Universidade Federal do Rio Grande do Sul
Programa de Pós-Graduação em Administração
Convênio UFRGS-URCAMP
Mestrado em Administração

MINERAÇÃO DE DADOS EM SUPERMERCADOS: O CASO DO SUPERMERCADO "TAL"

Por:

Lóren Pinto Ferreira Gonçalves
loren@alternet.com.br

Orientador:

Dr. Henrique Freitas

Porto Alegre, dezembro de 1999

SUMÁRIO

1. Tema e justificativa: mineração de dados e administração.....3
2. Objetivos.....7
3. A utilização de técnicas de mineração de dados como fonte de informação para a tomada de decisão.....8
   3.1 Dados, informação e conhecimento.....8
   3.2 Administração e decisão.....9
      3.2.1 Tomada de decisão e níveis organizacionais.....10
      3.2.2 Modelo de decisão de Simon.....11
      3.2.3 Racionalidade da decisão.....12
   3.3 Banco de dados.....12
   3.4 Mineração de dados e prospecção de conhecimento em banco de dados.....13
      3.4.1 O que é mineração de dados?.....14
      3.4.2 Por que utilizar mineração mineração de dados?.....14
      3.4.3 Ciclo Virtuoso de mineração de dados.....16
      3.4.4 Tarefas de mineração de dados.....17
      3.4.5 Técnicas de mineração de dados.....20
      3.4.6 Que técnica utilizar?.....24
      3.4.7 Metodologia de mineração de dados.....25
      3.4.8 Aplicações.....26
1. Método de Pesquisa.....28
2. Contexto da aplicação.....31
3. Cronograma.....33

Referências.....34

1. Tema e justificativa: mineração de dados e administração

O tema da presente pesquisa é mineração de dados, ou *data mining*, o qual é uma etapa na descoberta do conhecimento em bancos de dados que consiste no processo de analisar grandes volumes de dados sob diferentes perspectivas, a fim de descobrir informações úteis que normalmente não estão sendo visíveis. Para isto são utilizadas técnicas que envolvem métodos matemáticos, algoritmos e heurísticas[1] para descobrir padrões e regularidades entre os dados pesquisados (Brusso, 1998).

Num mundo globalizado, onde não mais existem fronteiras geográficas, onde as empresas competem mundialmente a informação torna-se um fator crucial na busca pela competitividade. A qualidade desta informação, em termos de confiabilidade, rapidez, formato, entre outros, é de extrema importância para o sucesso da empresa. Almeida (1995) afirma que o fato de uma empresa dispor de certas informações possibilita-lhe aumentar o valor agregado de seu produto ou reduzir seus custos em relação àquelas que não possuem o mesmo tipo de informação.

Freitas e Lesca (1992) dizem que as informações e o conhecimento compõem um recurso estratégico essencial para o sucesso da adaptação da empresa em um ambiente de concorrência. Segundo Oliveira (1997) toda empresa tem informações que proporcionam a sustentação para as suas decisões, entretanto apenas algumas conseguem otimizar o seu processo decisório e aquelas que estão neste estágio evolutivo seguramente possuem vantagem empresarial.

A utilização de ferramentas, técnicas e tecnologias apropriadas ao melhoramento da obtenção, tratamento, apresentação e disponibilização de informação é um fator que pode influenciar, ou pode até ser definitivo, no aumento da competitividade da organização. Por isso, torna-se importante o conhecimento de quaisquer recursos que possam ser utilizados para este fim.

Conforme Figueira (1998), a cada ano, companhias acumulam mais e mais informações em seus bancos de dados. Como conseqüência, estes bancos de dados passam a conter verdadeiros tesouros de informação sobre vários dos procedimentos dessas companhias. Toda esta informação pode ser usada para melhorar seus

[1] Heurísticas são processos ou regras de pesquisa e busca de soluções, conduzidos por processos de associações de idéias em geral incompletas, pela complexidade que os problemas tratados envolvem procurando simular ou substituir os processos de inferência dedutiva do raciocínio humano (Torres, 1995)

procedimentos, permitindo que a empresa detecte tendências e características disfarçadas, e reaja rapidamente a um evento que ainda pode estar por vir. No entanto, apesar do enorme valor desses dados, a maioria das organizações é incapaz de aproveitar totalmente o que está armazenado em seus arquivos. Esta informação preciosa está na verdade implícita, escondida sob uma montanha de dados, e não pode ser descoberta utilizando-se sistemas de gerenciamento de banco de dados convencionais. O autor diz ainda que a tecnologia tornou relativamente fácil este acúmulo de dados. A quantidade de informação armazenada está explodindo, e ultrapassa a habilidade técnica e a capacidade humana na sua interpretação.

Newing (1996) diz que as organizações estão acumulando vastas quantidades de dados, que registram as suas atividades, em bancos de dados, com a tendência recente de implementar uma arquitetura *data warehouse* aumentando a qualidade e a acessibilidade dos dados. Ao mesmo tempo, a informação é valorizada como nunca antes na história, e os dados armazenados nos *data warehouse*s são vasculhados por profissionais especializados, a procura de tendências e padrões. De compras através de cartões de crédito a imagens pixel-a-pixel de galáxias, bancos de dados são medidos hoje em gigabytes ou até em terabytes. A necessidade de transformar estes bytes de dados em informações significativas é óbvia, entretanto, a sua análise ainda é demorada, dispendiosa, pouco automatizada e sujeita a erros, mal entendidos e falta de precisão. A automatização dos processos de análise de dados, com a utilização de softwares ligados diretamente à massa de informações, se tornou uma necessidade (Figueira, 1998).

Newing (1996) salienta que isto está sendo feito com grandes custos, porém ele ressalta que a informação somente é valiosa se usada efetivamente.

Estes dados muitas vezes são mantidos mesmo depois de esgotados seus prazos legais de existência, como no caso de notas fiscais. Com o passar do tempo, este volume de dados passa a armazenar internamente o histórico das atividades da organização, com muitas informações implícitas que não podem ser descobertas e visualizadas sem a utilização de técnicas apropriadas.

Os usuários têm usado diversas ferramentas (*query*, servidores *OLAP*, ferramentas de inteligência competitiva, sistemas de informações gerenciais, entre outros) para examinar os dados que possuem, no entanto, a maioria dos analistas tem reconhecido que existem padrões, relacionamentos e regras escondidos nestas dados que

não podem ser encontrados utilizando estes métodos tradicionais. Para Newing (1996) a resposta é usar softwares de mineração de dados que utilizam algoritmos matemáticos avançados para examinar grandes volumes de dados detalhados. Senda mineração de dados o processo de extração de informação desconhecida de grandes bases de dados e utilização das mesmas para tomar decisões críticas de negócios, os softwares de mineração de dados são capazes de peneirar grandes volumes de dados para encontrar 'pepitas' de informação as quais produzem 'ouro' em forma de vantagem competitiva.

A mineração de dados é diferente dos sistemas onde a pessoa entra com uma hipótese e a verifica. Nos sistemas convencionais o usuário informa a hipótese e o sistema verifica se esta é verdadeira ou falsa, portanto é o usuário quem deve ter o 'insight' para verificar as normas ou regras, assim a descoberta das informações escondidas na base de dados fica dependente da racionalidade limitada do usuário. Na mineração de dados, ao contrário, o sistema retorna todas as regras encontradas e a pessoa faz uso dela da forma que achar melhor.

