

**Universidade Federal de Santa Catarina**  
**Programa de Pós-Graduação em**  
**Engenharia de Produção**

UM MODELO REPRESENTATIVO DE CONHECIMENTO  
PARA APLICAÇÃO DA MINERAÇÃO DE DADOS NO  
CADASTRO TÉCNICO URBANO

Dissertação de Mestrado

André Fabiano de Moraes

Florianópolis

2003

**UM MODELO REPRESENTATIVO DE CONHECIMENTO  
PARA APLICAÇÃO DA MINERAÇÃO DE DADOS NO  
CADASTRO TÉCNICO URBANO**

**Universidade Federal de Santa Catarina**  
**Programa de Pós-Graduação em**  
**Engenharia de Produção**

UM MODELO REPRESENTATIVO DE CONHECIMENTO  
PARA APLICAÇÃO DA MINERAÇÃO DE DADOS NO  
CADASTRO TÉCNICO URBANO

**André Fabiano de Moraes**

Dissertação apresentada ao  
Programa de Pós-Graduação em  
Engenharia de Produção da  
Universidade Federal de Santa Catarina  
como requisito parcial para obtenção  
do título de Mestre em  
Engenharia de Produção

Florianópolis

2003

André Fabiano de Moraes

UM MODELO REPRESENTATIVO DE CONHECIMENTO  
PARA APLICAÇÃO DA MINERAÇÃO DE DADOS NO  
CADASTRO TÉCNICO URBANO

Esta dissertação foi julgada e aprovada para a  
obtenção do título de **Mestre em Engenharia de  
Produção no Programa de Pós-Graduação em  
Engenharia de Produção** da  
Universidade Federal de Santa Catarina

Florianópolis, 21 de fevereiro de 2003.

Prof. Edson Pacheco Paladini, Dr.  
Coordenador do Curso

Banca Examinadora

---

Profª. Anita Maria da Rocha Fernandes, Dra.  
Universidade do Vale do Itajaí

---

Profª. Lia Caetano Bastos, Dra.  
Universidade Federal de Santa Catarina

**Orientadora**

---

Prof. Rogério Gonçalves Bittencourt, M.Eng.  
Universidade do Vale do Itajaí

---

Profª. Ana Benciveni Franzoni, Dra.  
Universidade Federal de Santa Catarina

Dedicatória  
A Deus pela oportunidade de aprender e viver.  
Em especial a Julia Rosa Vieira.  
Á todos os familiares.

## ***Agradecimentos***

À Universidade Federal de Santa Catarina,  
À Coordenação de Aperfeiçoamento  
de Pessoal de Nível Superior CAPES,  
À orientadora Prof. Lia Caetano Bastos,  
pelo acompanhamento pontual e competente,  
Aos professores do Curso de Pós-Graduação e  
À Prefeitura Municipal de Itajaí pela  
credibilidade retratada nesta pesquisa.

A todos que direta ou indiretamente  
contribuíram para a realização  
desta pesquisa.

*“Todo ser humano deve buscar o desconhecido.  
Pois será ignorante aquele que tiver medo e  
não procurar conhecer os mistérios da vida.”*

Albert Einstein

# SUMÁRIO

<b>LISTA DE FIGURAS.....</b>	<b>X</b>
<b>LISTA DE QUADROS .....</b>	<b>XI</b>
<b>LISTA DE ABREVIATURAS .....</b>	<b>XII</b>
<b>RESUMO .....</b>	<b>XIII</b>
<b>ABSTRACT .....</b>	<b>XIV</b>
<b>1 INTRODUÇÃO .....</b>	<b>1</b>
<b>1.1 Objetivos .....</b>	<b>2</b>
1.1.1 Objetivo Geral .....	2
1.1.2 Objetivos Específicos .....	2
<b>1.2 Justificativa.....</b>	<b>2</b>
<b>1.3 Limitação do Trabalho .....</b>	<b>3</b>
<b>1.4 Estrutura do Trabalho .....</b>	<b>3</b>
<b>2 CADASTRO TÉCNICO URBANO .....</b>	<b>4</b>
<b>2.1 Conceitos de Cadastro.....</b>	<b>4</b>
<b>2.2 Cadastro Técnico Multifinalitário Urbano.....</b>	<b>5</b>
<b>2.3 Objetivos do Cadastro Técnico Urbano .....</b>	<b>8</b>
<b>2.4 Vantagens do Cadastro Técnico Urbano.....</b>	<b>10</b>
<b>2.5 Desvantagens do Cadastro Técnico Urbano.....</b>	<b>11</b>
<b>3 SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS.....</b>	<b>12</b>
<b>3.1 Introdução .....</b>	<b>12</b>
<b>3.2 Conceitos de SIG.....</b>	<b>13</b>
<b>3.3 Conceitos de Espaço e Relações Espaciais .....</b>	<b>14</b>
<b>3.4 Utilização do SIG na Organização.....</b>	<b>15</b>
3.4.1 SIG no Nível Operacional.....	16
3.4.2 SIG no Nível Gerencial.....	17
3.4.3 SIG no Nível Estratégico .....	19
<b>4 MINERAÇÃO DE DADOS .....</b>	<b>22</b>
<b>4.1 Introdução .....</b>	<b>22</b>
<b>4.2 Níveis de Informação.....</b>	<b>22</b>
<b>4.3 Origem da Mineração de Dados .....</b>	<b>24</b>
<b>4.4 Mineração de Dados: um breve histórico.....</b>	<b>25</b>
<b>4.5 Conceitos de Mineração de Dados .....</b>	<b>26</b>
<b>4.6 Aplicabilidade .....</b>	<b>27</b>
<b>4.7 Descoberta de Conhecimento em Banco de dados (KDD) .....</b>	<b>27</b>
4.7.1 Origem do KDD e suas Fases .....	28
4.7.2 KDD e Data Mining.....	32
<b>4.8 Métodos e Técnicas Envolvidas na Mineração de Dados.....</b>	<b>32</b>
4.8.1 Técnicas de Inteligência Artificial.....	33
4.8.2 Técnicas Estatísticas.....	33
4.8.3 Indução de Regras .....	34

4.8.4	Regras de Associação.....	34
4.8.5	Regras de Classificação.....	35
4.8.6	Regras de Agrupamento (Clustering).....	37
<b>4.9</b>	<b>Critérios para Escolha da Ferramenta de Mineração de Dados .....</b>	<b>38</b>
<b>5</b>	<b>MODELO PARA UTILIZAÇÃO DA MINERAÇÃO DE DADOS.....</b>	<b>40</b>
<b>5.1</b>	<b>Introdução .....</b>	<b>40</b>
<b>5.2</b>	<b>Etapas de Aplicação.....</b>	<b>41</b>
5.2.1	Primeira Etapa – Aquisição da Base de Dados.....	41
5.2.2	Segunda Etapa – Aplicação de um Modelo Estatístico .....	42
5.2.3	Terceira Etapa – Pré-Processamento da Base de Dados .....	42
5.2.4	Quarta Etapa – Geração de Modelos.....	42
5.2.5	Quinta Etapa - Aplicação das Técnicas e das Ferramentas.....	42
5.2.6	Sexta Etapa – Apresentação e Interpretação dos Dados.....	43
<b>6</b>	<b>ESTUDO DE CASO.....</b>	<b>44</b>
<b>6.1</b>	<b>Introdução .....</b>	<b>44</b>
<b>6.2</b>	<b>Aplicação do Modelo Proposto na Base Cadastral .....</b>	<b>44</b>
6.2.1	Aquisição e Seleção da Base de Dados.....	44
6.2.2	Aplicação de um Modelo Estatístico Preliminar.....	45
6.2.3	Pré-Processamento dos Dados.....	47
6.2.4	Geração do Modelo e a Escolha da Ferramenta Utilizada .....	48
6.2.4.1	Weka 3.2.....	48
6.2.4.2	SQL Server 2000 .....	50
6.2.4.3	SIPINA .....	51
6.2.5	Aplicação das Técnicas e Ferramentas.....	52
6.2.6	Apresentação e Interpretação dos Dados .....	54
6.2.6.1	Representação Espacial .....	54
6.2.6.2	Representação do Conhecimento com SQL Server 2000 .....	55
6.2.6.3	Representação do Conhecimento SIPINA x WEKA.....	55
<b>7</b>	<b>CONCLUSÕES E RECOMENDAÇÕES.....</b>	<b>59</b>
<b>7.1</b>	<b>Conclusões .....</b>	<b>59</b>
<b>7.2</b>	<b>Recomendações .....</b>	<b>60</b>
	<b>BIBLIOGRAFIA .....</b>	<b>61</b>
	<b>ANEXOS .....</b>	<b>65</b>

## Lista de Figuras

Figura 1:	Relacionamento ilustrativo entre os sistemas.....	6
Figura 2:	Sistemática de atualização.....	7
Figura 3:	Interação dos diversos segmentos públicos com a administração municipal.....	8
Figura 4:	Tecnologias de Geoprocessamento.....	12
Figura 5:	Administração, controle e implantação do SIG.....	16
Figura 6:	Banco de dados representados em níveis de implementação.....	23
Figura 7:	Relacionamento multidisciplinar.....	25
Figura 8:	KDD.....	28
Figura 9:	O processo KDD.....	29
Figura 10:	Esquema resumido das técnicas da estatística.....	34
Figura 11:	Árvore de decisão conceitual.....	36
Figura 12:	Exemplo prático de uma árvore de decisão.....	37
Figura 13:	Dendograma utilizando a distância euclidiana.....	38
Figura 14:	Modelo proposto para atividades da mineração de dados.....	41
Figura 15:	Demonstração quantitativa do BD com relação ao acabamento dos imóveis.....	46
Figura 16:	Estado de conservação dos imóveis.....	47
Figura 17:	Avaliação da base de dados.....	48
Figura 18:	Ferramenta selecionada.....	49
Figura 19:	Formato do arquivo arff.....	50
Figura 20:	<i>Software</i> SQLServer 2000.....	51
Figura 21:	<i>Software</i> SIPINA for Windows.....	51
Figura 22:	Processo realizado pela ferramenta Weka utilizando algoritmo ID3.....	53
Figura 23:	Representação espacial inicial dos imóveis cadastrados no BD, caracterizados com boa conservação.....	54
Figura 24:	Representação espacial descoberta após a mineração dos dados, para os imóveis que apresentam boa conservação, com base nas regras.....	54
Figura 25:	Árvore de decisão montada pela ferramenta SQLServer 2000.....	55
Figura 26:	Árvore construída pela ferramenta SIPINA e as respectivas regras.....	57

## Lista de Quadros

Quadro 1:	Exemplo para Análise Espacial.....	15
Quadro 2:	Benefícios e Atividades do Nível Operacional.....	17
Quadro 3:	Benefícios e Atividades do Nível Gerencial.....	19
Quadro 4:	Benefícios e Atividades do Nível Estratégico.....	21
Quadro 5:	Exemplo da Padronização de Dados.....	31
Quadro 6:	Atributos do Banco de Dados Analisado.....	45
Quadro 7:	Algoritmos Identificados.....	49
Quadro 8:	Experiência Sipina X Weka.....	58

## Lista de Abreviaturas

BD	- Banco de Dados
BIC	- Boletim de Informações Cadastrais
CTMU	- Cadastro Técnico Multifinalitário Urbano
CTU	- Cadastro Técnico Urbano
DM	- <i>Data Mining</i>
DSS	- <i>Decision Support System</i>
DW	- <i>Data Warehouse</i>
EIS	- <i>Executive Information System</i>
GB	- <i>Giga Bytes</i>
GIS	- <i>Geographi Information Systems</i>
IA	- Inteligência Artificial
IBGE	- Instituto Brasileiro de Geografia e Estatística
JDBC	- <i>Java Database Connectivity</i>
KDD	- <i>Knowledge Discovery in Databases</i>
ODBC	- <i>Open Data Base Connectivity</i>
OLAP	- <i>On-line Analytic Processing</i>
OLTP	- <i>On-line Transaction Processing</i>
RDBMS	- <i>Relational Data Base Manager System</i>
SAD	- Sistema de Apoio a Decisão
SADs	- Sistemas de Apoio a Decisão
SGBDR	- Sistema Gerenciador de Banco de Dados Relacional
SGBDs	- Sistemas Gerenciadores de Bancos de Dados
SIG	- Sistema de Informação Geográfica
SQL	- <i>Structured Query Language</i>
TB	- <i>Tera Bytes</i>
TDIDT	- <i>Top Down Induction of Decision Trees</i>

## Resumo

A grande quantidade de informações existentes em um cadastro técnico urbano pode dificultar a atuação da administração municipal, se não for elaborada e selecionada adequadamente, de forma a permitir análises estratégicas para resolver os problemas de âmbito municipal. Novas técnicas computacionais que permitem a extração de conhecimento a partir de uma massa de dados têm surgido nos últimos anos, dentre elas a mineração de dados. Segundo Witten e Frank et. al. (1999), o nome *Data Mining* surgiu da semelhança entre a procura de informação importante para o mundo dos negócios escondidas em uma grande base de dados. O propósito deste trabalho é aplicar os métodos da mineração de dados em bases do cadastro técnico urbano, objetivando a descoberta de conhecimento de modo a auxiliar a administração municipal. A busca de novos padrões e relacionamentos entre diferentes variáveis utilizando algoritmos de mineração de dados, além das tradicionais ferramentas estatísticas, demonstra a conduta multidisciplinar entre diferentes áreas de atuação, proporcionando novos mecanismos para descoberta de conhecimento.

**Palavras-chave:** Mineração de Dados, Banco de Dados, Sistemas de Informações Geográficas e Cadastro Técnico Urbano.

## Abstract

*The great amount of existent information in an urban technical cadaster can hinder the performance of the municipal administration, if it be not elaborated and selected appropriately, in way to allow strategic analyses to solve the problems of municipal ambit. New technical computations that allow the knowledge extraction starting from a mass of data have been appearing in the last years, among them the data mining. According to Witten and Frank et. al. (1999), the name Data Mining it appeared of the likeness among the search of important information for the world of the business hidden in a great base of data. The purpose of this work is to apply the methods of the data mining in bases of the urban technical cadaster, objectifying the discovery of way knowledge to aid the municipal administration. The search of new patterns and relationships among different variables using algorithms of data mining, besides the traditional statistical tools, demonstrates the conduct several disciplines among different areas of performance, providing new mechanisms for knowledge discovery.*

**Key-words:** Data Mining, Database, Geographic Information Systems, Urban Technical Cadastre.

# 1 INTRODUÇÃO

O grande volume de informações operacionais e táticas pode ser prejudicial para administração municipal, se não forem elaboradas e selecionadas adequadamente, de forma a permitir análises estratégicas para resolver os problemas de âmbito municipal. Para que isto ocorra, identifica-se a necessidade de um estudo que encontre indicadores úteis ao planejamento. Sendo que, o cadastro técnico urbano por sua vez tem o papel fundamental de compor e compartilhar informações que alimente um sistema de informação estratégico municipal.

A administração, assim como os responsáveis pelo planejamento, necessitam de informações filtradas e tratadas em nível operacional e gerencial, de forma que possibilitem fazer análises rápidas e precisas por parte dos órgãos responsáveis pelo desenvolvimento de estratégias municipais. Para alcançar o conhecimento dos fatos e, de um modo geral, obter um planejamento detalhado baseado no cadastro técnico urbano, o cruzamento de informações se faz necessário, gerando indicadores que facilitam estrategicamente as ações. A partir deste momento, a aplicação de algoritmos de mineração de dados abre novos caminhos que possibilitam a descoberta de regras apontando conhecimento até então invisível.

A confiabilidade sobre os indicadores apontados depende dos recursos disponíveis para coleta dos dados e da metodologia aplicada. A busca por gestões estratégicas devem assegurar-se em dados, e processamento de informações que levem à estratégias preventivas, corretivas, preditas ou prospectivas.

A utilização da mineração de dados para descoberta de conhecimento esta sendo utilizada cada vez mais dentro das organizações em diferentes ramos e atividades. Devido esta abrangência algumas técnicas de mineração de dados são mais adequadas do que outras, dependendo muito dos dados e o objetivo final.

A proposta deste trabalho é apontar um modelo de aplicação para descoberta de conhecimento em bases de dados utilizadas pelo cadastro técnico urbano, utilizando técnicas de análise e mineração de dados. A busca de novos padrões e relacionamentos entre diferentes variáveis utilizando algoritmos de mineração de dados e modelos estatísticos, demonstra a interdisciplinaridade entre diferentes áreas de atuação, proporcionando novos mecanismos de descoberta.

## **1.1 Objetivos**

### **1.1.1 Objetivo Geral**

Utilizar algoritmos de mineração de dados através de um modelo de aplicação, de modo que atendam as necessidades dos especialistas e administradores municipais, visando auxiliar os processos de tomada de decisão.

### **1.1.2 Objetivos Específicos**

Têm-se como objetivos específicos:

- verificar os mecanismos utilizados para análise de dados junto aos departamentos responsáveis pelo cadastro técnico urbano;
  
- analisar a estrutura física da base de dados do cadastro técnico urbano investigada;
  
- propor um modelo para mineração de dados;
  
- avaliar a capacidade de estimar, prever e/ou anteceder causas relacionadas ao cadastro técnico urbano;
  
- aplicação do modelo proposto com a finalidade de identificar a eficiência da utilização da mineração de dados junto ao cadastro técnico urbano.

## **1.2 Justificativa**

Com base nos estudos realizados inicialmente em alguns departamentos que aplicam e utilizam-se do cadastro técnico urbano, detectou-se a necessidade de novos mecanismos de investigação, análise e predição sobre um conjunto de dados, devido ao grande volume de informações armazenadas nestes bancos de dados. Tais estudos deram origem ao levantamento das técnicas disponíveis para análise de dados, onde, dentre elas optou-se pela mineração de dados, que apresenta opções fundamentadas para auxiliar na tomada de decisões, como também na gestão do poder público auxiliando no cruzamento de informações. Ampliando assim, o conhecimento da área de estudo e principalmente dos dados coletados pelo cadastro técnico urbano.

### **1.3 Limitação do Trabalho**

O modelo proposto foi aplicado somente em uma base de dados cadastral, de forma que se possibilita compreender a aplicação da mineração de dados e conseqüentemente a interpretação das informações geradas.

### **1.4 Estrutura do Trabalho**

O referente trabalho está dividido em sete capítulos.

- No primeiro capítulo é apresentada uma introdução ao trabalho desenvolvido, aos objetivos e as limitações do mesmo.
- No segundo capítulo são discutidos os conceitos relativos ao cadastro técnico urbano e o cadastro técnico multifinalitário urbano.
- No terceiro capítulo são apresentados os conceitos de sistemas de informações geográficas e níveis de atuação.
- No quarto capítulo são apresentados os conceitos relativos à mineração de dados juntamente com seus métodos e técnicas.
- No quinto capítulo é discutido o modelo apresentado para a mineração de dados, as etapas de desenvolvimento para descoberta de conhecimentos com a aplicação da mineração de dados.
- No sexto capítulo se discute o estudo de caso e a aplicação prática do modelo proposto no quinto capítulo.
- No sétimo capítulo são apresentadas as conclusões e recomendações deste trabalho.
- E por último estão anexados: a bibliografia utilizada no desenvolvimento do trabalho o algoritmo APRIORI, as ferramentas investigadas, o questionário aplicado pelo IBGE no ano de 2001, o boletim de cadastramento BIC e o termo de pesquisa.

## 2 CADASTRO TÉCNICO URBANO

### 2.1 Conceitos de Cadastro

Cadastro é um conjunto de registros que têm informações relacionáveis. Na literatura portuguesa, segundo Ferreira (1986), cadastro é o registro público dos bens imóveis de determinado território, registro de clientes, censo, recenseamento. Quando se refere ao cadastro técnico reporta-se àquele com fundamentação teórica, para fins técnicos.

Cadastro técnico é um inventário público de todas as parcelas de terreno e dos imóveis de uma região. Constituído por Garcia (1974) e Henssen (1974):

- a) uma definição não ambígua das parcelas;
- b) registros relacionados que fornecem dados sobre as parcelas;
- c) uma parte cartográfica, composta por cartas que indicam a divisão em parcelas de uma área, juntamente com identificadores apropriados das parcelas;
- d) uma parte descritiva, que contém registros dos atributos físicos e abstratos relativos às parcelas identificadas nos mapas.

Burity (1998), definiu cadastro técnico urbano como um sistema de informação da terra atualizado e baseado em parcelas contendo um registro de interesses sobre a terra, uma descrição geométrica das parcelas e ligadas a outros registros que descrevem a natureza dos interesses, e a propriedade ou o controle desses interesses, e freqüentemente o valor da parcela e suas benfeitorias.

As prefeituras gerenciam os municípios através de informações provenientes do espaço urbano e rural, armazenadas em um cadastro técnico. Para que essa administração seja eficiente, é necessário dispor de informações que retratem de maneira real o espaço físico (Silva e Peixoto, 1996).

Na maioria dos municípios as informações referentes ao cadastro técnico urbano são coletadas utilizando-se o boletim de cadastro imobiliário (BIC). Exemplos desses boletim encontram-se no Anexo 4.

Segundo Hochheim (1996), o boletim de cadastro imobiliário (BCI) também denominado de Boletim de Informações Cadastrais (BIC), apresenta vários níveis de informações, desde a inscrição cadastral do imóvel; informações gerais para processamento (controle, inclusão/alteração/exclusão); localização do imóvel cadastrado; dados sobre o proprietário do imóvel; dados gerais sobre o imóvel; dados sobre o terreno; medidas do imóvel; dados sobre a edificação; outros dados (estado de conservação, existência de muro/passeio,...); e desenho/croqui do imóvel cadastrado. Todas as informações devem obrigatoriamente estar amparadas pela legislação vigente,

referente às finalidades tributárias. O boletim de informações cadastrais deve ser planejado em função das realidades e necessidades do município.

Segundo o BNDES (1999), a atualização do cadastro imobiliário, dentro de um processo mais amplo de diagnóstico do sistema tributário municipal, permite à prefeitura melhorar sua receita e corrigir as injustiças na cobrança de impostos.

## **2.2 Cadastro Técnico Multifinalitário Urbano**

Segundo Hochheim (1993), quando um cadastro técnico serve como um sistema básico de registros para uso de diversas pessoas e organizações responsáveis pela realização de diversos serviços, ele é dito multifinalitário. Sendo urbano, recebe a denominação de cadastro técnico multifinalitário urbano (CTMU). Um cadastro deste tipo tem múltiplos usos e diversos usuários.

Outra definição apresentada por Esteio (2003), diz que o Cadastro Técnico Multifinalitário visa a identificação das divisas de um imóvel juntamente com os proprietários, para a sua amarração à rede geodésica brasileira, garantindo assim a exata localização das divisas da propriedade, bem como a vinculação dos dados técnicos ao registro imobiliário a fim de proporcionar total embasamento técnico à garantia do direito de propriedade.

Segundo Dale et. al. (1990), cadastro técnico multifinalitário pode ser definido como um sistema de informações terrestres em escala grande, desenvolvido para servir organizações públicas e privadas além do cidadão comum, devendo suportar a transferência da terra através do registro de informações relevantes ao proprietário da mesma, também a taxaço da terra através do registro de detalhes referentes aos proprietários, aos ocupantes, às propriedades e seus respectivos valores. E a administração geral da terra pelo provimento de informações relativas à mesma de forma integrada, tornando possível formas complexas de análise e uma maior compreensão dos seus problemas.

Para Dale et. al. (1990), um cadastro técnico multifinalitário também deve ainda empregar uma unidade de terra proprietária, a parcela cadastral, como unidade fundamental da organização espacial; relacionar uma série de registros terrestres, tais como: culturas, valores e usos; próprios da parcela; ser o mais completo possível em termos da cobertura espacial e prover meios fáceis e eficientes de acesso aos dados.

Para tanto, é necessário que se tenha à disposição sistemas capazes de integrar toda e qualquer informação relativa à propriedade, em diferentes formatos e provenientes das mais variadas fontes. A figura 1, ilustra a hierarquia destes sistemas.

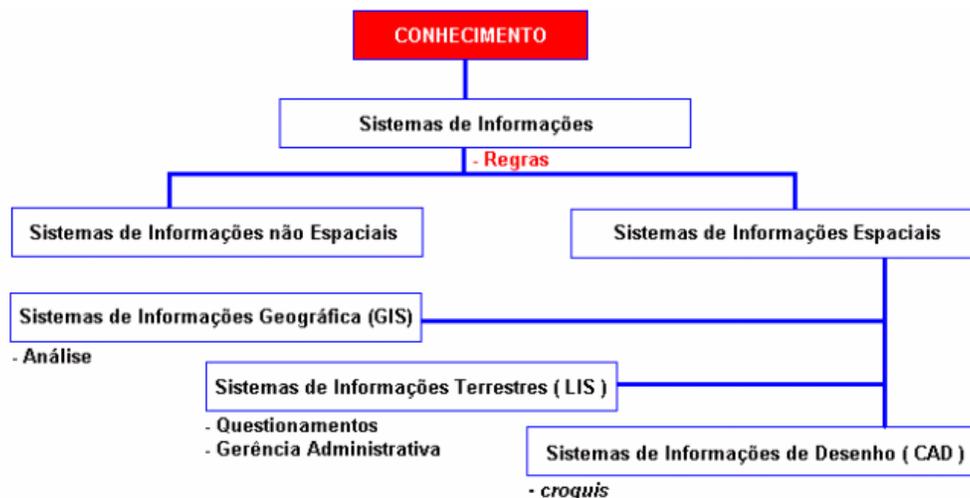


Figura 1 – Relacionamento ilustrativo entre os sistemas.  
 Fonte: Scholten et.al. (1990).

