do modelo segundo esta estratégia pressupõe o conhecimento prévio das possíveis classes e a correta classificação dos exemplos usados na modelagem. Um exemplo para este tipo de abordagem é em Análise de Risco, onde o objetivo seja classificar um cliente entre as classes de excelente, bom ou mau pagador.

**Modelos de Relacionamento entre Variáveis:** associa um item a uma ou mais variáveis de predição de valores reais, consideradas variáveis independentes ou exploratórias.

**Agregação (ou *clustering*):** Conforme citado por Lamboia e Pereira (2005), consiste na busca de similaridades entre os dados tal que permita definir um conjunto finito de classes ou categorias que os contenha e os descreva. A principal diferença entre essa abordagem e a Classificação é que na Agregação não se tem um conhecimento prévio sobre o número de classes possíveis nem a possível pertinência dos exemplos usados na modelagem. Este modelo agrupa informações homogêneas de grupos heterogêneos entre os demais e aponta o item que melhor representa cada grupo, também conhecido como *cluster*. Exemplos de aplicações para este método são a descoberta de grupos homogêneos de clientes de acordo com seus hábitos de consumo, ou a segmentação demográfica visando a determinação de características de um grupo social.

**Sumarização:** determina uma descrição compacta para um dado subconjunto. As funções de Sumarização são frequentemente usadas na análise exploratória de dados com geração automatizada de relatórios, sendo responsáveis pela descrição compacta de um conjunto de dados.

**Regras de Associação:** determinam relações entre campos de um banco de dados, identificam fatos que possam ser direta ou indiretamente associados a outros. Esta estratégia normalmente é usada em aplicações onde se buscam identificar itens que possam ser colocados juntos em um mesmo pacote de negociação. Por exemplo, a constatação de que vendas de fraldas e cerveja apresentam um comportamento ascendente às sextas-feiras podem identificar um tipo especial de cliente.

**Modelo de Dependência:** descreve dependências significativas entre variáveis. Estes modelos existem em dois níveis, estruturado e quantitativo. O nível estruturado especifica, geralmente em forma de gráfico, quais variáveis são localmente dependentes. O nível quantitativo especifica o grau de dependência, usando alguma escala numérica. Geralmente um modelo utiliza os dois tipos de dependências, um para encontrar a relação e outra para quantificá-la. Estes modelos são utilizados em áreas como desenvolvimento de probabilidades em sistemas especialistas da área médica, recuperação da informação e modelagem de genes humanos.

**Análise de Séries Temporais:** determina características sequenciais, como dados com dependência no tempo. Seu objetivo é modelar o estado do processo extraindo e registrando desvios e tendências no tempo. As séries são compostas por quatro padrões: tendência, variações cíclicas, variações sazonais e variações irregulares.

## 2.7 WEB MINING

Pode-se definir *Web Mining* como uma extensão do *Data Mining* aplicado a dados da Internet. É o uso de técnicas de *Data Mining* para descobrir e extrair automaticamente informações relevantes dos documentos e serviços ligados à *Web*. Segundo Liu (2007) *Web Mining* tem por objetivo extrair informações ou conhecimento útil através de estruturas da *Web*, como *hyperlinks*, conteúdo de páginas e dados de *logs* de usuários e, apesar de se utilizar de muitas técnicas do *Data Mining*, não é puramente uma aplicação deste devido à heterogeneidade e naturezas semi-estruturada e desestruturada dos dados da *Web*.

Ainda conforme citado por Liu (2007), pode-se dividir o *Web Mining* em três sub-áreas, conforme detalhado abaixo:

***Web Structure Mining:*** é a descoberta de conhecimento útil através de *hyperlinks* (ou *links*). Por exemplo, como exemplificado por Chiara (2003), muitos *links* apontando para um documento indicam sua popularidade, ou então muitos

*links* saindo de um documento indicam uma riqueza de tópicos cobertos pelo mesmo. É possível também identificar comunidades de usuários que compartilham os mesmos interesses. A criação de um *link* numa página indica que seu autor conferiu autoridade para a página que está sendo referenciada por esse *link*. Páginas em que chegam muitos *links* são chamadas de “Autoridades”. Por outro lado, são chamadas de *Hubs* as páginas que centralizam essas Autoridades.

Como exemplo do uso de *Web Structure Mining* tem-se a “Colocação” (ou *ranking*) em um processo de Recuperação de Informação. Verificando que um *link* é uma autoridade, ele é favorecido no *ranking*. Outra situação para uso do *Web Structure Mining* é que, descobrindo a estrutura que os *links* formam, pode-se estudar o fluxo de informações de um determinado *site*, fornecendo dicas de como melhorá-lo.

***Web Content Mining***: é a obtenção de conhecimento útil do conteúdo das páginas na Internet. Já vimos no item 2.3 que os dados da *Web* apresentam-se em formatos diversos, tais como textos, imagens, sons, vídeos, *hyperlinks*, etc. No entanto, o que recebe mais atenção é a mineração dentro de textos e hipertextos.

Do ponto de vista da “Recuperação da Informação” o objetivo é auxiliar o usuário na busca ou filtragem de informações. É o que fazem os principais mecanismos de buscas da internet ao tentar atender da melhor forma possível as solicitações feitas por usuários através de palavras-chaves. Já do ponto de vista dos bancos de dados, o objetivo é modelar os dados da *Web* e integrá-los de modo que consultas mais sofisticadas possam ser feitas. Um exemplo disso são os *sites* de consulta de preços de produtos, que buscam em diversos *sites* diferentes o preço do produto consultado.

***Web Usage Mining***: é a busca de conhecimento útil através da descoberta de padrões de acesso dos usuários obtidos de *logs* de acesso, onde ficam registrados todos os cliques feitos pelos usuários. Conforme citado por Chiara (2003), os servidores *Web* guardam as informações sobre a interação dos usuários sempre que é feito um acesso ao *site*. Estas informações podem ser usadas para definir padrões de comportamento, preferências e interesses dos usuários, possibilitando que as páginas sejam “personalizadas” antes de serem abertas pelo navegador.

### 3. CAPÍTULO 3 – DATA MINING EM EMPRESAS DE e-COMMERCE

Após um melhor entendimento de alguns conceitos fundamentais sobre *Data Mining* e *Web Mining*, como os dados estão estruturados, como é o armazenamento dessas informações nos *Data Warehouses*, como funcionam alguns algoritmos usados para o processamento dos dados, e também conceitos de Marketing Direto, *e-commerce* e Fidelização de Clientes, serão analisados casos práticos de como algumas empresas de *e-commerce* estão usando o *Data Mining* como uma ferramenta de grande ajuda em suas estratégias de marketing, e na captação e fidelização dos clientes.

