

DESCOBERTA DE PADRÕES RELEVANTES NA BASE DE DADOS DA FARMÁCIA HOMEOBEL CENTER

**SILVA, Leandro Santiago¹; LIMA, Rafael Sá²; MOREIRA, Fabiano C³;
MONTEIRO, Odlaniger Lourenço Damaceno⁴.**

¹ Graduando em Sistemas de Informação na área de Ciências Exatas e Tecnologia do CESUPA. leandrosantiago_@hotmail.com

² Graduando em Sistemas de Informação na área de Ciências Exatas e Tecnologia do CESUPA. rafaelfl@globo.com

³ Prof. Mestre em Ciência da Computação. E-mail: fabiano@plugplay.com.br

⁴ Prof. Mestre em Ciência da Computação. E-mail: lourenco@cesupa.br

DESCOBERTA DE PADRÕES RELEVANTES NA BASE DE DADOS DA FARMÁCIA HOMEOBEL CENTER

RESUMO

Este artigo visa apresentar o estudo de caso da Farmácia Homeobel Center, onde foi projetado uma avaliação do banco de dados com o intuito de identificar padrões que possibilitem a tomada de decisão. Esta análise foi realizada aplicando as etapas descritas do KDD (Knowledge Discovery in Databases) incluindo a principal fase, o DM (data mining), que extrai os resultados relevantes, onde nunca antes percebidos, para que assim estes possam ser validados pelos proprietários da farmácia.

Abstract

This article aims to show the study about the case of the Homeobel Pharmacy Center, where a database evaluation was projected with intention to identify standards that make possible the decision taken. This analysis was carried through applying the described stages of KDD (Knowledge Discovery in Databases) including the main phase, the DM (data mining), that extracts the excellent results, which were never perceived before, then these can be evaluated by the owners of the pharmacy.

1. INTRODUÇÃO

O mercado farmacêutico produz uma grande quantidade de informação que geralmente é coletada e armazenada constituindo-se em uma base de dados. As empresas do setor vêm tentando adequá-los nas suas organizações e procurando empregá-los de forma que estas possibilitem a melhoria do seu desempenho perante o mercado, buscando eficiência dos seus serviços. Após ter ciência de tal tipo de tecnologia, a farmácia Homeobel Center, uma das pioneiras no ramo farmacêutico artesanal em Belém-PA, optou por também adotar estas estratégias técnico-informacionais que promovem o melhoramento do desempenho estratégico da organização.

A prática utilizada foi a mineração de dados, que consiste na principal fase da descoberta de conhecimento em base de dados (KDD – Knowledge Discovery in Databases), onde este faz uso de uma extração de dados provida por um software que possui algoritmos que executam uma análise dos dados, onde não produzem padrões finais, mas sim, padrões intermediários. Com isso, o usuário pode encontrar padrões de interesse na visualização, visto que essa forma promove um conhecimento do domínio. Essa abordagem torna-se particularmente importante quando se observa que não há algoritmos genéricos para Mineração de Dados, e que pode fornecer uma forma eficiente de avaliar e validar o andamento do processo (Branco 2003).

Para a mineração dos dados da base, foi utilizado o software WEKA (*Waikato Environment for Knowledge Analysis*), disponível em <http://www.cs.waikato.ac.nz/>, o qual realizou um levantamento das localidades (bairros) dos consumidores em potencial do produto da empresa, a forma de pagamento mais freqüente em relação a cada localidade, e principalmente qual turno (manhã, tarde ou noite) há uma maior incidência

de compras na empresa, gerando assim, um levantamento sócio-econômico dos freqüentadores da empresa e conhecendo as regras de classificação, na qual diz respeito à descoberta de localização estratégica a fim de construir uma filial.

2. PROPOSTA DE LOCALIZAÇÃO ESTRATÉGICA COM ANÁLISE À BASE DE DADOS

Devido o aumento da área farmacêutica nos últimos dez anos, a concorrência tem se mostrado maior, e com o aparecimento de novas farmácias especializadas no ramo artesanal e de manipulação, a Homeobel Center, que é pioneira no mercado e possui a garantia de um bom produto e respeito dos seus clientes, viu-se na obrigação de construir uma nova filial em um ponto estratégico, perto das localidades onde seus clientes mais assíduos residem. De acordo com a localização da residência do cliente, cada ponto de venda realizará promoções a fim de atrair os clientes e assegurar a mutualidade da confiança de sua clientela.