A unidade básica de um cadastro é a parcela. Os limites de uma parcela são normalmente definidos pela demarcação física sobre o terreno e por uma descrição matemática, geralmente baseada em um sistema de coordenadas. A precisão e, conseqüentemente, o custo dos levantamentos cadastrais depende da precisão necessária às descrições dos limites. A precisão deve refletir fatores tais como o valor da terra, o risco e custo das disputas de terra e as necessidades de informação dos usuários do cadastro.

O produto gerado com a finalidade de representar, em escalas grandes, a delimitação rigorosa do parcelamento da propriedade territorial, denomina-se carta cadastral. (Oliveira, 1993).

De acordo com Joly (1997), as cartas cadastrais são a base da identificação das propriedades imobiliárias e comportam, entre outros, os limites das parcelas, as cercas, os muros, as construções e a situação do ambiente em geral. As cartas cadastrais assentam-se sobre uma base cartográfica normalmente obtida por técnicas fotogramétricas. A definição dos elementos que compõem esta base, bem como, o nível de detalhamento das parcelas, depende da escala das fotografias tomadas.

O estabelecimento de uma base cartográfica precisa se faz necessária para o esclarecimento da situação dos imóveis. Só assim é possível prover o cadastro dos componentes cartográfico, técnico, social, econômico e legal, a fim de torná-lo instrumento eficiente da administração municipal.

A constante atualização destas bases é fundamental para que o cadastro técnico atenda as suas finalidades básicas. A figura 2 ilustra a sistemática de atualização das bases cartográficas urbanas e mostra a sua importância no âmbito da administração

municipal.

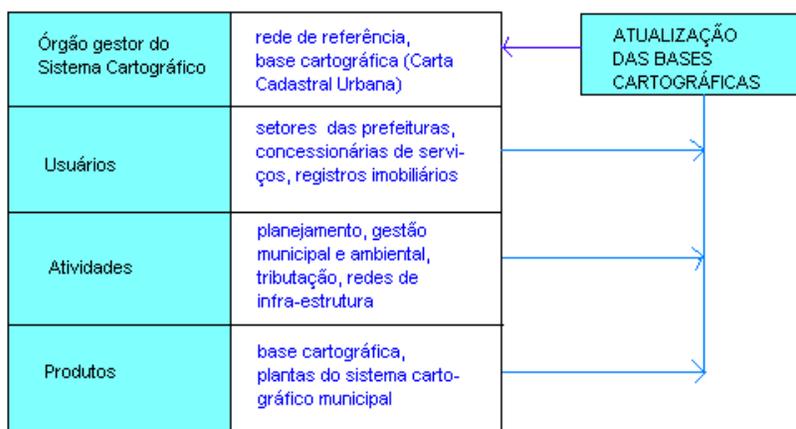


Figura 2 – Sistematização de atualização.

Fonte: Adaptada de Rosa (1994).

A carta cadastral é responsável pela geração de três plantas temáticas muito utilizadas pela administração municipal. Dentre elas, a planta de quadras utilizada como suporte na geração da planta genérica de valores e nos estudos de expansão urbana, a planta de parcelas ou lotes inclui as informações referentes às quadras e pode ser utilizada como instrumento de fiscalização de loteamentos irregulares, de planejamento dos serviços de infra-estrutura ou de base para estudos do zoneamento e a planta de edificações e benfeitorias inclui as informações das plantas descritas, sendo utilizada como instrumento de detecção e fiscalização de construções irregulares ou clandestinas.

As plantas que compõem o sistema cartográfico municipal contêm informações que permita a análise e elaboração de diagnósticos da realidade das parcelas, tornando possível o ajuste das ações a serem desenvolvidas nas diversas áreas urbanas. A figura 3, apresenta os principais setores da sociedade envolvidos com a administração pública que podem ser auxiliados através da utilização do cadastro técnico multifinalitário.



Figura 3: Interação dos diversos segmentos públicos com a administração municipal.

Fonte: Adaptado de Hochheim (1996).

Segundo Hochheim (1993), um cadastro técnico multifinalitário implica na colaboração entre os diferentes usuários do sistema, que devem definir, adquirir e atualizar os seus próprios dados e pelos quais são responsáveis. Assim os custos de sua implantação e manutenção podem ser repartidos entre seus usuários, o que contribui significativamente para viabilizar economicamente o cadastro técnico.

## 2.3 Objetivos do Cadastro Técnico Urbano

Conforme Hochheim (1996), um cadastro técnico urbano gera informações que podem ser associadas, através de chaves, a outras informações, diferentes temas, constituindo-se num sistema de informações de grande utilidade.

Possui como objetivos específicos:

- a) Permitir justa cobrança dos tributos. Um componente do cadastro técnico urbano é o cadastro fiscal, no qual são baseados os cálculos dos tributos municipais. Possibilitando a previsão da receita municipal e o cálculo da contribuição de melhoria.
- b) Inventariar terras. Permite um balanço anual das terras urbanas no que se

refere ao estoque de terras, transferência de propriedade, ocupação e uso.

- c) Gerar os dados espaciais para um sistema de informações. Torna-se assim, ferramenta de apoio para os órgãos encarregados do planejamento municipal e regional.
- d) Facilitar a economia nos processos de desapropriações legais. A planta cadastral permite elaboração de planos que envolvem obras de engenharia, com as vantagens:
  - i. Sigilo: sem planta há a necessidade de levantamentos topográficos, o que poderia favorecer a especulação imobiliária e interferências políticas, gerando gastos aos cofres públicos;
  - ii. Justa indenização: a avaliação das propriedades a indenizar pode ser feita de modo isento, baseada numa planta de valores genéricos devidamente atualizada.
- e) Gerar a base física para as operações de serviço público. É apoio para otimização dos serviços públicos para água e esgoto, abastecimento de luz e força, abastecimento de gás encanado, telefonia, coleta de lixo, transportes urbanos e segurança (polícia, bombeiros, ...)
- f) Mapeamento das instalações subterrâneas e aéreas. O cadastro deve ser feito quando da execução de uma obra subterrânea, após o fechamento da vala é praticamente impossível sua execução. São exemplos dessas instalações:
  - i. Redes de esgotos;
  - ii. Redes de abastecimento de água;
  - iii. Redes de transmissão de energia;
  - iv. Redes de iluminação pública;
  - v. Rede telefônica;
  - vi. Presença de árvores, abrigos;
  - vii. Pontos de contato das instalações de subsolo com a superfície.
- g) Estabelecer os limites municipais, de maneira exata e definitiva, evitando conflitos.
- h) Garantir a propriedade imobiliária
  - i. A planta cadastral posiciona espacialmente o imóvel, identificando seus vizinhos;

- ii. O registro cadastral descreve a situação do imóvel;
- iii. O registro de imóveis, baseado nas informações espaciais do cadastro, garante a propriedade do imóvel, em termos de documentação e situação espacial.

Para Hochheim (1996), o cadastro técnico urbano age como atividade no estabelecimento e manutenção dos padrões de desenvolvimento, das regulamentações técnicas e do código de obras.

As principais finalidades do cadastro são: garantir a propriedade, embasar o planejamento governamental, fornecer parâmetros para uma justa tributação / desapropriação / servidão e resguardar a função social da terra.

O cadastro técnico multifinalitário urbano pode ser utilizado na fiscalização para avaliar e tributar as parcelas, na legislação pode auxiliar na transferência de títulos, no gerenciamento do solo e do uso do solo pode auxiliar no planejamento e outras finalidades administrativas, no desenvolvimento sustentável e na proteção ambiental.

## **2.4 Vantagens do Cadastro Técnico Urbano**

Para Hochheim (1996), o cadastro técnico urbano facilita a provisão de informações para os tomadores de decisão, as funções de monitoramento e controle, a descentralização das funções de gestão, o processo de planejamento e a análise de estratégias e ações alternativas.

São vantagens do cadastro técnico urbano:

- Oferecer segurança, reduzir ou eliminar o risco de expulsão, (incentivo para investimentos na propriedade);
- Disponibilizar recursos: exigência para crédito a longo prazo (recursos financeiros significativos)
- Auxiliar nas negociações de terrenos tornando-as mais rápidas, seguras e menos onerosas.

Outras vantagens do Cadastro Técnico:

- Possibilitar a cobrança de impostos e taxas de forma eficiente e justa (justiça fiscal para os municípios);
- Fornecer informações para elaboração de planos diretores e projetos específicos de desenvolvimento;
- Gerar mapas em escalas menores (economia de tempo e dinheiro).

Todos os investimentos feitos pelo poder público na cidade localizam-se num espaço geográfico bem definido, que é necessário conhecer para seguir as diferentes tendências de implantação e suas possíveis repercussões, ao nível do emprego criado, das produções conseguidas, do escoamento dos produtos, dos efeitos ambientais, etc., ou seja, de um número enorme de variáveis que estão ligadas e voltadas ao bem estar das populações.

## **2.5 Desvantagens do Cadastro Técnico Urbano**

Para Hochheim (1996), não basta pensar em cobrir um território com um cadastro técnico sem pensar também nos mecanismos necessários para mantê-lo permanentemente atualizado. Um cadastro desatualizado tem seu interesse progressivamente reduzido pela diminuição do seu valor informativo.

### 3 SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS

#### 3.1 Introdução

Para Câmara (2000), na medida que avança a montagem de soluções corporativas de geoinformação, cresce a necessidade de técnicas de análise espacial, imprescindíveis para fazer bom uso desses grandes bancos de dados. Isso aponta para um cenário com grandes desafios e muita necessidade de atualização por todos. Ou seja, “se pensa que sabe tudo de geoprocessamento, pense de novo?”.

A figura 4 apresenta uma visão da integração entre as diferentes tecnologias de geoprocessamento disponíveis para os diversos profissionais. Entre elas têm-se bancos de dados geográficos e a análise espacial que representam o núcleo central de um sistema de informações geográficas.

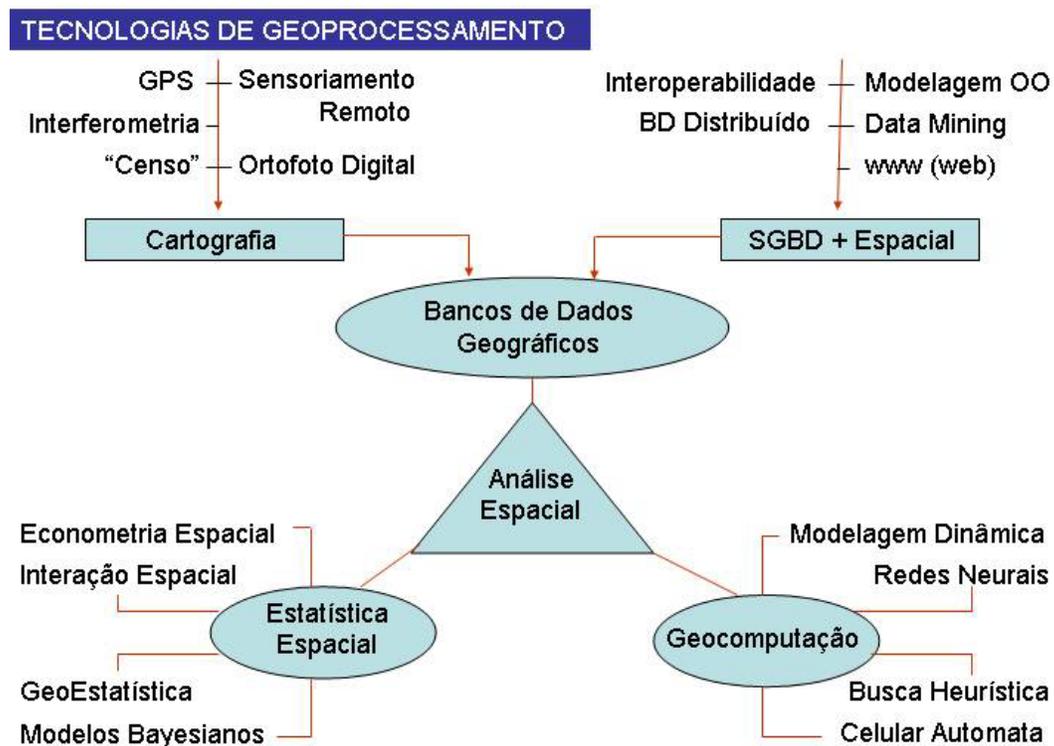


Figura 4 - Tecnologias de Geoprocessamento.

Fonte: Adaptada de Marília Sá Carvalho (2000).

### 3.2 Conceitos de SIG

Segundo Câmara (1995), o termo sistema de informação geográfica (SIG) é aplicado para sistemas que realizam o tratamento computacional de dados geográficos. Devido a sua ampla gama de aplicações, que inclui temas como agricultura, floresta, cartografia, cadastro urbano e redes de concessionárias (água e saneamento, energia elétrica e telefonia), há pelo menos três grandes maneiras de utilizar um SIG.

- ✓ como ferramentas para produção de mapas;
- ✓ como suporte para análise espacial de fenômenos;
- ✓ como um banco de dados geográfico, com funções de armazenamento e recuperação de informação espacial.

Ainda para Câmara (1995), um sistema de informação geográfica é um sistema de informação baseado em computador que permite capturar, modelar, manipular, recuperar, consultar, analisar e apresentar dados geograficamente referenciados. A tecnologia de SIG pode trazer enormes benefícios devido à sua capacidade de manipular a informação *espacial*<sup>1</sup> de forma precisa, rápida e sofisticada (Goodchild et. al., 1993). Na década de 80, o uso de SIG tornou-se comum em empresas, universidades e agências governamentais, e atualmente diversos profissionais o utilizam para as mais variadas aplicações. Essa diversidade de usos e aplicações fez surgir várias definições de SIG, tais como:

*“Um conjunto manual ou computacional de procedimentos utilizados para armazenar e manipular dados georeferenciados”  
(Aronoff, 1989)*

*“Conjunto poderoso de ferramentas para coletar, armazenar, recuperar, transformar e visualizar dados sobre o mundo real”  
(Burrough, 1986)*

*“Um sistema de suporte à decisão que integra dados referenciados espacialmente num ambiente de respostas e problemas”  
(Cowen, 1988)*

---

<sup>1</sup> O termo espacial é utilizado neste caso como sinônimo de georeferenciado, ou seja, referenciado a uma região do espaço geográfico.

*“Um banco de dados indexado espacialmente, sobre o qual opera um conjunto de procedimentos para responder a consultas sobre entidades espaciais” (Smith et. al., 1987)*

Estas definições de SIG refletem, cada uma à sua maneira, a multiplicidade de usos e visões possíveis desta tecnologia e apontam para uma perspectiva interdisciplinar de sua utilização. Com estes conceitos, é possível indicar as principais características de SIGs.

- a) Integrar, numa única base de dados, informações espaciais provenientes de dados cartográficos, dados de censo e cadastro urbano e rural, imagens de satélite, redes e modelos numéricos de terreno.
- b) Oferecer mecanismos para combinar as várias informações, através de algoritmos de manipulação e análise para consultar, recuperar, visualizar e plotar o conteúdo da base de dados geocodificados.

### **3.3 Conceitos de Espaço e Relações Espaciais**

Um aspecto fundamental nos dados tratados em um SIG é a natureza dual da informação: um dado geográfico possui uma localização geográfica (expressa como coordenada em um mapa) e atributos descritivos (que podem ser representados num banco de dados convencional).

Segundo Goodchild (1992), as aplicações de geoprocessamento lidam com dois grandes tipos de dados espaciais: geo-campos e geo-objetos (objetos geográficos)

- *geo-campos*: são variações espaciais contínuas. São usadas para grandezas distribuídas espacialmente, tais como tipo de solo, topografia e teor de minerais. Correspondem, na prática, a dados temáticos, imagens e modelos numéricos de terreno; e
- *geo-objetos (objetos geográficos)*: são individualizáveis e têm identificação. Este tipo de dado tem atributos não espaciais, armazenados em um banco de dados convencional, e pode estar associado a várias representações gráficas. Alguns exemplos já citados são: escolas, municípios e fazendas

Outro aspecto igualmente importante é que os dados geográficos não existem sozinhos no espaço: tão importante quanto localizá-los é descobrir e representar as relações entre os diversos dados. Alguns exemplos dos processos de análise espacial típico de um SIG estão representados no quadro 1.

Quadro 1: Exemplo para análise espacial

Análise	Pergunta Geral	Exemplo
Condição	“O que está...?”	“Qual a população desta cidade?”
Localização	“Onde está...?”	“Quais as áreas com declividade acima de 20%?”
Tendência	“O que mudou...?”	“Esta terra era produtiva há 5 anos atrás?”
Roteamento	“Por onde ir...?”	“Qual o melhor caminho para o metrô...?”
Padrões	“Qual o padrão...?”	“Qual a distribuição da dengue em Fortaleza?”
Modelos	“O que acontece se...?”	“Qual o impacto no clima se desmatar a Amazônia?”

Fonte: Adaptada de Maguire (1991).

### 3.4 Utilização do SIG na Organização

As atividades de gestão de uma organização podem ser estruturadas em três níveis conforme Antony (1995): estratégico, tático e operacional. No primeiro nível, o estratégico, inclui-se a elaboração de planos de longo prazo, a definição de objetivos para a organização e de estratégias para a prossecução dos objetivos. O nível tático considera as atividades de planejamento à médio prazo, a verificação da prossecução dos objetivos e metas e se as estratégias definidas estão a ser devidamente implementadas, a tomada de decisões referentes a ações de correção. O nível de gestão operacional inclui a elaboração de planos de curto prazo e o controle de execução de tarefas planejadas no sentido de verificar se estas estão a ser executadas eficientemente. O nível operacional corresponde às atividades operacionais propriamente ditas, isto é, às transformações que a organização tem por missão realizar utilizando os recursos disponíveis e seguindo as instruções, regras ou planos que tenham sido definidos.

Segundo Ferrari (1997), os sistemas de informações geográficas podem ser utilizados nesses três níveis, sendo que os benefícios provindos do seu uso serão igualmente distintos.

A figura 5 apresenta uma distribuição dos 3 níveis organizacionais sugerido para uma prefeitura ou órgão de planejamento, onde:

- SIG(a) encontra-se no nível operacional,
- SIG(b) encontra-se no nível gerencial (tático) e
- SIG(c) encontra-se no nível estratégico.

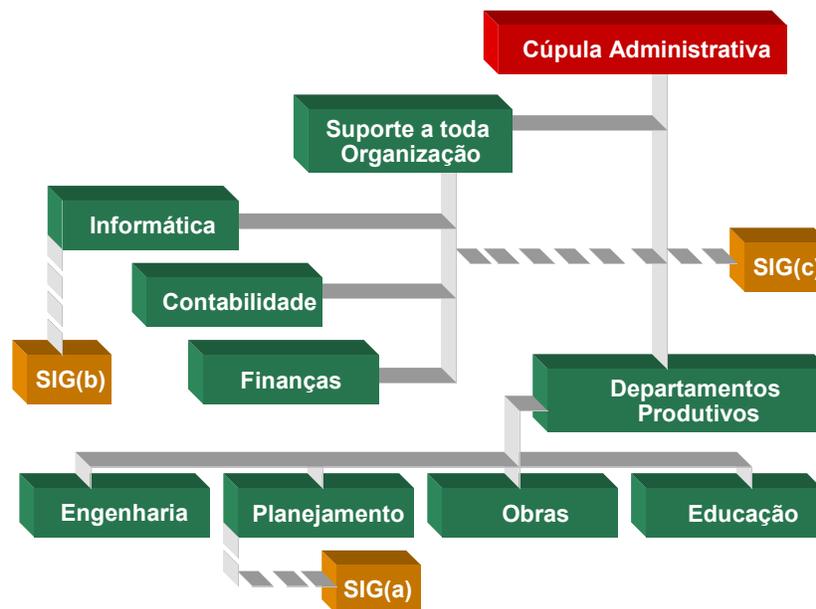


Figura 5 – Administração, controle e implantação do SIG.  
 Fonte: Adaptado de Ferrari (1997).

### 3.4.1 SIG no Nível Operacional

Para Ferrari (1997), as atividades do nível operacional são aquelas do dia-a-dia da organização. Além de rotineiras, geralmente as operações são volumosas e trabalhosas. Exemplos desse tipo de atividades: a compensação de cheques em um banco, uma linha de montagem em uma indústria, processamento de requisições de clientes, manutenção rotineira de prédios e equipamentos, etc.

Segundo Siroski, Yuoca e Davis (2002), são atividades do nível operacional de uma prefeitura: as liberações de alvarás de construção, os processos de aprovação de loteamentos, os processos de aprovação de desmembramento e regularização de lotes, a regularização de projetos para vias públicas, a definição de valores tributários, o controle de equipamentos urbanos, o controle operacional do transporte coletivo municipal, o controle da manutenção da rede de pavimentação, os projetos e controle de galerias e microdrenagem, o controle da sinalização viária, o cadastro de vigilância sanitária, o controle da arborização urbana, o acompanhamento de obras públicas, a distribuição de alunos nas escolas municipais, o controle e fiscalização dos lançamentos de efluentes sanitários, a manutenção de parques e praças, o licenciamento de uso de imóveis, a roteirização e o roteamento, a determinação de rotas para coleta de lixo, o projeto de expansão de rede de esgotamento pluvial, ser estimador de valor de lotes urbanos, a prestação de informações diretamente ao cidadão, delimitado de áreas de jurisdição (apoio à descentralização administrativa) e por último a geração de plantas de valores.

Assim como na automação de uma linha de montagem, o benefício imediato do uso de SIGs no suporte a atividades do nível operacional é a eficiência operacional, ou seja, a execução das atividades que já vêm sendo realizadas manualmente (sem apoio de SIGs) só que de maneira mais eficiente gastando menos recursos. O quadro 2 apresenta os benefícios advindos da utilização do SIG no nível operacional.

Quadro 2: Benefícios e Atividades do Nível Operacional

Benefícios	O papel do SIG	Exemplos de Usos
<ul style="list-style-type: none"> <li>- Ganho de produtividade</li> <li>- Redução ou eliminação de custos ou riscos</li> <li>- Qualidade na execução de tarefas.</li> </ul>	<ul style="list-style-type: none"> <li>- Automação de tarefas trabalhosas e rotineiras</li> <li>- Proporcionar maior qualidade e agilidade no armazenamento de informações essenciais para a execução das tarefas</li> <li>- Suporte ao planejamento monitoramento racionalização do trabalho</li> <li>- Suporte à visualização de informações e comunicação no trabalho</li> </ul>	<ul style="list-style-type: none"> <li>- Manutenção de mapas ou outras informações geográficas</li> <li>- Geração de mapas e outras informações para suporte a projetos de engenharia</li> <li>- Suporte ao gerenciamento de redes de infra-estrutura planejamento de manutenção preventiva, planejamento e monitoramento de tarefas, geração de esquemas de trabalho, suporte a situações de emergência, etc.</li> <li>- Otimização de rotas de coleta ou distribuição</li> </ul>

Fonte: Ferrari (1997).

### 3.4.2 SIG no Nível Gerencial

Para Ferrari (1997), as decisões de caráter tático são atividades típicas do nível gerencial. Qual a melhor localização para uma nova filial? Uma questão semelhante no setor governamental seria qual a região mais carente de um determinado serviço público?

Segundo Siroski, Yuoca e Davis (2002), são atividades do nível gerencial de uma prefeitura: o acompanhamento de endemias, a avaliação das diretrizes e normas para uso e ocupação do solo urbano e classificação viária, a delimitação de áreas sujeitas a inundações, o gerenciamento de bacias hidrográficas, a preservação de mananciais e fundos de vales, o registro e controle de poluição, a análise e gerenciamento do transporte coletivo municipal, a análise de acesso da população aos equipamentos urbanos, a preservação do patrimônio histórico, a análise de acidentes de trânsito, as

diretrizes viárias, as diretrizes para novos loteamentos, os estudos da distribuição e abrangência de equipamentos sociais urbanos, o gerenciamento de áreas verdes e de preservação ambiental, o controle de doenças transmissíveis; a cobertura vacinal; o controle de desnutrição; a administração de áreas de risco diversas: risco geológico, de inundação de saúde / sanitária, social; a administração dos lotes vazios; a análise de padrões de distribuição de crimes; a análise da distribuição espacial das atividades econômicas; a determinação da área de atuação de postos de saúde, escolas, lojas de atendimento; a determinação de melhor local para construção de escolas; a determinação de vetores de crescimento da cidade; a geração de mapas temáticos, padrão de edificação, pavimentação de ruas; a definição de áreas para depósitos de lixo e aterros sanitários e por último a identificação dos locais com maior índice de acidentes de trânsito.

Segundo Ferrari, (1997), no exemplo da filial, um sistema de informação pode ajudar a pessoa responsável a escolher uma boa localização para a nova loja. Os sistemas de informações geográficas podem dar suporte a decisões de caráter tático, comuns no planejamento urbano, no gerenciamento de franquias, entre outros. Em todos os exemplo, o benefício esperado é a eficácia administrativa: boas informações, bons planos, bom gerenciamento, boas decisões.

A segunda coluna do quadro 3 aponta o papel do SIG no nível gerencial na obtenção dos benefícios, provendo maior capacidade de manipulação, análise, integração e visualização de informações. O SIG não toma decisões. Mas o uso de um SIG, como uma ferramenta de trabalho, facilita a obtenção de boas decisões.