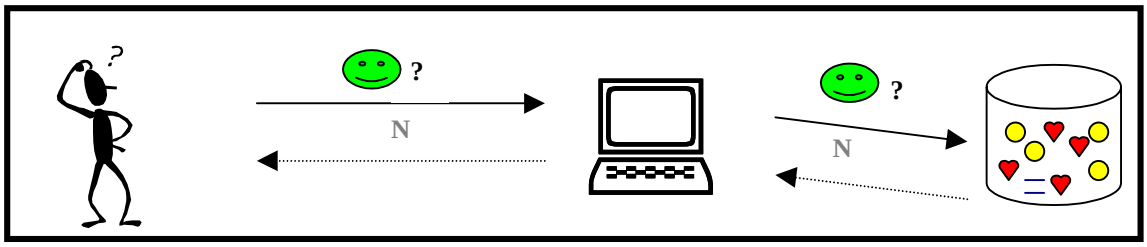

Figura 1 - Busca por informações em sistemas convencionais (SIG, OLAP, Query, etc.)

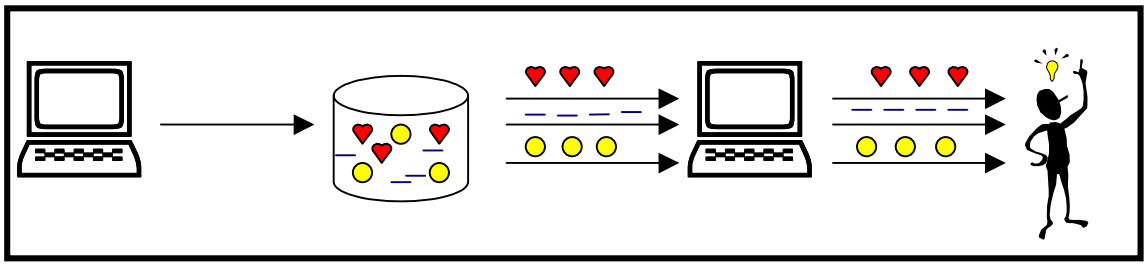

Figura 2 - Busca por informações em sistemas de mineração de dados

Neste contexto, as técnicas de mineração de dados ganham importância, pois ao encontrar padrões e relacionamentos ocultos nestes bancos de dados, trazem à tona conhecimentos não apenas sobre os objetos registrados, mas também, se a base de dados é um espelho fiel, sobre o mundo real registrado no banco de dados.

Para desenvolver este documento iremos tratar na seção dois dos objetivos desta pesquisa, na seção três trataremos do referencial teórico pertinente, na seção quatro do método de pesquisa que será utilizado, na seção cinco do contexto de aplicação, ou seja, das empresas que fazem parte da nossa amostra e das contribuições potenciais desta pesquisa, na seção seis é mostrado o cronograma de execução.

2. Objetivos

2.2.1. Objetivo Geral

- Aplicar técnicas e ferramentas de mineração de dados como fonte de informação para a tomada de decisão, mostrando o interesse de sua aplicação na administração de negócios e das organizações.

2.2.2. Objetivos Específicos

- Determinar algumas situações em que as técnicas de mineração de dados podem ser utilizadas e que técnica pode ser usada em cada uma destas situações;
- Identificar algumas das ferramentas de mineração de dados existentes;
- Analisar a possibilidade de utilização desta(s) ferramenta(s) nas bases de dados disponíveis;
- Avaliar, junto ao tomador de decisão, a pertinência e utilidade das informações obtidas;
- Identificar os recursos necessários à utilização de ferramentas de mineração de dados.

3. A utilização de técnicas de mineração de dados como fonte de informação para a tomada de decisão

Mineração de dados é uma etapa na descoberta do conhecimento em bancos de dados que consiste no processo de analisar grandes volumes de dados sob diferentes perspectivas, a fim de descobrir informações úteis que normalmente não estão sendo visíveis. Para isto são utilizadas técnicas que envolvem métodos matemáticos, algoritmos e heurísticas para descobrir padrões e regularidades entre os dados pesquisados (Brusso, 1998).

As informações obtidas através da mineração de dados serão utilizadas na tomada de decisão das empresas. É importante saber para que nível organizacional estas informações serão úteis, pois dependendo do nível mudam as características das informações.

Para compreendermos o assunto é preciso conhecermos alguns conceitos básicos, tais como: dados, informação e conhecimento.

3.1. Dados, informação e conhecimento

Suliman Jr. et al.. (1997) afirmam que os conceitos de dados, informação e conhecimento podem variar. Para os autores existe uma hierarquia de complexidade em que os dados constituem a parte mais simples dessa hierarquia e o conhecimento constitui a parte mais complexa.

Basicamente, se atribuímos algum significado especial a um dado, este se transforma em informação. Se os especialistas no domínio do problema elaboram uma norma (regra), a interpretação do confronto entre esta informação e essa regra constitui um conhecimento (Suliman Jr., 1997). Para Freitas et al.. (1997) a informação é considerada como um dado dotado de relevância e propósito, para cuja conversão é necessário conhecimento.

Nem todas as informações apresentam importância para uma tomada de decisão. Umas são mais importantes, mais relevantes do que outras. Sulliman Jr. (1997) define

relevância como grau de importância que uma informação possui para a tomada de decisão.

3.2. Administração e Decisão

Newman, Summer e Warren (apud Albertin, 1998) citam uma concepção de administração como um processo de aplicação de princípios e de funções para o alcance de objetivos.

Na abordagem dada por Dale (apud Albertin, 1998) para a administração, são apresentadas cinco funções essenciais, que são: planejamento, organização, pessoal, direção e controle.

A todo momento precisamos tomar decisões. Estas decisões vão desde as mais simples, como a escolha da roupa a ser vestida, até as mais difíceis, ou seja, aquelas que necessitam de um tratamento mais aprofundado, pois se não forem bem estruturadas podem trazer conseqüências desastrosas, como por exemplo a escolha de um curso no momento da inscrição para o vestibular. O mesmo acontece nas empresas. Como não podemos tomar decisões erradas, correndo o risco de prejudicarmos os negócios, é preciso que haja disponibilidade de informações de qualidade que auxiliem aos executivos no momento da tomada de decisão. As tecnologias da informação começam, então, a aparecer no intuito de propiciar o suporte necessário aos tomadores de decisão das empresas que buscam um diferencial em relação aos seus concorrentes.

Torres (1995) nos diz que por muito tempo as organizações têm feito uso das tecnologias de informação, mas de forma não administrada ou mal administrada. Ele afirma que grande parte das empresas ainda utiliza os recursos da informática voltados ‘para dentro’ das empresas, isto é, para resolver operações básicas, registrando e recuperando transações, processando documentos e dando apoio ao trabalho funcional.