#### 3.1 ESTUDO DE CASO: B2W Companhia Global do varejo

A B2W nasceu em 2006 com a fusão das empresas “Submarino” e “Americanas.com”, duas empresas que iniciaram suas atividades no mundo do *e-commerce* em 1999. Esta união possibilitou a criação de uma companhia de varejo com atuação por diversos tipos de canais de distribuição: televendas, televisão, catálogos, quiosques e Internet. A B2W possui em seu portfólio as marcas: Americanas.com, Submarino, Shoptime, Ingresso.com, Submarino Finance, B2W Viagens e Blockbuster.

Conforme informações obtidas no *site* da B2W (2011), este modelo de negócios permite vantagens de custo em relação aos varejistas tradicionais, já que atendem a uma base de clientes dispersa, operando em um único local, e usando tecnologia de forma intensiva. Os ganhos em escala permitem que sejam realizados maiores investimentos em infraestrutura de logística e em outras tecnologias que são usadas em empresas de comércio eletrônico.

A B2W oferece uma extensa seleção de produtos, composta de mais de 700.000 itens, em mais de 35 categorias, incluindo CDs, DVDs, aparelhos eletrônicos, brinquedos, itens de papelaria, câmeras e telefones celulares, perfumaria, etc. A empresa trabalha para garantir a satisfação do cliente, focada em

criar um processo de compras simples, entrega rápida e atendimento ágil. Prova disso é a Missão da empresa: “Realizar os sonhos de consumo das pessoas de forma conveniente”.

Conforme entrevista publicada na Revista INFO (2003) com o CIO do Submarino, Flávio Jansen, a empresa divulgou que conseguiu equilibrar suas receitas e despesas operacionais no ano anterior em grande parte devido ao investimento em Informática, com a implementação de um sistema de *Data Mining*. Segundo Jansen: “Trabalhamos muito para identificar o perfil de cada cliente, ver a frequência de compras, o tipo de produto que ele compra. Passamos a fazer ações de marketing direto tentando oferecer um produto mais próximo das preferências de cada um. Se uma pessoa gosta muito de DVD, oferecemos lançamentos nessa área em vez de um produto que não tem nada a ver com ela. Isso melhorou muito o resultado do marketing direto”. As ofertas da empresa são sempre enviadas por e-mail, e 20% das vendas da empresa são geradas pela mala-direta.

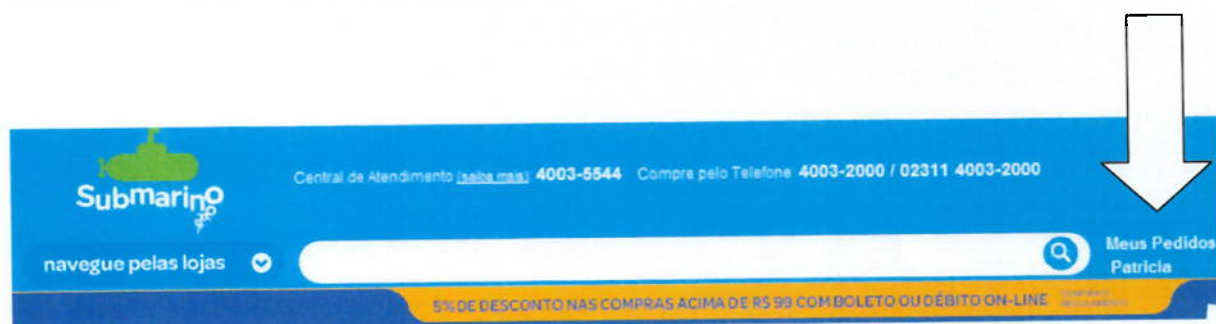
Ainda segundo Jansen, a ferramenta de *Data Mining* que foi adquirida pela empresa significou apenas 30% do trabalho. Os outros 70% consistem na correta modelagem do sistema, em identificar a melhor maneira de usá-la no banco de dados da empresa e analisar as informações obtidas. Jansen enfatiza que é um trabalho bastante complicado descrever um grupo de consumidores, porque não há uma característica dominante, existem muitas variáveis, como o produto, frequência de compra, valor médio da transação, se gosta ou não de barganhas. A empresa não utilizava os resultados logo no primeiro agrupamento, foram feitos aproximadamente 60 estudos, e com isso o refinamento do modelo. Houve muitos testes na base de dados sem enviar a mala-direta.

Outro trabalho importante mencionado é a limpeza da base de dados, que é a retirada de informações que podem distorcer os resultados, como por exemplo quando é feita alguma promoção com brinde, ou quando a empresa vende algum produto abaixo do preço de mercado. O produto vai aparecer nas preferências de muita gente apenas por causa da promoção.

A personalização do *site* do Submarino também é uma preocupação da empresa. De acordo com Jansen, o *site* terá ofertas baseadas no perfil da pessoa que está acessando. A equipe de televendas também tem acesso a essas

informações quando atendem um cliente, para poderem direcionar melhor as ofertas. Segundo ele: “Estudar o cliente é uma ferramenta muito poderosa. Conseguimos dar um atendimento melhor, atender as expectativas dele.”.