Quadro 3: Benefícios e Atividades do Nível Gerencial

Benefícios	O papel do SIG	Exemplos de Usos
<ul style="list-style-type: none"> <li>- Melhores ou novas informações.</li> <li>- Melhores decisões de carácter tático, planejamento, gerenciamento e alocação de recursos.</li> </ul>	<ul style="list-style-type: none"> <li>- Proporcionar agilidade na geração de informações para suporte a processos intelectuais: decisões planejamento, monitoramento...</li> <li>- Prover maior capacidade de manipulação de informações: análise espacial, modelagem simulações...</li> <li>- Mecanismos para integrar, visualizar e resumir informações complexas e diversas, mecanismos de comunicação.</li> </ul>	<ul style="list-style-type: none"> <li>- Suporte ao gerenciamento de redes de lojas ou franquias, análise espacial de demanda (consumidores, nível de renda, rede atual, concorrentes) e auxílio a decisões táticas (qual o melhor local para uma nova filial?)</li> <li>- Suporte ao planejamento e gerenciamento de equipamentos e serviços públicos, monitorar a demanda (demográfica), a qualidade dos serviços atuais e auxiliar o projeto de novos equipamentos.</li> <li>- Auxílio à elaboração e ao monitoramento demográfica, análise fiscal, definição da política de uso e ocupação do solo, política fiscal e de incentivos, suporte à visualização de informações e a discussões entre técnicos e políticos.</li> <li>- Identificação da distribuição espacial de doenças, crimes ou acidentes de trânsito.</li> </ul>

Fonte: Ferrari (1997).

### 3.4.3 SIG no Nível Estratégico

As atividades do nível estratégico são aquelas que contribuem diretamente para o cumprimento dos objetivos fundamentais da organização. Exemplos de objetivos estratégicos são: aumentar a satisfação dos clientes, melhorar a imagem da empresa junto a (potenciais) clientes e parceiros, aumentar a margem de lucro, aumentar a participação da empresa nos segmentos de mercado em que atua, expandir a participação da empresa a outros segmentos, etc. (Ferrari, 1997).

Segundo Siroski, Yuoca e Davis (2002), são atividades do nível estratégico de uma prefeitura: a reestruturação do trânsito; uma política da distribuição espacial das atividades econômicas; estudos do uso do solo e eficiência do sistema viário; análise da evolução da ocupação urbana e política de vazios urbanos; análise de aspectos demográficos; atualização do plano diretor da cidade; monitoramento do índice de qualidade de vida; planejamento de expansão de infra-estrutura, acompanhamento do índice de satisfação da população por área.

Para Ferrari (1997), uma empresa pública, os termos “clientes” e “parceiros comerciais” têm uma conotação ainda mais abrangente. A administração pública relaciona-se com outras instituições públicas, com agências de desenvolvimento, com setores do comércio e da indústria e com a população. A administração municipal, por exemplo, relaciona-se com o poder legislativo local, com companhias concessionárias, com os governos estadual e federal, com indústrias da região, com empresários interessados em investir na região, com a população em geral, com políticos, com agências de desenvolvimento.

A administração pública depende de um bom relacionamento com todos esses clientes e parceiros, depende de credibilidade em projetos, depende de apoio político, depende dos eleitores. Depende, enfim, de uma boa imagem onde a organização e a facilidade de visualização em busca de credibilidade para os projetos certamente é um importante benefício para o nível estratégico.

No quadro 4, cita-se exemplos do uso de SIGs em apoio a atividades do nível estratégico. Em todos os exemplos citados, os benefícios referem-se a uma boa imagem, à credibilidade, ao relacionamento com parceiros comerciais (compartilhamento de custos), e também ao aumento de receita (venda de dados, mapeamento e gerenciamento de arrecadação).

Quadro 4: Benefícios e Atividades do Nível Estratégico

Benefícios	Exemplos de Uso
<ul style="list-style-type: none"> <li>- Melhor imagem junto a “clientes” e “parceiros”</li> <li>- Novas fontes de receita, aumento de receita.</li> </ul>	<ul style="list-style-type: none"> <li>- Uso de SIGs em projetos para solução de problemas estratégicos ou seja, problemas com alto impacto político ou econômico no momento.</li> <li>- Uso de SIGs em projetos sociais para melhorar os serviços à população a satisfação da população com a administração pública e, conseqüentemente sua imagem.</li> <li>- Mapeamento e apoio ao gerenciamento da satisfação e das necessidades dos leitores e suporte a ações de marketing político.</li> <li>- Compartilhamento de base de dados e de custos operacionais entre prefeituras, concessionárias e outras instituições.</li> <li>- Fornecimento de uma boa infra-estrutura de informações espaciais ou serviços a setores da indústria ou a outros órgãos de administração pública.</li> <li>- Venda de dados, mapeamento e gerenciamento da arrecadação de impostos.</li> <li>- Além de prover subsídios técnicos, o uso de SIGs em projetos pode facilitar a comunicação com leigos e outorgar credibilidade.</li> </ul>

Fonte: Ferrari (1997).

## 4 MINERAÇÃO DE DADOS

### 4.1 Introdução

A partir da década de 80, surgiram programas para a extração de dados com duas funções diferenciadas, uma voltada para o ambiente operacional (OLAP - Banco de Dados Operacional) e outra para a análise dos dados de uma base para apoio à decisão (Banco de Dados de SAD) conforme apresentado a seguir na figura 6.

Passado mais dez anos, na década de 90, surgiu o *Data Warehouse*, que corresponde a grandes bancos de dados utilizados para análise e suporte à decisão. Com o surgimento do *Data Warehouse*, a utilização da mineração de dados tornou-se mais constante para descoberta de informações ocultas ao tomador de decisão. Esta técnica utiliza algoritmos específicos que processam grande quantidade de informações. A mineração de dados nas empresas apresenta uma evolução histórica, características dos dados, perguntas e respostas e a integração entre diferentes tecnologias de análise.

Para Groth (1998), a mineração de dados é o processo para automatizar a descoberta de informação. Embora existam muitas ferramentas para consultar e manipular dados, o usuário é abandonado quando precisa encontrar tendências e padrões úteis. A mineração de dados automatiza o processo de descobrimento destas tendências e padrões. No centro da mineração de dados está o processo de construção do modelo, sendo que, criar um modelo representativo baseado num conjunto existente de dados provou ser útil para compreender tais tendências, padrões, e correlações.

A mineração de dados é o processo de extração de informações válidas, desconhecidas, e de máxima abrangência a partir de grandes bases de dados, utilizando-as para tomar decisões. A mineração de dados vai muito além da simples consulta a um banco de dados, permite ao usuário explorar informações úteis a partir dos dados, descobrindo relacionamentos ocultos. Pode ser considerada uma forma de descobrimento de conhecimento em bancos de dados (KDD - *Knowledge Discovery in Databases*), área de pesquisa envolvendo inteligência artificial e banco de dados.

As técnicas de mineração de dados permitem o processamento em vários tipos de dados armazenados a fim de encontrar dados relevantes para a administração, possibilitando a transformação dos dados em informações as informações em ação e a ação em valores ou resultados.

### 4.2 Níveis de Informação

A figura 6, apresenta um esquema de implementação de banco de dados aplicado nos diferentes níveis de trabalho (operacional, gerencial e estratégico), podendo-se conseqüentemente extrair diferentes níveis de conhecimento.

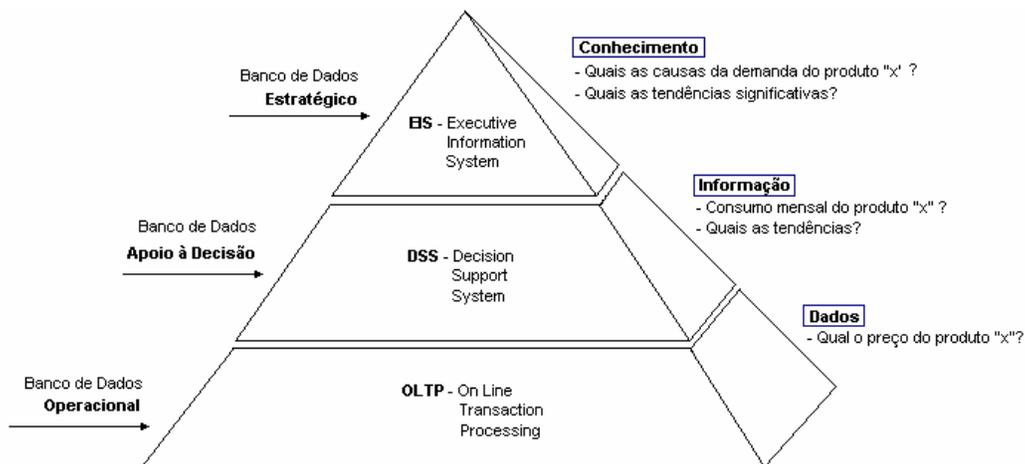


Figura 6 –Banco de dados representados em níveis de implementação.

Fonte: Adaptado de Inmon (1997) e Kimball et. al. (1998).

Segundo Inmon (1997, p.237), os sistemas EIS (*Executive information Systems*), são sistemas de informações executivas e constituem uma das mais poderosas formas de computação. Por meio dos EIS o analista executivo pode localizar problemas com precisão e detectar tendências que são de vital importância para a gerência. O processamento EIS é feito sob medida para ajudar o executivo a tomar decisões. O EIS se transforma na janela do executivo para a corporação. O processamento EIS examina uma ampla perspectiva e recolhe os aspectos que são relevantes para o funcionamento do negócio. Alguns dos usos tradicionais do EIS são: Análise e investigação de tendências, mensuração e rastreamento de indicadores de fatores críticos, análise prospectiva, monitoramento de problemas e análise da concorrência.

Little em 1970 definiu DSS (*Decision Support System*) ou também chamado de (SAD – Sistema de Apoio a Decisão), como “um modelo baseado nos procedimentos de forma a processar dados e julgamentos para auxiliar o gestor nas suas tomadas de decisão”. Para que um sistema tenha sucesso deverá ser robusto, simples, fácil de controlar, adaptativo, completo e de fácil comunicação.

A evolução dos SADs inicia em 1960 com arquivos mestres e relatórios, em 1965 com a complexidade de manutenção de desenvolvimento, em 1970 com banco de dados como única fonte de dados para todo processamento, em 1975 processamento de transações *online* – OLTP e por último em 1980 com a utilização de PCs e tecnologias L4G (linguagem de quarta geração). Inmon (1997).

Ainda para Inmon (1997, p.29-30), o processamento de transações *online* OLTP apresenta o tempo de resposta, quase sempre, um fator de missão crítica. Quando o tempo de resposta no OLTP fica ruim, os negócios começam imediatamente a se ressentir.

### 4.3 Origem da Mineração de Dados

Segundo DW Brasil (2001), a mineração de dados é uma ramificação de três diferentes áreas. A primeira e mais tradicional é a estatística clássica. Sem a estatística não seria possível o surgimento da mineração de dados, visto que serve como base para maioria das tecnologias que contemplam análise de dados. A estatística clássica envolve conceitos como distribuição normal, variância, análise de regressão, desvio simples, análise de conjuntos, análises de discriminantes e intervalos de confiança, todos usados para estudar dados e os relacionamentos entre eles.

A segunda área que deu origem a mineração de dados é a inteligência artificial (IA). Essa área, que é construída a partir dos fundamentos da heurística, em oposição à estatística, tenta imitar a maneira como o homem pensa na resolução dos problemas estatísticos. Em função disso, requer um poder de processamento que era impraticável até os anos 80, quando os computadores começaram a oferecer um bom poder de processamento a preços mais acessíveis. Isto permitiu adotar os conceitos de IA em alguns produtos de ponta, como por exemplo módulos de otimização de consultas para SGBDs.

A terceira e última área é chamada de aprendizagem de máquina (*machine learning*) ou mecanismo de aprendizagem, que pode ser descrita como a união entre a estatística e a inteligência artificial. As técnicas de IA, foram sendo absorvidas pela *machine learning*, que foi capaz de se valer das crescentes taxas de preço/performance oferecidas pelos computadores nos anos 80 e 90, conseguindo mais e mais aplicações devido às suas combinações entre heurística e análise estatística.

*Machine Learning* procura fazer com que os programas de computador “aprendam” com os dados analisados, fazendo com que tomem decisões diferentes baseadas nas características dos dados avaliados. Eles utilizam a estatística para os conceitos fundamentais e adicionam heurísticas da inteligência artificial para alcançar os seus objetivos. A figura 7 apresenta o relacionamento multidisciplinar que deu origem a mineração de dados.



Figura 7 - Relacionamento multidisciplinar.

Fonte: Adaptada de OWG – *Smart Business Smart Solution* (2001).

#### 4.4 Mineração de Dados: um breve histórico

O surgimento da mineração de dados apresenta uma trajetória passando por quatro fases importantes de evolução, até chegar ao amadurecimento das técnicas que hoje são colocadas a disposição por muitas ferramentas de análise de dados.

A primeira fase em (1960 a 1970), é a fase da coleta de dados tendo como característica principal os dados históricos e estáticos, onde o foco da questão era responder perguntas do tipo, “*Qual foi o total de receita da empresa nos últimos cinco anos?*”. Tendo como tecnologia recursos computacionais limitados tais como computadores de grande porte, fitas e discos.

A segunda fase em (1970 a 1980), é a do acesso aos dados, apresentando como característica principal os dados históricos e dinâmicos, acessados ao nível de registros, onde o foco da questão era responder perguntas do tipo, “*Qual foi o número de unidades vendidas no mês de julho em São Paulo?*”. Apresentando como tecnologia, melhores recursos computacionais comparados aos recursos citados na primeira fase, a exemplo dos bancos de dados relacionais totalmente baseados na linguagem padrão de manipulação de dados SQL (*Structured Query Language*).

A terceira fase é do *data warehouse* (década de 1990), apresentando como característica principal os dados históricos e dinâmicos acessados de várias dimensões, onde o foco da questão era responder perguntas do tipo, “*Quais são as unidades vendidas no mês de julho, nos últimos dez anos nos diferentes estados do Brasil?*”. Apresentando como tecnologia, metodologias para selecionar, limpar, agrupar, visualizar as informações armazenadas em bancos de dados relacionais.

A quarta e última fase (a partir de 2000) é a da mineração de dados fortemente relacionada ao desenvolvimento de *data warehouse*. Tem como característica principal a previsão e conhecimento no auxílio da tomada de decisão, onde o foco da questão é responder questões do tipo, “*Que acontecerá no próximo mês com as vendas em São Paulo? Porque?*”. A tecnologia utilizada envolve algoritmos avançados para descobrir conhecimento nos dados armazenados em grandes conjuntos de dados.

#### 4.5 Conceitos de Mineração de Dados

Para Harrison (1998), a mineração de dados traz idéias e técnicas de uma variedade de campos que possuem vocabulários muito diferentes. Os estatísticos, os pesquisadores da inteligência artificial, os administradores de banco de dados e o pessoal de *marketing* usam palavras diferentes para dizer a mesma coisa, e usam as mesmas palavras para dizer coisas diferentes. Ainda para os estatísticos, economistas e outros pesquisadores, a “mineração de dados” é um termo pejorativo. Refere-se à prática da seletividade tentando encontrar dados que apoiarão uma hipótese em particular.

Segundo a empresa especializada em soluções de inteligência empresarial OWG *Smart Business* (2001), conceitua-se a mineração de dados como a detecção automática de tendências e associações “escondidas” nos dados. Faz parte de um processo maior chamado “*knowledge discovery*”. Pode também ser descrito como a aplicação das técnicas da *machine learning* às aplicações de negócio.

Conforme as publicações do DCC (departamento de ciência da computação) e do laboratório de computação e heurística da universidade federal de Minas Gerais (2001), a mineração de dados é um conjunto de técnicas que envolvem métodos matemáticos, algoritmos e heurísticas para descobrir padrões e regularidades em grandes conjuntos de dados.

A mineração de dados para Harrison (1998), é a exploração e análise, por meios automáticos ou semi-automáticos, de grandes quantidades de dados para descobrir modelos e regras significativas. Nenhum dos algoritmos de mineração foram criados com propósitos comerciais. O emprego comercial da mineração de dados usa técnicas emprestadas da estatística, da ciência da computação e da inteligência artificial. A escolha de uma combinação específica de técnicas a serem aplicadas em uma certa situação depende da natureza da tarefa da mineração de dados a ser executada e da natureza dos dados disponíveis.

## 4.6 Aplicabilidade

Segundo Harrison (1998, p.177-186) e Carvalho (2001, p.17), a mineração de dados se aplica bem a tarefas como classificação, estimativas, previsões, agrupamento por afinidades, reunião e descrição. Algumas dessas tarefas são executadas melhor com a abordagem ‘de cima para baixo’, chamadas de ‘teste hipotético’. Em um teste hipotético, um banco de dados contendo comportamentos passados é usado para verificar ou desaprovar noções preconcebidas, idéias e intuições acerca da relação entre os dados.

A mineração de dados pode ser aplicada em várias áreas ou setores, dentre eles destacam-se o ramo de alimentos, a segmentação de mercados, o planejamento da produção industrial, a previsão do volume de vendas, a previsão do mercado financeiro, governamental, telefonia, cartões de crédito, bancos, televisão por assinatura, comércio eletrônico e vários outros segmentos que possuam grandes massas de dados, visto que, a mineração de dados necessita disto para calibrar os algoritmos e extrair dos dados conclusões confiáveis.

Para o centro de computação paralela da universidade de Belfast (2003), a aplicação da mineração de dados também pode ser realizada com sucesso nas seguintes áreas.

- Medicina - efeitos colaterais de drogas, análise de custom de hospitais, análise genética, predição etc.
- Marketing – análise de produto, padrões de compra, predição de vendas etc.
- Áreas científicas – aquisição de conhecimento, pesquisas etc.
- Engenharia - criando sistemas especialistas, diagnóstico de automóvel, descoberta de falhas etc.

## 4.7 Descoberta de Conhecimento em Banco de dados (KDD)

Para o mesmo Robin et.al.(1999), o aproveitamento da informação permite ganho de competitividade: “*conhecimento é poder (e poder = \$\$!)*”. Recursos de análise de dados tradicionais são inviáveis para acompanhar esta evolução. A exemplo do processo iterativo de criação, teste e refinamento de hipóteses.

*“Morrendo de sede por conhecimento em um oceano de dados”.*

São exemplos de volume de informações e dos problemas de análise de dados para tomada de decisão em alguns setores:

- BD da *Wal-Mart*: 20 milhões de transações por dia, no total de 600 milhões de transações mês.
- *Data Warehouse* da Mobil: 100 TB (Tera Bytes);

- BD da NASA: coleta de informações dos satélites 50 GB por hora e 36 TB (Tera Bytes) de informações mês.

Para a maioria dos exemplos, acima citados, que envolvem uma grande análise de dados é preciso de uma solução inteligente como:

- Ferramentas de automatização das tarefas repetitivas e uma sistemática de análise de dados;
- Ferramentas de auxílio para as tarefas cognitivas da análise;
- Integração das ferramentas em sistemas apoiando o processo completo de descoberta de conhecimento para tomada de decisão.

#### 4.7.1 Origem do KDD e suas Fases

Segundo Robin et.al.(1999), a informatização dos meios produtivos proporcionou a geração de grandes volumes de dados através:

- das transações eletrônicas;
- dos novos equipamentos científicos e industriais para observação e controle;
- dos dispositivos de armazenamento em massa (exemplo: coletores de dados, e outros);

A administração e o conhecimento sobre esse grandes volumes de dados deu origem ao chamado KDD (*Knowledge Discovery in Databases*), conforme ilustrado na figura 8.

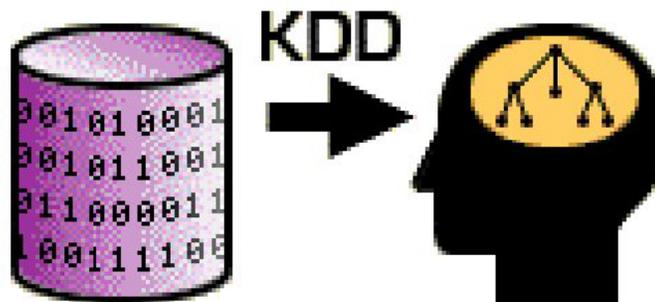


Figura 8 – KDD.  
Fonte: Robin et.al. (1999).

---

Para Back et.al.(2000), o KDD (*Knowledge Discovery in Databases*), teve um rápido crescimento com a primeira definição dada em 1989. O processo KDD tem natureza iterativa e interativa. É dito iterativo, por ser composto de uma série de etapas seqüenciais, podendo haver retorno a etapas anteriores, isto é, as descobertas realizadas (ou a falta delas) podem levar a novas hipóteses de descoberta. Nesse caso, o usuário pode decidir pela retomada dos processos de mineração, ou uma nova seleção de atributos, por exemplo, para validar hipóteses que surgiram durante o processo, por isso é dito interativo.

Para Robin et. al. (1999), o KDD contempla: a compreensão do domínio e dos

objetivos da tarefa, a criação do conjunto de dados envolvendo as variáveis necessárias, as operações como identificação de ruídos (*outliers*) e como tratar a falta de dados em alguns campos, redução de dimensionalidade e combinação de atributos, escolha e execução do algoritmo de acordo com a tarefa a ser cumprida, interpretação dos resultados com possível retorno aos passos anteriores e por último a consolidação através da incorporação e documentação do conhecimento e comunicação aos interessados.

Para Brusso (1998), o processo KDD é cooperativo entre humanos e computadores. Os humanos projetam as bases de dados, descrevem problemas e definem objetivos. Os computadores processam os dados, procuram por padrões que coincidem com as metas estabelecidas.

A figura 9 apresenta uma ilustração do KDD com suas fases, dentre elas, a seleção das informações, o pré-processamento das informações, a transformação dos dados, o processo de mineração de dados (*Data Mining*), a interpretação e avaliação dos resultados e por último o conhecimento adquirido.

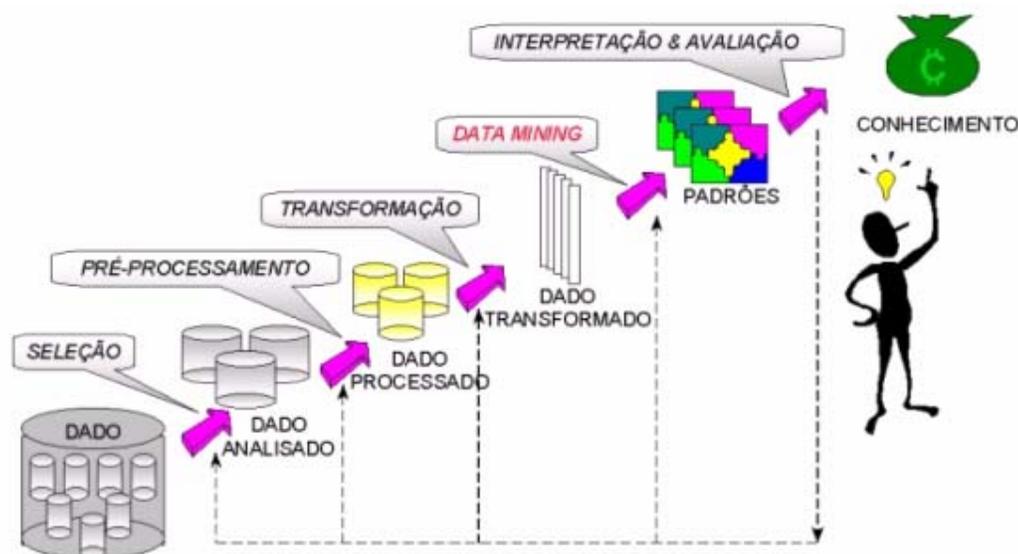


Figura 9 - O processo KDD e suas fases.

Fonte: Adaptado de Fayad (1996).

a) **Seleção:** consiste na compreensão do domínio e dos objetivos da tarefa, criação do conjunto de dados envolvendo as variáveis necessárias. (Robin et.al., 1999)

Para Back et.al. (2000) a seleção de dados deve:

- ✓ definir os objetivos do estudo/hipóteses;
- ✓ definir necessidade de dados para testar as hipóteses;

- ✓ localizar os dados, definir modelo computacional para realizar as consultas às bases de dados;
- ✓ realizar a redução de dimensões;
- ✓ combinar os atributos;
- ✓ requer analista com conhecimento, definindo quais são as informações relevantes para atingir os objetivos;
- ✓ a seleção dos atributos relevantes pode ser automática ou manual;
- ✓ separar dados para treinamento, para testes e para evolução;

Para Robin et.al.(1999), o analista humano com seu conhecimento, decide quais informações relevantes, *a-priori*, podem ser analisadas. A exemplo de informações disponibilizadas em um banco de dados, a qual pode conter vários atributos como: nome, identificação, idade, sexo, estado civil, endereço, renda, residência própria, entre outros.

b) **Pré-Processamento:** são operações que permitem identificação ruídos. *Outliers*<sup>2</sup>, falta de dados em alguns campos, entre outras (Robin et.al., 1999).

Para Back et.al.(2000) o pré-processamento deve:

- obter atividades que visam gerar uma representação conveniente para os algoritmos de mineração;
- eliminar registros incompletos ou inconsistentes, bem como completar registros incompletos quando possível.

c) **Transformação:** tem por objetivo a redução de dimensionalidade e a combinação de atributos. (Robin et.al., 1999).