> *“Esse tipo de uso é necessário, porém é importante que a visão das possibilidades de utilização dessas tecnologias seja ampliada e contemple o novo universo que cada vez mais mudará as relações de competitividade em todos os segmentos da economia.” (Torres, 1995)*

3.2.1. Tomada de decisão e níveis organizacionais

As decisões são tomadas em todos os níveis organizacionais, porém elas são diferenciadas. No nível operacional as decisões a serem tomadas são, em grande maioria programadas ou programáveis, ou seja, por tratarem de problemas rotineiros elas podem ser formalizadas em manuais, estatutos, etc. Já no nível estratégico, onde encontram-se os executivos mais preocupados com aspectos externos à organização (tais como: concorrência, governo, mercado, etc.), as decisões são, na maioria das vezes, não programadas ou não-programáveis, a decisão deverá partir da análise da pessoa responsável pela mesma, não há regras e nem documentos formalizando a ação que deve ser tomada em tal situação. Neste ponto encontra-se atualmente o maior potencial de utilização das tecnologias de informação, pois devido à alta competitividade empresarial aquele que obtiver informações relevantes em tempo hábil e formato adequado estará à frente dos seus concorrentes.

Segundo Freitas et al.. (1997) a hierarquia entre os três níveis pode ser representada por meio da pirâmide organizacional que também representa a abrangência e importância das decisões dentro da organização, que aumentam na medida em que a decisão acontece em seus níveis superiores. A pirâmide transmite a idéia da hierarquia dentro da empresa (Figura 3), onde os elementos colocados em posições superiores são os responsáveis pelas decisões chamadas estratégicas.

À medida em que o nível de decisão se desloca para os níveis superiores da pirâmide aumenta a incerteza e o risco. As decisões no nível estratégico são geralmente tomadas em uma situação de incerteza e risco.

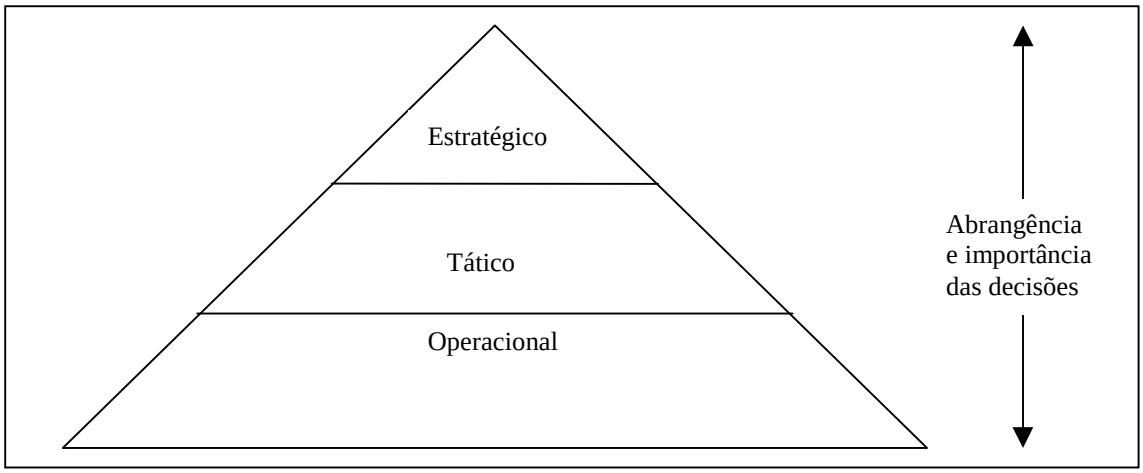

Figura 3 – Modelo da pirâmide (adaptado por Kendall & Kendall, 1991 e Le Moigne, 1974), conforme Freitas et al. (1997)

3.2.2. Modelo de decisão de Simon

Simon (apud Freitas et al.., 1997) propõe um modelo de decisão dividido em três grandes fases com uma constante revisão entre si, conforme demonstrado na Figura 4.

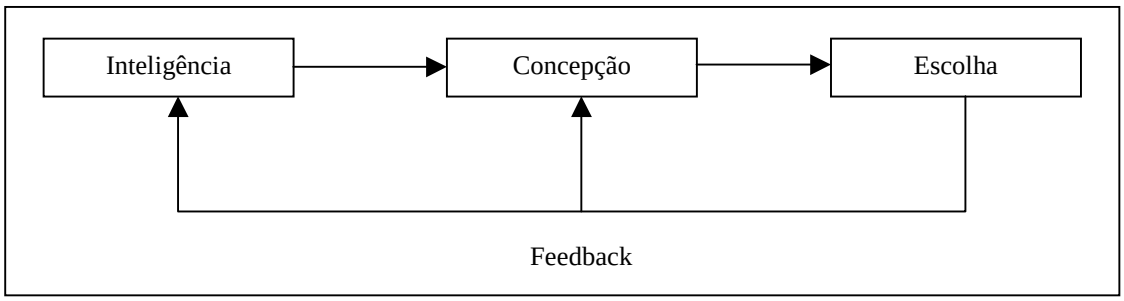

Figura 4 – Processo decisório baseado em Simon (1957), conforme Freitas et al. (1997)

As fases propostas por Simon (apud Freitas et al., 1997) têm as seguintes funções:

- inteligência ou investigação – onde identifica-se qual é o problema;
- desenho ou concepção – onde identifica-se quais são as soluções possíveis;
- escolha – onde identifica-se qual é a melhor alternativa dentre as verificadas no item anterior;

- feedback – eventos em que fases já vencidas do processo sejam resgatadas.

3.2.3. Racionalidade da decisão

A racionalidade, segundo Feitas et tal. (1997), se ocupa da seleção de alternativas que mais se encaixem em algum sistema de valores e são, até certo ponto, uma aceitação do razoável. Alter (1996) também nos fala sobre a escolha de uma alternativa satisfatória ao invés de uma alternativa ótima. Segundo este autor esta idéia é consistente com a teoria da limitação da racionalidade, onde as pessoas decidem num período de tempo limitado, baseado em informações limitadas a com uma habilidade limitada para processar estas informações. Se pensarmos nas últimas decisões que tivemos de tomar certamente concluiremos que não obtivemos todas as informações relevantes.

3.3. Banco de dados

Freedman (1995) define banco de dados e base de dados como qualquer área eletrônica de dados, ou seja, qualquer coleção de dados armazenada eletronicamente. Paladino (1986) dá definições diferentes para estes termos. Para ele banco de dados é o conjunto de informações cobrindo um domínio particular de conhecimento, uma ampla coleção de bibliotecas de dados armazenados para controle e base de dados como o conjunto de informações, armazenadas de forma sistemática, facilitando a recuperação e atualização dos itens, necessários para uma série de aplicações automáticas.

Neste trabalho trataremos os termos banco de dados e bases de dados como sinônimos, assim como Freedman (1995).

3.4. Mineração de dados e prospecção de conhecimento em banco de dados

A prospecção de conhecimento em bancos de *dados (Knowledge Discovery in Databases - KDD*) é um processo que envolve a automação da identificação e do reconhecimento de padrões (Suliman Jr., 1997). Os autores garantem que o desenvolvimento desse processo é possível através de uma geração de ferramentas e técnicas para análise de dados em bancos de dados. A mineração de dados é um passo no processo de prospecção, que consiste na utilização de algoritmos que produzem uma enumeração particular de padrões.