A farmácia Homeobel Center é uma empresa que já está no mercado a mais de vinte anos, porém não possui uma infra-estrutura tecnológica tão qualificada quanto poderia ter. No entanto, procurando melhorar os seus serviços, e como a demanda por seus medicamentos é muito grande, há uma necessidade da criação de uma filial, onde esta foi estudada para garantir o aproveitamento do investimento, construindo as instalações da mesma em uma localidade onde haja uma grande procura pelos seus serviços, assim como a digitalização concisa dos dados já existentes na base de dados, que atualmente mostra-se de forma desorganizada. Foi efetuada uma reestruturação dos dados para se obter a extração das informações que não são utilizadas em prol da organização, devido à falta de conhecimento por parte dos proprietários, pois na mesma, até mesmo a parte de maquinário mostra-se obsoleta.

A partir da criação de uma nova filial, formar-se-ão três lojas que serão localizadas em pontos estratégicos da cidade de Belém, nas quais aumentarão seu nível competitivo dando um passo a frente em relação as demais organizações do ramo, e fazendo com que a mesma não chegue a ter a diminuição da clientela devido o difícil acesso, problema no qual geravam grandes reclamações por parte de seus clientes.

Até o presente momento, na área farmacêutica empresarial, não foi constatado nenhuma participação tecnológica neste ramo da área de manipulação. Este projeto tem o intuito de promover a técnica para as outras organizações e despontar a farmácia em questão como pioneira no ramo com a utilização da técnica visando a melhoria de serviços e promoção dos incentivos para utilização da combinação da área tecnológica com o ramo farmacêutico.

3. PROCESSO DE DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS

Inicialmente, o KDD *“consiste em um processo não trivial de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis, em conjunto de dados”* (Fayyad, 1996), ou seja, é um processo capaz de descobrir e recuperar um conhecimento em banco de dados. Este processo foi proposto em 1989 para referir-se às

etapas que produzem conhecimentos a partir dos dados e, principalmente, à etapa de mineração dos dados, que é a fase que transforma dados em informações. Este processo envolve encontrar e interpretar padrões nos dados, de modo iterativo e interativo, através da repetição dos algoritmos e da análise de seus resultados.

Dado, informação e conhecimento são elementos fundamentais para a comunicação e a tomada de decisão nas organizações, mas seus significados não são tão evidentes (Davenport, 1998).

Os dados são elementos brutos, sem significado, que possuem estrutura fundamental sobre a qual um sistema de informação é construído, ou seja, são fatos, imagens ou sons que podem ou não ser úteis para uma determinada tarefa. Porém, quando a eles são atribuídos valores, transformam-se em informações, estas por sua vez possuem uma interpretação lógica ou natural dada a ele por seu usuário já que possui o conteúdo e a forma apresentados de maneira útil a ser empregado em um processo de tomada de decisão. A informação só se torna conhecimento a partir da sua utilização, isso acontece quando se consegue perceber o contexto e fazer uso daquela informação ou dado disponível. Pode ser considerado como a informação processada pelos indivíduos, já que adquirimos conhecimento por meio do uso da informação nas nossas ações. Logo, não se pode desvincular a ação e o indivíduo, pois ele está estritamente relacionado com a percepção do mesmo, que codifica, decodifica, distorce e usa a informação de acordo com suas características pessoais, ou seja, de acordo com seus modelos mentais (Davenport, 1998).

