

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE  
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIA EM SAÚDE**

**ALTAIR VON STEIN JUNIOR**

**DESCOBERTA DE REGRAS POR MEIO DE PROCESSO DE KDD PARA A  
CLASSIFICAÇÃO DE MICROÁREAS HOMOGÊNEAS DE RISCO**

**CURITIBA**

**2008**

**ALTAIR VON STEIN JUNIOR**

**DESCOBERTA DE REGRAS POR MEIO DE PROCESSO DE KDD PARA A  
CLASSIFICAÇÃO DE MICROÁREAS HOMOGÊNEAS DE RISCO**

**Dissertação de mestrado apresentado ao Programa de Pós-Graduação em Tecnologia em Saúde da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Tecnologia em Saúde.**

**Orientador: Prof. Dra. Andreia Malucelli**

**Co-orientador: Prof. Dr. Laudelino Cordeiro Bastos**

**CURITIBA**

**2008**

Dados da Catalogação na Publicação  
Pontifícia Universidade Católica do Paraná  
Sistema Integrado de Bibliotecas – SIBI/PUCPR  
Biblioteca Central

S819d  
2008 Stein Junior, Altair Von  
Descoberta de regras por meio de KDD para classificação de microáreas  
homogêneas de risco / Altair Von Stein Junior ; orientadora, Andreia Malucelli ;  
co-orientador, Laudelino Cordeiro Bastos. – 2008.  
106 f. : il. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná,  
Curitiba, 2008.  
Bibliografia: f. 72-75

1. Saúde. 2. Exploração de dados. 3. Sistemas especialistas (Computação).  
4. Inteligência artificial. 5. Qualidade de vida. I. Malucelli, Andreia. II. Bastos,  
Laudelino Cordeiro. III. Pontifícia Universidade Católica do Paraná. Programa  
de Pós-Graduação em Tecnologia em Saúde. IV. Título.

CDD 20. ed. – 613



Pontifícia Universidade Católica do Paraná  
Centro de Ciências Biológicas e da Saúde  
Programa de Pós-Graduação em Tecnologia em Saúde

ATA DA SESSÃO PÚBLICA DE DEFESA DE DISSERTAÇÃO DE MESTRADO  
DO PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIA EM SAÚDE  
DA PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ

DEFESA DE DISSERTAÇÃO Nº 082

Aos 21 dias do mês de outubro de 2008 realizou-se a sessão pública de defesa da dissertação “**Descoberta de regras por meio do processo de KDD para classificação de microáreas homogêneas de risco**”, apresentada por **Altair Von Stein Junior** como requisito parcial para a obtenção do título de Mestre em Tecnologia em Saúde, – Área de Concentração – **Informática em Saúde** perante uma Banca Examinadora composta pelos seguintes membros:

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Andreia Malucelli,  
PUCPR (Orientadora)

Andreia Malucelli  
assinatura

APROVADO  
parecer (aprov/ reprov.)

Prof. Dr. Laudelino Cordeiro Bastos,  
PUCPR (Co-orientador)

[Assinatura]

APROVADO

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Marcia Regina Cubas,  
PUCPR

[Assinatura]

APROVADO

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Maria Teresinha Arns Steiner,  
UFPR

[Assinatura] APROVADO

Conforme as normas regimentais do PPGTS e da PUCPR, o trabalho apresentado foi considerado APROVADO (aprovado/reprovado), segundo avaliação da maioria dos membros desta Banca Examinadora. Este resultado está condicionado ao cumprimento integral das solicitações da Banca Examinadora registradas no Livro de Defesas do Programa.

Prof. Dr. Munir Antonio Gariba,  
Diretor do PPGTS PUCPR

[Assinatura]



## AGRADECIMENTOS

Agradeço a Deus pelas oportunidades colocadas em minha vida.

A minha família que suportou os problemas e incômodos sem queixas nem revolta.

