

1 Introdução

A digitalização da sociedade juntamente com a difusão da Internet e a globalização trazem a necessidade de maior agilidade e rapidez para os processos de negócio, principalmente no que se refere aos sistemas de tomada de decisão. Além disso, podemos observar um aumento significativo na quantidade de dados armazenados pelas empresas, pois as informações assumiram um papel indispensável e vital para a sobrevivência das empresas no mercado.

Entretanto, é necessária a existência de processos que possibilitem extrair destes dados o conhecimento necessário para a gestão e a tomada de decisões, caso contrário são apenas um conjunto de dados que não trazem nenhum benefício para a corporação. Nesse sentido, surgem os sistemas do tipo *data warehouse* – ou armazém de dados – para auxiliar no armazenamento dos dados e facilitar a análise para a tomada de decisões. Estes passam a ser realidade, principalmente, nas grandes corporações.

Data warehouse é um termo amplamente utilizado nos dias de hoje para definir um ambiente de sistemas de informação que tem por objetivo auxiliar os usuários de níveis gerenciais. É uma “fotografia” dos dados em determinado período de tempo, integrados por uma combinação de tecnologias e orientados a assunto. Composto por uma coleção de tecnologias de suporte a análise de decisão, tem por responsabilidade o armazenamento dos dados que ao longo do tempo se proliferaram. A análise dos dados contidos em uma *data warehouse* é realizada por vários tipos de ferramentas, sendo OLAP (*On Line Analytical Processing*) a mais simples e mais utilizada.

A tecnologia em questão compreende Sistemas Gerenciadores de Banco de Dados com visões relacionais ou multidimensionais dos dados, facilidades para acesso dos dados, arquitetura cliente-servidor, interfaces gráficas de usuários, dentre outras.

Com o aumento do volume de dados nas empresas e a diminuição do tempo para resposta às mudanças, o uso da *data warehouse* se fez necessário para atender à necessidade de integração dos sistemas voltados para o nível gerencial com dados que nunca ou raramente são alterados. Os sistemas transacionais do nível operacional lidam com os dados “ativos” espelhando as operações de negócio da empresa. Os dados estáveis da *data warehouse* ajudam a compor séries históricas, apoiando o trabalho de análise ao longo de determinado período.

Geralmente, os dados armazenados em uma *data warehouse* são apenas consultados, sendo agregados para fins de estudo e tomadas de decisão. As aplicações que utilizam esses dados são classificadas como sistemas de informação gerencial, sistemas de informação para executivos ou sistemas de suporte à decisão. Além disso, é possível a manipulação de grandes quantidades de dados armazenadas a fim de encontrar padrões implícitos entre os mesmos, possibilitando a previsão de informações/situações futuras.

1.1. Motivação

O mercado de *business intelligence*, ao qual as *data warehouse* pertencem, está em grande e constante expansão fomentando a pesquisa e desenvolvimento de ferramentas melhores e mais sofisticadas.

Cabe ressaltar que a chave do processo de busca de conhecimento é a mineração de dados (ou *data mining*), entendida como tarefas e técnicas dedicadas à exploração de grandes quantidades de dados à procura de regras de associação, relacionamentos entre variáveis, classificação de itens, dentre outras atividades.

Do ponto de vista de negócio, os padrões que, eventualmente, estão escondidos entre uma grande quantidade de dados podem ser extremamente valiosos e úteis. Nas empresas, cada vez mais os diferentes setores procuram por ferramentas que possam auxiliá-los a maximizar os ganhos e minimizar as perdas.

Por exemplo, a aplicação prática de mineração de dados no cenário de CRM (*Customer Relationship Management*) é bastante interessante, uma vez que as empresas estão aumentando o foco no cliente, conhecendo-o individualmente com o intuito de oferecer-lhe mais produtos, cada vez mais adequados e até mesmo personalizados.

Dentro deste contexto, a eficiência e eficácia da adesão de novos clientes e da fidelidade dos antigos dependem do sucesso do processo de descobrimento de padrões, incluindo a mineração de dados.

1.2. Objetivos

Este trabalho tem por objetivo estudar e propor melhorias em algoritmos de recomendação de produtos com o intuito de auxiliar na descoberta de padrões de consumo e, assim, melhorar a relação das empresas com seus clientes.

Para atingir o objetivo central, duas etapas serão seguidas:

- Estudar o processo de descoberta de conhecimento em *data warehouses*;
- Executar testes com grandes volumes de dados.

1.3. Trabalhos Relacionados

Para encaminhar o problema de recomendação de produtos, inúmeras soluções têm sido propostas. Esta seção apresenta brevemente alguns trabalhos relacionados que serviram de inspiração e fontes de informação para o desenvolvimento desta e dissertação.