Segundo Suliman Jr. et al.. (1997) a expressão prospecção de conhecimento em bancos de dados foi inventada em 1989 para se referir ao processo de encontrar conhecimento em dados e para enfatizar o alto nível das aplicações de métodos de mineração de dados em particular.

A expressãa mineração de dados é comumente usada por estatísticos, analistas de dados e pela comunidade *MIS* (*Management Information Systems*), gerentes de sistemas de informações, enquanto KDD é mais usada pelos pesquisadores de inteligência artificial (Suliman, 1997).

As fases do processo de prospecção de conhecimento em banco de dados, segundo Suliman Jr. et al.. (1997), são identificadas como:

a) Desenvolver a compreensão do domínio da aplicação, o conhecimento anterior relevante e os objetivos do usuário final;
b) Criar um conjunto-alvo de dados em que a prospecção deverá ser efetuada;
c) Realizar a redução e projeção de dados, reduzindo o número efetivo de variáveis consideradas ou encontrar representações não variáveis para os referidos dados;
d) Escolher as tarefas de mineração de dados: decidindo se o objetivo do processo KDD é a classificação, regressão, clusterização ou outro;
e) Escolher os algoritmos de mineração de dados, selecionando métodos para uso na busca de padrões nos dados;
f) Mineração de dados;
g) Interpretação dos padrões obtidos;
h) Consolidação do conhecimento.

3.4.1. O que é mineração de dados?

Mineração de dados (também conhecida como Descoberta de Conhecimento em Bases de Dados) tem sido definida como a extração não trivial da informação importante, implícita, previamente desconhecida, de dados. Ela usa o aprendizado de máquina, técnicas estatísticas e de visualização para descobrir e apresentar o conhecimento em uma forma facilmente compreensível pelos humanos.

Berry e Linoff (1997) definem mineração de dados como a exploração e análise, por meio automático ou semi-automático, de grandes quantidades de dados no intuito de descobrir padrões e regras.

Mineração de dados tem sido descrita como a interseção entre a inteligência artificial, aprendizagem da máquina e tecnologias de bancos de dados. Muitas vezes a meta é construir automaticamente um modelo de software que prediga um valor de saída dado um conjunto de valores de entrada. Uma variedade de técnicas podem ser usadas, e cada uma tem sua estrutura própria.

3.4.2. Por que utilizar mineração de dados?

> *"Nós estamos construindo sistemas para colecionar dados, mas o próximo desafio é interpretar estes dados, e isto é o que mineração de dados faz".*
>
> *Paul Lalley*[2]

Vivemos em um mundo em que um dos fatores mais fortes de competitividade para qualquer empresa, em qualquer ramo de negócios, é o uso da informação e da tecnologia da informação (Torres, 1995).

Kotler (1998) afirma que na sociedade da informação, de hoje, o desenvolvimento de informações confiáveis pode proporcionar à empresa um salto

[2] apud Newing (1996)

sobre suas concorrentes, para ele a administração deve desenvolver e administrar informações para conhecer as mudanças de desejos dos consumidores, dos canais de distribuição, as novas iniciativas dos concorrentes, etc.

A tecnologia mineração de dados visa explorar grandes bancos de dados para obter, de forma automática, valiosas informações que poderão causar diferenciais efetivos no negócio.

Segundo Feldens, Moraes e Pavan (199?) os sistemas de mineração de dados são capazes de aprender e apoiar a realização de descobertas a partir de bases de dados. Estes sistemas podem auxiliar o processo, analisando volumes muito grandes de dados, evidenciando relacionamentos difíceis de se perceber, muitas vezes revelando situações inesperadas, trazendo à tona problemas com a qualidade de serviço/produto ou das próprias informações, possíveis de erros nas bases de dados e até fraudes.

Com o rápido crescimento da informatização, da automação dos processos, e por conseqüência, da quantidade de informações armazenadas, o desenvolvimento de ferramentas eficientes de mineração de dados se tornou um desafio importante em diversas áreas de pesquisa, como em bancos de dados, estatística, inteligência artificial e aprendizado de máquina, entre outros (Figueira, 1998).

Harrison (1998) afirma que os pequenos varejistas usam o conhecimento do cliente para inspirar sua fidelidade. Uma empresa pequena constrói seus relacionamentos com os clientes atendendo as suas necessidades, lembrando suas preferências e aprendendo através das interações passadas como servi-los melhor no futuro (Berry, 1997).

Como uma grande empresa pode realizar tais procedimentos se as interações existentes geralmente ocorrem com funcionários diferentes? Então, como a empresa poderá atender suas necessidades, lembrar-se de suas preferências e aprender com as interações passadas? O que pode substituir a intuição da pessoa que conhece os clientes por nome, fisionomia e voz, e lembra-se dos seus hábitos e preferências? Através da aplicação inteligente da tecnologia da informação, mesmo a maior empresa pode vir a ficar próxima dos seus clientes. Conforme Harrison (1998) o *data warehouse* fornece a memória para a empresa, mas ele salienta que memória sem inteligência tem pouco uso.

> *A inteligência nos permite vasculhar nossa memória observando padrões, inventando regras, tendo novas idéias para fazer previsões sobre o futuro.(Harrison, 1998)*

Um exemplo da utilização de mineração de dados: a rede americana Wal-Mart descobriu que as pessoas que vão as suas lojas às quintas-feiras para comprar fraldas Huggies tendem a adquirir dezenove itens adicionais. Assim, toda quinta-feira a Wal-Mart altera a disposição dos produtos de suas lojas para assegurar que os compradores de Huggies encontrem os tais dezenove produtos (Menconi, 1998).

3.4.3. Ciclo Virtuoso de mineração de dados

Berry e Linoff (1997) afirmam que o ciclo virtuoso de mineração de dados reconhece que mineração de dados é um passo num processo que requer ganho de conhecimento através do entendimento crescente dos consumidores, mercados, produtos e competidores para os processos internos. Este é um processo contínuo que traz resultados a toda hora.

O ciclo virtuoso de mineração de dados é composto por quatro estágios:

- identificação do problema do negócio;
- utilização de técnicas de mineração de dados para transformar dados em informações;
- ação a partir da informação;
- medição dos resultados.

O ciclo virtuoso de mineração de dados, conforme Harrison (1998) está demonstrado na Figura 5.