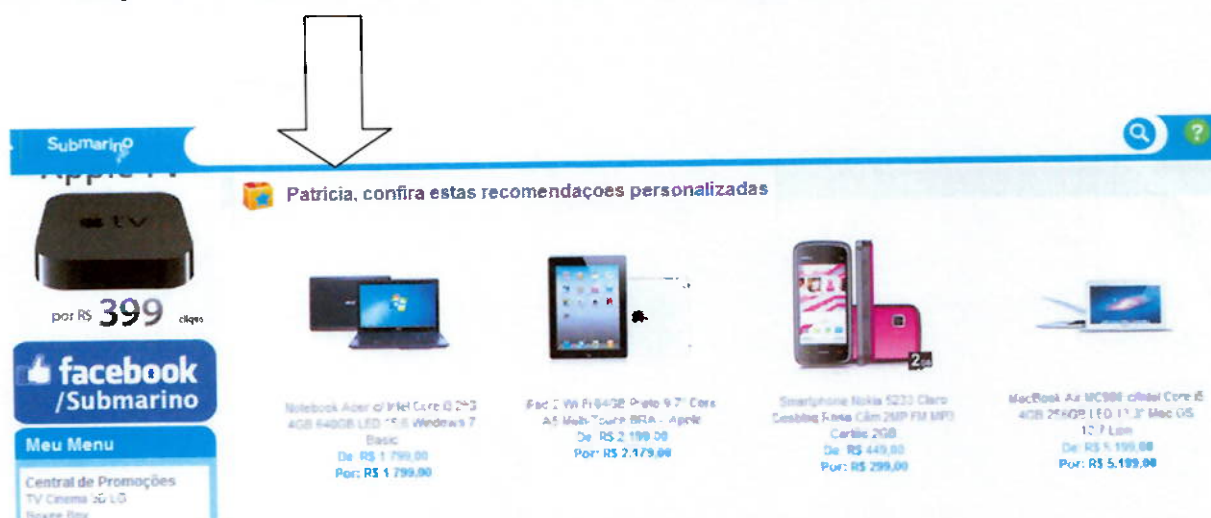
Na Figura 6 observa-se que o *site* Submarino (2012) utiliza *cookies* para identificar os usuários que acessam os seus *sites*. *Cookie* é um pequeno arquivo de texto que o *site* grava na máquina em que o usuário está utilizando. Sua função principal é manter a persistência de sessões HTTP, e são utilizados para identificar usuários e armazenar preferências.



**Figura 6 – Utilização de cookies para identificar o usuário**

Fonte: Site <http://www.submarino.com.br/>, 2012.

Além de identificar o usuário, o *site* também utiliza os dados armazenados de navegações anteriores para fazer recomendações de produtos, baseando-se em produtos similares que foram pesquisados pelos usuários, e nas regras de associação já determinadas, como pode-se ver na Figura 7.



**Figura 7 – Recomendações de produtos**

Fonte: Site <http://www.submarino.com.br/>, 2012.

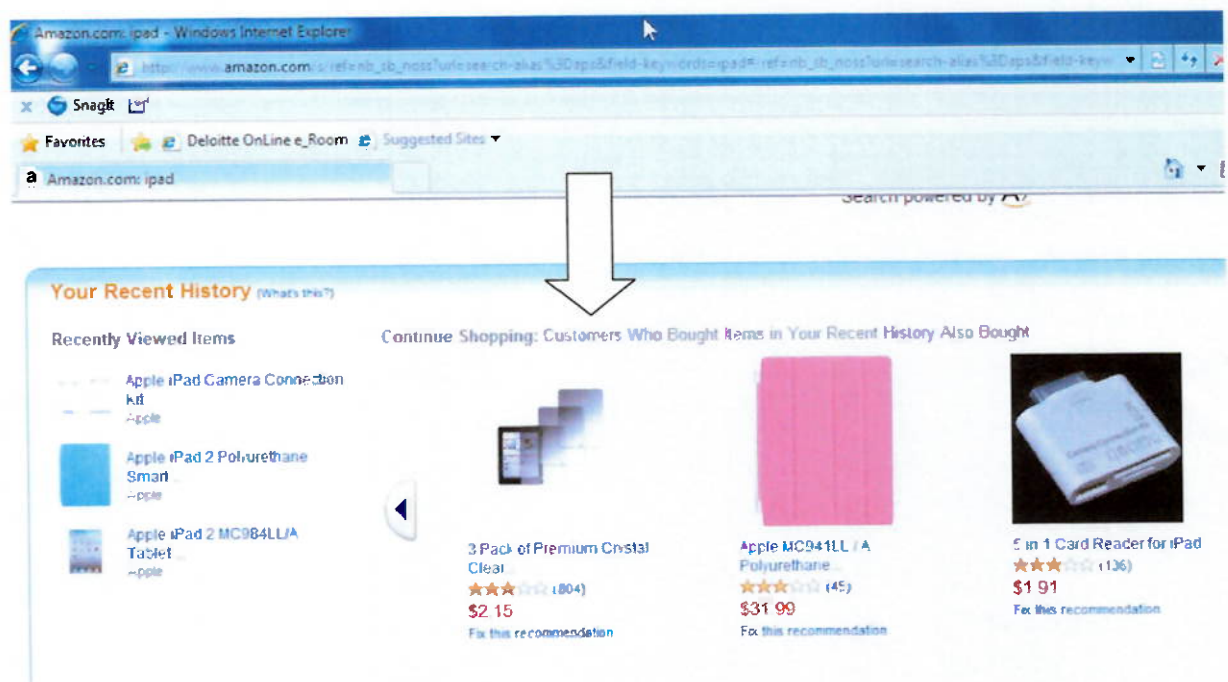
### 3.2 ESTUDO DE CASO: Amazon.com

A empresa americana Amazon.com foi criada em 1994, e em pouco tempo surpreendeu o mundo com a velocidade com que suas vendas cresciam, baseadas somente nos meios virtuais. A Amazon foi a pioneira em vender exclusivamente pela Internet, e em 1998 ela já valia mais do que a sua maior concorrente Barnes & Noble e todas as outras livrarias independentes juntas. A Amazon processava pedidos para todo o 1 milhão de títulos que oferecia, mas tinha somente 2.000 deles no seu depósito. Ela obtinha a maior parte dos livros diretamente das editoras ou de um atacadista, que enviavam seus livros para o depósito da Amazon, onde eram reempacotados e enviados aos clientes (Ghemawat, 2000 apud Vasconcellos et al, 2001).

A idéia inicial de seu fundador, Jeff Bezos, era a de utilizar a Internet para montar uma empresa praticamente sem infraestrutura física. Conforme citado por Spector (2000) apud Vasconcellos et al, (2001), o próprio Jeff Bezos dizia "jamais tornaremos a Amazon.com divertida e atraente como são as grandes livrarias de concreto. Jamais será possível ouvir o rangido das encadernações e sentir o cheiro dos livros, nunca teremos café com leite gostoso e sofás macios na Amazon.com. Mas poderemos fazer coisas realmente diferentes que surpreenderão as pessoas e tornarão sua visita encantadora e divertida."

Quando iniciou a Amazon, a visão de longo prazo de Bezos já incluía uma sucessão de passos para a expansão. Ele começou vendendo apenas livros, mas já antecipava que passaria a oferecer CDs, DVDs e fitas de VHS, que eram populares nos anos 90. Assim que o *site* entrou no ar, a livraria já permitia que os leitores publicassem comentários – positivos ou negativos – sobre os livros. Hoje em dia isso é trivial, mas na época muita gente achou que Bezos era maluco por permitir comentários negativos, pois não era exatamente algo que ajudaria a vender livros. Com o tempo, isso fez com que os consumidores achassem que podiam confiar na Amazon. A livraria tornou-se o primeiro lugar onde muitos deles buscam novas publicações (Revista INFO, 2011).