Agradeço em especial a Dr<sup>a</sup>. Deborah Ribeiro Carvalho pela cordialidade, apoio, incentivo e tempo cedido para as orientações que foram fundamentais. Agradeço ainda pela sua sinceridade e sua marcante disposição em colaborar com as pessoas.

Agradeço também em especial o Prof. Dr. Emerson Paraíso pelas suas orientações e colaborações também fundamentais.

À Prof<sup>a</sup>. Dra. Andreia Malucelli, orientadora, pelo constante apoio, incentivo, orientações e contribuições para o desenvolvimento deste estudo.

Ao Prof. Dr. Laudelino Cordeiro Bastos, co-orientador, pelas orientações e contribuições realizadas.

A todos os professores da PUCPR que opinaram e apontaram os possíveis caminhos a serem trilhados para a conclusão deste estudo.

Aos profissionais e especialistas envolvidos neste estudo, pela compreensão, disponibilidade e apoio.

Ao Adriano Lopes, bibliotecário da PUCPR pela paciência e dedicação oferecida.

Aos meus amigos que me apoiaram e entenderam os dias de mau humor.

A todos que direta ou indiretamente colaboraram na execução deste estudo.

O meu sincero agradecimento.

**“Dados, dados por toda a parte  
mas nenhum pensamento para pensar”**

**Jessé H. Shera**

## RESUMO

Para melhorar a atenção básica e otimizar os recursos disponíveis, as equipes de saúde identificam em sua área de atuação, áreas consideradas homogêneas de risco à saúde para atuarem sobre esses riscos objetivando contribuir para a qualidade de vida da população. Entretanto, apesar da importância da identificação de microáreas de risco, este processo ocorre por meio de uma análise subjetiva das equipes de saúde sobre um conjunto de dados referentes às condições epidemiológicas, demográficas, socioeconômicas e físicas do ambiente. Neste contexto, por meio de uma pesquisa exploratória metodológica, o objetivo deste estudo foi identificar, com o auxílio de técnicas computacionais, regras referentes às condições do ambiente físico capazes de contribuir para a classificação de microáreas de risco. Como resultados foram obtidos: um conjunto de atributos essenciais para classificar uma microárea em relação ao risco contido no ambiente físico; uma base de dados referente às condições das microáreas da cidade de Curitiba; e uma árvore de decisão para classificar uma microárea em relação ao risco contido no ambiente físico das microáreas. O conjunto de atributos poderá ser utilizado para especificar um sistema de informações que monitore as condições das microáreas e colabore com o planejamento e distribuição de ações para a melhoria da qualidade de vida da população de uma determinada região. A base de dados obtida apresenta o perfil de 46% das microáreas da cidade, assim como, apresenta a percepção dos agentes comunitários de saúde em relação ao risco contido nestas microáreas. A árvore de decisão resultante foi convertida em 68 regras, as quais foram analisadas sob a perspectiva de desempenho e qualidade. Após análise, as regras foram divididas em dois subconjuntos, um com as regras consideradas inconsistentes e outro com as regras que confirmaram o conhecimento dos especialistas. A análise da relação entre os conjuntos mostrou que as regras do segundo conjunto, apesar de terem desempenho inferior computacionalmente, foram as mais interessantes, na opinião dos especialistas, para classificar uma microárea em relação ao risco contido no ambiente físico. As regras obtidas podem otimizar e diminuir a subjetividade na interpretação dos dados para a classificação de microáreas, assim como, são capazes de prever a qual classe de risco, entre baixo, médio ou alto, a microárea pertence quando consideradas as características do ambiente físico. Com a classificação em baixo, médio ou alto risco, é possível melhorar a priorização de recursos, assim como as medidas de prevenção de doenças.