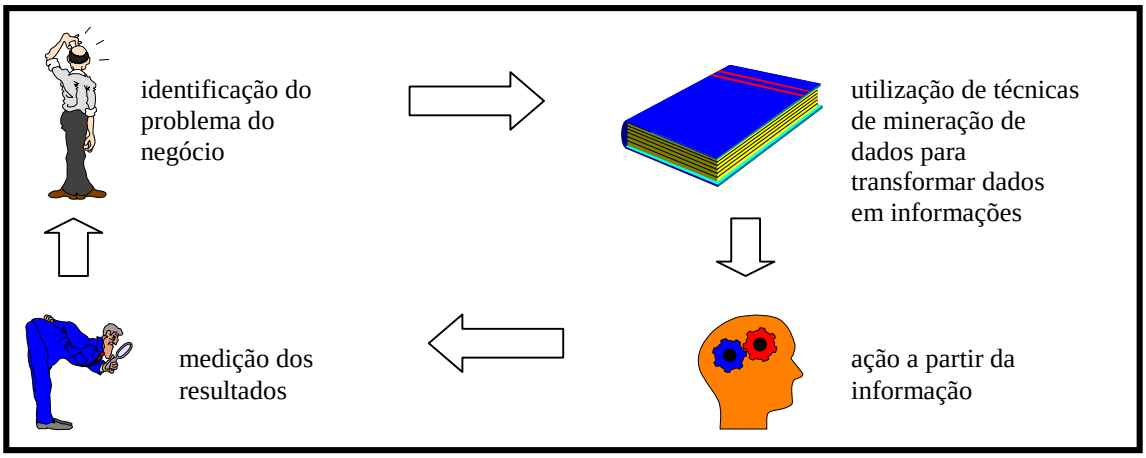

Figura 5 - Ciclo virtuoso de mineração de dados, adaptado de Harrison (1998)

3.4.4. Tarefas de mineração de dados

A mineração de dados pode desempenhar uma série limitada de tarefas, e apenas sob certas circunstâncias (Berry, 1997).

Harrison (1998) identificou seis tarefas de mineração de dados. São elas: classificação, estimativa, previsão, agrupamento por afinidade ou análise de cesta de supermercado, segmentação ou *clustering* e descrição.

a. Classificação

Segundo Berry e Linoff (Berry, 1997) a classificação é a tarefa mais comum de mineração de dados. Ela consiste em examinar os aspectos de um objeto e ligá-lo a uma das classes predefinidas. Para o nosso propósito, os objetos a serem classificados são geralmente representados por registros nos bancos de dados e a ação da classificação consiste em atualizar cada registro preenchendo um campo com o código da classe. Conforme os autores a classificação trata com valores discretos.

A tarefa de classificação é caracterizada por uma boa definição das classes, adquirida em um conjunto de exemplos pré-classificados (Harrison, 1998).

Alguns exemplos de classificação, segundo Harrison (1998):

- atribuir palavras-chave a artigos jornalísticos;
- classificar pedidos de créditos como de baixo, médio e alto risco;
- determinar que número telefônico corresponde ao fax;
- esclarecer pedidos de seguro fraudulentos.

b. Estimativa

Harrison (1998) afirma que a estimação trabalha com resultados contínuos. Dado algum dado de entrada, nós usamos a estimativa para estipular um valor a uma variável contínua desconhecida, tal como renda, altura ou limite de cartão de crédito.

Alguns exemplos de estimativa, segundo Harrison (1998):

- estimar o número de filhos em uma família;
- estimar a renda total de uma família;
- estimar o valor em tempo de vida de um cliente;
- estimar a probabilidade com que alguém responderá ao pedido de transferência de saldo.

c. Previsão

Previsão é o mesmo que classificação ou estimativa, exceto pelo fato de que os registros são classificados de acordo com algum comportamento futuro previsto ou valor futuro estimado (Harrison, 1998).

Na previsão a única forma de checar a acuracidade da classificação é esperar e ver (Berry, 1997).

Alguns exemplos de previsão, conforme Harrison (1998):

- previsão da quantia de dinheiro que um cliente utilizará caso seja oferecido a ele um certo limite de crédito;
- previsão de quais clientes sairão nos próximos seis meses;

- previsão de quais assinantes de telefone usariam um serviço extra, como segmentação por telefone, CPA, ou redirecionamento de ligação.

d. Agrupamento por afinidade ou análise da cesta de supermercado

Conforme Berry e Linoff (1997) a tarefa do agrupamento por afinidade é determinar quais são as coisas que vão juntas numa cesta ou carrinho de supermercado, por exemplo. Para eles o agrupamento por afinidade pode também ser usado para identificar oportunidades de vendas de pacotes de produtos ou de produtos e serviços.

As cadeias de varejo usam esta técnica para planejar a disposição dos produtos nas prateleiras das lojas ou em um catálogo, de modo que os itens geralmente adquiridos na mesma compra sejam vistos próximos entre si (Harrison, 1998).

e. Segmentação ou *Clustering*

Segmentação ou *clustering* é a tarefa de segmentar uma população heterogênea em um número maior de subgrupos homogêneos ou clusters. No *clustering* não há classes predefinidas (Berry, 1997).

Na segmentação os registros são agrupados de acordo com a semelhança. Depende de quem está analisando determinar qual o significado que cada um dos segmentos resultantes terá (Harrison, 1998).

A segmentação é freqüentemente um dos primeiros passos na análise de mineração de dados (FAQ, 1996).

f. Descrição

Segundo Berry e Linoff (1997), às vezes o objetivo da mineração de dados é simplesmente descrever o que está acontecendo em uma base de dados complicada no

intuito de aumentar nosso entendimento sobre as pessoas, produtos ou processos que produziram os dados.

Conforme Harrison (1998) a divergência de gênero na política americana é um exemplo de como uma simples descrição “o número de mulheres que apóiam os democratas é maior do que o de homens” pode provocar grande interesse e estudos por parte de jornalistas, sociólogos, economistas e cientistas políticos, sem contar os próprios candidatos.

3.4.5. Técnicas de mineração de dados

Há muitas técnicas diferentes de mineração de dados. A técnica a ser usada é determinada pelo tipo de informação que você está tentando determinar através dos dados. As técnicas devem ser aplicadas nas áreas corretas e nos dados corretos.

Berry e Linoff (1997) salientam que nenhuma técnica resolve todos os problemas de mineração de dados. A familiaridade com uma variedade de técnicas é necessária para encontrar o melhor caminho para resolver estes problemas.

Harrison (1998) identificou as seguintes técnicas de mineração de dados: análise de seleção estatística, MBR, algoritmos genéticos, detecção de agrupamentos, análise de vínculos, árvores de decisão e indução de regras e redes neurais artificiais.

a. Análise de seleção estatística

É uma forma de agrupamento usada para encontrar grupos de itens que tendem a ocorrer em conjunto em uma transação ou seleção estatística. Como técnica de agupamento, é útil quando desejamos saber quais itens ocorrem ao mesmo tempo ou em uma seqüência particular. A informação resultante pode ser usada para vários objetivos, como planejar a arrumação de lojas, criar “pacotes” de produtos, entre outros (Harrison, 1998).

b. Raciocínio baseado em memória (MBR)

O MBR (Memory Based Reasoning) ou raciocínio baseado em memória é uma técnica que usa exemplos conhecidos como modelo para fazer previsões sobre exemplos desconhecidos (Harrison, 1998).

Berry e Linoff (1997) dizem que o MBR procura os vizinhos mais próximos nos exemplos conhecidos e combina seus valores para atribuir valores de classificação ou de previsão.

Segundo Harrison (1998) uma das maiores vantagens do MBR é a habilidade de ser executado em virtualmente qualquer fonte de dados, mesmo sem modificações. Para ele os dois elementos-chave do MBR são a função de distância usada para encontrar os vizinhos mais próximos para fazer uma previsão. O autor identifica outra vantagem principal do MBR: sua habilidade de aprender sobre novas classificações simplesmente introduzindo novos exemplos no banco de dados.

c. Algoritmos Genéticos

Os algoritmos genéticos aplicam mecanismos de seleção genéticos e naturais para uma busca usada para encontrar conjuntos de parâmetros ótimos que descreve uma função preditiva. É usado para mineração de dados direta (Berry, 1997).