Na Figura 8 está sendo apresentado um claro exemplo da aplicação de *Data Mining* no site da Amazon: é a sessão “Consumidores que compraram este item também compraram...”, que mostra sugestão de produtos comprados por consumidores que compraram o item pesquisado. Para que esta sugestão fosse feita, provavelmente foi aplicado um algoritmo de Associação onde foi encontrado um padrão de comportamento no qual muitas pessoas compraram os itens sugeridos, juntamente com o item pesquisado.

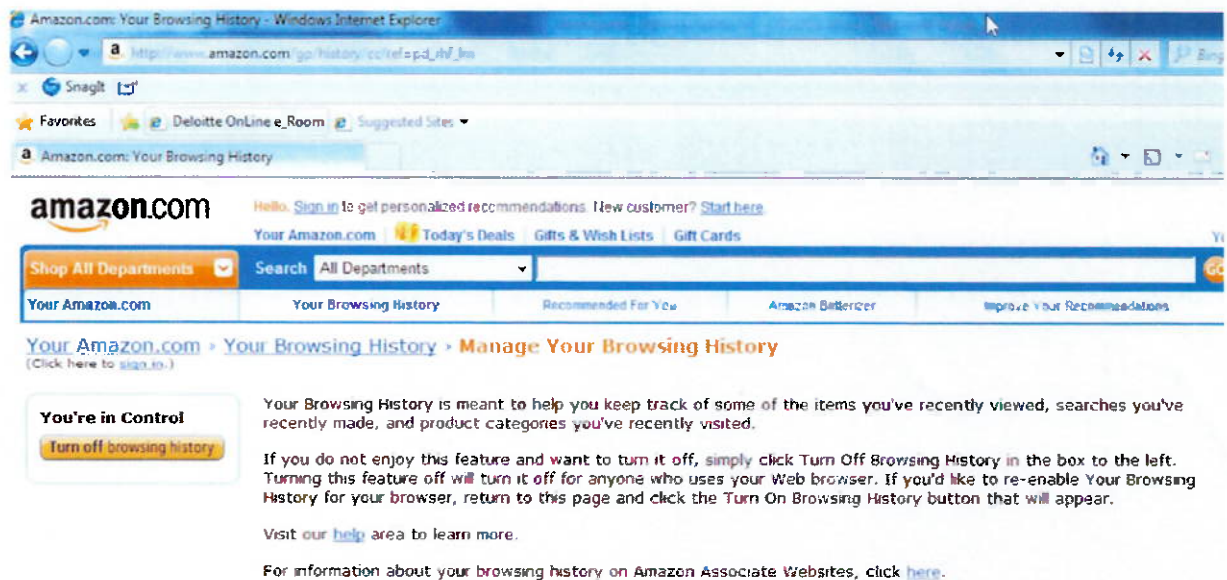


**Figura 8 – Itens sugeridos pelo site**

Fonte: Site <http://www.amazon.com/>, 2012.

Outro *link* interessante observado na Figura 9 é o item “*Your Recent History*”, ou “Seu Histórico Recente”. Ao clicar no *link* que contém a explicação para este item, o site abre uma página informando que esta sessão traz um histórico dos itens que foram visitados recentemente, pesquisas que foram feitas e categorias de produtos que foram visitadas. Além disso, o site fornece a possibilidade de “desligar” esta funcionalidade, e ainda informa que esta ação passará a valer para todos os usuários que utilizam o mesmo navegador que está sendo utilizado. Para armazenar o histórico de cada consumidor, a Amazon se utiliza dos *logs* de navegação que são gerados durante o tempo que seus usuários navegam pelo site, e armazena essas

informações para serem usadas nos próximos “contatos” de cada usuário com o *site* da Amazon.



**Figura 9 – Sessão “Seu Histórico Recente”**

Fonte: Site <http://www.amazon.com/>, 2012.

Conforme citado em Kurtzman (2004), a Amazon possuía em 2004 cerca de 41 milhões de clientes ativos que poderiam ser alvos de campanhas via e-mail. Com um sistema que fazia “a criação” das campanhas por e-mail automaticamente, eram enviadas em torno de 1000 diferentes campanhas todos os dias, sem que nenhum funcionário estivesse acompanhando. O próprio sistema evitava enviar campanhas que tiveram baixa taxa de cliques, ou altas taxas de cancelamento. O sistema também evitava enviar muitas campanhas para os mesmos clientes, mesmo se estes estivessem como “clientes alvo” de muitas campanhas. Este sistema, porém, trouxe para a Amazon um problema: algumas promoções enviadas pelas campanhas automáticas promoveram uma venda de produtos maior do que a capacidade de seus estoques, o que exigia que houvesse um alinhamento entre o sistema de envios de campanhas, e o reabastecimento de produtos nos estoques.

Outro ponto de atenção citado por Kurtzman (2004) é em relação à sugestão de produtos relacionados quando o usuário faz uma pesquisa no *site*. Em algumas vezes dois produtos possuem o mesmo nome, ou similar, mas não tem

absolutamente nada a ver um com o outro. Neste caso a empresa precisa ter um mecanismo de filtro que tenha a sensibilidade de não apresentar resultados inapropriados.

No *site* da Amazon existe uma sessão chamada "Goldbox", que tem o objetivo apresentar aos clientes produtos que eles não estão cientes de que a Amazon vende. A empresa reduz a margem de lucro desses produtos que provavelmente um cliente não vai comprar, a fim de levá-lo a comprar itens de outras novas categorias que ele não está habituado. Segundo Katzman (2004), o propósito desta ação é encorajar novos comportamentos por parte dos clientes. E sabe-se que se houverem novos comportamentos de consumo, haverá novas vendas.

### 3.3 ESTUDO DE CASO: WalMart

De acordo com dados do *site* do Walmart (2012), a rede foi fundada nos Estados Unidos, em 1962, com a abertura da primeira loja de descontos Walmart em Rogers, no Arkansas. Em 1972 a empresa abriu seu capital na bolsa de valores, o que acelerou sua expansão, contando com 276 lojas no final dos anos 70. Em 1983 a empresa abriu sua primeira loja com a marca "Sam's Club", e em 1988 o primeiro "Supercenter". Em 1989 já existiam 1.402 lojas Walmart e 123 Sam's Club. As vendas subiram de \$1 bilhão em 1980, para \$26 bilhões.