3.1 ETAPAS DO KDD:

Segundo Fayyad, 1996, o KDD possui sete etapas, que serão ilustradas a seguir:

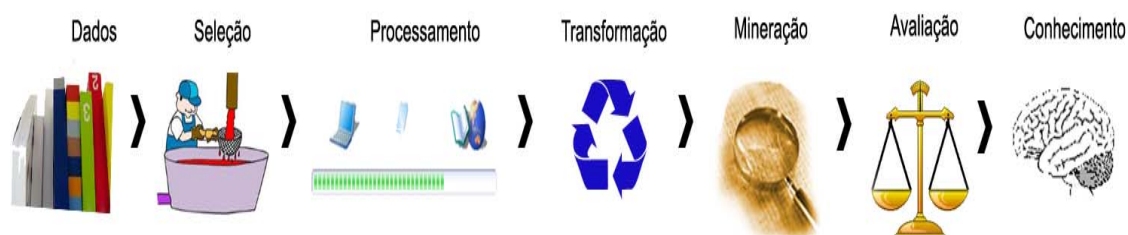


Figura 1 – As etapas do processo de KDD (adaptado de FIGUEIRA, 1998).

O processo de KDD (Figura 1) começa com o entendimento do domínio da aplicação e a relevância do conhecimento em relação às metas a serem atingidas. Em seguida, é feita a seleção dos conjuntos de dados a serem utilizados durante o processo, isto é, um agrupamento organizado de dados, que será o alvo da prospecção. A etapa da limpeza dos dados (*data cleaning*) vem a seguir, por meio de um pré-processamento dos dados, visando a adequá-los aos algoritmos, pois o ambiente de extração, freqüentemente apresenta ruídos, distorções, etc. Para diminuí-los, devem-se realizar

operações básicas, tais como remover os ruídos ou valores incoerentes, coletar as informações necessárias para o modelo e decidir sobre estratégias para controlar campos de dados perdidos. Segundo FAYYAD, 1996, essa etapa pode tomar até 80% do tempo necessário para todo o processo, devido às dificuldades bem conhecidas de integração de bases de dados heterogêneas. Após esta eliminação, efetuamos a redução e projeção dos dados, encontrando características úteis para representá-los, dependendo do objetivo da tarefa. Deve-se também usar a redução de dimensionalidade ou método de transformação para reduzir o número efetivo de variáveis em consideração. Prosseguindo no processo, chega-se à fase de *Data Mining* especificamente, que começa com a escolha dos algoritmos a serem aplicados. Essa escolha depende fundamentalmente do objetivo do processo de KDD: classificação, segmentação, agrupamento por afinidades, estimativas, árvores de decisão, etc. De modo geral, na fase de *Data Mining*, ferramentas especializadas procuram padrões nos dados. Essa busca pode ser efetuada automaticamente pelo sistema ou interativamente com um analista responsável pela geração de hipóteses.

Em geral, o processo de busca é iterativo, de forma que os analistas revêm o resultado, formam um novo conjunto de questões para refinar a busca em um dado aspecto das descobertas, e realimentam o sistema com novos parâmetros. Ao final do processo, o sistema de *Data Mining* gera um relatório das descobertas, que passa então a ser interpretado pelos analistas de mineração. Somente após a interpretação das informações obtidas, encontra-se o conhecimento, que segundo Dilly *apud* Oliveira (2001), significa validar o conhecimento extraído da base de dados, identificarem padrões e interpretá-los transformando-os em conhecimentos que possam apoiar as decisões.

4. ESTUDO DE CASO: ANÁLISE DA BASE DE DADOS DA FARMÁCIA HOMEOBEL CENTER VISANDO LOCALIZAÇÃO ESTRATÉGICA PARA CONSTRUÇÃO DE FILIAL.

Na farmácia Homeobel Center é utilizado um SGBD (Sistema de Gerenciamento de Banco de Dados) Firebird, onde o mesmo é desenvolvido e mantido por uma comunidade de desenvolvedores e pela Fundação FireBird. O FireBird é um produto de código livre, desenvolvido sobre licença da Interbase License. A origem do FireBird é a versão Open Source do Interbase® 6.0 de Julho de 2000, quando a Borland liberou os códigos fontes de seu produto. Encontra-se disponível em <http://www.firebird.com.br/>. Entretanto, foi necessário a digitalização da maior parte dos dados, pois estes apresentavam somente nos chamados “cadernões”. Depois de realizar a digitalização, foi necessário fazer uma análise dos campos presentes, para retirar os ruídos. Desta forma, pode-se adquirir um banco mais completo para o devido fim. Após a retirada desses dados, a base foi convertida para um formato compatível com o software Weka utilizado no processo de MD (Mineração de Dados), utilizando o formato .CSV.