Harrison (1998) diz que os algoritmos genéticos usam os operadores de seleção, cruzamento e mutação para desenvolver sucessivas gerações de soluções.

d. Detecção de agrupamentos

Harrison (1998) define esta técnica como a construção de modelos que encontram registros de dados semelhantes. Segundo o mesmo essas reuniões por semelhança são chamadas grupos (clusters).

Segundo Berry e Linoff (1997) a detecção de agrupamentos trata-se de mineração de dados indireto, uma vez que a meta é encontrar similaridades não conhecidas previamente.

Reunir dados é uma boa maneira de começar qualquer análise. Agrupar por semelhança pode fornecer o ponto de partida para saber o que há nos dados e descobrir como usá-los melhor (Harrison, 1998).

e. Análise de vínculos

Harrison (1998) afirma que a análise de vínculos segue as relações entre registros para desenvolver modelos baseados em padrões nas relações.

O mesmo autor salienta que como técnica de mineração de dados, a análise de vínculos não é muito compatível com a tecnologia de bancos de dados relacionais. A maior área onde é aplicada, segundo ele, é a área policial, onde pistas são ligadas entre si para solucionar crimes.

Conforme Berry e Linoff (1997) as poucas ferramentas disponíveis enfocam com maior freqüência a visualização dos vínculos ao invés de analisar os padrões.

f. Árvores de decisão e indução de regras

As árvores de decião são usadas para a mineração de dados direta (Harrison, 1998), particularmente para a classificação.

Berry e Linoff (1997) dizem que as árvores de decisão são um modelo poderoso produzido por uma classe de técnicas que inclui árvores de regressão e de classificação e indução qui-quadrado automática.

Harrison (1998) identificou como uma das principais vantagens das árvores de decisão a facilidade de explicação de seu modelo, devido a sua forma de regras explícitas.

g. Redes Neurais

Harrison (1998) diz que as redes neurais são provavelmente a técnica de mineração de dados mais comum, talvez sinônimo de mineração de dados para algumas pessoas.

Segundo Almeida (1995) as redes neurais têm sua origem em pesquisas neurológicas, e seu modelo de base é o cérebro humano. Como no cérebro humano, as redes neurais possuem neurônios interconectados de modo que os dados os percorram. Esses neurônios transmitem informação através de sinapses ou conexões.

O conceito-chave das redes neurais é a utilização de dados na criação de bases de conhecimentos. As redes neurais, ao contrário dos sistemas especialistas não precisam de um especialista para a criação da sua base de conhecimentos. Não trabalha com regras, sua aquisição é feita automaticamente a partir de exemplos coletados em bancos de dados (Almeida,1995).

Conforme Kotler (1998) o software de redes neurais, projetado conforme os padrões das células do cérebro humano pode, realmente, "aprender" a partir de grandes conjuntos de dados. Ao examinar repetidamente milhares de registros de dados, o software pode desenvolver um modelo estatístico poderoso descrevendo os relacionamentos e os padrões de dados importantes - nada que um pesquisador humano tenha tempo (ou capacidade visual) de fazer de maneira rigorosa e consistente.

Nas redes neurais não há uma codificação de programas a fim de introduzir o conhecimento sobre um problema. Por um processo iterativo (processo de aprendizado) as redes neurais lêem os exemplos fornecidos sobre um problema e criam assim um modelo de resolução. Elas são bem adaptadas a dois tipos de tarefas: reconhecimento de formas e generalização (Almeida, 1995).

Kotler (1998) afirma que a IBM desenvolveu um conjunto de seis programas de computador chamados ‘*Data Mining’*, que pode analisar imensos conjuntos de dados e revelar conglomerados, relacionamentos e assim por diante. Usando o ‘*Data Mining’*, a empresa de venda por catálogo Land's End ficou em condições de identificar cerca de 5.200 segmentos de consumidores, baseados em diferentes padrões de compra. Como resultado, a Land's End passou a melhorar o direcionamento de suas listagens de nomes

de consumidores para os segmentos de maior probabilidade de interesse por suas ofertas.

A vantagem dos sistemas especialistas é que a maioria deles não exige computadores grandes e poderosos . Dentro dos últimos cinco anos, as redes neurais e a tecnologia de inteligência artificial tornaram-se, finalmente, adaptáveis aos microcomputadores. A Nielsen é uma das grandes empresas de pesquisa de marketing que está desenvolvendo o seu próprio sistema especialista para uso em microcomputadores. Um de seus produtos mais novos, o Spotlight, ajuda as empresas a determinar suas participações de mercado em uma fração do tempo anteriormente necessário. "Estamos falando de minutos em vez de dias", afirma o diretor de vendas e serviços aos clientes da Nielsen (Kotler, 1998).

Kotler (1998) afirma que os vendedores, que obtiveram grandes benefícios da automação de vendas em anos recentes, têm também reduzido semanas inteiras no processo de vendas usando o poder dos sistemas especialistas. Na Wells Fargo Alarm Services, era norma os vendedores demorarem até dez dias para fechar um negócio, desde a preparação das propostas até o acompanhamento pós-venda. Hoje, cada um dos cinqüenta gerentes e 185 vendedores da empresa usam um microcomputador IBM ThinkPad equipado com sistema especialista. Em vez de retornarem ao escritório após uma visita, os vendedores da Wells Fargo ligam seus micros e apertam uma tecla. O programa do computador começa a fazer perguntas relacionadas às necessidades dos consumidores, o sistema prepara, automaticamente, uma fatura de materiais, uma proposta de preço e o contrato de venda, tudo em menos de vinte minutos.

3.4.6. Que técnica utilizar?

Harrison (1998) diz que não há uma técnica que resolva todos os problemas de mineração de dados. A familiaridade com as técnicas é necessária para proporcionar a melhor abordagem de acordo com os problemas apresentados. O autor afirma que a escolha das técnicas de mineração de dados dependerá da tarefa específica a ser executada e dos dados disponíveis para análise.

A técnica a ser escolhida dependerá do tipo de dados que temos e do tipo de informação que estamos tentando determinar (FAQ, 1996).