Em janeiro de 2012, a rede possui 10.020 lojas em 28 países, emprega 2,2 milhões de pessoas e vende para 176 milhões de clientes ao ano. Conforme citado no *site* do Walmart: "Este é o exemplo perfeito de como crescer sem perder de vista seus valores. Nosso maior valor sempre foi, e sempre será, o serviço ao cliente".

Segundo Hays (2004), no ano de 2004 o Walmart possuía 460 terabytes de dados armazenados em seus computadores e, segundo especialistas, naquela época a Internet possuía menos da metade dessa quantidade de dados. Com 3.600 lojas somente nos Estados Unidos, e aproximadamente 100 milhões de clientes por semana, o Walmart possuía acesso a todos os tipos de informações, que eram coletados em suas lojas, e depois processados e armazenados. Todas as

informações são preciosas para o Walmart, e são a base para as reuniões de vendas que ocorrem semanalmente na empresa. Em alguns casos o Walmart compartilha estas informações com fornecedores através de uma rede extranet, para que eles possam, por exemplo, acompanhar como estão as vendas de seus produtos.

A empresa também se empenha em manter seus dados em segredo. Alguns de seus sistemas são desenvolvidos pelos próprios empregados do Walmart, com o objetivo de manter os concorrentes afastados. Empresas fornecedoras de equipamentos e softwares precisam assinar acordos de confidencialidade.

O Walmart usa sua imensa base de dados para conseguir obter eficiência em todos os níveis da sua operação, seja dentro das lojas, seja internamente para saber a pontualidade de seus fornecedores. O objetivo da empresa é evitar os dois piores pesadelos de uma empresa de varejo: ter muito estoque, ou nenhum.

Conforme exemplificado por Hays (2004), com as informações das vendas das últimas semanas e meses, o Walmart negocia com cada um dos fornecedores as metas de vendas para o ano seguinte. Caso estes fornecedores não atinjam as metas, ou tenham algum outro tipo de problema na entrega ou reabastecimento, podem esperar uma negociação ainda mais difícil para o ano seguinte.

Os dados do Walmart não são utilizados apenas nas operações de vendas e negociações com fornecedores, mas também para prever e direcionar futuros negócios da empresa, como a abertura de novas lojas e qual o número de novos compradores esperados para elas.

Historicamente numa loja convencional, o foco do Walmart era nos produtos que eram vendidos, e não para quem eram vendidos. Porém a empresa descobriu nas vendas *on line* um importante meio de aprender mais sobre os clientes, pois era necessário que eles fornecessem informações pessoais para completar a transação, além de registrar quais produtos eram comprados. Conforme citados por executivos da empresa, o mais importante não é saber o que uma pessoa compra individualmente, mas sim os produtos que são comprados em conjunto com outros. Por exemplo, sabendo que produtos são normalmente comprados juntos, o Walmart pode também estabelecer regras para os preços deste conjunto de produtos, a fim

de que todo o "conjunto" tenha um valor abaixo do mercado, e não necessariamente que todos os preços isoladamente sejam baixos.

### 3.4 OUTROS EXEMPLOS DE APLICAÇÃO

A seguir serão apresentados alguns outros exemplos de empresas que não são necessariamente empresas de *e-commerce*, mas que conseguiram obter ótimos resultados com a aplicação de *Data Mining* em suas bases de dados.

Conforme levantado em Lamboia e Pereira (2005):

Armazenando e analisando a movimentação financeira de seus clientes, o Banco Itaú conseguiu aumentar sua taxa de retorno nas malas diretas em 30%, e reduzir a conta com serviços de correio a um quinto do valor.

A Sprint, um dos líderes do mercado americano de telefones de longa distância, desenvolveu com a análise de *Data Mining* em seu *Data Warehouse*, um método capaz de prever com 61% de segurança se um consumidor trocaria de companhia telefônica dentro de um período de dois meses. Com um marketing agressivo conseguiu evitar a deserção de 120.000 clientes e uma perda de 35 milhões de dólares em faturamento.

A empresa SERPRO no Brasil, utilizando o seu *Data Mining* e *Data Warehouse*, conseguem hoje cruzar e analisar informações em cinco minutos, o que anteriormente demandava quinze dias de trabalho.

Podemos também citar algumas outros casos de aplicações comerciais do *Data Mining*, conforme levantado por Bertoli (2012):

O *Bank of America* pode agora elaborar uma detalhada visão demográfica de hábitos bancários e propriedades financeiras de grupos selecionados em seus usuários. Consultas no *Data Warehouse* de 800 Gb levam 30 segundos cada, comparadas com a espera de 3 meses para o acesso a dados de 100 fitas magnéticas dos tradicionais mainframes e a subsequente distribuição aos

requisitantes, são realmente rápidas. O sistema utiliza os dados do banco inteiro e de suas 30 unidades comerciais.

O *Canadian Imperial Bank of Commerce* (CIBC), sediado em Toronto, Canadá, utiliza o sistema SAS como um *Data Warehouse* que fornece informações gerenciais para o suporte à decisão. A análise da mineração inclui modelagem estatística e consultas relativas aos padrões de transações do consumidor em diferentes ramos, para ajudar no planejamento da empresa.

A *MasterCard International* processa diariamente cerca de 12 milhões de transações e utiliza mineração de dados para extrair todos os tipos de estatísticas sobre os portadores de cartões. Este processamento está incluído nas vendas do *Data Warehouse* de transações para os mais de 20.000 parceiros comerciais. Isto permite a visualização de diferentes classes de portadores de cartões e a análise de como estes utilizam seus cartões, com o objetivo de desenvolver promoções especializadas e detecções de fraude.

Utilizando as técnicas da mineração de dados, a PUC-RJ, depois de examinar milhares de alunos, identificou o seguinte padrão de comportamento: se o candidato é do sexo feminino, trabalha e teve aprovação com boas notas no vestibular, então não efetivava a matrícula. De acordo com os costumes do Rio de Janeiro, uma mulher em idade de vestibular, se trabalha é porque precisa, e neste caso deve ter feito inscrição para ingressar na universidade pública gratuita. Se teve boas notas provavelmente foi aprovada na universidade pública onde efetivará matrícula. Claro que há exceções: pessoas que moram em frente à PUC-RJ, pessoas mais velhas, de alto poder aquisitivo ou que voltaram a estudar por outras razões além de ter uma profissão, etc. Mas a grande maioria obedece à regra anunciada.