Os campos utilizados foram: DS_BAIRRO_RES, FORM_PGTO, LOC_COMP, TURN_COMP, no que diz respeito respectivamente ao bairro de residência do cliente, a forma de pagamento, o local de compra (presente ou domicílio) e por fim, o turno da compra. A quantidade de registros utilizados a partir da base é de 1.218 (um mil duzentos e dezoito).

Foi constatado que esses campos utilizados eram de suma importância para/com o propósito de descoberta de localização estratégica para construção de uma filial, o que faz-se necessário conhecer quais e onde essas localizações relevantes se fixam para construção da mesma, provendo em quais localidades as compras se fazem presente ou sob domicílio. Assim, estariam aptos a realizar o processo de mineração dos dados desta base.

4.1 MINERAÇÃO DE DADOS

Com intuito de descobrir dados relevantes sobre a criação de uma filial da respectiva organização em um ponto estratégico e prevendo uma visão mais ampla do negócio, foi necessária a utilização da base de dados já existente na organização para que esta pudesse ser aplicada no software WEKA, almejando obter resultados de valores que possam auxiliar a tomada de decisão estratégica.

Com base do mesmo, foi utilizado a ferramenta Weka, que teve seu início em 1993, utilizando a plataforma Java, na Universidade de Waikato, Nova Zelândia, onde posteriormente seria adquirida por uma empresa no final de 2006. O Weka encontra-se licenciado ao abrigo da General Public License, sendo portanto, possível estudar e alterar o respectivo código fonte.

Este software tem como objetivo agregar algoritmos provenientes de diferentes abordagens/paradigmas na sub-área da inteligência artificial dedicada ao estudo da aprendizagem por parte de máquinas. Essa sub-área pretende desenvolver algoritmos e técnicas que permitam um computador "aprender" (no sentido de obter novo conhecimento) quer indutiva quer dedutivamente.

O Weka procede à análise computacional e estatística dos dados fornecidos, recorrendo a técnicas de data mining, tentando, indutivamente, a partir dos padrões encontrados, gerar hipóteses para soluções e nos extremos inclusive teorias sobre os dados em questão.

A regra de classificação é uma das tarefas mais referenciadas na literatura de MD. Neste tipo de tarefa, o objetivo é descobrir um relacionamento entre um atributo meta (cujo valor ou classe será previsto) e um conjunto de atributos previsores. O sistema deve descobrir este relacionamento a partir de exemplos com classe conhecida. O relacionamento descoberto será usado para prever o valor do atributo meta (ou a classe) para exemplos cujas classes são desconhecidas (Fertig *et al.*, 1999).

Segundo Romão et al. (2000), regras do tipo SE-ENTÃO são também chamadas de regras de produção, constituem uma forma de representação simbólica e possuem a seguinte forma:

SE <antecedente> ENTÃO <conseqüente>

O antecedente é formado por expressões condicionais envolvendo atributos do domínio da aplicação existentes no banco de dados. O conseqüente é formado por expressões que indicam a previsão de algum valor para um atributo meta, obtido em função dos valores encontrados nos atributos que compõem o antecedente.

Portanto, a tarefa é descobrir regras de classificação capazes de prever o valor de um atributo meta a partir dos valores de atributos antecedentes, permitindo o planejamento de ações futuras. Na tarefa de classificação, os dados são divididos em duas bases de dados, mutuamente exclusivas, chamadas base de treinamento e base de testes. O algoritmo de mineração de dados gera regras a partir da base de treinamento e as avalia sobre a base de testes.

O algoritmo de PRISM foi desenvolvido por Jadzia Cendrowska (Cendrowska, 1987) baseado no algoritmo de indução de regras ID3. A motivação na sua criação foi de superar as limitações na representação de árvores de decisão. PRISM produz regras modulares (independentes) que são de mais fácil compreensão. A geração de regras de classificação é realizada, diretamente, a partir do conjunto de treinamento. Não há a necessidade da etapa intermediária de construção de uma árvore de decisão.

5. RESULTADOS OBTIDOS

Sobre a análise do algoritmo de regras de classificação, foi utilizado o algoritmo PRISM, que consta no software Weka, onde o mesmo se mostrou mais eficiente e apto para o que foi proposto anteriormente, fazendo assim possível o levantamento da descoberta dos padrões a seguir.