3.5.7. Metodologia de mineração de dados

Segundo Berry e Linoff (1997) há três variações de metodologia para mineração de dados. Os autores apresentam cada uma das variações passo a passo:

a. Teste de hipóteses

1º. Criar hipóteses;
2º. Definir os dados necessários para testar as hipóteses;
3º. Alocar os dados;
4º. Preparar os dados para análise;
5º. Desenhar os modelos computacionais e as questões dos bancos de dados para confrontar as hipóteses com os dados;
6º. Avaliar os resultados dos modelos e das questões;
7º. Agir baseado nos resultados da mineração de dados;
8º. Medir os efeitos das ações tomadas;
9º. Reiniciar o processo de mineração de dados tirando vantagem dos novos dados gerados a partir das ações tomadas.

b. Descoberta de conhecimento direto

1º. Identificar fontes de dados preclassificados;
2º. Preparar os dados para análise;
3º. Selecionar técnicas apropriadas de descoberta de conhecimento baseado em características dos dados e na meta da mineração de dados;
4º. Dividir os dados em conjuntos de formação, teste e avaliação;

5º. Usar o conjunto de dados de formação para construir um modelo computacional;
6º. Afinar o modelo aplicando-o ao conjunto de dados de teste;
7º. Avaliar a acuracidade do modelo aplicando-o ao conjunto de dados de avaliação;
8º. Agir baseado nos resultados da mineração de dados;
9º. Medir o efeito das ações tomadas;
10º. Reiniciar o processo de mineração de dados tirando vantagem dos novos dados gerados através das ações tomadas.

c. Descoberta de Conhecimento Indireto

1º. Identificar fontes de dados disponíveis;
2º. Preparar os dados para análise;
3º. Selecionar técnicas apropriadas de descoberta de conhecimento indireto baseado em características dos dados e na meta da mineração de dados;
4º. Usar a técnica selecionada para descobrir estruturas escondidas nos dados;
5º. Identificar alvos potenciais para a descoberta de conhecimento indireto;
6º. Gerar novas hipóteses a serem testadas.

3.4.8. Aplicações

A mineração de dados tem se mostrado extremamente valioso em um grande número de aplicações, incluindo segmentação de mercado, detecção de fraude em cartões de crédito, análises financeiras e de investimentos, detecção e predição de erros em grandes negócios, análise de informações, ferramentas inteligentes e limpeza em bases de dados (Greenfeld, 1996).

Feldens (199?) apresentou algumas das aplicações atuais para a mineração de dados, além das citadas acima, são elas: marketing e melhoria do processo industrial.

É necessário entender como acontece o processo de tomada de decisão, saber que decidimos baseados em informações, que temos uma racionalidade limitada, o que é mineração de dados, para que serve, etc.

Através dos conhecimentos teóricos obtidos estaremos aptos a determinar em que situações, ou seja, para que tipo de tomada de decisão, poderemos utilizar as técnicas e ferramentas de mineração de dados e que técnica pode ser utilizada em cada uma destas situações, assim poderemos aplicar estas ferramentas nas bases de dados disponíveis e dar início ao nosso trabalho de pesquisa.

4. Método de pesquisa

Este trabalho enquadra-se como um estudo de caso. Yin (apud Oliveira, 1999) diz que o estudo de caso é um tipo de pesquisa empírica que investiga um fenômeno contemporâneo dentro de seu contexto de vida real, especialmente quando os limites entre o fenômeno e o contexto não estão claramente evidentes.

Em geral o estudo de caso é a estratégia preferida quando questões do tipo "como" e "por que" são colocadas, quando o investigador tem pouco controle sobre os eventos e quando o foco é em um fenômeno contemporâneo entre algum contexto da vida real (Yin, 1994).

Benbasat et tal. (1987) afirmam que o estudo de caso múltiplo é necessário quando a intenção da pesquisa é descrição, construção ou teste de teoria. Nesta pesquisa a intenção é a descrição da aplicação das técnicas de mineração de dados como fonte de informação para a tomada de decisão.

Yin (apud Campomar, 1991) apresenta uma explicação de como os estudos de casos podem ser feitos onde, depois de definir-se claramente o problema a ser pesquisado, deixando claro que o uso do estudo de casos é estratégia adequada para resolver este problema, deverá (ão) ser:

- desenhada a estrutura de coleta de dados e a apresentação das perguntas principais, decidindo-se por um único ou múltiplos casos;
- definido pela conveniência e oportunidade e não para aumentar a possibilidade de inferências.
- decidido se o estudo será de natureza global, abrangendo todos os elementos do caso como um todo, ou de natureza encaixada, abrangendo vários níveis dentro do caso.
- preparado um protocolo relacionando as atividades a serem realizadas e os procedimentos a serem seguidos.
- determinados os instrumentos para a coleta de dados, os quais, normalmente, poderiam ser literatura, documentos de arquivo, entrevistas (com decisão sobre estrutura e disfarce), observação (participativa ou não), experiências e, mesmo, artefatos.

- feitas as análises principalmente por analogias, contendo comparações com teorias, modelos e outros casos.
- as conclusões deverão ser específicas, com possíveis inferências e explicações permitindo que as generalizações sejam usadas como base para novas teorias e modelos.
- claramente expostas as limitações gerais inerentes ao método e as específicas que aparecem em cada pesquisa.

A pesquisa será composta por três casos porque este é um número suficiente para verificarmos a aplicação das técnicas de mineração de dados, tendo em vista que serão três bases de dados diferentes e de empresas com clientes diferentes.

O foco será empresas do setor de supermercados devido à riqueza de seus dados e ao fato deste setor viver em forte competição. As bases de dados a serem utilizadas serão aquelas que guardam a movimentação diária realizada pelos clientes, pois para minerar dados, é necessário arquivar toda e qualquer transação realizada pelo cliente (Menconi, 1998).

Os entrevistados serão os tomadores de decisão das empresas que compõem a amostra.

A empresa que formará a amostra será escolhida por conveniência, em função da disponibilidade de sua(s) base(s) de dados.

Serão realizadas pesquisas bibliográficas e na Internet para a aquisição do conhecimento necessário sobre mineração de dados, suas técnicas e ferramentas existentes.

Serão contactados supermercados para a obtenção da(s) base(s) de dados a serem submetidas à aplicação das técnicas de mineração de dados.

A escolha da(s) ferramenta(s) a ser(em) utilizada(s) também será por conveniência, dependendo da sua disponibilização por parte de seus proprietários ou a existência de softwares shareware ou freeware.

Após escolhidas a amostra e a(s) ferramenta(s) que serão utilizadas se dará o processo de aplicação desta(s) ferramenta(s) sobre as bases de dados disponíveis. A metodologia de mineração de dados que será aplicada é a que Berry e Linoff (1997) chamaram de descoberta de conhecimento indireto, visto que serão primeiramente identificadas a(s) base(s) de dados disponíveis, depois os dados serão preparados para a

análise, serão e scolhidas as técnicas apropriadas, estas técnicas serão utilizadas e no final gerarmos e testarmos novas hipóteses.

A aplicação das ferramentas de mineração de dados poderão ocorrer nas instalações da empresa ou fora dela, dependendo da vontade do responsável pela empresa.

O passo posterior será a avaliação das informações obtidas através da etapa anterior.

De posse dos dados obtidos os mesmos serão apresentados aos tomadores de decisão da empresa que faz parte da amostra desta pesquisa, e será realizada uma entrevista com os mesmos, no intuito de possibilitar a verificação do grau de satisfação destas pessoas com relação às informações geradas/descobertas por esta aplicação.

5. Contexto de aplicação

Segundo a Abras (Associação Brasileira de Supermercados) não há outro setor da atividade econômica no País que tenha crescido tanto, do zero ao estágio atual, em prazo tão curto. Em mais de quatro décadas, o auto-serviço brasileiro impôs-se como a forma mais moderna, econômica e racional de se adquirir uma infinidade de produtos.