Conforme notícia do *site* Call to Call (2006), a empresa "Baú da Felicidade", que pertencia ao Grupo Sílvio Santos, e no ano de 2011 foi vendida para a rede de lojas "Magazine Luiza", que possui 121 lojas nos estados de São Paulo, Minas Gerais e Paraná e uma carteira de 3 milhões de clientes, está concluindo um projeto de *Business Intelligence* com o objetivo de alinhar o planejamento e operação do orçamento da empresa, de acordo com o desempenho de vendas, previsão de receitas e liquidez efetiva dos carnês e mercadorias vendidas nas lojas. O novo sistema permite um conhecimento detalhado do negócio, com estudos gráficos de

tendências e com a associação de dados a documentos internos ou externos que atestam sua origem.

As bases de dados da empresa foram integradas em um *Data Warehouse*, onde ficarão centralizadas todas as informações pertinentes à tomada de decisão dos negócios, foram construídos *Datamarts* capazes de modelar a informação de acordo com as necessidades de cada usuário.

Também segundo o *site* Call to Call (2006), a AES Eletropaulo, maior distribuidora de energia elétrica em consumo e faturamento da América Latina, com uma marca de 6 milhões de clientes e 16 milhões de pessoas atendidas, desenvolveu um sistema de apoio à decisão voltado não só para a operação eficiente da empresa, mas principalmente, para a abertura de novas frentes de negócios a partir de sua gigantesca base de dados. O projeto é baseado no uso de ferramentas estatísticas, de GIS (Sistema de Informação Geográfica) e de *Data Mining*. As aplicações são utilizadas para fins que vão desde a detecção de ligações clandestinas até a análise de tendências de inadimplências.

Segundo um executivo da empresa, as concessionárias do setor elétrico têm feito utilização muito restrita de seus dados, limitando-se a utilizá-los para efeito de operação e administração do próprio negócio. "Com o apoio do *Data Mining* e do GIS podemos transformar esses dados em inteligência de negócio, e até oferecê-los para outras corporações dedicadas ao varejo de massa, o que irá transformar a nossa mina de dados em uma verdadeira mina de receitas", disse ele.

De acordo com empresa especializada, a Eletropaulo vem demonstrando ser um dos exemplos mais bem sucedidos no país de uso de *Data Mining* em criação de valor dentro de empresas de *utilities*. A utilização desta ferramenta já é bastante utilizada nas empresas de telefonia, onde a alta taxa de infidelidade dos clientes acabou precipitando um verdadeiro "boom" de *Data Mining*, visando o conhecimento máximo do assinante para a sua fidelização e para estratégias consistentes de marketing.

Outro exemplo interessante relatado pelo *site* Call to Call (2006) é o do portal Globo.com, um dos maiores portais de acesso e conteúdo do país, que colocou em prática uma política de relacionamentos com clientes no conceito de CRM Analítico,

baseada em tecnologia estatística. O CRM Analítico é uma tecnologia que usa não apenas os mecanismos clássicos de atenção e interação com o consumidor (como atendimento personalizado e marketing “one-to-one”, que é o marketing individualizado) mas também se utiliza de ferramentas estatísticas (ou de *Data Mining*) para conhecer o perfil do consumidor e prever seu comportamento futuro.

Segundo um executivo da empresa, o CRM Analítico vai se tornando uma ferramenta indispensável à medida que se multiplica o universo de assinantes de acesso à Internet da Globo.com, assim como o volume de acessos ao conteúdo do portal. Com o apoio da ferramenta de *Data Mining* a empresa consegue extrair informações sobre seus usuários cadastrados e apoiar o departamento de Marketing na definição de campanhas para aquisição. No que se refere à fidelidade do assinante, a estatística é empregada para acompanhar o relacionamento, permitindo pontuar e premiar assinantes e orientar ofertas de novos produtos e *upgrades*.

Além disso, o CRM Analítico amplia a eficiência da exploração dos ativos de dados da Globo.com e, de quebra, promove uma expressiva redução de custos de desenvolvimento, armazenamento e manutenção de bases de dados. Isto ocorre porque as estatísticas abrangem também os níveis de demanda da infraestrutura, as situações típicas de falha, descontentamento ou de congestionamento da rede. Assim, a Globo.com consegue otimizar o gerenciamento do tráfego, economizando em suporte e programando melhor os investimentos em expansão ou ajuste na rede.

Ainda segundo o executivo: “Quanto aos resultados em marketing de relacionamento e marketing direto, eles dependem, em grande medida, da adequada formatação do perfil dos grupos de usuários e assinantes. E, para tanto, são mobilizados parâmetros que relacionam dados geográficos, demográficos e psicográficos com hábitos de navegação e histórico de uso dos serviços”.

E finalizando, é apresentado o exemplo da empresa Sky, também conforme o *site* de notícias Call to Call (2006). A empresa de TV por assinatura está se utilizando da tecnologia de *Business Intelligence*, apoiada pela plataforma de *Data Mining*, para ancorar as suas ações de marketing direto e de relacionamento, orientadas para atividades como fidelização de clientes, ampliação da base instalada e planejamento de campanhas.

A empresa, que iniciou há cerca de 2 anos a implantação de um *Data Warehouse* com informações sobre todas as interações comerciais e de marketing com os clientes, está conseguindo ampliar em até 100% o retorno de suas campanhas, a partir da manipulação desses dados de acordo com modelos analíticos. Segundo uma gerente da empresa, com o suporte do *Data Mining* a empresa já está conseguindo fazer análises preditivas de fenômenos críticos, como o risco de inadimplência, de acordo com o perfil dos grupos de usuários, ou as taxas prováveis de "churn" (que significa a perda de clientes sofrida por uma empresa) relacionadas a eventos futuros. "Um exemplo típico de aplicação é verificar o risco de perda de clientes, em caso de cancelamento de uma programação, mas pode-se até mesmo estudar eventos externos à empresa, como o impacto da variação do Índice de Intenção de Consumo (IIC) sobre a inadimplência futura, de acordo com a área geo-econômica", comenta a executiva.

Antes de partir para a constituição dos modelos analíticos, a equipe de tecnologia da informação da SKY fez uma imersão completa em todas as instâncias administrativas do negócio-fim da empresa e só então passou a desenvolver óticas de análise adequadas às necessidades do negócio. "Por meio do cruzamento de dados que vão do CEP do assinante ao tipo de aparelho que ele usa, podemos avaliar, com alto nível de certeza, a sua receptividade a campanhas como a compra de pacotes de pay-per-view ou de novos pontos de acesso no interior das residências", comemora a gerente.