5.1 PADRÕES RELEVANTES

A tabela 1 mostra o resultado obtido por uma mineração realizada através de classificação, induzida pelo algoritmo de PRISM, onde observamos um diferencial que se torna relevante, pois quando se trata de descoberta de localização estratégica, tem que se avaliar o quanto o negócio irá produzir lucros para a empresa, de forma que a mesma não venha a construir uma filial com uma estrutura grande em locais que geralmente as compras são realizadas de forma à domicílio, fazendo assim necessário nesses locais, a construção de uma filial bem menor, apenas para servir de base de entrega de produtos através de domicílio. Contudo, foram geradas as seguintes regras:

1. Se a forma de pagamento for do tipo à vista, e o bairro de residência for “Centro”, e o turno de compra for noturno, então o local de compra é presente (local).
2. Se a forma de pagamento for à vista, e bairro de residência for “Telégrafo”, e o turno de compra for à tarde, então local de compra é presente (local).
3. Se turno de compra é pela parte da manhã, e bairro de residência for “Fátima”, e forma de pagamento for “cheque”, então local de compra é a domicílio.
4. Se o bairro de residência for o “Comércio”, e com forma de pagamento à vista, no turno de compra à tarde, então local de compra é a domicílio.

Tabela 1 – Classificador induzido pelo algoritmo PRISM com seleção Local de Compra

<p>1. If FORM_PGTO = A VISTA and DS_BAIRRO_RES = CENTRO and TURN_COMP = NOITE then LOCAL</p>
<p>2. If FORM_PGTO = A VISTA and DS_BAIRRO_RES = TELEGRAFO and TURN_COMP = TARDE then LOCAL</p>
<p>3. If TURN_COMP = MANHA and DS_BAIRRO_RES = FATIMA and FORM_PGTO = CHEQUE then DOMICILIO</p>
<p>4. If DS_BAIRRO_RES = COMERCIO and FORM_PGTO = A_VISTA and TURN_COMP = TARDE then DOMICILIO</p>

Com análise a tabela 1, foi possível constatar que os presentes bairros de “Fátima” e “Comércio” tipicamente realizam suas compras com entrega em domicílio, tendo em vista essa informação, de acordo com as necessidades da empresa em questão, estes bairros foram avaliados como não possuidores de uma clientela que valha investir em uma construção de filial de grande porte, pois as análises mostram-se desfavoráveis com relação aos compradores das localidades ditas.

Por outro lado, verificou-se que os bairros do “Telégrafo” e do “Centro” possuem consumidores em potencial, salvo que a clientela local, tipicamente realiza consumação com pagamento à vista. Com base nas informações observadas na tabela 1, foi possível realizar uma análise e chegar a uma decisão no quesito da possível construção da filial perante os locais estratégicos citados. A construção de uma filial de grande porte seria um investimento com mais retorno se construída nas redondezas dos bairros do “Telégrafo” e “Centro”, pois são consumidores em potencial a compras locais, já os consumidores dos bairros de “Fátima” e “Comércio”, são clientes que preferem suas compras entregues nas suas residências, sendo assim, será mais proveitoso para a empresa, a criação de uma farmácia de menor porte, com objetivo de ser mais uma distribuidora que uma farmácia de grande potencial, ressaltando-se que se for investir na criação de uma loja de grande porte na localidade, só seriam despesas a mais, pois a consumação local é fraca e não compensa o investimento.

6. CONSIDERAÇÕES FINAIS

Este artigo visou a descrição da descoberta em base de dados, fazendo alusão a todas suas etapas do KDD (Knowledge Discovery in Databases), tendo o seu foco principal na etapa de DM (data mining), na qual possibilita a extração de resultados que

ajudam a tomada de decisão. Isto foi possível devido a utilização adequada de uma ferramenta para mineração de dados (Weka) na qual pode-se aplicar uma base para efetuar a busca de dados relevantes.