**Palavras-chaves:** Ambiente socioeconômico, Área de risco, Inteligência artificial.

## ABSTRACT

During many years, the health teams had given cares to the health in the fragmented form, adopting the illness as main focus. This model of assistance caused a low covering of the illnesses prevention programs and promotion to the health, for the lack of constraint and compromise between the health teams and the community. One of the strategies in order to change this situation in the country was presented by the Health department, which have implanted the Health Family Program, in 1994. In order to improve the performance of the health teams and to optimize the resources, the teams identified in its performance area, considered areas homogeneous of risk to the health, classifying them in order to act on these risks. Doing that, they hope to contribute for the quality of life of the population. However, the identification of risk areas, is carried through by the nurse and the communitarian agents of health by doing a subjective analysis on a data set. This set of data refers to the epidemiologist, demographic, socioeconomic and physical conditions of the environment in question and is gotten by means of sources of available data as of the Brazilian Institute of Geography and Statistics and of the System of Information of the Primary Care. However, the current sources not always allow to catch the local reality, this fact occurs due to the characteristics of the population to be added the familiar level, and thus, they occult some characteristics of the collective one. In this context, by one exploratory research, this study aims to identify important characteristics to evaluate the risk contained in the physical environment of the micro-areas, as well as, to construct a database and to apply knowledge discovery techniques to identify stereotypes able to contribute in the classification of micro-areas of risk. As result, we have obtained a set of attributes, considered for specialists as essential in order to classify an micro-area in relation to the risk contained in the physical environment. With the essential attributes a database was constructed and was applied the data mining technique. As result of data mining step a decision tree was obtained in order to classify the micro-area in relation to the ambient risk. This decision tree was converted into 68 rules, which had been analyzed under the perspective of performance and quality. The rules had been divided in two sets, those considered inconsistent (Group A) and the ones that confirm the knowledge of the specialists (Group B). The relation between the two sets showed that the rules of the Group B, although they have inferior computational performance, for the specialists they had been most interesting to classify the micro-area in relation to the risk contained in the physical environment. It's possible also, optimize and diminish the subjectivity in the interpretation of the data for the classification of micro-areas by using a set of characteristics and rules gotten during the research. The rules gotten for the knowledge discovery in databases process are able to predict which risky level, among low, average or high, the micro-area belongs when considered the characteristics of the physical environment.

**Keywords:** Socioeconomic Factors, Risk zone, Artificial intelligence.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Representação da divisão de um território. ....	19
Figura 2 – Processo de descoberta de conhecimento.....	24
Figura 3 - Relação entre o processo de KDD e a etapa de MD.....	26
Figura 4 – Modelo de árvore de decisão .....	28
Figura 5 – Representação esquemática do percurso metodológico .....	31
Figura 6 – Distribuição da rede municipal de saúde de Curitiba .....	33
Figura 7 – Áreas de abrangência envolvidas no estudo .....	39

## LISTA DE TABELAS

Tabela 1 – Comparação dos resultados obtidos entre as bases original e pré-processada.....	53
Tabela 2 – Medidas para avaliação do desempenho do conjunto de regras obtido na etapa de mineração de dados.....	55
Tabela 3 – Medidas para avaliação do Conjunto A.....	61
Tabela 4 – Medidas para avaliação do conjunto B.....	65
Tabela 5 - Comparação entre o desempenho dos conjuntos A e B .....	66

## LISTA DE QUADROS

Quadro 1-Atributos com valores constantes retirados por serem considerados redundantes. .	40
Quadro 2-Relação dos atributos e valores modificados.....	41
Quadro 3 – Conjunto de atributos para classificar uma microárea em relação ao ambiente físico.....	46
Quadro 4 - Atributos considerados irrelevantes para caracterizar as microáreas.....	47
Quadro 5 – Comparação das regras obtidas no processamento entre a base original e a base pré-processada.....	54
Quadro 6 – Mediana do quão surpreendente foram as regras para os especialistas e visão geral das médias de avaliação para cada regra.....	57
Quadro 7 – Regras consideradas inconsistentes pelos especialistas.....	60
Quadro 8 - Atributos indicados como inconsistentes .....	62
Quadro 9 – Regras que confirmam o conhecimento dos especialistas .....	64
Quadro 10 – Relação entre o atributo e sua posição na regra .....	67
Quadro 11 – Atributos mais distantes do início das regras.....	69
Quadro 12 – Atributos não encontrados nas regras .....	69