De posse desses estudos analíticos, as diversas diretorias da SKY podem planejar não só suas ações de marketing de relacionamento, mas até desenvolver estratégias financeiras adequadas aos cenários previstos para o mercado. Ainda segundo a executiva, os estudos de *Data Mining* podem interferir até mesmo em avaliações sobre a grade de programação, oferecendo subsídios sobre os conteúdos mais sintonizados com os diferentes perfis de assinantes; ou ainda, orientar estratégias de atendimento para o pessoal de suporte técnico, além de otimizar a política de telemarketing. "Em um mercado altamente competitivo - como é o caso do nosso - o *Data Mining* pode ser uma ferramenta até para avaliar os pontos de vulnerabilidade da concorrência, a partir da análise de novos clientes egressos dessas empresas. Com isso, as ferramentas de conhecimento deixam, de fato, de servir apenas como material para análise histórica do relacionamento com

os clientes, passando a ocupar uma posição central nas decisões de médio e longo prazos da SKY", conclui a gerente.

## 4. CAPÍTULO 4 – CONSIDERAÇÕES FINAIS

### 4.1 CONCLUSÕES

Diariamente somos milhões de pessoas acessando a Internet. Isso representa um enorme volume de dados deixados na *Web*, seja em formato de textos, imagens, arquivos de áudio, vídeo, etc. A produção de informações resultantes de acessos à Internet, compras eletrônicas, acessos à redes sociais, e-mails, transações comerciais e profissionais, etc, está crescendo muito rapidamente. Isto é uma verdadeira mina de ouro para as empresas que conseguirem armazenar, organizar e interpretar tais informações.

Conforme publicado por Álvares (2012) em matéria na Revista Exame, segundo o Sebrae/SP, é dez vezes mais barato manter um cliente do que conseguir novos. Conforme citado: "a fidelização é uma forma de identificar oportunidades de faturamento não evidentes em um primeiro momento. Uma empresa que conhece um cliente a ponto de conhecer sua família, por exemplo, consegue ter componentes que envolvem escolhas mais complexas e, conseqüentemente, pode manter uma relação mais rentável".

Com os estudos de casos apresentados pudemos comprovar que o *Data Mining* é uma poderosa ferramenta para auxiliar as empresas a conhecerem melhor seus clientes, e em todos os casos gerou aumento nas vendas devido ao melhor direcionamento das estratégias de marketing, ou de uma melhor reorganização dos produtos e na apresentação de seus *sites*. No entanto, a utilização do *Data Mining* não é uma tarefa fácil: o investimento financeiro é bastante alto, pois inclui a aquisição de computadores robustos e com grande capacidade de processamento e armazenamento, as próprias licenças de utilização do software de *Data Mining* escolhido, e o principal, que é a contratação de pessoas altamente capacitadas e conhecedoras deste tipo de ferramenta. Muitas vezes a implantação de um sistema de *Data Mining* pode levar meses e até anos para ser concluída, e o retorno do investimento feito não é percebido a curto prazo. E o pior também pode acontecer, se os profissionais responsáveis não forem realmente conhecedores do assunto, ou

se por algum outro motivo o resultado não for satisfatório, a empresa pode ver todo o seu investimento ser desperdiçado.

Outro fator importante percebido com este trabalho é que os benefícios do *Data Mining* vão muito além da fidelização dos clientes propriamente dita, ele está ajudando as empresas a serem mais eficientes operacionalmente, otimizando o gerenciamento de problemas de infraestrutura e suporte, e auxiliando a fazerem um melhor planejamento dos recursos que serão investidos em expansão da rede, como vimos no caso do portal Globo.com, ou então como ferramenta para gerir o relacionamento com seus fornecedores, como faz o Walmart, evitando inclusive problemas de entregas e reabastecimento.

Apesar dos muitos exemplos apresentados, sabemos que o uso do *Data Mining* ainda está engatinhando no mercado brasileiro. Conforme dados levantados pelo Sebrae/SP, e citados por Álvares (2012), apenas 47% das micro e pequenas empresas cadastram seus clientes, e somente 16% delas atualizam estes dados. Isto quer dizer que ainda há muito trabalho a ser feito. É uma tarefa complexa obter, organizar, processar, entender e conseguir converter tantas informações em reais benefícios para o negócio. Além dos fatores anteriormente já citados, como altos investimentos em pessoas, equipamentos de *hardware*, *software*, e tempo, ainda temos desafios que se tornarão cada vez maiores com o passar dos anos, que são as questões relacionadas à viabilidade deste crescimento do volume de informações, tais como a largura de banda de transmissão, meios de armazenamento e velocidade de processamento.

No contraponto de tudo isso temos uma questão muito mais polêmica e merecedora de atenção: a invasão da privacidade dos clientes. Até onde uma empresa pode ir ao desvendar os hábitos de seus clientes? Até que ponto as pessoas querem ter seu comportamento conhecido? Qual o limite entre “surpreender um cliente” e “invadir sua privacidade”? Possivelmente nenhuma dessas questões tenha resposta neste momento, e ainda serão o foco de muitos debates e discussões. O próprio Sistema Legal e Jurídico ainda não está preparado para lidar com esse tipo de questão, simplesmente porque são problemas que não existiam há 15 ou 20 anos atrás.

A única certeza que podemos ter daqui pra frente é que as empresas que conseguirem dominar e gerir a informação, e o conhecimento decorrente dela, possuirão um grande diferencial competitivo em relação às demais. E isso pode significar um futuro corporativo longo e lucrativo, ou apenas o esquecimento.