Em [11] são discutidas as oportunidades de se construir um melhor relacionamento com clientes, aumentando a frequência das compras e conseqüentemente aumentando o lucro. Para que seja possível identificar padrões e possibilitar a recomendação de produtos e serviços é de suma importância para as operações de venda manter o histórico das transações com o máximo de detalhes possível, como quantidade, valores, tipo de pagamento, as respostas aos estímulos de marketing, etc. Suas análises ficam facilitadas com a utilização de técnicas de mineração de dados, como regras de associação, que buscam automaticamente padrões nos dados, gerando modelos de comportamento. Neste trabalho, cada cliente é classificado com uma pontuação (*score*) que representa a probabilidade de resposta positiva ao estímulo posto no modelo, auxiliando o profissional de marketing a focar mais precisamente no público alvo de suas campanhas.

Santos [12] descreve a aplicação das técnicas de mineração de dados na gestão de relacionamento com clientes. Como estudo de caso, foi apresentado um sistema desenvolvido com o intuito de apoiar as decisões na área de vendas, auxiliando na recomendação de produtos que viessem a ser de interesse do consumidor em questão. O usuário escolhe um ou mais produtos que serão confrontados com um ou mais produtos também à escolha do usuário e, assim, o sistema tenta determinar regras de associação entre os itens selecionados. Ao concluir, o autor comenta que as regras de associação são uma técnica de mineração de dados de simples implementação, mas fornece conhecimentos importantes sobre os hábitos dos clientes.

Adomavicius e Tuzhilin [14] fazem uma pesquisa sobre o estado da arte de sistemas de recomendação e chegam à conclusão de que, mesmo tendo avançado muito nessa última década, esse tipo de sistema ainda apresentam limitações que precisam ser superadas. Dentro desse contexto, ações como melhoria na modelagem dos itens e usuários, incorporação de informação contextual no processo de recomendação suporte à sugestões por multicritérios e a criação de um sistema mais flexível e menos invasivo são apontadas como sendo as principais.

Burke [15] define três tipos de sistemas de recomendação: (i) sistemas colaborativos são aqueles em que as recomendações são baseadas nas preferências dos outros usuários; (ii) sistemas baseados no conteúdo são aqueles que classificam as informações baseados em aprendizado de máquina e sugerem baseados nas escolhas anteriores do usuário; (iii) sistemas baseados no conhecimento (*knowledge-based systems*) dos usuários oferecem produtos que atendem aos requisitos do usuário, como cidade, preferências por comida, entre outros. Além disso, é esclarecido que sistemas de recomendação precisam ser inicializados com grandes quantidades de dados, pois caso contrário as previsões e sugestões não serão eficientes como esperado. Com isso, confronta-se o problema de que, quando há muitos usuários com hábitos conhecidos, o sistema não é tão útil; mas, se estiver com uma grande quantidade de itens avaliados, pode não ser preciso para um usuário em particular. Sistemas baseados no conteúdo apresentam um problema parecido: bons classificadores “aprendem” depois que muitos itens foram avaliados e alguns procuram resolver esse problema fazendo uso de algoritmos classificadores baseados na similaridade com os dados da vizinhança, causando a limitação de sugerir itens que sejam similares aos itens previamente avaliados. Nessa linha, e baseado em uma série de exemplos de protótipos, o autor defende os sistemas de recomendação baseados em conhecimento como sendo mais valiosos que os outros, pois sua recomendação não depende das avaliações de usuários e suas tomadas de decisão são independentes do gosto do mesmo. Mas, além disso, discute em cima de um de seus exemplos (*FindMe*) que sistemas híbridos de recomendação podem ser criados combinando diferentes técnicas.

Middleton et al. [16] defendem que os sistemas de recomendação podem ajudar na melhoria das buscas na Internet, onde existe uma grande quantidade de páginas com dados não-estruturados e onde, neste cenário, as ferramentas de busca são efetivas apenas quando as páginas atendem às chaves das consultas explicitamente. No decorrer do artigo, os autores descrevem o sistema *Quickstep*, criado utilizando ontologias durante o processo de *profiling* e, baseados nos testes, concluem que os resultados dos usuários que usaram o sistema com ontologias foram mais precisos, recomendando itens que atendiam mais aos interesses dos usuários e que não eram diretamente sugeridos.

1.4. Organização do Trabalho

Este trabalho está organizado da seguinte forma. O Capítulo 2 apresenta o conceito de Mineração de Dados, cenário no qual esta pesquisa se enquadra. O Capítulo 3 detalha melhor a técnica *Market Basket Analysis*, foco deste estudo. O Capítulo 4 descreve o protótipo criado para testar a técnica de *Market Basket Analysis* e apresenta

os resultados dos testes. Por fim, o Capítulo 5 apresenta as conclusões e indica algumas linhas de pesquisa a serem desenvolvidas a partir do conteúdo exposto neste trabalho.