## LISTA DE GRÁFICOS

Gráfico 1 – Frequência dos valores dos atributos referentes aos equipamentos sociais na base de dados.....	48
Gráfico 2 - Frequência dos valores dos atributos referentes às áreas de lazer na base de dados .....	48
Gráfico 3 – Frequência dos valores dos atributos referentes ao comércio na base de dados...	48
Gráfico 4 – Frequência dos valores dos atributos referentes à hidrografia na base de dados..	49
Gráfico 5 – Frequência dos valores dos atributos referentes ao tipo de solo na base de dados	49
Gráfico 6 – Frequência dos valores dos atributos referentes às características da comunidade na base de dados .....	50
Gráfico 7 – Frequência dos valores dos atributos referentes aos vetores na base de dados ....	50
Gráfico 8 - Frequência dos valores dos atributos referentes às características de infra-estrutura na base de dados .....	51
Gráfico 9 – Distribuição da frequência do valor do atributo utilizado como classe alvo.....	52

## LISTA DE ABREVIATURAS

<b>ACS</b>	- Agentes comunitários de saúde
<b>CAPS</b>	- Centros de Atendimento Psicossocial
<b>CMUM</b>	- Centro Municipal de Urgência Médica
<b>CNS</b>	- Conselho Nacional de Saúde
<b>DM</b>	- <i>Data mining</i>
<b>ERP</b>	- Estimativa rápida participativa
<b>ESF</b>	- Equipes de Saúde da Família
<b>IBGE</b>	- Instituto Brasileiro de Geografia e Estatística
<b>IDH</b>	- Índice de Desenvolvimento Humano
<b>Int</b>	- Valor do intervalo da amostragem
<b>KDD</b>	- <i>Knowledge Discovery in Databases</i>
<b>m</b>	- Média
<b>max</b>	- Valor Máximo
<b>MD</b>	- Mineração de dados
<b>Min</b>	- Valor Mínimo
<b>MS</b>	- Ministério da Saúde
<b>PSF</b>	- Programa da Saúde da Família
<b>PUCPR</b>	- Pontifícia Universidade Católica do Paraná
<b>SIAB</b>	- Sistema de Informação da Atenção Básica
<b>SMS</b>	- Secretaria Municipal de Saúde
<b>SUS</b>	- Sistema Único de Saúde
<b>TDIDT</b>	- <i>Top- Down Induction of Decision Trees</i>
<b>US</b>	- Unidade de Saúde
<b>WEKA</b>	- <i>Waikato Environment for Knowledge Analysis</i>