Para Back et.al., (2000) a transformação deve:

- agrupar informações;
- definir tipo de dados : Numérico contínuo, categórico;
- normalizar informações (feminino=0,masculino=1, indefinido= 0.5);
- agrupar faixas de valores;

O quadro 6 exemplifica a definição e como proceder para padronizar alguns dados antes de submeter-los a uma análise mais detalhada.

---

<sup>2</sup> **Outliers:** por definição, são observações infreqüentes; aponta dados que não aparecem na distribuição característica do resto dos dados. Estes podem refletir propriedades genuínas do fenômeno subjacente (variável), ou devido a medida de erros ou outras anomalias que não deveriam ser modeladas. (StatSoft, 1998)

Quadro 5: Exemplo da Padronização de Dados

Atributo	Tipo de dados	Valores	Representação
Idade	Numérico contínuo	18 – 74	Normalizado para [0;1]
Sexo	Catégorico	Masculino, Feminino, Desconhecido	1, 0, 0.5
Estado Civil	Catégorico	Casado, Solteiro, Divorciado, Desconhecido	1, 0, 0.5
Proprietário da Casa	Catégorico	Sim,Não Desconhecido	1, 0, 0.5
Produtos de esporte (\$)	Numérico contínuo	\$0 a \$ 1500	Normalizado para [0;1]
Equip. de exercício (\$)	Numérico contínuo	\$0 a \$ 2500	Normalizado para [0;1]
Eletrodomésticos (\$)	Numérico contínuo	\$0 a \$ 5000	Normalizado para [0;1]
Entretenimento (\$)	Numérico contínuo	\$0 a \$ 2500	Normalizado para [0;1]
Produtos para o lar (\$)	Numérico contínuo	\$0 a \$ 5000	Normalizado para [0;1]
Total (\$)	Numérico contínuo	\$0 a \$ 12500	Normalizado para [0;1]

Fonte: Robin et.al.(1999).

d) **Mineração de Dados (*Data Mining*)**: fase responsável pela escolha e execução do algoritmo de acordo com a tarefa. (Robin et.al., 1999).

Para Back et.al. (2000), o DM deve:

- selecionar uma técnica / algoritmo / ferramenta apropriada baseado nas características dos dados selecionados e no objetivo do processo;
- aplicar a técnica / algoritmo / ferramenta de aprendizado;
- obter o conjunto de padrões;

e) **Interpretação e Avaliação**: nessa fase procede-se a interpretação dos resultados com possível retorno aos passos anteriores. (Robin et.al., 1999).

Para Back et.al.(2000), a interpretação e avaliação deve:

- com a análise dos resultados, devem ser interpretados os resultados para
- descobrir informações anteriormente desconhecidas e úteis, utilizado para suportar tomadas de decisão;
- deve ser tomado o cuidado com a falta de interpretação ou interpretação incorreta;
- validar o modelo (aplicando dados de teste, por exemplo), e verificar a necessidade de retornar as fases anteriores;

f) **Conhecimento**: Procede-se a consolidação, incorporação e documentação do conhecimento e comunicação aos interessados. (Robin et.al., 1999).

Para Back et.al. (2000), alguns tipos de descobertas podem ter conhecimento de:

- dependências: quando se descobre uma relação entre um atributo e outro;
- descrição de conceitos (aprendizado supervisionado): quando se descobre que dado o valor de um atributo, se define a que classe os indivíduos pertencem (ex.: é animal, tem asas, pode voar -> pássaro);
- identificação de classes (cluster): detectar padrões e montar as classes;
- fórmulas: algoritmos de regressão para obter uma relação matemática entre atributos (linear, quadrática, polinomial, etc.);
- detecção de desvios (*Outliers*): quando se detecta anomalias, desvios, definir os dados que estão fora do padrão (ex.: descobrir fraudes como homem fazendo cesariana);

#### 4.7.2 KDD e Data Mining

Para Back et.al. (2000), o processo KDD trata da descoberta de conhecimento útil dos dados, enquanto a mineração de dados (*Data Mining*) é uma das fases do KDD relacionada a aplicação de algoritmos para extração de padrões dos dados.

Segundo Robin et.al. (1999), a mineração de dados é um passo do processo de KDD que produz um conjunto de padrões sob um custo computacional aceitável. Ainda para Robin et.al. (1999), o KDD (*Knowledge Discovery in Databases*):

- utiliza algoritmos de mineração de dados para extrair padrões classificados como “conhecimento”;
- incorpora também tarefas como escolha do algoritmo adequado, processamento e amostragem de dados e interpretação de resultados;

#### 4.8 Métodos e Técnicas Envolvidas na Mineração de Dados

Alguns especialistas da inteligência artificial, estatística, física e outras áreas afins que trabalham no domínio de dinâmicas não lineares, desenvolveram nos últimos tempos um novo conjunto de métodos lógicos. Esses métodos exigem máquinas de alto desempenho. A mineração de dados utiliza esses métodos para que, a partir de um conjunto de dados, seja possível descobrir uma representação otimizada da sua estrutura. Os principais métodos são de:

✓ Classificação	✓ Associação	✓ Agrupamento
-----------------	--------------	---------------

Carvalho (2001) resume as metodologias adotadas na mineração de dados como: a descoberta não supervisionada de relações, teste de hipótese e a modelagem matemática dos dados.

Para Carvalho (2001, p.21-27), as três possíveis metodologias de mineração de dados necessitam basicamente das mesmas técnicas para sua realização:

- ✓ Classificação;
- ✓ Estimativa;
- ✓ Análise de Afinidade;
- ✓ Análise de Agrupamento;

- ✓ Previsão;
- ✓ Implementação de um protocolo aplicável de mineração de dados.

As técnicas normalmente são utilizadas em conjunto para estudar os dados e achar tendências e padrões nos mesmos. Existem inúmeras técnicas de mineração de dados, dentre elas, apresenta-se as principais, sendo:

- Técnicas de inteligência artificial;
- Técnicas estatísticas (Árvore de decisão, correlação, etc.);
- Indução de regras;
- Regras de Associação;
- Regras de Classificação;
- Regras de Agrupamentos.

#### **4.8.1 Técnicas de Inteligência Artificial**

Para Rich (1988, p.19), embora as técnicas de inteligência artificial devam ser projetadas tendo em vista as restrições impostas pelos problemas de I.A., há certo grau de independência entre os problemas e as técnicas de resolução. Tanto é possível resolver problemas de I.A. sem utilizar técnicas de I.A., como aplicar as técnicas da I.A. à solução de problemas de outras áreas. A fim de tentar caracterizar as técnicas de I.A. da forma mais independente possível, apresenta-se três importantes métodos de IA:

- Busca – Fornece um meio de resolver problemas para os quais não haja uma abordagem mais direta disponível, bem como uma estrutura dentro da qual qualquer técnica direta que estiver disponível pode ser utilizada.
- Utilização do conhecimento – Fornece um meio de resolver problemas complexos explorando as estruturas dos objetos que estão envolvidos.
- Abstração – Fornece um meio de separar características e variações importantes de outras irrelevantes que poderiam, de outro modo, inutilizar qualquer processo.

#### **4.8.2 Técnicas Estatísticas**

Segundo Costa Neto (1977), um estudo estatístico completo que recorra às técnicas da estatística indutiva irá envolver também, direta ou indiretamente, tópicos de estatística descritiva, cálculo de probabilidades e amostragem. Logo, para se relatar a importância da estatística, todos esses assuntos devem ser abordados em maior ou menor grau, dentro de uma seqüência conforme indicado na figura 10.

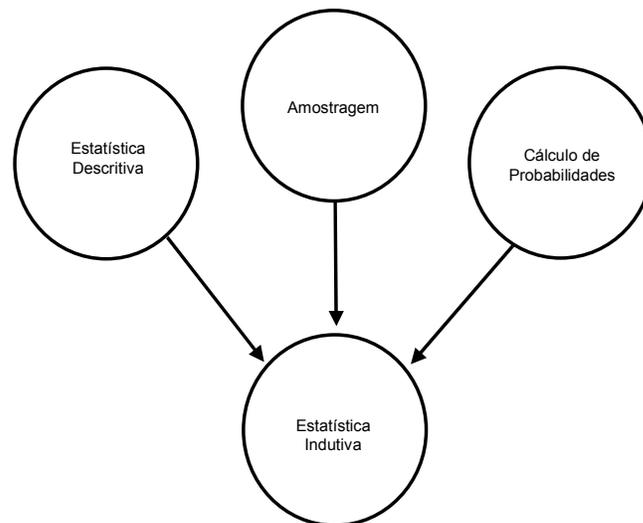


Figura 10 – Esquema resumido das técnicas da estatística.  
Fonte: Costa Neto (1977).

---

### 4.8.3 Indução de Regras

A Indução de Regras, ou *Rule Induction*, tem como objetivo a detecção de tendências em um grupo de informações armazenadas em uma base de dados, ou de “regras” sobre o dado. As regras são, apresentadas na forma de “listas” ou “relações”. Vários algoritmos e índices são utilizados para executar esse processo, tais como: **Gini**, o **C 4.5** e o **CHAID** também chamado por Harrison (1998, p. 197) de indução qui-quadrado automática. Na Indução de Regras, a grande maioria dos processos é realizada pela máquina, e uma pequena parte é feita pelo usuário.

### 4.8.4 Regras de Associação

A prática da mineração de dados tem por objetivo a predição (processo que envolve os atributos do banco de dados para prever valores de uma variável) ou a descrição (determinação de padrões para a descrição dos dados). As regras de associação se encaixam na descrição de dados para futuras predições relevantes na tomada de decisões. Pode-se dizer que este processo está dividido em duas etapas: na primeira são descobertos os conjuntos de itens freqüentes, e na segunda etapa são descobertas as associações a partir destes conjuntos.

*“O grande problema da análise da base de dados é encontrar todas as regras de associações que satisfaçam as especificações de suporte mínimo e da confiabilidade que o usuário determina”.*

Para se fazer associação dos dados conceitualmente, é usada uma tabela onde cada item tem uma coluna própria e cada transação uma linha. Como exemplo temos a associação binária onde, será atribuído o valor 1 para cada item da transação caso ele ocorra e 0 caso contrário.

O algoritmo Apriori é um dos algoritmos mais conhecidos e usados para a mineração de dados. Basicamente, ele percorre todos os dados da base de transações para encontrar os grupos de itens que aparecem com frequência, gerando os itens candidatos e depois verifica se os mesmos possuem o grau de suporte maior ou igual ao determinado pelo usuário. Sendo que ele "acumula" os itens encontrados eliminando os itens que já não satisfazem o percentual a partir da primeira passagem pela base. Para uma melhor compreensão verificar no Anexo 1 o algoritmo Apriori.

Para as regras de associação pode-se identificar o grau de suporte e a confiança para cada regra encontrada pelo algoritmo. Ainda para o Laboratório Central de Informática -UPF (2000), várias são as aplicações das regras de associação existentes entres elas tem-se:

- previsão e diagnóstico de alarmes em telecomunicação;
- análise de admissão em cursos universitários;
- avaliação de solicitação de créditos bancários, seguros, cartões de crédito;
- "descoberta" do perfil do cliente;
- uso de cartões de crédito para prever possíveis produtos de interesse do cliente;
- reconhecer a possibilidade de "choques" de tratamentos anteriores com atuais para pacientes médicos;
- diagnóstico de exames cancerígenos que necessitam de muito tempo para avaliação.

#### **4.8.5 Regras de Classificação**

Segundo Filho (1999), regra de classificação é a técnica que define agrupamento de itens em classes, determinando um modelo de comportamento para cada classe de registros. Neste processo, é escolhida uma variável como atributo meta, sendo que, a regra de classificação avalia o comportamento desta variável em relação as demais.

Segundo Quinlan (1979) o modelo de árvore de decisão é o mais utilizado pela regra de classificação.

Segundo Harrison (1998, p.197), as árvores de decisão apresentam um modelo poderoso produzido por uma classe de técnicas que inclui árvores de regressão e de classificação (CART). Uma das principais vantagens das árvores de decisão é que o

modelo é bem explicável, uma vez que tem a forma de regras explícitas. Isso permite às pessoas avaliarem os resultados, identificando atributos-chave no processo. As próprias regras podem ser expressas facilmente como declarações lógicas, em uma linguagem SQL.

Para Morales et. All (2000), árvores de decisão apresentam um método de aprendizagem supervisionado que constrói árvores de classificação a partir de exemplos. Onde os algoritmos mais utilizados são o ID3, C4.5 ambos de (Quinlan) e o CART de (Breiman). A figura 11 exemplifica conceitualmente o processo utilizado na montagem de uma árvore de decisão.

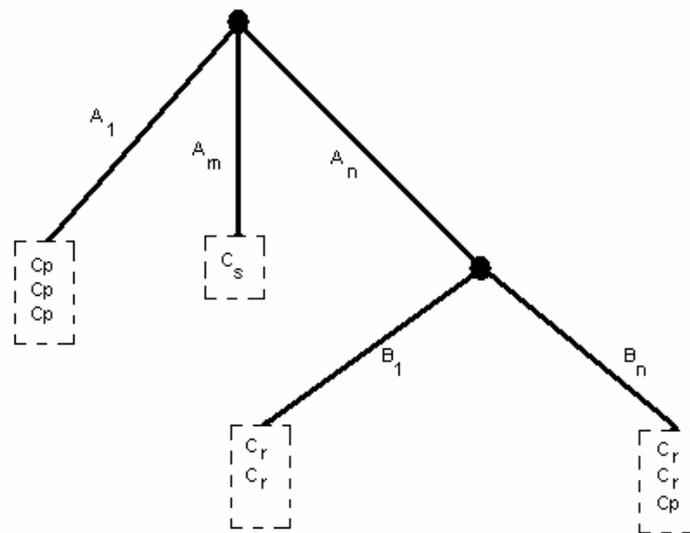


Figura 11 – Árvore de decisão conceitual.  
Fonte Página dedicada ao algoritmo ID3 (1997).

---

O algoritmo ID3 foi apresentado por Quinlan em 1979 para a indução de modelos de classificação, também chamados de árvore de decisão. Dado um conjunto de experiências, cada experiência tendo a mesma estrutura, gera-se um ramo da árvore.(atributo/classe). O problema se concentra em determinar uma árvore de decisão com base em um conjunto de experiências.

Para Morales et. al. (2000), o algoritmo ID3 é simples e potente, constrói uma árvore de decisão sob as seguintes premissas. Cada vértice (nó) corresponde a um atributo, e cada aresta da árvore um valor possível do atributo. Uma folha da árvore corresponde ao valor esperado da decisão segundo os dados de treino utilizados. A explicação de uma determinada decisão está na trajetória da raiz a folha representativa desta decisão. Cada vértice é associado ao atributo mais informativo que ainda não tenha sido considerado. Para medir o nível de informação de um atributo (critério utilizado na seleção do nó de uma árvore de decisão) se utiliza o conceito de entropia da

teoria da informação o qual não fará parte deste conteúdo. Menor o valor da entropia, menor a incerteza e mais utilidade tem o atributo para a classificação. Para cada atributo previsor, deve-se elaborar uma tabela onde as linhas contenham valores do atributo previsor, as colunas devem conter valores do atributo meta (classe), e cada célula deve conter o número de tuplas para a respectiva combinação de valores do atributo previsor e classe. Segundo artigo dedicado ao algoritmo ID3 - PDA (1997) o algoritmo de construção da árvore de decisão (ID3), a cada nó de árvore marcado como nó folha (nó que não é ramificado) deve ser submetido a pergunta. Abaixo é mostrada a estrutura em que o nó folha é avaliado pelo algoritmo e também na figura 12 é exemplificada a aplicação do algoritmo.

**IF** condição nó folha = TRUE **THEN**

    marcar como folha

**ELSE**

    escolher atributo de acordo com a medida.

    expandir nó.

**END**

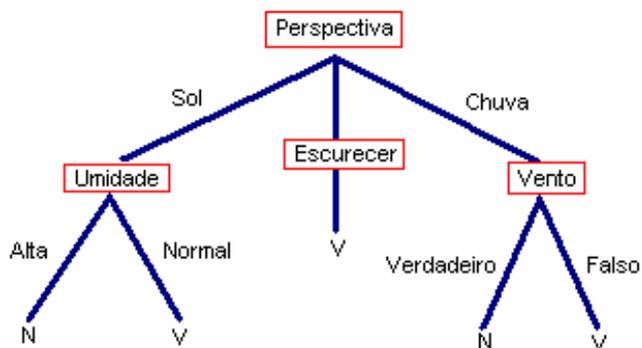


Figura 12 - Exemplo prático de uma árvore de decisão.  
Fonte: Adaptado de Quinlan (1993) pág.19.

#### 4.8.6 Regras de Agrupamento (Clustering)

Conforme Filho (1999), o método de *clustering* é adequado para situações onde

não se conhece a priori as classes de dados, a classificação pode ser útil no processo de *clustering*. Para Johnson (1998), a análise de *clusters* tem como objetivo descobrir agrupamentos naturais de itens.

Segundo Stuker et. al. (2001), com a utilização do cluster é possível encontrar e separar objetos em grupos similares e também encontrar e separar variáveis segundo os valores obtidos pelos objetos.grupamentos.

Dentre os principais procedimentos utilizados para cluster, tem-se:

- ✓ Ligação simples ou do vizinho mais próximo;
- ✓ Ligação completa ou do vizinho mais longe;
- ✓ Centróide;
- ✓ Média das distâncias e Método de Ward.

A figura 13 é uma ilustração da representação visual das regras de agrupamento, utilizando um dendograma para medir a distância euclidiana, sendo possível identificar os agrupamentos de dados.

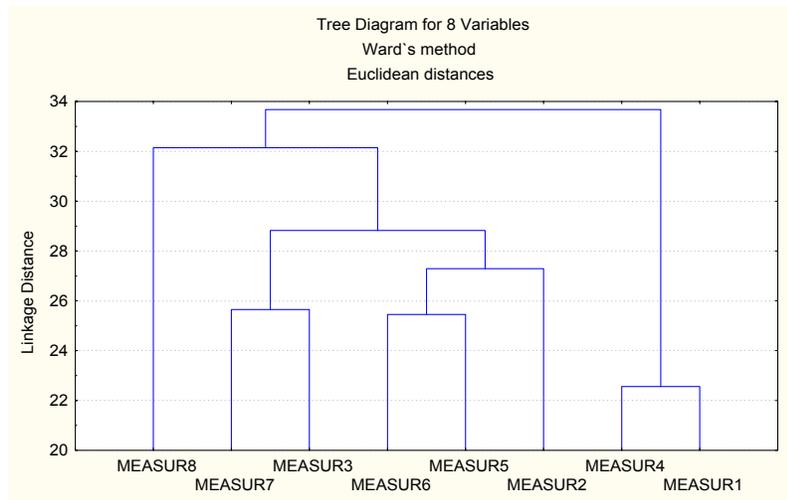


Figura 13 - Dendograma utilizando a distância euclidiana.

Fonte: StatSoft (1998).

## 4.9 Critérios para Escolha da Ferramenta de Mineração de Dados

Para obter os benefícios da mineração de dados, nada melhor que optar pelas ferramentas que disponibilizam de forma fácil à interação entre as técnicas e a descoberta de novos conhecimentos.

Várias ferramentas disponibilizam as técnicas de mineração de dados tendo como característica as diferentes etapas da mineração de dados, algumas contemplam um

número limitado de métodos e algoritmos e outras trazem inovações dos algoritmos discutidos anteriormente. No Anexo 2 encontra-se disponível uma relação e também o endereço de algumas conceituadas ferramentas disponíveis no mercado.

Para Harrison (1998, p.199), como o processo de mineração de dados é cíclico, várias tarefas devem ser executadas, alterando-se entre testes de hipóteses e descoberta de conhecimentos. A fase de descoberta de conhecimento é mais bem atendida pela segmentação, árvore de decisão, redes neurais, análise de vínculos, raciocínio baseado em memória e análise de seleção estatística não-dirigida. Sendo que, a escolha da técnica de mineração de dados e conseqüentemente da ferramenta dependerá da tarefa específica a ser executada e dos dados disponíveis para análise conforme exemplificado na tabela n1. A abordagem proposta por Harrison (1998), divide-se em dois passos:

- a) traduzir o problema de negócio a ser resolvido em séries de tarefas de mineração de dados;
- b) compreender a natureza dos dados disponíveis em termos de conteúdo e tipos de campos de dados e estrutura das relações entre os registros.

Tabela 1: Relacionando Método e Técnicas de Mineração de Dados

Técnica	Tarefa	Classificação	Estimativa	Previsão	Agrupamento por afinidade	Segmentação	Descrição
Estatística padrão	✓		✓	✓	✓	✓	✓
Análise de seleção estatística				✓	✓	✓	✓
Raciocínio baseado em memória	✓			✓	✓	✓	
Algoritmos genéticos	✓			✓			
Detecção de grupos						✓	
Análise de vínculos	✓			✓	✓		
Árvores de decisão	✓			✓		✓	✓
Redes Neurais	✓		✓	✓		✓	

Fonte: Thomas Harrison (1998).

## 5 MODELO PARA UTILIZAÇÃO DA MINERAÇÃO DE DADOS

### 5.1 Introdução

Este capítulo trata especificamente do modelo de aplicação de mineração de dados para esta pesquisa, que por sua vez pode auxiliar na seleção de informações originadas do cadastro técnico urbano, bem como o cruzamento de outras informações que possam ser submetidas à mineração de dados conforme a necessidade e o interesse do especialista.

Através desta pesquisa é possível constatar a utilização dos recursos da mineração de dados junto aos sistemas de informações geográficas, juntamente com a possibilidade de realizar novos experimentos com as informações espaciais, tais como: acompanhamento, auditoria, preservação e prevenção utilizando a base do cadastro técnico urbano, unificando diferentes áreas de atuação e proporcionando ao tomador de decisão mecanismos confiáveis de análise de dados.

O modelo proposto nessa pesquisa contempla seis etapas para se alcançar resultados, ou seja, conhecimento. Conforme ilustrado na figura 14, tem-se:

- Primeira etapa: aquisição da base de dados juntamente com a seleção inicial dos atributos relevantes;
- Segunda etapa: aplicação de um modelo estatístico;
- Terceira etapa: pré-processamento na base de dados;
- Quarta etapa: a geração de um modelo de dados adequado para análise dos algoritmos;
- Quinta etapa: aplicação das técnicas juntamente com as ferramentas de mineração de dados;
- Sexta etapa: apresentação e interpretação dos resultados.

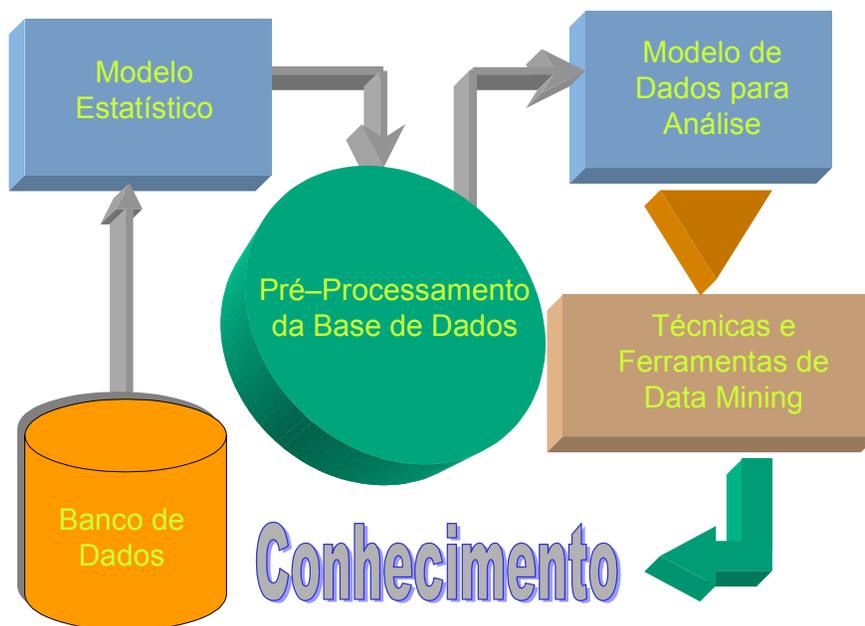


Figura 14 – Modelo proposto para atividades da mineração de dados.

## 5.2 Etapas de Aplicação

As etapas que envolvem a aplicação da mineração de dados propostas para a realização desta pesquisa destacam-se principalmente pela primeira e última etapa, ou seja, pela escolha cuidadosa da base de dados a ser investigada e a interpretação dos resultados obtidos. Para melhor compreensão descreve-se a seguir o procedimento de cada etapa dentro do modelo apresentado.

### 5.2.1 Primeira Etapa – Aquisição da Base de Dados

A primeira etapa refere-se à aquisição da base de dados a ser analisada, bem como a determinação das variáveis que serão investigadas pela mineração de dados, (atributo meta). Nesta etapa procura-se adotar alguns cuidados iniciais com relação a estrutura original do banco de dados, bem como avaliação da estrutura (física e lógica), isto é, para que não ocorra a perda de informações neste processo de extração. A aquisição da base de dados envolve atividades relativas a compreensão do sistema de armazenamento

da base de dados e a extração das informações da mesma. A determinação do atributo meta, esta relacionada com o objetivo central da pesquisa onde se busca levantar todas as informações associadas a ele. Podendo-se complementar a escolha deste atributo na etapa seguinte (segundo etapa) através dos resultados encontrados pelo modelo estatístico aplicado.