Os supermercados conquistaram a condição de maiores abastecedores de alimentos e artigos de higiene e limpeza longe de quaisquer subsídios ou favores oficiais, graças a uma vocação para o crescimento impulsionada por investimentos contínuos e crença no desenvolvimento do Brasil (Abras, 199?).

Aqui será descrito o contexto do supermercado que faze parte da amostra dessa pesquisa, ou seja, da empresa dona da(s) base(s) de dados que será(ão) utilizada(s). Como a mesma ainda não está definida torna-se impossível o preenchimento deste item, no momento.

As contribuições potenciais dessa pesquisa são as seguintes:

- Servir de fonte para consultas sobre a descoberta do conhecimento em bases de dados, pois este ainda é um assunto novo e, portanto, possui pouca bibliografia, principalmente em português;
- A contribuição principal dessa pesquisa para o mestrando é o aprimoramento de seus conhecimentos sobre sistemas de informações e tomada de decisão, que é a sua área de interesse. Outra contribuição importante é o contato com novas ferramentas de alta tecnologia que podem facilitar e melhorar a descoberta de informações importantes;
- Para as empresas que farão parte da amostra desta pesquisa serão fornecidas as informações descobertas, que estavam escondidas em suas bases de dados e que, portanto, não estavam acessíveis aos tomadores de decisão, e que após este trabalho poderão ser utilizadas para obter vantagem competitiva;
- Esta pesquisa proporcionará o conhecimento das tarefas, técnicas e ferramenta(s) que auxiliam na busca de informações relevantes nas grandes quantidades de dados que as empresas estão armazenando, bem

como exemplos de informações que só podem ser obtidas através desta aplicação.

6. Cronograma

Atividades	Meses												
	Dez	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1. Aprovação do projeto	X												
2. Definição da amostra	X	X											
3. Revisão de literatura	X	X	X	X	X	X	X						
4. Obtenção das bases de dados	X	X	X	X	X	X	X						
5. Obtenção da(s) ferramenta(s)	X	X	X	X	X	X	X						
6. Estudo da(s) ferramenta(s)		X	X	X	X	X	X						
7. Aplicação da(s) ferramenta(s)			X	X	X	X	X	X					
8. Definição dos procedimentos de análise dos dados			X	X	X	X	X	X					
9. Tabulação dos dados				X	X	X	X	X					
10. Análise dos dados				X	X	X	X	X					
11. Elaboração do Instrumento para a entrevista				X	X	X	X	X					
12. Demonstração dos resultados						X	X	X	X	X			
13. Realização da entrevista						X	X	X	X	X			
14. Análise dos dados da entrevista							X	X	X	X	X	X	
15. Relatório					X	X	X	X	X	X	X	X	X
16. Defesa da dissertação													X

8. Referências bibliográficas

ABRAS (Associação Brasileira de Supermercados) [25 de novembro de 1999] Disponível na World Wide Web <http://www.abrasnet.com.br>.

ALBERTIN, Luiz. **Administração de informática: funções e fatores críticos de sucesso**. São Paulo: Atlas, 1998.

ALTER, S. **Information systems: a managerial perspective**. Menlo Park. CA: Benjamin e Cummings, 2ª ed. 1996.

ALMEIDA, Fernando C. **Desvendando o uso de redes neurais em problemas de administração de empresas**. RAE, São Paulo, v. 35, n. 1, p. 46-55, Jan./Fev. 1995.

BENBASAT, Izak, GOLDSTEIN, David, MEAD, Melissa. **The case research strategy in studies of information systems**. Mis quarterly, Sep. 1987.

BERRY, Michael J. A., LINOFF, Gordon. **Mineração de dados techniques: for marketing, sales and customer support**. USA: Wiley Computer Publishing, 1997.

BRUSSO, Marcos José. **O paralelismo na mineração de regras de associação**. Porto Alegre: UFRGS, 1998 (Trabalho Individual I, Programa de Pós-Graduação em Computação).

CAMPOMAR, Marcos C. **Do uso de "estudo de caso" em pesquisas para dissertações e teses em administração**. RAE, São Paulo, v. 26, n. 3, p. 95-97, Jul./Set. 1995.

FAQ (Frequently Asked Questions). mineração de dados, 1996. [20 de dezembro de 1998] Avaiable from World Wide Web <http://www.rpi.edu/faq.html>.

FELDENS, Miguel Artur. **Knowledge discovery in databases**. [20 de dezembro de 1998] Avaiable from World Wide Web <http://www.ufrgs.br/~feldens>

_____, MORAES, Rodrigo Leal, PAVAN, Altino, CASTILHO, José Mauro Volkmer. **Mineração de dados na gestão hospitalar**. Porto Alegre: UFRGS. [20 de dezembro de 1998] Disponível na World Wide Web <http://www.inf.ufrgs.br/~feldens/datamining.html>.

FIGUEIRA, Rafael. **Mineração de dados e bancos de dados orientados a objetos**. Rio de Janeiro: UFRJ, 1998 (Dissertação, Mestrado em Ciência da Computação).

FREEDMAN, Alan. **Dicionário de informática: o guia ilustrado completo**. São Paulo: Makron Books, 1995.

FREITAS, Henrique, LESCA, Humbert. **A inovação e a informação: ser competitivo na era do conhecimento... também no Brasil**. Análise, Porto Alegre, v.3,nº 2, 1992.

_____, BECKER, João Luiz, KLADIS, Constantin Metaxa, HOPPEN, Norberto. **Informação e decisão: sistemas de poio e seu impacto**. Porto Alegre: Ortiz, 1997.

GREENFELD, Norton. **Mineração de dados**. Unix Review, p. 9-14, may. 1996.

HARRISON, Thomas H. **Intranet *data warehouse***. São Paulo: Bekerley Brasil, 1998.

KOTLER, Philip. **Administração de marketing: análise, planejamento, implementação e controle**. 5ª ed. São Paulo: Atlas, 1998.

MENCONI, Darlene. **A mineração de informações**. Info Exame. São Paulo, Ano 12, nº 144, p. 98-93, mar. 1998.

NEWING, Rod. **Mineração de dados**. Management Accounting. p. 34-35. oct. 1996.

OLIVEIRA, Djalma de Pinho Rebouças. **Sistemas de informações gerenciais: estratégicas, táticas operacionais**. 4ª ed. São Paulo: Atlas, 1997

OLIVEIRA, Mirian. **Um método para obtenção de indicadores visando a tomada de decisão na etapa de cencepção do processo construtivo: a percepção dos principais intervenientes**. Porto Alegre: UFRGS, Dissertação de Mestrado, PPGA/EA, 1999.

PALADINO, Enzo. **Novo dicionário de informática**. Rio de Janeiro: Ciência Moderna Computação, 1986.

SULIMAN JR., Alberto, SOUZA, Jano Moreira. **Prospecção de Conhecimento em Bancos de Dados**. Developers Magazine, Rio de Janeiro, Ano1, Nº 6, p. 38-39,fev. 1997.

TORRES, Norberto. **Competitividade empresarial com a tecnologia da informação**. São Paulo: Makron Books, 1995.

YIN, Robert K. **Case study research: design and methods**. Second edition. Vol. 5. Sage Publications, 1994.