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>14</b>
1.1 OBJETIVO .....	15
1.2 OBJETIVOS ESPECÍFICOS .....	15
1.3 ORGANIZAÇÃO DO TRABALHO .....	16
<b>2 FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>17</b>
2.1 SISTEMA ÚNICO DE SAÚDE .....	17
2.2 ATENÇÃO À SAÚDE E O TERRITÓRIO.....	18
<b>2.2.1 Reconhecendo o território.....</b>	<b>20</b>
<b>2.2.2 Estimativa rápida participativa.....</b>	<b>20</b>
2.2.2.1 Organização dos detalhes .....	21
2.2.2.2 Sistematização da área de trabalho.....	22
2.2.2.3 Análise dos dados.....	22
2.3 DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS.....	23
<b>2.3.1 Etapas do processo de KDD.....</b>	<b>24</b>
2.3.1.1 Pré-processamento .....	25
2.3.1.2 Mineração de dados.....	26
2.3.1.2.1 Tarefa de classificação.....	27
2.3.1.2.2 Árvore de decisão.....	28
2.3.1.3 Pós-processamento .....	30
<b>3 METODOLOGIA .....</b>	<b>31</b>
3.1 CARACTERIZAÇÃO DO CAMPO DE ESTUDO .....	32
3.2 SUJEITOS ENVOLVIDOS NAS PESQUISAS .....	34
3.3 CONSIDERAÇÕES ÉTICAS DO ESTUDO.....	34
3.4 ETAPAS PARA O DESENVOLVIMENTO DO ESTUDO .....	35
<b>3.4.1 Etapa 1 – Identificação de atributos para classificar as microáreas .....</b>	<b>35</b>
3.4.1.1 Condições ambientais.....	36
3.4.1.2 Condições de infra-estrutura .....	37
<b>3.4.2 Etapa 2 - Construção da base de dados .....</b>	<b>38</b>
<b>3.4.3 Etapa 3 – Aplicação do processo de KDD .....</b>	<b>40</b>
3.4.3.1 Pré-processamento .....	40
3.4.3.2 Mineração de dados.....	42

3.4.3.3 Pós-processamento .....	43
<b>4 RESULTADOS E DISCUSSÃO .....</b>	<b>45</b>
4.1 RESULTADOS DA ETAPA DE IDENTIFICAÇÃO DE ATRIBUTOS .....	45
4.2 RESULTADOS DA CONSTRUÇÃO DA BASE DE DADOS .....	47
4.3 RESULTADOS DA ETAPA DE KDD .....	53
<b>4.3.1 Resultados da etapa de Pré-processamento.....</b>	<b>53</b>
<b>4.3.2 Resultados da etapa de Mineração de dados.....</b>	<b>54</b>
<b>4.3.3 Resultados do Pós-processamento .....</b>	<b>55</b>
4.3.3.1 Avaliação do Conjunto A .....	58
4.3.3.2 Atributos apontados como a origem das inconsistências nas regras.....	62
4.3.3.3 Avaliação do Conjunto B.....	63
4.3.3.4 Comparação entre os Conjuntos A e B.....	65
4.3.3.5 Resultados referentes aos atributos .....	66
<b>5 CONCLUSÕES.....</b>	<b>70</b>
5.1 TRABALHOS FUTUROS .....	71
<b>REFERÊNCIAS.....</b>	<b>72</b>
<b>APÊNDICE A - INSTRUMENTO PARA VALIDAÇÃO DE ATRIBUTOS.....</b>	<b>76</b>
<b>APÊNDICE B – INSTRUMENTO PARA COLETA DE DADOS SOBRE AS CONDIÇÕES DAS MICROÁREAS.....</b>	<b>80</b>
<b>APÊNDICE C - INSTRUMENTO PARA OBTER O QUÃO SURPREENDENTE FORAM AS REGRAS PARA O ESPECIALISTA .....</b>	<b>84</b>
<b>ANEXO A – TERMOS DE CONSENTIMENTO LIVRE E ESCLARECIDO .....</b>	<b>102</b>
<b>ANEXO B - APROVAÇÃO DO COMITÊ DE ÉTICA EM PESQUISA (PESQ.1).....</b>	<b>105</b>
<b>ANEXO C - APROVAÇÃO DO COMITÊ DE ÉTICA EM PESQUISA (PESQ.2) .....</b>	<b>107</b>

## 1 INTRODUÇÃO

Durante muitos anos, as equipes de saúde prestaram cuidados à saúde de forma curativista<sup>i</sup> e fragmentada, adotando a doença como foco principal. Este modelo de assistência refletiu, dentre outros fatores, na elevação dos custos do atendimento, pelo uso abusivo dos recursos tecnológicos como substituição ao exame clínico para diagnosticar doenças. O atendimento focado na doença também causou a diminuição na cobertura dos programas de prevenção de doenças e promoção à saúde, pelo pouco comprometimento e vínculo entre as equipes de saúde e a comunidade. Uma das estratégias para a mudança desta situação no país, foi apresentada pelo Ministério da Saúde (MS), que implantou o Programa da Saúde da Família (PSF) em 1994.