### **5.2.2 Segunda Etapa – Aplicação de um Modelo Estatístico**

A segunda etapa refere-se ao modelo estatístico a ser adotado para filtragem das informações, inicialmente é realizada uma pré-análise para obter-se uma visão sobre o conjunto de dados, ou seja, um conhecimento mínimo das informações a serem manipuladas e que serão submetidas à mineração de dados. Recomenda-se a utilização de um modelo estatístico que obtenha informações quantitativas, tais como: percentual, média, mediana, moda e outras que sejam essenciais aos especialistas e que facilitem a interpretação dos dados nas próximas etapas.

### **5.2.3 Terceira Etapa – Pré-Processamento da Base de Dados**

A terceira etapa refere-se ao pré-processamento dos dados, onde procura-se adotar um formato ao conjunto de dados inicial, obedecendo a um padrão com base no algoritmo a ser utilizado, sendo assim possível interpretar os dados sem qualquer ruído ou inconsistência que eventualmente possa existir na base de dados original. Nesta etapa também é identificada a ausência de dados, erros de preenchimento e anomalias diversas que não condizem com o foco de investigação determinado na primeira etapa. Faz parte dessa também a compreensão da codificação utilizada na base de dados.

### **5.2.4 Quarta Etapa – Geração de Modelos**

Na quarta etapa devido a grande variedade de ferramentas e a forma com que cada algoritmo está implementado pelas mesmas, detecta-se a necessidade de criar modelos de dados específicos para a realização de cada experimento. Ou seja, a cada processo de mineração onde o algoritmo escolhido procura por associações, classificações ou agrupamentos entre os dados deve-se gerar um modelo único para cada ferramenta, evitando a perda do seu padrão original e conseqüentemente dúvidas sobre a investigação.

### **5.2.5 Quinta Etapa - Aplicação das Técnicas e das Ferramentas**

Nesta etapa ocorre a aplicação propriamente dita do algoritmo de mineração de dados de acordo com a ferramenta adotada. Sendo que, a escolha da ferramenta pode influenciar muito na sua utilização, proporcionando ou não a facilidade de uso.

### **5.2.6 Sexta Etapa – Apresentação e Interpretação dos Dados**

Nesta etapa faz-se a apresentação e interpretação dos resultados obtidos pela mineração de dados. Busca-se utilizar ferramentas que proporcionem uma melhor compreensão do conhecimento adquirido, entre elas têm-se geradores de regras, visualizadores de árvores, os diagramas de redes e tabelas de associação.

Para consolidar a importância desta etapa cita-se Harrison (1998, p.215), que confirma a representação visual dos dados e das regras como visualizadores de árvores, os visualizadores de regras de associação, os mapas de dados tridimensionais, os sistemas de informações geográficas (GIS) e os diagramas de grupos que estão entre os recursos que podem ajudar na compreensão das relações complexas.

Tendo em vista o papel fundamental da apresentação dos resultados obtidos, bem como facilitar a interpretação do conhecimento encontrado sem a necessidade de profunda experiência técnica. Esta etapa apresenta-se como a mais importante para expressar também o valor que a mineração de dados pode ter no processo de aquisição de novos conhecimentos.

## **6 ESTUDO DE CASO**

### **6.1 Introdução**

Devido à rápida evolução das estruturas urbanas e conseqüentemente ao crescimento das cidades, o surgimento de aglomerados humanos normalmente sem planejamento prévio, apresenta um contexto na maioria das vezes imprevisível e inexplicável, tornando-se impossível administrar tanta informação sem controle. O objetivo deste estudo de caso é demonstrar a aplicação da mineração de dados, utilizando como fonte alguns dados do cadastro técnico urbano, em especial a base de dados de uma prefeitura municipal. Com os resultados alcançados e conseqüentemente o conhecimento descoberto com a utilização dos métodos, proporciona-se ao especialista novos mecanismos de investigação e acompanhamento dos dados.

### **6.2 Aplicação do Modelo Proposto na Base Cadastral**

A base de dados utilizada foi extraída do BIC (Boletim de Informações Cadastrais) da Prefeitura Municipal de Itajaí, localizada no Estado de Santa Catarina – Brasil. As informações constantes no banco de dados foram coletadas através de convênio da Prefeitura Municipal de Itajaí com a Universidade do Vale do Itajaí – UNIVALI, utilizando o BIC como mecanismo principal de coleta de dados junto aos contribuintes, conforme Anexo 4.

Após o estudo realizado no sistema de armazenamento (físico e lógico) dos dados adotado pela prefeitura municipal, se fez a extração dos dados relevantes ao cadastro técnico urbano, proporcionando um rápido entendimento sobre o conjunto de dados. O número de registros existentes na base de dados originados da extração para este estudo de caso totalizou 49.413 cadastros, armazenados em meio digital (texto).

#### **6.2.1 Aquisição e Seleção da Base de Dados**

O processamento de volumosos bancos de dados além de ser um processo automatizado onde o computador é responsável direto pela tarefa, o especialista é parte integrante deste procedimento, pois é o mesmo que define os objetivos a serem alcançados e o que fazer com o conhecimento adquirido. Neste caso o auxílio de um especialista da área de cadastro técnico urbano ou de planejamento, além de filtrar as informações que deverão atender aos requisitos técnicos também contribuem significativamente na validação dos resultados.

Após o contato com alguns especialistas, procedeu-se a seleção dos atributos que serão submetidos aos algoritmos de mineração, um dos passos mais delicados devido a grande diversificação de informações que podem existir em um sistema cadastral. No entanto, foram selecionados alguns atributos que fazem parte do BIC (Boletim de Informações Cadastrais), que possibilitam identificar os dados de cada imóvel cadastrado junto à prefeitura. Desta forma pode se obter as características dos lotes e dos imóveis. Para melhor compreender estes atributos selecionados veja a quadro 7 e também o BIC conforme Anexo 4.

Quadro 6: Atributos do Banco de Dados Analisado

DESCRIÇÃO	TIPO	OPÇÕES
TIPO IMPOSTO	NUMÉRICO	1 ou 2
ISENTO IPTU	NUMÉRICO	0,1 ou 2
ISENTO TAXAS	NUMÉRICO	0 ou 1
PASSEIO	NUMÉRICO	1,2,3,4 ou 5
TOPOGRAFIA	NUMÉRICO	1,2,3 ou 4
SITUACAO LOTE	NUMÉRICO	1,2,3 ou 4
BENFEITORIA	NUMÉRICO	1,2,3 ou 4
TIPO IMOVEL	NUMÉRICO	1,2,3,...,24
SITUACAO CONSTRUCAO	NUMÉRICO	1,2 ou 3
CONSERVACAO	NUMÉRICO	1,2,3 ou 6
ESPECIE	NUMÉRICO	1,2,3,4 ou 5
UTILIZACAO	NUMÉRICO	1,2,3 ou 4
ACABAMENTO	NUMÉRICO	1,2,3,4,5 ou 6
PATRIMONIO	NUMÉRICO	1,2,3,4 ou 5
PAVIMENTOS	NUMÉRICO	Valor
INCRA	NUMÉRICO	0 ou 1
FRENTE	NUMÉRICO	Valor
FUNDOS	NUMÉRICO	Valor
AREA CONSTRUIDA	NUMÉRICO	Valor

Fonte: Boletim de Informações Cadastrais (BIC)

### 6.2.2 Aplicação de um Modelo Estatístico Preliminar

Para obter-se informações preliminares da base de dados optou-se pela aplicação de um modelo estatístico que possibilita a visualização sintetizada da massa de dados através dos campos sumarizados. Esse procedimento é mais representativo para atributos com valores quantitativos, pois além de fornecer informações sobre os mesmos possibilita agir sobre os registros detectados com valores nulos ou incompletos. Sendo que, para os registros que não apresentaram valores (falta de preenchimento), optou-se por não considerá-los para análise estatística, visto não apresentar prejuízos.

Dentre os (49.413) registros integrantes da base de dados original, detecta-se (87) registros com valores nulos e que não atenderam as especificações constantes no boletim de cadastramento da prefeitura. Resultando assim um total de (49.325)

registros.

A informação contida na figura 15 sintetiza os dados quantitativos originados da variável “acabamento”, sendo esta variável escolhida para exemplificar inicialmente uma característica técnica dos imóveis para o ano de 2001.

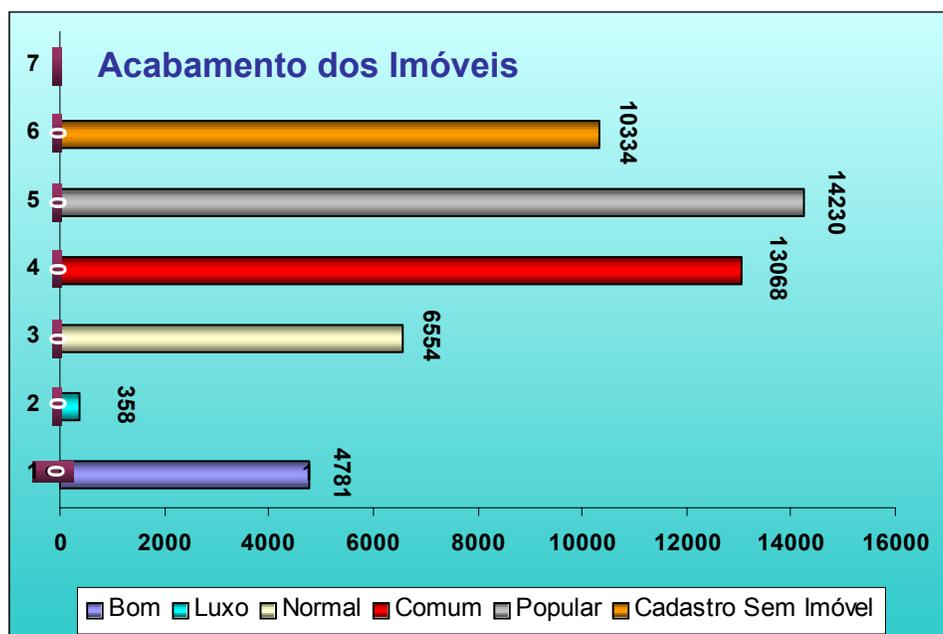


Figura 15 – Demonstração quantitativa do BD com relação ao acabamento dos imóveis.

Outra característica principal na qual o cadastro técnico urbano pode interagir, é auxiliando a administração na fiscalização, tendo como ponto forte a conservação dos imóveis. Conforme levantamento realizado junto ao município de Itajaí em Santa Catarina, identificou-se uma grande carência quanto a este fato onde existem inúmeras ações preventivas que podem ser facilmente acionadas, mediante a utilização das técnicas de mineração de dados.

Na figura 16 apresenta-se em percentuais a distribuição dos dados da variável “conservação”, escolhida para este processo de mineração como atributo “meta”, sendo que, o estado regular e a má conservação dos imóveis totalizam 29% dos casos, contra 50% dos casos com boa conservação.

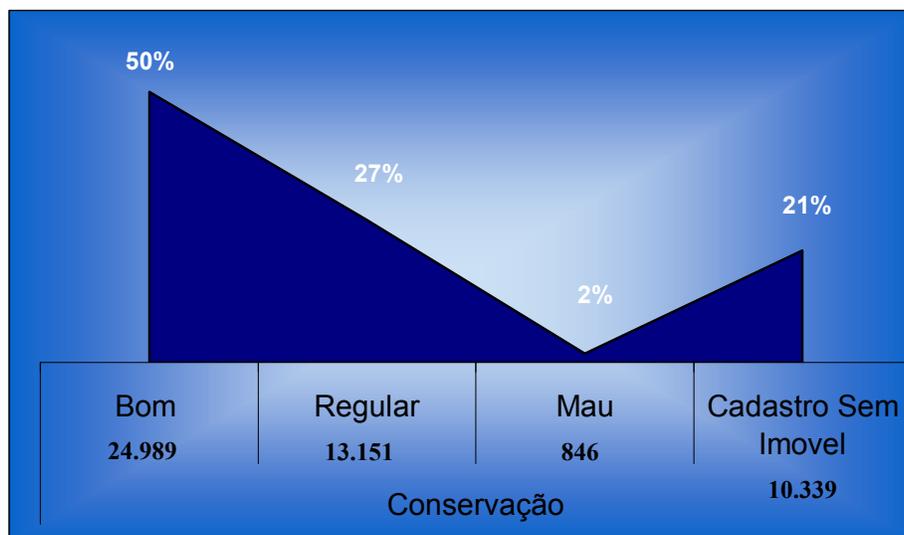


Figura 16 - Estado de conservação dos imóveis.

### 6.2.3 Pré-Processamento dos Dados

Após a análise estatística sobre os 49.413 registros, o pré-processamento se fez necessário para realizar a filtragem dos dados, que foi apontado pelo modelo estatístico, optando-se pela sua eliminação física.

Com a utilização de uma ferramenta de consulta a banco de dados o pré-processamento sobre o conjunto de dados foi realizado de maneira simplificada. A figura 17 apresenta a utilização da ferramenta de banco de dados relacional SQL Server 2000, na qual foi possível também avaliar a consistência das informações armazenadas mediante alguns procedimentos básicos de consulta aos dados. O código SQL descrito no exemplo tem a finalidade de verificar o preenchimento obrigatório de todos os campos que serão aplicados na mineração de dados.

A informação resultante da consulta descarta todos os registros que por algum motivo obtiveram valores nulos para os campos requeridos. A informação selecionada pela consulta indica que se pode aplicar a mineração de dados mediante a necessidade do especialista.

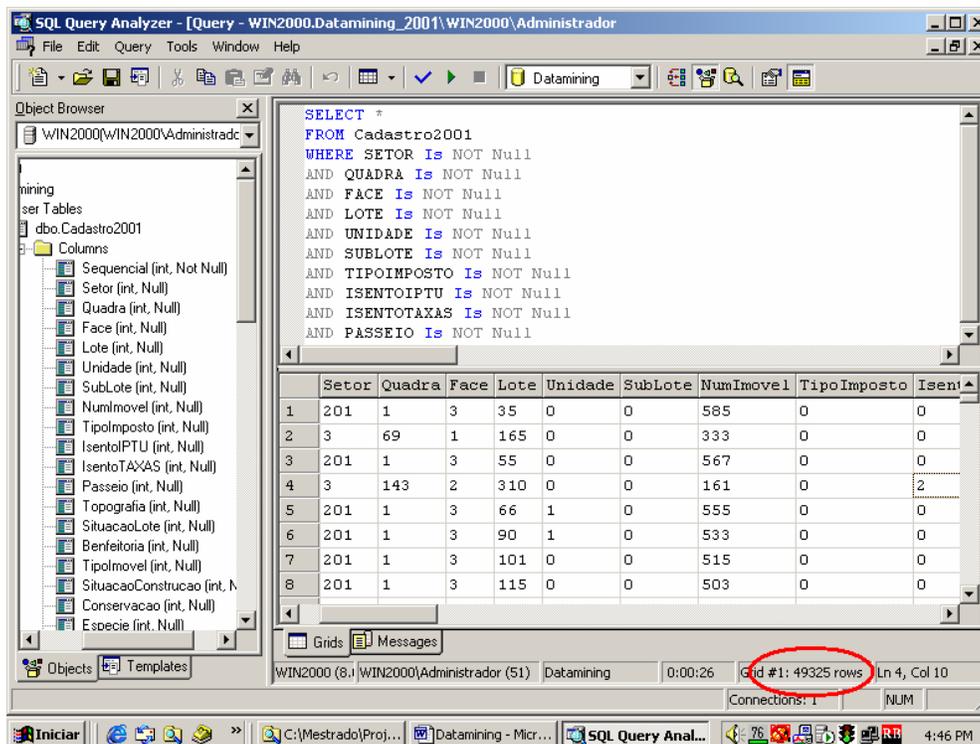


Figura 17 - Avaliação da base de dados.

Fonte: Software – SQL Server 2000.

## 6.2.4 Geração do Modelo e a Escolha da Ferramenta Utilizada

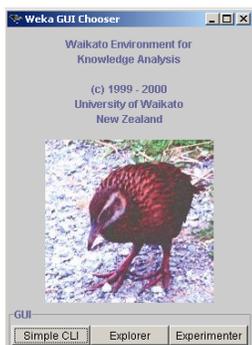
Neste momento gera-se o modelo adequado a mineração de dados a partir da escolha da ferramenta a ser utilizada. Para este estudo de caso três ferramentas foram aplicadas: Weka 3-2, SQL Server2000 e por último o SIPINA for Windows 2-5.

### 6.2.4.1 Weka 3.2

A ferramenta Weka mostrada na figura 18 possui código aberto possibilitando o gerenciamento dos recursos implementados. Possui também a capacidade de processar grande quantidade de registros não apresentando limitação.

Segundo (Witten, 1999), a ferramenta desenvolvida pela Universidade de *Waikato* na Nova Zelândia. “*Weka*” representa o *Waikato* um ambiente para análise de conhecimento. O termo *Weka* adotado originou-se de um pássaro natural só encontrado nas ilhas da Nova Zelândia. O software desenvolvido na linguagem Java (orientado a objeto), tecnologia disponível para a maioria das principais plataformas operacionais. A qual permite prover uma interface uniforme para muitos algoritmos de aprendizagem, juntamente com métodos de pré e pós-processamento para avaliar o resultado de

esquemas de aprendizagem em qualquer *dataset* (conjunto de dados). O enfoque principal do sistema *Weka* está em ser um classificador com os algoritmos de filtro. Porém, também inclui implementação de algoritmos para aprender com associação de regras, como também agrupar dados para os quais nenhum valor de classe é especificado.



**WEKA**  
Versão 3-2  
(09/08/2001)

<http://www.cs.waikato.ac.nz/~ml/>



Figura 18 – Ferramenta selecionada.

Os algoritmos existentes e utilizados pela classificação dos dados são apresentados no quadro 8.

#### Quadro 7: Algoritmos Identificados

<b>Algoritmos de Classificação</b>		
ZeroR	LWR	ClassificationViaRegression
DecisionStump	m5.M5Prime	CostSensitiveClassifier
DecisionTable	NaiveBayesSimple	CVPParameterSelection
HyperPipes	NaiveBayes	FilteredClassifier
JB1	OneR	LogitBoost
JBk	Prism	MetaCost
Id3	SMO	MultiClassClassifier
j48.J48	VotedPerceptron	MultiScheme
j48.PART	VR	RegressionByDiscretization
KernelDensity	AdaBoostM1	Stacking
kstar.KSTAR	AdditiveRegression	ThresholdSelector
LinearRegression	AttributeSelectedClassifier	UserClassifier
Logistic	Bagging	
<b>Algoritmo de Associação</b>		
APRIORI		
<b>Algoritmo de Cluster</b>		
EM – COBWEB		

Fonte: Software – Weka versão 3-2.

Sendo que, antes de realizar a transformação do conjunto de dados gerando um modelo para o formato padrão da ferramenta, é recomendável submeter o mesmo conjunto ao módulo do programa da ferramenta escolhida, neste caso, o *software* Weka chamado *Experiment Environment* (ambiente de experimento), que dispõem o conjunto de dados de cada estudo, realizando a integridade entre atributos e dados, interrompendo o processamento caso encontre alguma incoerência. Este recurso é muito valioso para que o algoritmo escolhido tenha sucesso na leitura do arquivo de origem.

O formato do arquivo ARFF pressupõem a utilização do programa *Weka* para sua utilização. Este arquivo possui linhas começando com um símbolo % são comentários, após os comentários vem o nome da relação, identificado por uma linha começando com @*relation* nome do arquivo. A figura 19 exemplifica o formato do arquivo padrão interpretado pela ferramenta.

```

@relation algoritmo
@attribute bairro {arraialdcunha, bairronaocadastrad, barradorio, cabecudas, canha}
@attribute isentoiptu {nao, sim, recadastramento}
@attribute isentotaxas {nao, sim}
@attribute passeio {cimento, ceramica, terra, pedra, outros}
@attribute topografia {aonivel, acimanivel, abaixonivel, alagado}
@attribute situacaolote {encravado, esquina, meioquadra, beco}
@attribute benfeitoria {normal, smuro-spasseio, sempasseio, semmuro}
@attribute tipoimovel {apto, casa, barraco, deposito-aberto, galpao-aberto, comercio,
@attribute situacaoconstrucao {frente, fundos, cadastrosemimovel}
@attribute conservacao {bom, regular, mau, cadastrosemimovel}
@attribute especie {alvenaria, madeira, mista, barraco, cadastrosemimovel}
@attribute utilizacao {utiliz-propria, alugada, fechada, cadastrosemimovel}
@attribute acabamento {bom, luxo, normal, comum, popular, cadastrosemimovel}
@attribute patrimonio {particular, condominio, func-prefeitura, isento-iptu, isento-ip
@attribute incra {nao, sim}
@attribute frente real
@attribute fundos real
@attribute areaconstruida real
@data
arraialdcunha,nao,nao,cimento,aonivel,meioquadra,normal,posto-servico,frente,bom,alvenaria,util
bairronaocadastrad,nao,nao,cimento,aonivel,esquina,normal,comercio,frente,bom,alvenaria,utiliz-
bairronaocadastrad,nao,nao,cimento,aonivel,meioquadra,normal,casa,frente,bom,mista,utiliz-prop
bairronaocadastrad,nao,nao,terra,aonivel,meioquadra,smuro-spasseio,casa,frente,bom,madeira,util
bairronaocadastrad,nao,nao,terra,aonivel,meioquadra,smuro-spasseio,casa,frente,regular,madeira,
bairronaocadastrad,nao,nao,terra,aonivel,meioquadra,smuro-spasseio,casa,frente,bom,mista,utiliz
bairronaocadastrad,nao,nao,terra,aonivel,meioquadra,smuro-spasseio,casa,frente,bom,mista,utiliz
bairronaocadastrad,nao,nao,terra,aonivel,meioquadra,smuro-spasseio,deposito-aberto,frente,regul

```

Figura 19 - Formato do arquivo arff, utilizado para importar os dados no software Weka.

Fonte: Software – Weka versão 3-2.

#### 6.2.4.2 SQL Server 2000

Outra importante ferramenta utilizada nesta etapa foi o *software* (SQLServer 2000 - *Analisy Services*) figura 20, desenvolvida pela Microsoft para o ambiente *Windows*. Possibilitando realizar a importação dos dados sem muito trabalho, pois utiliza para este processo o ODBC (drive padrão de conexão de dados), assim realizando a conexão com

a base de dados para este estudo de caso.



Figura 20 – Software SQLServer 2000.

---

#### 6.2.4.3 SIPINA

A figura 21 apresenta a ferramenta Sipina (*knowledge Discovery in Databases. University of Lyon (França)*), desenvolvida para o ambiente *Windows* foi uma escolha devido a facilidade de manuseio e por trabalhar também com o algoritmo ID3, podendo assim efetuar o comparativo entre as diferentes informações sobre a base de dados estudada. Para melhorar compreender as regras geradas pelo Sipina, sugere-se a utilização do boletim de informações cadastrais no Anexo 4 para consulta, observando os valores para cada atributo descrito na regra. Embora seja uma ferramenta de distribuição gratuita a desvantagem encontrada na mesma foi a limitação com relação ao número máximo de 16.384 atributos aceitos em um experimento.

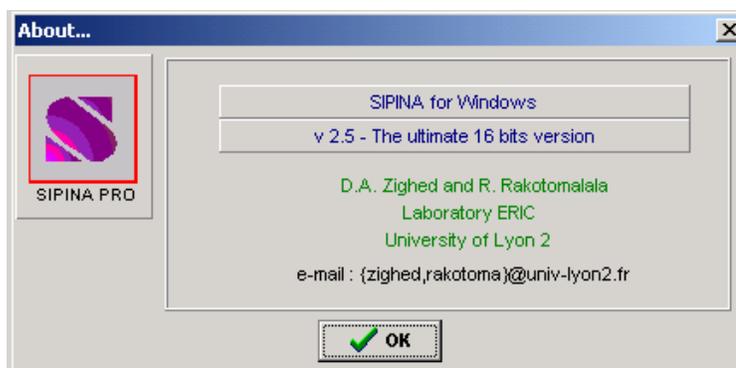


Figura 21 – Software SIPINA for Windows.

---

Para o *software* SIPINA a geração do formato do arquivo pode seguir as mesmas especificações adotadas para o drive ODBC conforme comentado para o software SQLServer.

### 6.2.5 Aplicação das Técnicas e Ferramentas

A partir do conjunto de dados determinado e formatado, elaborou-se diversos estudos, buscando ao mesmo tempo obter resultados verídicos e verificar o comportamento do algoritmo ID3, na versão programada para linguagem JAVA.

Num primeiro estudo, foi utilizado o arquivo completo, para verificar o comportamento do algoritmo ID3 e conhecer a base de dados do cadastro técnico urbano.

Alguns procedimentos foram realizados para que a aplicação do algoritmo ID3 pudesse apresentar resultados de forma clara e objetiva, ou seja, conhecimento ao analista. Nesta etapa, detectou-se que os dados já formatados no padrão exigido pela ferramenta adotada, ainda não se encontravam compatíveis para que o algoritmo escolhido pudesse executar o experimento a procura de regras desconhecidas, impossibilitando assim comparar tais regras com a realidade até então vivenciada pelos especialistas do cadastro técnico urbano.

Após algumas experiências foi possível encontrar o padrão adequado para o conjunto de dados a ser minerado pelo algoritmo ID3, sendo que, na ferramenta Weka 3.2 o respectivo algoritmo identificou melhor a base estudada com os campos de característica nominal, ou seja, valores descritivos.