A base aplicada é de uma organização farmacêutica, e busca a extração de informações implícitas que auxiliam na tomada de decisão final. A finalidade da aplicação da base em um software de data mining é a busca e análise de regras que possam gerar possíveis localizações estratégicas para criação de uma organização no âmbito comercial, que foram obtidos através das informações provenientes do cadastro dos clientes. Com base nessas regras, os empresários da farmácia podem avaliar seus clientes sob diferentes pontos de vista, como por exemplo, averiguar quais locais compram mais, e qual tipo de compra por localidade dos mesmos, além de verificarem dados, como o tipo de compra, à vista, cheque ou cartão, definidos por bairros da cidade de Belém.

Visando a aplicação de tecnologia conjuntamente com a área farmacêutica, os proprietários aceitaram a proposta do projeto com intuito de expandir os seus negócios aliado ao que está acontecendo de novo no mercado, assim, com as informações relevantes, a possibilidade da percepção de seus negócios a curto prazo alavancarem e a longo prazo a utilização de outras novas tecnologias a fim do reconhecimento perante a clientela já estabelecida e aquisição de futuros clientes que reconheçam o trabalho de qualidade que pode ser ofertado da união de tecnologia e negócio.

O trabalho abre caminho para uma série de análises em Data Mining, partindo-se de um Data Warehouse elaborado e melhorado, outras técnicas podem ser utilizadas, entre elas, merecem destaque as regras de Associação e Agrupamento.

7. REFERÊNCIAS

BIANCCA, Nardelli Schenatz, **Utilização de Data Mining em um sistema de informação gerencial para o diagnóstico da formação de professores da graduação**. Dissertação (Mestrado em Engenharia de Produção) – Centro de Tecnologia, Universidade Federal de Santa Catarina, Florianópolis, 2005. Disponível em: <<http://teses.eps.ufsc.br/defesa/pdf/16068.pdf>>. Acesso em: 10 de março de 2007.

CARVALHO, Luís Alfredo Vidal de, **Datamining: a mineração de dados no marketing, medicina, economia, engenharia e administração**, São Paulo: Editora Érica, 2001.

CORRÊA, Adriana Cristina Giusti, SCHIABEL, Homero, **Descoberta de conhecimento em base de imagens mamográficas**. Departamento de Computação da UNIRP, Centro Universitário de Rio Preto, São José do Rio Preto. Disponível em: <<http://www.sbis.org.br/cbis9/arquivos/546.pdf>>. Acesso em: 26 de março de 2007.

COSTA, Marcelo Senos, **Mineração de dados utilizando redes neurais**. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Departamento de Computação, Universidade Estadual de Londrina, Londrina, 2005.

FREITAS, Alex A., **Data mining and knowledge discovery with evolutionary algorithms**, New York: Editora Springer Verlag, 2002.

NAVEGA, Sergio, **Princípios essenciais do data mining**. Disponível em: <<http://www.intelliwise.com/reports/i2002.pdf>>. In: Centro Nacional de Desenvolvimento do Gerenciamento da Informação - CENADEM, 2002, São Paulo. Anais do Infoimagem. Acesso em: 05 abril de 2007.

SANTIAGO, Dayana Gonçalves, PEREIRA, Rogério Ramos, MOREIRA, Fabiano Cordeiro, **Comparativo entre algoritmos genéticos e algoritmos meméticos na busca por padrões seqüenciais. Estudo de caso: mineração de dados no SEBRAE Pará**, Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Área de Ciências Exatas e Tecnologia, Centro Universitário do Pará, Belém, 2007.

TRAUTENMÜLLER, Patrícia, HIRA, Adilson Yujji, ZUFFO, Marcelo Knörich, **Ambiente de visualização de informações para acompanhamento da evolução clínica de pacientes oncológicos**. Departamento de Engenharia de Sistemas Eletrônicos da Escola Politécnica da USP, São Paulo. Disponível em: <<http://www.sbis.org.br/cbis/arquivos/787.pdf>>. Acesso em: 20 de abril de 2007.

Wikipedia. **Firebird (servidor de base de dados)**. Disponível em: <http://pt.wikipedia.org/wiki/Firebird_%28servidor_de_base_de_dados%29>. Acesso em: 10 de maio de 2007.

Wikipedia. **Weka**. Disponível em: <<http://pt.wikipedia.org/wiki/Weka>>. Acesso em: 07 de maio de 2007.