## 4.2 TRABALHOS FUTUROS

Após a apresentação deste cenário fascinante sobre o universo do *Data Mining* e o enorme volume de dados que a Internet produz todos os dias, fica a vontade de se aprofundar muito mais em todas essas questões, e também de acompanhar o desenrolar de tantos desafios que as empresas tem pela frente, como a evolução da *Web*, dos computadores, processadores e softwares, as questões relacionadas à privacidade dos clientes, o avanço da legislação dos países para acompanhar estas mudanças, o crescimento das Redes Sociais, que despeja na *Web* uma quantidade de dados muito significativa, tema que não foi abordado neste trabalho mas que é merecedor de ser analisado pois pode ser uma fonte de dados extremamente valiosa para a aplicação de *Data Mining*.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ÁLVARES, D. **5 dicas para fidelizar seus clientes**. Revista Exame online. 2011. Disponível em: < [http://exame.abril.com.br/pme/noticias/5-dicas-para-fidelizar-seus-clientes?utm\\_source=newsletter&utm\\_medium=e-mail&utm\\_campaign=news-diaria.html](http://exame.abril.com.br/pme/noticias/5-dicas-para-fidelizar-seus-clientes?utm_source=newsletter&utm_medium=e-mail&utm_campaign=news-diaria.html) >. Acesso em: 26/01/2012.
- BORTOLI, J. **Data Mining (Mineração de Dados)**. Disponível em: < <http://www.guiafar.com.br/portal/index.php/pt/artigos/tecnologia-da-informacao/159-data-mining-mineracao-de-dados?showall=1> >. Acesso em: 03 de Janeiro de 2012.
- CHIARA, R. **Aplicação de Técnicas de Data Mining em Logs de Servidores Web**. 2003. 176 p. Dissertação (Mestrado) - Universidade de São Paulo, São Carlos, 2003.
- FACCHINI, C. **Comércio eletrônico deve crescer 30% e faturar R\$20bi em 2011 – Seção de Economia/Empresas IG**. Disponível em: < <http://economia.ig.com.br/empresas/comercioservicos/comercio+eletronico+deve+crescer+30+e+faturar+r+20+bi+em+2011/n1238184906202.html> >. Acesso em: 13 de novembro de 2011.
- HAYS, C.L. **What Wal-Mart Knows About Customer's Habits**. 2004. The New York Times Journal. Disponível em < [http://www.nytimes.com/2004/11/14/business/yourmoney/14wal.html?\\_r=1](http://www.nytimes.com/2004/11/14/business/yourmoney/14wal.html?_r=1) >. Acesso em 15 de janeiro de 2012.
- KOTLER, P. **Administração de Marketing: a edição do novo milênio**. São Paulo: Prentice Hall, 2000. 765p.
- KURTZMAN, C. **Attack of the Amazons: Data Mining at Amazon.com**. 2004. Disponível em: < <http://www.adastra.com/apogee/amazon.html> >. Acesso em: 03 de Janeiro de 2012.
- LAMBOIA, F.; PEREIRA, L.M. **Data Mining**. 2005. 30 p. Trabalho de curso (Bacharelado em Informática) – Universidade Estadual do Oeste do Paraná, Cascavel, 2005.
- LIU, B. **Web Data Mining: exploring hyperlinks, contents and usage data**. Berlin: Springer, 2007. 532p.
- MAGALHÃES, A.S. **E-commerce e e-banking no Brasil: uma perspectiva do usuário**. 2007. 273p. Dissertação (Mestrado) – Universidade de São Paulo, 2007.
- POTTER, R.; TURBAN, E.; RAINER, K. **Administração de Tecnologia da Informação**. 3. Ed. São Paulo:Campus, 2005. 618p.

PÓVOA, M. **A grande oportunidade na web para esta década**. Disponível em: <<http://www.tiinside.com.br/13/01/2012/a-grande-oportunidade-na-web-para-esta-decada/os/257801/news.aspx>>. Acesso em: 16 de janeiro de 2012.

**RELATÓRIO WEBSHOPPERS - 24ª Edição**. São Paulo, 2011. E-bit – Inteligência de Comércio Eletrônico. Disponível em <<http://www.webshoppers.com.br/>>. Acesso em: 15 de janeiro de 2012.

REVISTA INFO. São Paulo, 2003. Editora Abril. Mensal. Disponível em: <[http://info.abril.com.br/edicoes/203/arquivos/1608\\_1.shl](http://info.abril.com.br/edicoes/203/arquivos/1608_1.shl)>. Acesso em: 05 de dezembro de 2011.

REVISTA INFO. São Paulo, 2011. Editora Abril. Mensal. Disponível em: <<http://info.abril.com.br/noticias/mercado/como-jeff-bezos-fez-fortuna-com-a-amazon-10112011-7.shl?1>>. Acesso em: 20 de janeiro de 2012.

SFERRA, H.H; CORRÊA, A.M.C.J. **Conceitos e aplicações de Data Mining**. Revista de Ciência e Tecnologia. V11, N° 22. P 19-34. 2003.

Site B2W/Institucional. Disponível em: <<http://www.b2winc.com/>>. Acesso em: 05 de dezembro de 2011.

Site call to call. **Baú adota o Business Intelligence**. 2006. Disponível em <[http://www.calltocall.com.br/site/web/noticias.asp?id\\_noticia=197](http://www.calltocall.com.br/site/web/noticias.asp?id_noticia=197)>. Acesso em: 27/11/2011.

\_\_\_\_\_. **Eletropaulo foca estratégias com uso de ferramentas estatísticas, de GIS e de data mining**. 2006. Disponível em <[http://www.calltocall.com.br/site/web/noticias.asp?id\\_noticia=411](http://www.calltocall.com.br/site/web/noticias.asp?id_noticia=411)>. Acesso em: 27/11/2011.

\_\_\_\_\_. **Globo.com adota CRM para melhor atendimento aos clientes**. 2006. Disponível em: <[http://www.calltocall.com.br/site/web/noticias.asp?id\\_noticia=675](http://www.calltocall.com.br/site/web/noticias.asp?id_noticia=675)>. Acesso em: 27/11/2011.

\_\_\_\_\_. **Sky consegue prever comportamento do consumidor**. 2006. Disponível em: <[http://www.calltocall.com.br/site/web/noticias.asp?id\\_noticia=128](http://www.calltocall.com.br/site/web/noticias.asp?id_noticia=128)>. Acesso em: 27/11/2011.

Site eCommerceOrg – Tudo sobre Comércio Eletrônico. Disponível em: <<http://www.e-commerce.org.br/>>. Acesso em: 15 de janeiro de 2012.

Site Submarino. Disponível em: <<http://www.submarino.com.br/>>. Acesso em: 05 de dezembro de 2011.

Site Walmart Institucional. Disponível em <<http://walmartstores.com>>. Acesso em 15 de janeiro de 2012.

THOMÉ, A.C.G. **Redes Neurais – Uma ferramenta para KDD e Data Mining**. 50p. Artigo.

VASCONCELLOS, E; MURITIBA, S.N; OLIVEIRA, P.M. **Modelo de Análise dos Graus de Utilização do Negócio Eletrônico em Empresas Virtuais e Tradicionais: Um Estudo de Caso das Livrarias Amazon.com e Barnes & Nobles.** V SEMEAD, São Paulo, p. 1-13, junho 2001.