Com base nos dados estatísticos coletados na primeira fase do método proposto, os dados avaliados apresentaram características de domínio dos especialistas. Após a aplicação do algoritmo de mineração junto ao modelo de dados filtrado, o conhecimento encontrado apresentou um novo conceito sobre a visão de domínio. Conforme exemplificadas logo abaixo, as regras são interpretadas com base na quantidade de registros classificados pelo algoritmo, sendo atribuído um grau de precisão para as respectivas regras. A figura 22 apresenta a forma na qual a ferramenta *Weka* processa o algoritmo gerando as regras para análise.

```
IF espécie = alvenaria THEN
| IF acabamento = luxo THEN
| | IF patrimônio = particular THEN
| | | Estado de Conservação = BOM => com 0,76% Precisão
```

```

*** Run information ***
Scheme: weka.classifiers.Id3
Relation: contactcad
Instances: 49325
Attributes: 5
especie
conservacao <----| Variável Analisada
incra
patrimonio
acabamento

Test mode: 10-fold cross-validation

*** Classifier model (full training set) ***
especie = alvenaria
| acabamento = luxo
| | patrimonio = particular: bom
| | patrimonio = condominio: null
| | patrimonio = isentoiptu: bom
| | patrimonio = funcprefeitura: bom
| | patrimonio = isentoiptutx: bom
| acabamento = bom
| | patrimonio = particular
| | | incra = nao: bom
| | | incra = sim: bom
| | patrimonio = condominio: null
| | patrimonio = isentoiptu: bom
| | patrimonio = funcprefeitura: null
| | patrimonio = isentoiptutx: bom
    
```

```

=== Run information ===
Scheme: weka.classifiers.Id3
Relation: contactcad
Instances: 49325
Attributes: 5
especie
conservacao <----| Variável Analisada
incra
patrimonio
acabamento

Test mode: 10-fold cross-validation

*** Classifier model (full training set) *** Id3 ***
especie = alvenaria
| acabamento = luxo
| | patrimonio = particular: bom
| | patrimonio = condominio: null
| | patrimonio = isentoiptu: bom
| | patrimonio = funcprefeitura: bom
| | patrimonio = isentoiptutx: bom
| acabamento = bom
| | patrimonio = particular
| | | incra = nao: bom
| | | incra = sim: bom
| | patrimonio = condominio: null
| | patrimonio = isentoiptu: bom
| | patrimonio = funcprefeitura: null
| | patrimonio = isentoiptutx: bom
    
```

```

Time taken to build model: 8.32 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      38885      78.8343 %
Incorrectly Classified Instances    10436      21.1576 %
Kappa statistic                    0.649
Mean absolute error                 0.1478
Root mean squared error             0.2721
Relative absolute error              47.0635 %
Root relative squared error         68.6643 %
Unclassified Instances              4          0.0081 %
Total Number of Instances           49325

=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.892    0.288    0.761     0.892   0.821      bom
0.476    0.095    0.646    0.476  0.548      regular
0        0        0        0      0          mau
1        0        1        1      1          cadastrosemimovel

===== Confusion Matrix =====
a  b  c  d  -- classified as
22291 2692 0  2 | a = bom.....**possui o total de 24985 registros - 24989 = 4 rows
6896 6255 0  0 | b = regular.....**possui o total de 13151 registros
112 733 0  1 | c = mau.....**possui o total de 846 registros
0  0  0 10339 | d = cadastrosemimovel..**possui o total de 10339 registros

**** Os resultados apresentados na matrix confusion, demonstra o número de vezes que
ocorreu a combinação dos dados para cada situação de "conservação do imóvel".
    
```

Figura 22 - Processo realizado pela ferramenta Weka utilizando o algoritmo ID3.

## 6.2.6 Apresentação e Interpretação dos Dados

### 6.2.6.1 Representação Espacial

O resultado gerado pelo algoritmo ID3, pode ser interpretado pelo especialista simplesmente com base na interpretação das regras ou se preferir com uma representação espacial através de um SIG, conforme as figuras 23 e 24, nelas são mostradas a situação espacial inicial com base nos itens armazenados no banco de dados sem a aplicação do algoritmo ID3 e após a sua aplicação. Neste momento é possível identificar facilmente a dimensão do conhecimento através das regras e da representação espacial.

Para um sistema de informação geográfica as interpretações das regras podem ser facilmente implementadas e representadas espacialmente desde que se possa a partir da regra gerada representar o conjunto de dados que a originou. Em determinados casos a possibilidade de representar espacialmente, às regras obtidas pelos algoritmos agrada pelo fato de simplificar o processo de amostragem, facilitando a descoberta de conhecimento para a tomada de decisão.



Figura 23 - Representação espacial inicial dos imóveis cadastrados no BD, caracterizados com boa conservação.

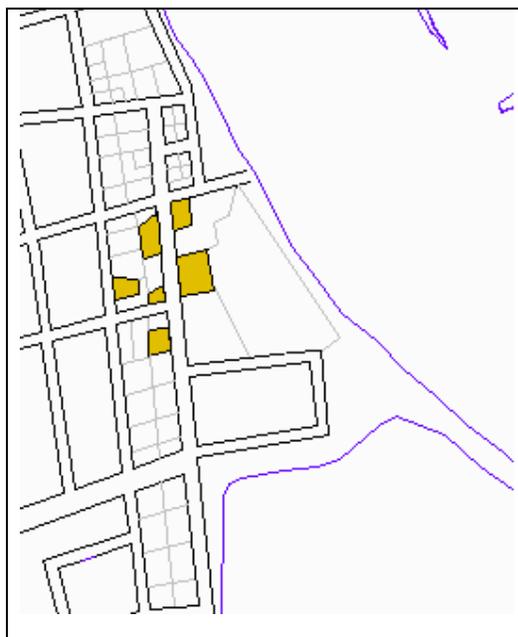


Figura 24 - Representação espacial descoberta após a mineração dos dados, para os imóveis que apresentam boa conservação, com base nas regras.

### 6.2.6.2 Representação do Conhecimento com SQL Server 2000

Outra avaliação efetuada sobre os 49.325 registros contidos na base de dados, foi realizada com a ferramenta (SQLServer 2000 - *Analisis Services*), esse procedimento possibilitou realizar comparativos de grande similaridade entre o conhecimento descoberto no processamento da ferramenta *Weka*, juntamente com o resultado obtido na árvore de decisão montada pela ferramenta SQLServer 2000, conforme representado na figura 25.

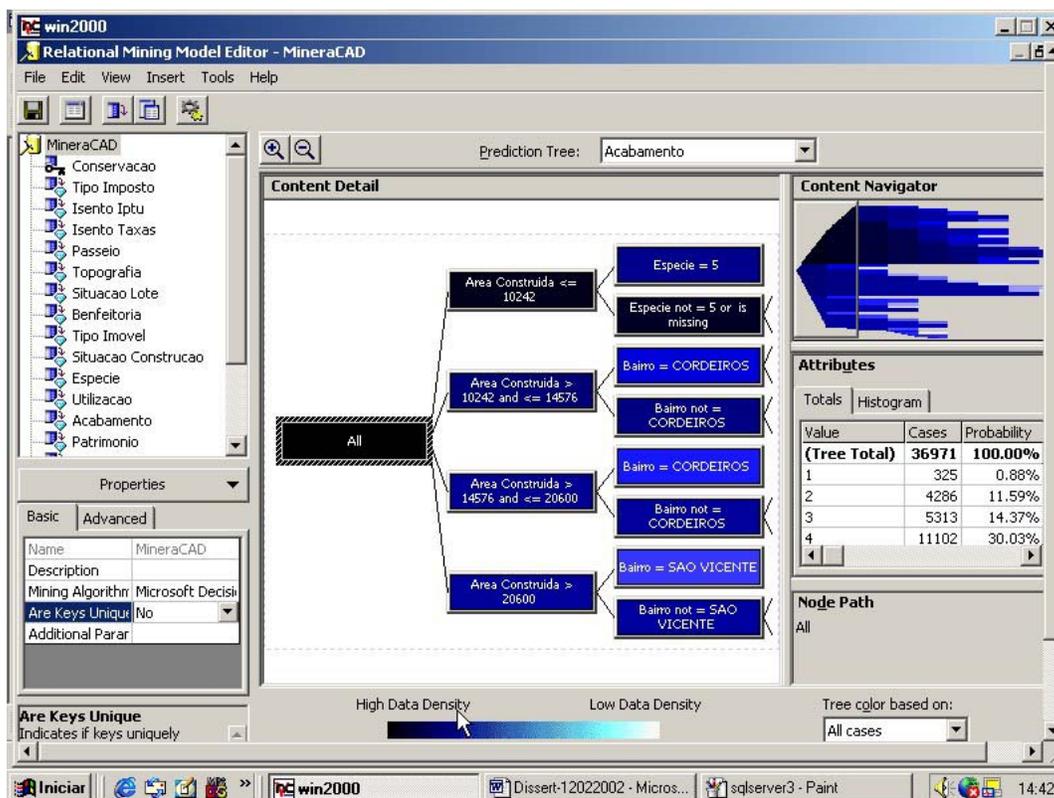


Figura 25 – Árvore de decisão montada pela ferramenta SQLServer 2000.

Fonte: Software - SQL Server 2000.

### 6.2.6.3 Representação do Conhecimento SIPINA x WEKA

Como intuito de interpretar melhor as regras geradas pela ferramenta Weka e também pelo software SQLServer, optou-se pela aplicação de uma terceira ferramenta para esclarecer algumas informações apresentadas anteriormente.

Neste experimento, procura-se comparar as regras geradas por ambas as

ferramentas relatadas anteriormente. Cada análise realizada sobre as regras proporcionou uma amplitude de investigação e interpretação, da base de dados focada como objeto de estudo. Para aplicação do algoritmo ID3 em cada ferramenta, utilizou-se a mesma quantidade de registros armazenada na base de dados, sendo que, devido algumas limitações da ferramenta *Sipina for Windows*, a redução de 49.325 para 15.000 registros foi adotada para que ambas as ferramentas pudessem executar a mineração de dados com sucesso.

Na figura 26, pode-se identificar a árvore de decisão montada pelo algoritmo ID3, através da ferramenta *Sipina*, e juntamente as regras que podem ser interpretadas pelos especialistas, sem que os mesmos tenham domínio da tecnologia utilizada, mais sim, um bom conhecimento das atividades que envolvem o seu trabalho e conseqüentemente o seu negócio. No quadro 9 encontra-se o resultado do mesmo experimento aplicado com a ferramenta *Weka*. Através deste experimento, o qual se pode concluir que além das regras apresentarem as mesmas características em ambas as ferramentas, a sua representação varia na forma de sua visualização final. Na ferramenta *Sipina*, a possibilidade de analisar a montagem da árvore conforme a figura 26, é um recurso muito interessante, pois facilita a interpretação das regras. Uma desvantagem detectada na utilização de ambas as ferramentas foi à identificação da complexidade que acompanham as regras, conseqüentemente dificultando a sua interpretação. Na ferramenta *Weka 3.2* a visualização da árvore não está disponível, mas em contrapartida a presença da matriz confusão.

if ACABAMENTO=1 and ESPECIE=1 and PATRIMONIO=1 then  
 CONSERVACAO=BOM with <0.92#1692#0.086297#1.000000>{1557;132;0;3}  
 if ACABAMENTO=2 and ESPECIE=1 and PATRIMONIO=1 then  
 CONSERVACAO=BOM with  
 <0.76#2805#0.081768#1.000000>{2119;674;0;12}  
 if ESPECIE=2 and PATRIMONIO=1 then  
 CONSERVACAO=BOM with  
 <0.57#1217#-0.005703#0.091719>{695;509;0;13}  
 if ESPECIE=3 and PATRIMONIO=1 then  
 CONSERVACAO=REGULAR with  
 <0.67#1599#0.153647#1.000000>{396;1075;0;128}  
 if ESPECIE=4 then  
 CONSERVACAO=CADASTROSEMIMOVEL with  
 <1.00#2428#0.426599#1.000000>{1;1;2425;1}  
 if ACABAMENTO=4 and ESPECIE=1 and PATRIMONIO=1 then  
 CONSERVACAO=BOM with  
 <0.73#3072#0.078239#1.000000>{2254;810;0;8}  
 if ACABAMENTO=3 and ESPECIE=1 and PATRIMONIO=1 then  
 CONSERVACAO=BOM with  
 <0.90#1705#0.084729#1.000000>{1543;155;0;7}  
 if ACABAMENTO=4 and ESPECIE=1 and PATRIMONIO=2 then  
 CONSERVACAO=BOM with <0.70#10#0.000195#0.575593>{7;3;0;0}  
 if ACABAMENTO=1 and ESPECIE=1 and PATRIMONIO=3 then  
 CONSERVACAO=BOM with <0.64#67#0.000592#0.725009>{43;21;0;3}  
 if ACABAMENTO=4 and ESPECIE=1 and PATRIMONIO=3 then  
 CONSERVACAO=BOM with <0.84#32#0.001363#0.988906>{27;5;0;0}  
 if ACABAMENTO=3 and ESPECIE=1 and PATRIMONIO=2 then  
 CONSERVACAO=BOM with <0.89#27#0.001298#0.994802>{24;3;0;0}  
 if ACABAMENTO=6 and ESPECIE=1 and PATRIMONIO=1 then  
 CONSERVACAO=BOM with <0.98#164#0.008639#1.000000>{161;2;0;1}  
 if ESPECIE=3 and PATRIMONIO=3 then  
 CONSERVACAO=REGULAR with  
 <0.57#47#0.003510#0.996402>{11;27;0;9}  
 if ACABAMENTO=1 and ESPECIE=1 and PATRIMONIO=2 then  
 CONSERVACAO=BOM with  
 <0.83#30#0.001235#0.981372>{25;5;0;0}  
 if ESPECIE=2 and PATRIMONIO=2 then  
 CONSERVACAO=BOM with  
 <0.77#13#0.000409#0.769677>{10;3;0;0}  
 if ACABAMENTO=3 and ESPECIE=1 and PATRIMONIO=3 then  
 CONSERVACAO=BOM with  
 <0.88#24#0.001117#0.987251>{21;3;0;0}  
 if ACABAMENTO=6 and ESPECIE=1 and PATRIMONIO=2 then  
 CONSERVACAO=BOM with <1.00#4#0.000200#0.801837>{4;0;0;0}  
 if ACABAMENTO=2 and ESPECIE=1 and PATRIMONIO=3 then  
 CONSERVACAO=BOM with <0.58#19#-0.000060#0.364084>{11;7;0;1}  
 if ESPECIE=2 and PATRIMONIO=3 then  
 CONSERVACAO=REGULAR with  
 <0.67#21#0.001993#0.990903>{7;14;0;0}  
 if ESPECIE=3 and PATRIMONIO=2 then  
 CONSERVACAO=REGULAR with <0.86#7#0.000920#0.970746>{1;6;0;0}  
 if ACABAMENTO=6 and ESPECIE=1 and PATRIMONIO=3 then  
 CONSERVACAO=BOM with <0.75#8#0.000226#0.627833>{6;2;0;0}  
 if ACABAMENTO=2 and ESPECIE=1 and PATRIMONIO=2 then  
 CONSERVACAO=BOM with <1.00#7#0.000349#0.941144>{7;0;0;0}  
 if ACABAMENTO=5 and ESPECIE=1 then  
 CONSERVACAO=REGULAR with <1.00#2#0.000282#0.785361>{0;2;0;0}

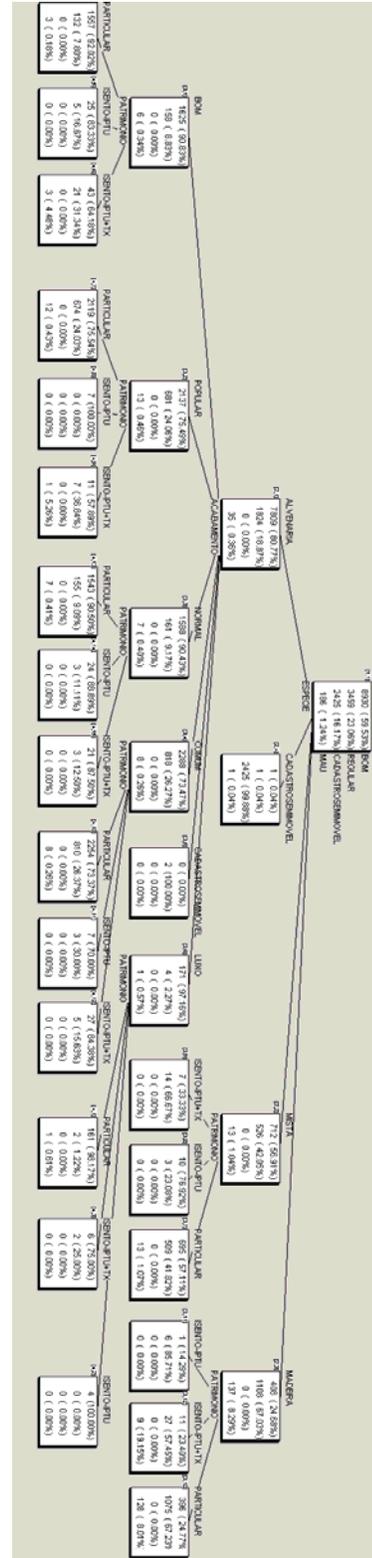


Figura 26 – Árvore construída pela ferramenta Sipina com as respectivas regras.

## Quadro 8: Experiência Sipina x Weka

Resultado da experiência com a ferramenta Weka	
==== Run information =====	
Scheme: weka.classifiers.Id3	patrimônio = func-prefeitura: null
Instances: 15000	patrimônio = isento-iptu: regular
Attributes: 5	patrimônio = isento-iptu+tx: regular
especie conservacao incra patrimonio acabamentoo	acabamentoo = cadastrosemimovel: null
	especie = mista
	acabamentoo = bom
Test mode: 10-fold cross-validation	patrimônio = particular: bom
==== Classifier model (full training set) ==== Id3	patrimônio = condominio: null
especie = alvenaria	patrimônio = func-prefeitura: null
acabamentoo = bom	patrimônio = isento-iptu: bom
patrimônio = particular	patrimônio = isento-iptu+tx: bom
incra = nao: bom	incra = nao: bom
incra = sim: bom	incra = sim: bom
patrimônio = condominio: null	patrimônio = condominio: null
patrimônio = func-prefeitura: null	patrimônio = func-prefeitura: null
patrimônio = isento-iptu: bom	patrimônio = isento-iptu: null
patrimônio = isento-iptu+tx: bom	patrimônio = isento-iptu+tx: bom
acabamentoo = luxo	acabamentoo = normal
patrimônio = particular: bom	patrimônio = particular: bom
patrimônio = condominio: null	patrimônio = condominio: null
patrimônio = func-prefeitura: null	patrimônio = func-prefeitura: null
patrimônio = isento-iptu: bom	patrimônio = isento-iptu: bom
patrimônio = isento-iptu+tx: bom	patrimônio = isento-iptu+tx: regular
acabamentoo = normal	acabamentoo = comum
patrimônio = particular: bom	patrimônio = particular: bom
patrimônio = condominio: null	patrimônio = condominio: null
patrimônio = func-prefeitura: null	patrimônio = func-prefeitura: null
patrimônio = isento-iptu: bom	patrimônio = isento-iptu: regular
patrimônio = isento-iptu+tx: bom	patrimônio = isento-iptu+tx: regular
acabamentoo = comum	acabamentoo = popular
incra = nao	patrimônio = particular: bom
patrimônio = particular: bom	patrimônio = condominio: null
patrimônio = condominio: null	patrimônio = func-prefeitura: null
patrimônio = func-prefeitura: null	patrimônio = isento-iptu: null
patrimônio = isento-iptu: bom	patrimônio = isento-iptu+tx: regular
patrimônio = isento-iptu+tx: bom	acabamentoo = cadastrosemimovel: null
incra = sim: regular	especie = barraco: null
acabamentoo = popular	especie = cadastrosemimovel
patrimônio = particular: bom	patrimônio = particular: cadastrosemimovel
patrimônio = condominio: null	patrimônio = condominio: null
patrimônio = func-prefeitura: null	patrimônio = func-prefeitura: null
patrimônio = isento-iptu: bom	patrimônio = isento-iptu: cadastrosemimovel
patrimônio = isento-iptu+tx: bom	patrimônio = isento-iptu+tx: cadastrosemimovel
acabamentoo = cadastrosemimovel: regular	
especie = madeira	Time taken to build model: 7.23 seconds
acabamentoo = bom	==== Stratified cross-validation ===== Summary =====
patrimônio = particular: regular	Correctly Classified Instances 12042 80.28 %
patrimônio = condominio: null	Incorrectly Classified Instances 2956 19.7067 %
patrimônio = func-prefeitura: null	Kappa statistic 0.6183
patrimônio = isento-iptu: null	Mean absolute error 0.1431
patrimônio = isento-iptu+tx: bom	Root mean squared error 0.2682
acabamentoo = luxo: bom	Relative absolute error 50.5514 %
acabamentoo = normal	Root relative squared error 71.2907 %
patrimônio = particular: regular	UnClassified Instances 2 0.0133 %
patrimônio = condominio: null	Total Number of Instances 15000
patrimônio = func-prefeitura: null	==== Detailed Accuracy By Class =====
patrimônio = isento-iptu: bom	TP Rate FP Rate Precision Recall F-Measure Class
patrimônio = isento-iptu+tx: regular	0.951 0.392 0.781 0.951 0.858 bom
acabamentoo = comum	0.327 0.05 0.662 0.327 0.437 regular
patrimônio = particular: regular	0 0 0 0 0 mau
patrimônio = condominio: null	1 0 0.999 1 1 cadastrosemimovel
patrimônio = func-prefeitura: null	
patrimônio = isento-iptu: regular	==== Confusion Matrix =====
patrimônio = isento-iptu+tx: regular	a b c d <-- classified as
acabamentoo = popular	8488 440 0 1   a = bom
patrimônio = particular: regular	2329 1130 0 0   b = regular
patrimônio = condominio: null	48 137 0 1   c = mau
patrimônio = func-prefeitura: null	0 0 0 2424   d = cadastrosemimovel

## 7 CONCLUSÕES E RECOMENDAÇÕES

### 7.1 Conclusões

Com base no método adotado foi possível cadenciar os procedimentos para mineração de dados gerando assim um modelo de aplicação, sendo possível desmistificar uma grande variedade de recursos com determinado grau de complexidade que existem a disposição dos especialistas para auxiliar na procura de informações que tragam conhecimento aos tomadores de decisão.

A possibilidade de analisar os dados e juntamente o seu relacionamento com dinamismo é um dos grandes benefícios que a mineração de dados proporciona. Com base nas regras geradas pelo algoritmo ID3 comprovou-se que a implementação e conseqüentemente a representação espacial pode ser realizada em sistemas de informações geográficas com sucesso.

Com relação ao objetivo geral desta pesquisa, procurou-se comprovar que a aplicação dos métodos e conseqüentemente das técnicas de mineração, podem ser positivamente aplicadas no cadastro técnico urbano, levando em consideração as abordagens discutidas nos capítulos anteriores.

Com aplicação e os resultados obtidos no processamento do algoritmo ID3, junto à ferramenta Weka, gerou algumas regras que puderam ser apresentadas, a exemplo levado em consideração somente os imóveis que apresentaram um bom estado de conservação. Demonstrando assim, novas formas para avaliar e conceituar a conservação dos respectivos imóveis e/ou cadastros. Ainda conclui-se através desta pesquisa a fácil interação com diferentes ferramentas de mineração de dados que proporcionam tecnologias distintas, mas que utilizam a princípio algoritmos baseados no ID3 apresentado por Quinlan em 1979, C.45, CART e diversos outros que auxiliam na descoberta de conhecimentos.

Através deste trabalho confirma-se que as tecnologias de mineração de dados não podem ser ignoradas no que diz respeito ao gerenciamento de um grande volume de informação, independentemente dos recursos financeiros de apoio a decisão e as estratégias de atuação, pois as mesmas devem ser baseadas em dados e fatos concretos, para que os resultados esperados sejam alcançados.

## 7.2 Recomendações

Com base no estudo de caso apresentado detectou-se a possibilidade de encontrar novos relacionamentos entre as informações armazenadas pelo cadastro técnico urbano de um município e as informações obtidas pelo questionário do (Censo Demográfico/2000) conforme Anexo 3, realizado periodicamente pelo IBGE. Estas informações disponíveis junto ao Instituto Brasileiro de Geografia e Estatística, são dados sócio-econômicos representados por setores<sup>1</sup>, ao contrário do cadastro técnico urbano onde é possível obter informações técnicas individualizadas de cada cadastro como também de toda demarcação territorial de um município.

Para a continuidade desse trabalho sugere-se a inclusão de novos atributos, principalmente um estudo detalhado sobre o relacionamento entre o cadastro técnico urbano e o levantamento censitário realizado pelo IBGE, objetivando uma melhor extração do conhecimento. A possibilidade de interagir diferentes soluções tecnológicas de mineração de dados que possam ser integradas num contexto, que permite alcançar o conhecimento baseado em fatos reais e facilitando ainda mais as decisões por parte dos responsáveis.

Para melhorar a representação das regras geradas pelo algoritmo ID3 através da ferramenta *Weka*, recomenda-se a implementação de novos componentes devido a limitação dos gráficos fornecidos pela versão 3.2, ou seja, componentes gráficos que possibilitem a representação do conhecimento em diferentes amostragens, a exemplo de facilidade de interpretação com a visualização da árvore de decisão e conseqüentemente do conhecimento encontrado.

E por último recomenda-se a criação de grupos de profissionais nas instituições para discutir e adotar a mineração de dados como forma de inteligência organizacional, para que avaliem e desenvolvam novos mecanismos de investigação de dados, pois o conhecimento é um fator determinante na conquista de novas metas e desafios.

---

<sup>1</sup> **Setor:** Neste caso é visto como a demarcação geográfica onde teremos o agrupamento de informações para uma determinada localização.

## BIBLIOGRAFIA

- ANTONY, R. N., *Planning and control systems: a framework for analysis*. Harvard University Press, 1995.
- ARONOFF, S. *Geographical Information Systems: A management Perspective*. Ottawa, WDI Publications, 1989.
- BALBUINO FILHO, Nicolau. **Direito Imobiliário Registral**. São Paulo: Ed. Saraiva, (p.21-29), 2001.
- BACK, Alexandre Carlos de. Neves, Cloves das. Savi. Elvis Felipe. et all Trabalho da Disciplina de Data Mining - **KDD Descoberta de Conhecimento em Bases de Dados**. Universidade Federal de Santa Catarina – UFSC. 2000.
- BERRY, Michael J.A.(1.997), “*Data Mining Techniques*”, Wiley Computer Publishing.
- BNDES **Ação Administrativa para Recadastramento Imobiliário**. 1999.  
<http://federativo.bndes.gov.br/dicas/F18%20-%20Cadastro%20Municipal.htm>  
 Acesso em 06/11/2002.
- Brand, Estelle & Gerritsen, Rob. *Data Mining and Knowledge Discovery. Exclusive Ore Inc.* (1998). Acesso em <http://www.xore.com>
- BURROUGH, P. A.; *Principles of geographic information system for land resources assessment*. Oxford. Oxford University Press, 1986.
- BERSON, Alex. *Data Warehousing, Datamining, and OLAP*. ISBN 0-07-006272-2. USA, McGraw-Hill, 1998.
- BRUSSO, Marcos José. **O paralelismo na mineração de regras de associação**. Porto Alegre: UFRGS, 1998.
- BURITY, E.F. & SILVA, J.L.N. **Cadastro : Proposta de Padronização de Terminologia**. COBRAC – UFSC. Florianópolis 1998.
- CÂMARA, Gilberto. **Tese de doutorado**. <http://www.dpi.inpe.br/gilberto>, 1995.
- CÂMARA, Gilberto. Artigo – As roupas novas do Imperador (parte II), **Revista GEOinfo**, Ano 3. .13 (p.26), 2000.
- CARVALHO, Luís Alfredo Vidal de, 1960 – **Datamining a mineração de dados no marketing, medicina, economia, engenharia e administração**. Editora Érica, São Paulo, 2001.

- COWEN, D. J. *Gis versus CAD versus DBMS: What are the differences. Photogrammetric Engineering and Remote Sensing*, 1988.
- DALE, Peter F. e MCLAUGHLIN, John D. *Land Information Management: an introduction with special reference to cadastral problems in third world countries*. Clarendon Press, 1990. 265p.
- DCC Departamento de Ciência da Computação - Publicações e Linhas de Pesquisa do Laboratório de Computação Heurística.  
<http://www.lch.dcc.ufmg.br/node8.html>, 2001. Acesso em 21/07/2001.
- DW Brasil - *Decision warehouse. Datamining*. Acesso em 22/07/2001.  
<http://www.dwbrasil.com.br/html/dmining.html>
- ESTEIO ENGENHARIA E AEROLEVANTAMENTOS S.A. **Definição de Cadastro Técnico Multifinalitário (CTM)**.  
[http://www.esteio.com.br/servicos/so\\_cadastro.htm](http://www.esteio.com.br/servicos/so_cadastro.htm). Acesso em 12/02/2003.
- FELDENS, Miguel Artur. *Knowledge discovery in databases*. 1997.  
<http://www.ufrgs.br/~feldens> Acesso em [20 de dezembro de 1998]
- FERRARI, Roberto. **Viagem ao SIG: planejamento estratégico, viabilização, implantação e gerenciamento de sistemas de informação geográfica**. Curitiba: Sagres, 1997.
- FERREIRA, A. B. H. **Novo Dicionário da Língua Portuguesa**. Editora Nova Fronteira. 1986.
- FILHO, Adhemar Maria do Vale. WEBBER, Celso Kopp. CUNHA, Fernanda dos S. Cunha. et. al. Seminário da Disciplina de **Mineração de Dados**, Universidade Federal de Santa Catarina – UFSC. 1999.
- GARCIA, Jose A.G. *The use of orthophotographic techniques in a modern cadastre including a data bank. in Cadastre: various functions, characteristics, techniques and the planning of a land records system*. Canadá, National Council, 1974.
- GOEBEL, Michael & Gruenwald, Le. *A Survey of Data Mining And Knowledge Discovery Software Tools*. 1999.
- GOODCHILD, M; HAINING, R.P.; WISE, S. *Integrating GIS and spatial data analysis problems and possibilities. International Journal of Geographical Information System*, 1992.
- GOODCHILD, M; PASKS, B; STEYART, L. *Environmental Modelling with GIS*. Oxford, Oxford University Press, 1993.

- GROTH, Robert. *Data Mining: a hands-on approach for business professionals*. ISBN 0-13-756412-0. New Jersey, Prentice Hall, 1998
- GRUPO DE SISTEMAS INTELIGENTES. **Mineração de Dados**. <http://www.din.uem.br/~ia/mineracao/geral/index.html>, 1998. Acesso em 01/01/2002.
- HARRISON, Thomas H., **Intranet Data Warehouse Ferramentas e Técnicas para utilização de Data Warehouse na Intranet**. Ed. Berkeley, 1998.
- HENSSEN, Johan L. G. *General aspects of cadastre and land registration, in Cadastre: various functions, characteristics, techniques and the planning of a land records system*. Canadá National Council, 1974.
- HENSSEN, Johan L. G. *Cadastre: indispensable for development* TC Journal, 1990.
- HOCHHEIM, Norberto. **Um método para análise probabilística da viabilidade econômica do cadastro técnico urbano**. Florianópolis, Universidade Federal de Santa Catarina. Tese para concurso de professor Titular, Dezembro,1993.
- HOCHHEIM, Norberto. **Cadastro Técnico Urbano**. Florianópolis, Universidade Federal de Santa Catarina. Notas de Aula, 1996.
- INMON, William H. **Como construir o data warehouse**. Rio de Janeiro. Editora Campus, 1997. Pág. 237 a 253.
- JOHNSON, Richard A., WICHERN, Dean W. *Applied multivariate statistical analysis*. Prentice-Hall, 4. ed., New Jersey, 1998.
- JOLY, Fernand. *A Cartografia*. Papirus, 1997. 136p.
- KDNuggets – **Página destinada ao KDD e a Mineração de dados**. Acesso em 01/01/2000. <http://www.kdnuggets.com>
- KIMBALL, Ralph. REEVES, Laura. ROSS, Margy. Warren Thornthwaite. et. al. *The warehouse lifecycle toolkit*. 1998.
- LCI - Laboratório Central de Informática UPF. **Mineração Sobre Regras De Associação**. <http://lci.upf.tche.br/~6276/si/apriori.html>, 2000. Acesso 21/07/2001.
- OLIVEIRA, Cêurio. **Dicionário Cartográfico**. IBGE, 1993. 645p.
- OWG *Smart Business. Smart Solutions*. Empresa especializada em soluções de Inteligência Empresarial (**Business Intelligence**). <http://www.dwbrasil.com.br/html/dmining.html>, Acesso 22/07/2001

- PÁGINA DEDICADA AL ALGORITMO ID3.  
<http://www.geocities.com/SiliconValley/Network/3764/>, 1997. Acesso 21/07/2001
- QUINLAN, J. Ross. *C4.5 Programs for Machine Learning*, ed. Morgan Kaufmann, 1993.
- RICH, Elaine. **Inteligência Artificial**. São Paulo, McGraw-Hill, 1988. Pág. 06 a 19.
- ROBIN, Jacques, Bezerra, Ricardo. et. all. Disciplina de **Mineração de dados** 99/1.  
<http://www.di.ufpe.br/~compint/aulas-IAS/prolog-991/Welcome.html>, 1999.  
Acesso em 01/05/2000.
- RODRIGUES, P. H. e VILAÇA, S. **Subsídios para a utilização de geoprocessamento em sistemas municipais de informação**. Rio de Janeiro: Revista Administração Municipal, v. 41, n. 211, p. 51-60, abr./jun. 1994.
- SCHOLTEN, Henk. e STILLWELL, John. *Geographical Information Systems: the emerging requirements. In: Geographical Information Systems for Urban and Regional Planning*. 1990. 12p.
- SILVA, I. da; PEIXOTO, R. S. S. L.. **Cadastro - evolução e perspectivas**. Palestra proferida no 2º Congresso Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis. 1996.
- SMITH, T.R.; MENOS, S.; STAR, J.; ESTES, J.E. et. al.; *Requirements and Principles for the Implementation and Construction of Large-Scale Geographical Information Systems*. International Journal of Geographical Information Systems, 1987.
- STATISTICA for Windows (*Computer program manual*) versão 6.0. StatSoft, Inc. 1998. <http://www.statsoft.com>. Acesso em 01/05/2000.
- TUBINO, D. F. **Manual de planejamento e controle da produção**. São Paulo: Atlas, 2000.
- WITTEN, I. H. Franck, Eibe. *Data Mining: Practical machine learning tools and techniques with Java implementations*, 1999.

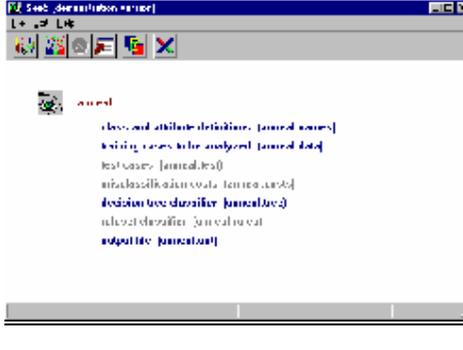
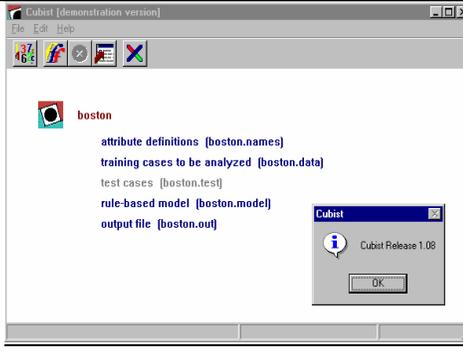
## **ANEXOS**

- **Anexo 1:** Algoritmo APRIORI
- **Anexo 2:** Ferramentas avaliadas
- **Anexo 3:** Questionário aplicado pelo IBGE – 2001
- **Anexo 4:** Boletim de cadastramento BIC
- **Anexo 5:** Termo Projeto de Pesquisa

**Anexo 1: Algoritmo – Procedure Apriori**

```
L1 = { frequent 1- itemsets };
for (k=2; Lk - 1 ≠ ∅; k++) do
  Ck = apriori_gen(Lk - 1);
  forall transactions t ∈ D do
    Ct = subset(Ck, t);
    forall candidates c ∈ Ct do
      c.count++;
    od
  od
  Lk = { c ∈ Ck | c.count ≥ min_sup };
od
Answer = ∪k Lk;
end
```

## Anexo 2: Ferramentas de Investigação

<p>SEE5 Release 1.11</p> <p><a href="http://www.rulequest.com/see5-public.zip">http://www.rulequest.com/see5-public.zip</a></p>	 <p>The screenshot shows the SEE5 software interface. The title bar reads 'SEE5 - Shell (Boston.names)'. The main window contains a list of rules and their associated data files:</p> <ul style="list-style-type: none"> <li>class and attribute definitions (boston.names)</li> <li>training cases to be analyzed (boston.data)</li> <li>test cases (boston.test)</li> <li>initial classification rules (boston.rules)</li> <li>decision tree classifier (boston.dec)</li> <li>initial classifier (boston.initial)</li> <li>output file (boston.out)</li> </ul>
<p>CUBIST <i>Release</i> 1.08</p> <p><a href="http://www.rulequest.com/cubist-public.zip">http://www.rulequest.com/cubist-public.zip</a></p>	 <p>The screenshot shows the CUBIST software interface. The title bar reads 'Cubist [demonstration version]'. The main window contains a list of rules and their associated data files:</p> <ul style="list-style-type: none"> <li>attribute definitions (boston.names)</li> <li>training cases to be analyzed (boston.data)</li> <li>test cases (boston.test)</li> <li>rule-based model (boston.model)</li> <li>output file (boston.out)</li> </ul> <p>A small dialog box titled 'Cubist' is open in the foreground, displaying 'Cubist Release 1.08' and an 'OK' button.</p>
<p>SIPINA for Windows 2.5</p> <p><a href="http://eric.univ-lyon2.fr/~ricco/ricco.html">http://eric.univ-lyon2.fr/~ricco/ricco.html</a></p>	 <p>The screenshot shows the SIPINA for Windows 2.5 software interface. The title bar reads 'SIPINA for Windows 2.5'. The main window contains a dialog box for file selection, with the title bar reading 'SIPINA for Windows 2.5'. The dialog box has a list of files and a 'Select' button.</p>
<p>PolyAnalyst 4.2</p> <p><a href="http://www.megaputer.com/html/polyanalyst4.0.html">http://www.megaputer.com/html/polyanalyst4.0.html</a></p>	 <p>The screenshot shows the PolyAnalyst 4.2 software interface. The title bar reads 'PolyAnalyst'. The main window contains a 3D map of a city area, with various data visualization options and a 'Load' button.</p>

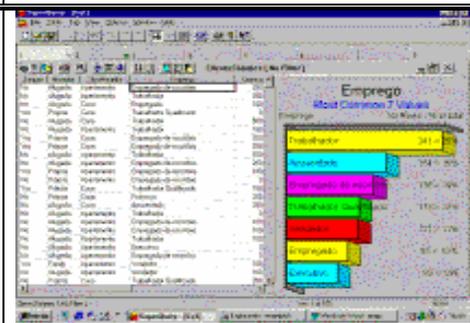
XpertRule *version 1.47*

<http://www.attar.com/index.htm>



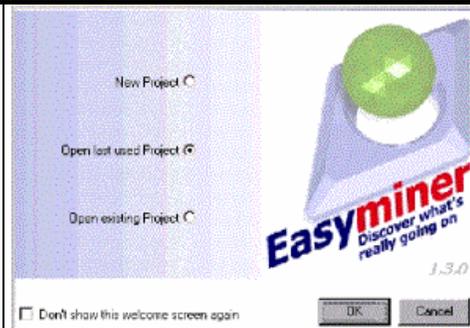
SuperQuery *version 1.52*

<http://www.azmy.com>



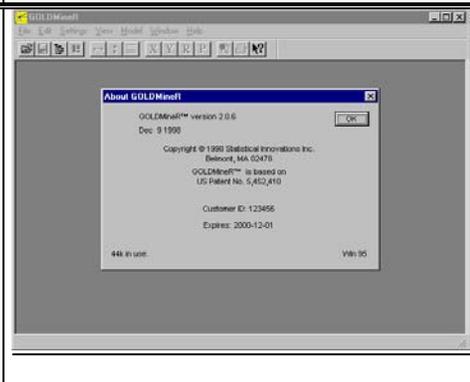
Easyminer *version 1.3.5*

<http://www.mineit.com/products/easyminer/evaluation/downloads/>



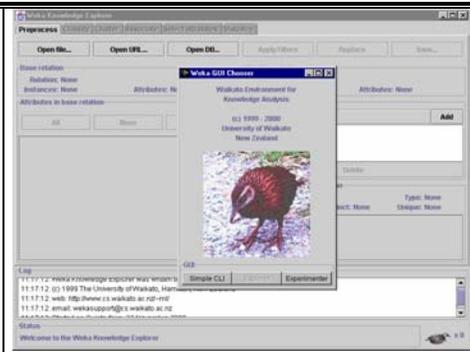
GOLDMINER *version 2.0.6*

<http://www.spss.com>



WEKA – WAIKATO  
Weka Knowledge Explorer

<http://www.cs.waikato.ac.nz/~ml/>



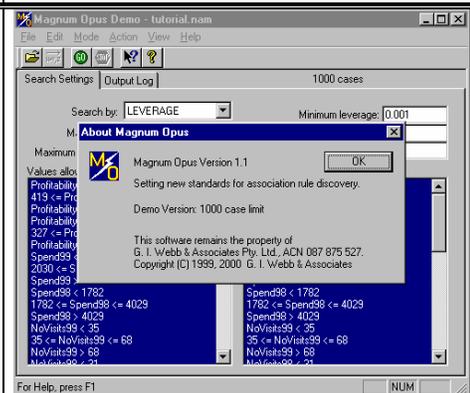
Miner3D.web

<http://miner3D.com/m3Dweb/help/index.html>



MAGNUM

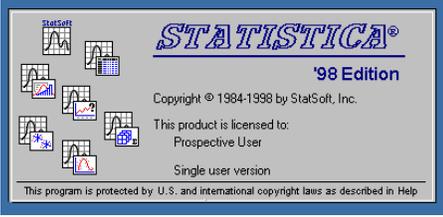
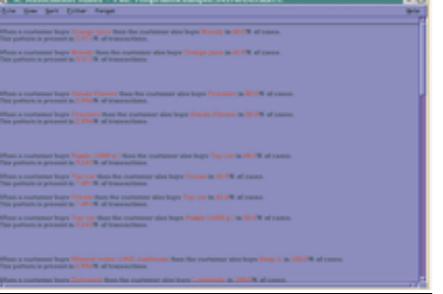
<http://www.rulequest.com/MODemo.zip>



CLUSTER

<http://stelanet.eps.ufsc.br/aran/DataMining/>



<p style="text-align: center;">STATISTICA</p> <p><a href="http://www.statsoft.com">http://www.statsoft.com</a></p>	 <p>The image shows the Statistica '98 Edition software box. It features the StatSoft logo, the product name 'STATISTICA' in a stylized font, and the version '98 Edition'. Below the name, it states 'Copyright © 1984-1998 by StatSoft, Inc.' and 'This product is licensed to: Prospective User'. It also mentions 'Single user version' and a copyright notice: 'This program is protected by U.S. and international copyright laws as described in Help'.</p>
<p style="text-align: center;">SQL Server2000</p> <p><a href="http://www.microsoft.com/brasil/sql">http://www.microsoft.com/brasil/sql</a></p>	 <p>The image is a banner for Microsoft SQL Server 2000 Analysis Services. It features a background image of a modern building with a grid pattern. The text 'Microsoft SQL Server 2000' is prominently displayed in the center, with 'Analysis Services' written in red below it.</p>
<p style="text-align: center;">Intelligent Miner - IBM</p> <p><a href="http://www.ibm.com">http://www.ibm.com</a></p>	 <p>The image is a screenshot of the Intelligent Miner software interface. It shows a window titled 'Intelligent Miner' with a menu bar (File, Edit, View, Help) and a main area displaying a list of data mining tasks. Each task entry includes a name, a description, and a status indicator.</p>
<p style="text-align: center;">Oracle9i Data Mining</p> <p><a href="http://www.oracle.com">http://www.oracle.com</a></p>	 <p>The image is an advertisement for Oracle9i Data Mining. It features the text 'Oracle9i Complete e-Business Intelligence Infrastructure' at the top. Below this, there is a red diamond icon labeled 'Data Mining'. To the right, there are two server racks: one labeled 'ORACLE 9i UNBREAKABLE' and another labeled 'IAS 9i CACHE'. The Oracle logo is visible at the bottom right.</p>

## Anexo 3: Questionário aplicado pelo IBGE – 2001

FAÇA OS ALINHAMENTOS CONFORME O MODELO: 0 1 2 3 4 5 6 7 8 9 NÃO RISCUE. SE NECESSÁRIO, APAGUE COM A BORRACHA APROPRIADA. ESCRIBA SOBRENTE COM A LÁPIS/RETA INDICADA. PREENCHA A QUADRICULA DESTA FORMA: X

**censo Demográfico 2000**  
CD 1.01  
Questionário Básico

MUNICÍPIO: \_\_\_\_\_  
AGÊNCIA: \_\_\_\_\_  
USO EXCLUSIVO DO IBGE

1 047521988 01

**1 IDENTIFICAÇÃO** 1.01 NÚMERO DO QUESTIONÁRIO

1.02 UF	1.03 MUNICÍPIO	1.04 DISTRITO	1.05 SUBDISTRITO	1.06 SETOR	1.07 Nº DA PÁGINA DA FOLHA DE COLETA	1.08 Nº NA FOLHA DE COLETA
1.09 Nº NA FOLHA DE DOMICÍLIO COLETIVO	1.10 TOTAL DE HOMENS	1.11 TOTAL DE MULHERES	1.12 QUESTIONÁRIO DO DOMICÍLIO 1 - ÚNICO 2 - EM CONTINUAÇÃO 3 - TEM OUTROS		1.13 TOTAL DE QUESTIONÁRIOS UTILIZADOS NO DOMICÍLIO	

LOCALIDADE:  
LOGRADOURO:  
(Endereço completo)

**2 CARACTERÍSTICAS DO DOMICÍLIO**

2.01 - ESPÉCIE  
1 - PARTICULAR PERMANENTE → Signa questão 2.02  
2 - PARTICULAR IMPROVISADO → Passe para a Lista de Moradores  
3 - COLETIVO →

2.02 - TIPO  
1 - CASA    2 - APARTAMENTO    3 - CÔMODO

2.03 - ESTE DOMICÍLIO É:  
1 - PRÓPRIO - JÁ PAGO → Signa questão 2.04  
2 - PRÓPRIO - AINDA PAGANDO →  
3 - ALUGADO →  
4 - CEDIU POR EMPREGADOR →  
5 - CEDIU DE OUTRA FORMA →  
6 - OUTRA CONDIÇÃO →  
Passe para o questionário 2.05

2.04 - O TERRENO EM QUE SE LOCALIZA ESTE DOMICÍLIO É:  
1 - PRÓPRIO    2 - CEDIU    3 - OUTRA CONDIÇÃO

2.05 - A FORMA DE ABASTECIMENTO DE ÁGUA UTILIZADA NESTE DOMICÍLIO É:  
1 - REDE GERAL  
2 - POÇO OU SARCENTE (na propriedade)  
3 - OUTRA

2.06 - A ÁGUA UTILIZADA NESTE DOMICÍLIO CHEGA:  
1 - CANALIZADA EM PELO MENOS UM CÔMODO  
2 - CANALIZADA SÓ NA PROPRIEDADE OU TERRENO  
3 - NÃO CANALIZADA

2.07 - QUANTOS BANHEIROS EXISTEM NESTE DOMICÍLIO?  
(Considere somente os que tenham chuveiro ou banheira e aparelho sanitário)  
Se 0 (zero) siga questão 2.08. Caso contrário, passe para o questionário 2.08.  
Se 8 banheiros ou mais registre 8 (oito).

2.08 - NESTE DOMICÍLIO, TERRENO OU PROPRIEDADE HÁBITE BANHEIRO UTILIZADO PELOS MORADORES?  
1 - SIM → Signa questão 2.09  
2 - NÃO → Passe para o questionário 2.10

2.09 - O ESCOAMENTO DESTES BANHEIRO DO BANHEIRO É LIGADO A:  
1 - REDE GERAL DE ESGOTO OU PLUVIAL  
2 - FOGGA BÉPICA  
3 - FOGGA RUDIMENTAR  
4 - VALA  
5 - RIO, LAGO OU MAR  
6 - OUTRO ESCOAMENTO

2.10 - O LIXO DEBTE DOMICÍLIO:  
1 - É COLETADO POR SERVIÇO DE LIMPEZA  
2 - É COLOCADO EM SACAMBIA DE SERVIÇO DE LIMPEZA  
3 - É QUEIMADO (na propriedade)  
4 - É ENTERRADO (na propriedade)  
5 - É JOGADO EM TERREIRO BALÇO OU LOGRADOURO  
6 - É JOGADO EM RIO, LAGO OU MAR  
7 - TEM OUTRO DESTINO  
Passe para a Lista de Moradores

**IBGE**  
Instituto Brasileiro de Geografia e Estatística



PREENCHA OS ALGARISMOS CONFORME O MODELO:

0 1 2 3 4 5 6 7 8 9

NÃO RISCAR. SE NECESSÁRIO, APAGUE COM A BORRACHA APROPRIADA. ESCRIBA SOMENTE COM A LÁPISERA INDICADA.

PREENCHA A QUADRICULA DESTA FORMA. X

3

<b>2ª PESSOA</b>		NOME:	
4.01 - SEXO	1 - MASCULINO	2 - FEMININO	MARQUE A QUADRICULA, SE A PRÓPRIA PESSOA PRESTAR AS INFORMAÇÕES
4.02 - QUAL É A RELAÇÃO COM A PESSOA RESPONSÁVEL PELO DOMÍLIO?			
01 - CÔNJUGE, COMPANHEIRO(a)	02 - NETO(a), BISNETO(a)	03 - PENIONISTA	
03 - FILHO(a), ENTEADADO(a)	04 - IRMÃO, IRMÃ	04 - EMPREGADO(a) DOMÉSTICO(a)	
04 - PAI, MÃE, SOCORO(a)	07 - OUTRO PARENTE	05 - PARENTE DO(a) EMPREGADO(a) DOMÉSTICO(a)	

ATENÇÃO: Registre o mês e ano de nascimento (Questão 4.03) e a idade em 31 de julho de 2000 (Questão 4.04). Se não for possível obter o mês e ano, registre todos os dígitos, registre a idade presumida (Questão 4.05) em anos (se a idade for maior ou igual a 1 ano), ou em meses (se a idade for menor que 1 ano), de acordo em branco as questões 4.03 e 4.04.

4.03 - QUAL É O MÊS E ANO DO SEU NASCIMENTO?	4.04 - QUAL ERA A SUA IDADE EM 31 DE JULHO DE 2000?	4.05 - QUAL É A SUA IDADE PRESUMIDA?
MÊS ANO	1 ANO OU MAIS EM MESES	1 ANO OU MAIS EM MESES
2	anos 4 meses	2 anos 4 meses

4.06 - SABE LER E ESCRIVER?	1 - SIM	2 - NÃO
-----------------------------	---------	---------

<b>3ª PESSOA</b>		NOME:	
4.01 - SEXO	1 - MASCULINO	2 - FEMININO	MARQUE A QUADRICULA, SE A PRÓPRIA PESSOA PRESTAR AS INFORMAÇÕES
4.02 - QUAL É A RELAÇÃO COM A PESSOA RESPONSÁVEL PELO DOMÍLIO?			
01 - CÔNJUGE, COMPANHEIRO(a)	04 - NETO(a), BISNETO(a)	03 - PENIONISTA	
03 - FILHO(a), ENTEADADO(a)	05 - IRMÃO, IRMÃ	04 - EMPREGADO(a) DOMÉSTICO(a)	
04 - PAI, MÃE, SOCORO(a)	07 - OUTRO PARENTE	05 - PARENTE DO(a) EMPREGADO(a) DOMÉSTICO(a)	

ATENÇÃO: Registre o mês e ano de nascimento (Questão 4.03) e a idade em 31 de julho de 2000 (Questão 4.04). Se não for possível obter o mês e ano, registre todos os dígitos, registre a idade presumida (Questão 4.05) em anos (se a idade for maior ou igual a 1 ano), ou em meses (se a idade for menor que 1 ano), de acordo em branco as questões 4.03 e 4.04.

4.03 - QUAL É O MÊS E ANO DO SEU NASCIMENTO?	4.04 - QUAL ERA A SUA IDADE EM 31 DE JULHO DE 2000?	4.05 - QUAL É A SUA IDADE PRESUMIDA?
MÊS ANO	1 ANO OU MAIS EM MESES	1 ANO OU MAIS EM MESES
2	anos 4 meses	2 anos 4 meses

4.06 - SABE LER E ESCRIVER?	1 - SIM	2 - NÃO
-----------------------------	---------	---------

<b>4ª PESSOA</b>		NOME:	
4.01 - SEXO	1 - MASCULINO	2 - FEMININO	MARQUE A QUADRICULA, SE A PRÓPRIA PESSOA PRESTAR AS INFORMAÇÕES
4.02 - QUAL É A RELAÇÃO COM A PESSOA RESPONSÁVEL PELO DOMÍLIO?			
01 - CÔNJUGE, COMPANHEIRO(a)	06 - NETO(a), BISNETO(a)	03 - PENIONISTA	
03 - FILHO(a), ENTEADADO(a)	05 - IRMÃO, IRMÃ	04 - EMPREGADO(a) DOMÉSTICO(a)	
04 - PAI, MÃE, SOCORO(a)	07 - OUTRO PARENTE	05 - PARENTE DO(a) EMPREGADO(a) DOMÉSTICO(a)	

ATENÇÃO: Registre o mês e ano de nascimento (Questão 4.03) e a idade em 31 de julho de 2000 (Questão 4.04). Se não for possível obter o mês e ano, registre todos os dígitos, registre a idade presumida (Questão 4.05) em anos (se a idade for maior ou igual a 1 ano), ou em meses (se a idade for menor que 1 ano), de acordo em branco as questões 4.03 e 4.04.

4.03 - QUAL É O MÊS E ANO DO SEU NASCIMENTO?	4.04 - QUAL ERA A SUA IDADE EM 31 DE JULHO DE 2000?	4.05 - QUAL É A SUA IDADE PRESUMIDA?
MÊS ANO	1 ANO OU MAIS EM MESES	1 ANO OU MAIS EM MESES
2	anos 4 meses	2 anos 4 meses

4.06 - SABE LER E ESCRIVER?	1 - SIM	2 - NÃO
-----------------------------	---------	---------

**IBGE**  
Instituto Brasileiro de Geografia e Estatística

FAÇA OS ALFABETISMOS CONFORME O MODELO

0 1 2 3 4 5 6 7 8 9

NÃO RESOLVE, SE NECESSÁRIO, APAGUE COM A BORRACHA APROPRIADA. ESCRVA SOMENTE COM A LÁPIS/ESBIA INDICADA.

PREENCHA A QUADRICULA DESTA FORMA X

4

<b>5ª PESSOA</b>		NOME:
4.01 - SEXO	1 - MASCULINO    2 - FEMININO	<input type="checkbox"/> MARQUE A QUADRICULA, SE A PRÓPRIA PESSOA PRESTAR AS INFORMAÇÕES

4.02 - QUAL É A RELAÇÃO COM A PESSOA RESPONSÁVEL PELO DOMÍLIO?		
<input type="checkbox"/> 02 - CÔNJUGE, COMPANHEIRO(a)	<input type="checkbox"/> 05 - NETO(a), BENEFICIA(a)	<input type="checkbox"/> 08 - AGREGADO(a)
<input type="checkbox"/> 03 - FILHO(a), ENTEADO(a)	<input type="checkbox"/> 06 - IRMÃO, IRMÃ	<input type="checkbox"/> 09 - PENSIONISTA
<input type="checkbox"/> 04 - PAI, MÃE, SÓCIO(a)	<input type="checkbox"/> 07 - OUTRO PARENTE	<input type="checkbox"/> 10 - EMPREGADO(a) DOMÉSTICO(a)
<input type="checkbox"/> 11 - PARENTE DO(a) EMPREGADO(a) DOMÉSTICO(a)		

ATENÇÃO: Registre o mês e ano do nascimento (Questão 4.03) e a idade em 31 de julho de 2000 (Questão 4.04). Se não for possível obter o mês e ano, esgotados todos os esforços, registre a idade presumida (Questão 4.05) em anos (se a idade for maior ou igual a 1 ano), ou em meses (se a idade for menor que 1 ano), destacando em branco as questões 4.03 e 4.04.

4.03 - QUAL É O MÊS E ANO DO SEU NASCIMENTO?		4.04 - QUAL ERA A SUA IDADE EM 31 DE JULHO DE 2000?		4.05 - QUAL É A SUA IDADE PRESUMIDA?	
MÊS	ANO	1 ANO OU MAIS	EM MESES	1 ANO OU MAIS	EM MESES
<input type="text"/>	<input type="text"/>	2 <input type="text"/> anos	4 <input type="text"/> meses	2 <input type="text"/> anos	4 <input type="text"/> meses

4.06 - SABE LER E ESCRVER?	
1 - SIM    2 - NÃO	

<b>6ª PESSOA</b>		NOME:
4.01 - SEXO	1 - MASCULINO    2 - FEMININO	<input type="checkbox"/> MARQUE A QUADRICULA, SE A PRÓPRIA PESSOA PRESTAR AS INFORMAÇÕES

4.02 - QUAL É A RELAÇÃO COM A PESSOA RESPONSÁVEL PELO DOMÍLIO?		
<input type="checkbox"/> 02 - CÔNJUGE, COMPANHEIRO(a)	<input type="checkbox"/> 05 - NETO(a), BENEFICIA(a)	<input type="checkbox"/> 08 - AGREGADO(a)
<input type="checkbox"/> 03 - FILHO(a), ENTEADO(a)	<input type="checkbox"/> 06 - IRMÃO, IRMÃ	<input type="checkbox"/> 09 - PENSIONISTA
<input type="checkbox"/> 04 - PAI, MÃE, SÓCIO(a)	<input type="checkbox"/> 07 - OUTRO PARENTE	<input type="checkbox"/> 10 - EMPREGADO(a) DOMÉSTICO(a)
<input type="checkbox"/> 11 - PARENTE DO(a) EMPREGADO(a) DOMÉSTICO(a)		

ATENÇÃO: Registre o mês e ano do nascimento (Questão 4.03) e a idade em 31 de julho de 2000 (Questão 4.04). Se não for possível obter o mês e ano, esgotados todos os esforços, registre a idade presumida (Questão 4.05) em anos (se a idade for maior ou igual a 1 ano), ou em meses (se a idade for menor que 1 ano), destacando em branco as questões 4.03 e 4.04.

4.03 - QUAL É O MÊS E ANO DO SEU NASCIMENTO?		4.04 - QUAL ERA A SUA IDADE EM 31 DE JULHO DE 2000?		4.05 - QUAL É A SUA IDADE PRESUMIDA?	
MÊS	ANO	1 ANO OU MAIS	EM MESES	1 ANO OU MAIS	EM MESES
<input type="text"/>	<input type="text"/>	2 <input type="text"/> anos	4 <input type="text"/> meses	2 <input type="text"/> anos	4 <input type="text"/> meses

4.06 - SABE LER E ESCRVER?	
1 - SIM    2 - NÃO	

ASSINATURA DO ENTREVISTADO:



IBGE  
Instituto Brasileiro de Geografia e Estatística

## Anexo 4: Boletim de cadastramento BIC

PREFEITURA MUNICIPAL DE		U1 PARA USO DO PROCESSAMENTO		U2 INSCRIÇÃO CADASTRAL	
01 CÓDIGO DO MUNICÍPIO		02 Nº DE ARQUIVAMENTO		03	
DISTRITO		SETOR		QUADRA	
LOTE		UNIDADE			
03 INFORMAÇÕES GERAIS					
04 CONTROLE		05 COMANDO INCLUSÃO		06 INSCRIÇÃO ANTERIOR	
ALTERAÇÃO COM EMISSÃO		ALTERAÇÃO SEM EMISSÃO		CANCELAMENTO	
04 LOCALIZAÇÃO DO IMÓVEL					
TIPO, NOME DO LOGRADOURO					
07 CÓDIGO DO LOGRADOURO		08 SEÇÃO		09 NÚMERO	
10 COMPLEMENTO					
11 BAIRRO		12 LOTEAMENTO		13 QUADRA	
14 LOTE		15 TOTAL DE ITENS		9	
05 NOME DO PROPRIETÁRIO ENDEREÇO PARA CORRESPONDÊNCIA					
16 NOME DO PROPRIETÁRIO OU DETENTOR					
17 TIPO					
18 NOME DO LOGRADOURO		19 NÚMERO			
20 COMPLEMENTO					
21 BAIRRO					
22 NOME DO MUNICÍPIO		23 CEP		24 SIGLA / 25	
TOTAL DE ITENS 6					
06 INFORMAÇÕES GERAIS SOBRE O IMÓVEL					
26 OCUPAÇÃO DO LOTE		CONSTRUÇÃO PARALIZADA		CONSTRUÇÃO EM ANDAMENTO	
NÃO CONSTRUÍDO		EM DEMOLIÇÃO		NATUREZA TEMPORÁRIA	
RUINAS		CONSTRUÍDO		EM REFORMA	
27 BEM/IMÓVEL / PATRIMÔNIO		PÚBLICO		PARTICULAR	
28		RELIGIOSO			
29 UTILIZAÇÃO TERRENO / USO		PRESTAÇÃO DE SERVIÇOS		SERVIÇO PÚBLICO	
RESIDENCIAL		INDUSTRIAL		RELIGIOSO	
COMERCIAL		MISTA		RURAL	
30		31		32	
33		34		35	
36		37		38	
39		40		41	
42		43		44	
45		46		47	
TOTAL DE ITENS 6					
07 INFORMAÇÕES SOBRE O TERRENO					
43 SITUAÇÃO					
MEIO DE QUADRA		ESQUINA/MAIS DE UMA FRENTE		VILA	
COND. HORIZ		ENCRAVADO		GLEBA	
AGLOMERADO					
44 TOPOGRAFIA PLANO					
ACLIVE		DECLIVE		IRREGULAR	
45 PEDOLOGIA INUNDÁVEL		BREJO/MANGUE		ROCHOSO	
ARENOSO		NORMAL		46	
TOTAL DE ITENS 9					
08 MEDIDAS DO IMÓVEL					
TESTADA PRINCIPAL		48		5	
TESTADA 2		50		7	
CÓDIGO DO LOGRADOURO		51		5	
SEÇÃO DO LOGRADOURO		52		3	
TESTADA 3		54		0	
CÓDIGO DO LOGRADOURO		55		8	
SEÇÃO DO LOGRADOURO		56		6	
TESTADA 4		58		2	
CÓDIGO DO LOGRADOURO		59		0	
SEÇÃO DO LOGRADOURO		60		4	
PROFUNDIDADE		61		2	
ÁREA DO LOTE		62		0	
ÁREA DO LOTE DE VILA/COND. HOR		63		9	
ÁREA CONSTRUÍDA DA UNIDADE		64		7	
TOTAL DE UNIDADES NO LOTE		65		5	
ÁREA TOTAL CONSTRUÍDA		66		3	
TOTAL DE ITENS		72		8	

09 INFORMAÇÕES SOBRE A EDIFICAÇÃO										
73 TIPO										
CASA	APARTAMENTO	SALA COMERCIAL	LOJA	GALPÃO	TELHEIRO	CASA MISTA	ESPECIAL			
15	31	40	58	66	74	86	87			
78 ESTRUTURA										
ALVENARIA	MADEIRA			METALICA			CONCRETO			
11	20			38			46			
79 COBERTURA										
PALHA ZINCO	TELHA DE CIMENTO AMANTO	TELHA DE BARRO	LAJE	METALICA ESPECIAL						
19	27	35	43	83						
80 PAREDES										
SEM	ALVENARIA	SOLO CIMENTO	MADERA BENEFICIADA	MADERA BRUTA	MISTA					
10	36	86	87	89	90					
81 FORRO										
SEM	MADEIRA	LAJE	CHAPAS COMPENSADO	ESTUQUE PLASTICO						
17	25	41	86	87						
82 REVESTIMENTO DA FACHADA PRINCIPAL										
SEM	EMBOCO	REBOCO	MATERIAL CERAMICO	MADEIRA	PEDRA A VISTA	CONCRETO	ESPECIAL			
14	22	30	49	57	86	87	89			
83 INSTALACAO SANITARIA										
SEM	EXTERNA		INTERNA SIMPLES	INTERNA COMPLETA	MAIS DE UMA INTERNA					
11	20		86	87	89					
84 PISO										
TERRA BATIDA	CIMENTO	CERAMICA	TABUAS	TACO	MATERIAL PLASTICO	CARPET	ESPECIAL			
16	24	32	86	87	89	90	91			
86 ESQUADRIAS										
SEM	FERRO		MADEIRA	ALUMINIO	VENEZIANA			TOTAL DE ITENS		
13	21		30	86	87			87		
10 INFORMAÇÕES OPCIONAIS										
88 ESTADO DE CONSERVACAO										
01 - NOVA ÓTIMA	02 - BOM	03 - REGULAR	04 - MAU	89 Nº DE PESSOAS DE 0 A 6 ANOS		91 Nº DE PESSOAS DE 7 A 14 ANOS	92 Nº DE PESSOAS DE 14 A 21 ANOS	93 Nº DE PESSOAS ACIMA DE 21 ANOS	96 LANÇAMENTO ENGRÓBADO	
4				6	4	2		0	10	
89 MURO PASSEIO										
01 - COM MURO E PASSEIO					97 VALOR VENAL					
02 - COM MURO SEM PASSEIO										
03 - COM PASSEIO SEM MURO										
04 - SEM MURO SEM PASSEIO										
2						98 TOTAL DE ITENS				
11 CROQUIS										
ANOTAÇÕES					CADASTRADOR					
					DATA					

11/2 DR. FERNANDES

**BCI - BOLETIM DE CADASTRO IMOBILIÁRIO**

PROJETO CIATA  
SISTEMA MODULAR

PREFEITURA MUNICIPAL DE <h1 style="margin: 0;">SOMBRIO</h1>		<b>01 PARA USO DO PROCESSAMENTO</b> 01 CÓDIGO DO MUNICÍPIO    02 Nº DE ARQUIVAMENTO		<b>02 INSCRIÇÃO CADASTRAL</b> 03 _____ DISTRITO   SETOR   QUADRA   LOTE   UNIDADE	
<b>03 INFORMAÇÕES GERAIS</b> 05 COMANDO    06 INSCRIÇÃO ANTERIOR INCLUSÃO    ALTERAÇÃO COM EMISSÃO    ALTERAÇÃO SEM EMISSÃO    CANCELAMENTO					
<b>04 LOCALIZAÇÃO DO IMÓVEL</b> 07 CÓDIGO DO LOGRADOURO    08 SEÇÃO    09 NÚMERO    10 COMPLEMENTO 11 BAIRRO    12 LOTEAMENTO    13 QUADRA    14 LOTE					
<b>05 NOME DO PROPRIETÁRIO / ENDEREÇO PARA CORRESPONDÊNCIA</b> 16 NOME DO PROPRIETÁRIO OU DETENTOR 17 TIPO    18 NOME DO LOGRADOURO    19 NÚMERO 20 COMPLEMENTO    21 BAIRRO 22 NOME DO MUNICÍPIO    23 CEP    24 SIGLA UF					
<b>06 INFORMAÇÕES GERAIS SOBRE O IMÓVEL</b> 26 OCUPAÇÃO DO LOTE    NÃO CONSTRUÍDO    CONSTRUÇÃO EM ANDAMENTO    CONSTRUÍDO 27 BEM IMÓVEL/PATRIMÔNIO    PÚBLICO    PARTICULAR    RELIGIOSO 29 UTILIZAÇÃO    TERRENO S/USO    RESIDENCIAL    COMERCIAL    PRESTAÇÃO DE SERVIÇOS    INDUSTRIAL    OUTROS 38 ANO DE AQUISIÇÃO    40 IMUNE / ISENTO IPTU NÃO    IMUNE    ISENTO    41 ISENTO TSU NÃO    SIM					
<b>07 INFORMAÇÕES SOBRE O TERRENO</b> 43 SITUAÇÃO    MEIO DA QUADRA    ESQUINA / MAIS DE UMA FRENTE    ENCRAVADO    GLEBA 44 TOPOGRAFIA    PLANO    IRREGULAR    45 PEDOLOGIA    INUNDÁVEL    NORMAL					
<b>08 MEDIDAS DO IMÓVEL</b>					
TESTADA PRINCIPAL	48		5		
TESTADA 2	50		7		
CÓDIGO DO LOGRADOURO	51		5		
SEÇÃO DO LOGRADOURO	52		3		
TESTADA 3	54		0		
CÓDIGO DO LOGRADOURO	55		8		
SEÇÃO DO LOGRADOURO	56		6		
TESTADA 4	58		2		
CÓDIGO DO LOGRADOURO	59		0		
SEÇÃO DO LOGRADOURO	60		4		
ÁREA DO LOTE	62				0
ÁREA CONSTRUÍDA DA UNIDADE	64				7
TOTAL DE UNIDADES NO LOTE	65				5
ÁREA TOTAL CONSTRUÍDA	66				3

PREFEITURA MUNICIPAL DE ITAJAI						SERVICO: <input type="checkbox"/> 1 - INCLUSAO <input type="checkbox"/> 2 - ALTERACAO <input type="checkbox"/> 3 - EXCLUSAO			
IMPOSTO PREDIAL E TERRITORIAL URBANO									
BOLETIM DE INFORMACOES CADASTRAIS									
(65)	(63)	(66)	(64)	(67)	(68)	INSCRICAO C.P.D. D			
SETOR	QUADRA	FACE	LOTE	UNID.	SL				
1 NOME DO CONTRIBUINTE						2 NUM. IMOVEL			
RUA DO IMOVEL						3 CGC/CPF			
4 COMPLEMENTO DO IMOVEL						5 COD. RUA			
BAIRRO OU LOTEAMENTO DO IMOVEL						6 COD. BAIRRO			
7 RUA PARA CORRESPONDENCIA						8 NUM. IMOVEL COR			
9 COMPLEMENTO CORRESPONDENCIA									
10 CIDADE						11 U F			
						12 CEP			
TIPO IMPOSTO									
13 1-PREDIAL		14 ISENTO IPTU 0=NAO 1=SIM 2=RECADAST.		15 ISENTO TAXAS 0=NAO 1=SIM		16 U.U.T. (Vir.Venal Terreno)	17 U.U.C. (Vir.Venal Constr.)		
20 PASSEIO <input type="checkbox"/> - OPCAO		25 TOPOGR. <input type="checkbox"/> - OPCAO		26 SITUACAO LOTE <input type="checkbox"/> - OPCAO		27 BENFEIT. <input type="checkbox"/> - OPCAO	28 AREA DE PRESERV. <input type="checkbox"/> - OPCAO		
1-CIMENTO 2-CERAMICA 3-TERRA 4-PEDRA 5-OUTROS		1-AO NIVEL 2-ACIMA NIVEL 3-ABAIXO NIVEL 4-ALAGADO		1-ENCRAVADO 2-ESQUINA 3-MEIO QUADRA 4-BECO		1-SEM MURO 2-SEM PASSEIO 3-SEM MURO E S/PASSEIO 4-NORMAL	0-NAO 1-SIM		
<b>INFORMACOES SOBRE A CONSTRUCAO</b>									
40 TIPO DO IMOVEL <input type="checkbox"/> - OPCAO		41 SITUACAO <input type="checkbox"/> - OPCAO		42 CONSERVACAO <input type="checkbox"/> - OPCAO		43 ESPECIE <input type="checkbox"/> - OPCAO			
1-APTO 2-CASA 3-BARRACO 4-DEP. ABERTO 5-GALPAO AB. 6-COMERCIO 7-INDUSTRIA 8-OFICINA 9-HOTEL C/REST. 10-HOSPITAL 11-POSTO SERV.		12-CLUBE 13-ESCOLA 14-TEMPLO 15-SERV. PUB. 16-CINEMA 17-TEATRO 18-OUTROS 19-DEP. FECH. 20-GALPAO FECH. 21-ESPECIAIS 22-CAD. SEM IMOV. 23-REST. OU BAR 24-HOTEL S/REST		1-FRENTE 2-FUNDOS 3-CAD. SEM IMOVEL		1-BOM 2-REGULAR 3-MAU  6-CAD. S/ IMOVEL		1-ALVENARIA 2-MADEIRA 3-MISTA 4-BARRACO 5-CAD. S/ IMOVEL	
44 UTILIZ. <input type="checkbox"/> - OPCAO		45 ACABAMEN. <input type="checkbox"/> - OPCAO		46 PATRIMONIO <input type="checkbox"/> - OPCAO		47 PAVIMENTOS			
1-UTILIZ. PROPRIA 2-ALUGADA 3-FECHADA 4-CAD. S/ IMOVEL		1-LUXO 2-BOM 3-NORMAL 4-COMUM 5-POPULAR 6-CAD. S/ IMOVEL		1-PARTICULAR 2-CONDOMINIO 3-FUNCOE. PREFEITURA 4-ISENTO IPTU 5-ISENTO IPTU+ TX.					
48 INCRA <input type="checkbox"/> - OPCAO		49 FRENTE		50 FUNDOS		51 AREA TERR.			
0-NAO 1-SIM						52 AREA CONST.			
68 HISTORICO (1)						69 Nr. Reparc.			
69 HISTORICO (2)									
69 HISTORICO (3)									

## Anexo 5: Termo Projeto de Pesquisa

Itajaí SC, 22 de Março de 2001

**Para:** Prefeitura Municipal de Itajaí  
Departamento de Informática  
Sr. Paulo M. Klappoth

**Assunto:** Projeto de Pesquisa

Venho através desta solicitar a atenção de Vossa Senhoria com relação a proposta do projeto de pesquisa a ser realizado em conjunto com a Prefeitura Municipal de Itajaí, voltado para análise do cadastro técnico urbano do município.

Com objetivo de elaborar um projeto de pesquisa voltado para o município, mais especificamente para auxiliar seus decisores apontando novos caminhos, sendo o mesmo realizado através do **Programa de Pós - Graduação da Engenharia de Produção e Sistemas da Universidade Federal de Santa Catarina – UFSC**, apresento a proposta de estudo a qual baseia-se nas informações específicas do cadastro técnico urbano, conforme detalhamento em anexo.

O referente estudo abordará temas relacionados as informações técnicas cadastrais obtendo desde o perfil dos imóveis e situação atual, até a ligação com o cadastro sócio econômico do município caso exista interesse do mesmo.

O objetivo principal deste estudo é a aplicação das técnicas de mineração de dados (*Data Mining*), para a descoberta de novos conhecimentos. Os frutos obtidos durante a pesquisa serão apresentados unicamente ao município, sendo de interesse para este pesquisador somente a metodologia aplicada e os algoritmos utilizados na obtenção dos resultados. Com certeza todo trabalho de pesquisa e desenvolvimento a ser realizado não acarretará nenhuma despesa aos cofres públicos, visto que, para a realização desta pesquisa a necessidade de informações coletadas anteriormente pela prefeitura, é a base para o sucesso do projeto.

Agradeço antecipadamente a atenção dispensada e para maiores esclarecimentos segue endereço para contato logo abaixo.

Atenciosamente,

André Fabiano de Moraes  
Bacharel em Ciência da Computação – Univali  
Mestrando na Engenharia de Produção e Sistemas – UFSC

email – [afdm@eps.ufsc.br](mailto:afdm@eps.ufsc.br)  
email – [afdm@terra.com.br](mailto:afdm@terra.com.br)  
Telefone – (xxx) 047 346-1161 ou 9981-2970

*Handwritten notes in blue ink:*  
24/04/2001