Segundo BRASIL (2000) “A Saúde da Família é entendida como uma estratégia de reorientação do modelo assistencial, operacionalizada mediante a implantação de equipes multiprofissionais em unidades básicas de saúde”. Com base nesse conceito, o PSF busca reorientar a atenção à saúde, resgatando as ações de prevenção de doenças e promoção à saúde. Para que isso ocorra, a equipe de saúde torna-se responsável por uma área de atuação delimitada e realiza o planejamento de ações a serem desenvolvidas procurando atender as necessidades da população local.

No sentido de melhorar o desempenho e otimizar os recursos disponíveis, as equipes identificam na sua área de atuação, áreas consideradas homogêneas de risco à saúde de forma a atuarem sobre esses riscos, com o objetivo de contribuir para a qualidade de vida da população.

Entretanto, a identificação de áreas de risco, realizada pelas equipes de saúde, ocorre por meio de análise subjetiva embasada em um conjunto de dados. Esses dados são referentes às condições epidemiológicas, demográficas, socioeconômicas e físicas do ambiente em questão e são obtidos por meio de fontes de dados disponibilizadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE), o Sistema de Informação da Atenção Básica (SIAB), dentre outros. No entanto, estas fontes nem sempre permitem captar a realidade local, pois, os dados disponíveis são referentes às características das famílias, conseqüentemente, dados agregados a nível familiar ocultam algumas características da coletividade. Exemplificando, segundo Pereira e Barcellos (2006) ao identificar pelo SIAB, que a fonte principal de abastecimento de

---

<sup>i</sup> Prática curativista: modelo de assistência que reflete uma reação a uma doença instalada. Ou seja, os indivíduos devem ter a doença instalada para que ocorra uma ação terapêutica.

água de uma família ocorre por meio de um córrego, não há indicação se este córrego é contaminado ou não.

Neste contexto, este estudo visa identificar características importantes para avaliação do risco contido no ambiente físico das microáreas, formar uma base de dados e aplicar técnicas de descoberta de conhecimento em bases de dados<sup>ii</sup> (KDD) para extrair regras capazes de contribuir na classificação de áreas homogêneas de risco. As características e regras obtidas poderão ser utilizadas para modelar um sistema de informações referente às características do ambiente físico contribuindo para a qualidade da informação.

Para atingir ao objetivo proposto foi utilizada a tarefa de classificação (discutida na seção 2.2), pela qual se buscou descobrir, por meio de uma base de dados, classes diferentes de atributos para classes pré-definidas. A base de dados utilizada é referente as condições do ambiente físico de microáreas e contém dados coletados por meio de entrevista estruturada aplicada aos ACS de Curitiba, Paraná.

## 1.1 OBJETIVO

O objetivo geral é identificar, com o auxílio de técnicas computacionais, regras referentes ao ambiente físico capazes de contribuir para a classificação de microáreas homogêneas de risco.

## 1.2 OBJETIVOS ESPECÍFICOS

- a) relacionar um conjunto de atributos pertencentes ao ambiente físico que contribuam para categorizar o risco contido em uma microárea;
- b) elaborar uma base de dados com as características referentes ao ambiente físico de microáreas;

---

<sup>ii</sup> Do inglês, Knowledge discovery from databases.

- c) extrair regras, por meio de KDD, capazes de classificar microáreas em baixo, médio ou alto risco.

### 1.3 ORGANIZAÇÃO DO TRABALHO

Este trabalho, além do capítulo já apresentado referente à introdução, está organizado em mais quatro capítulos.

O capítulo 2 apresenta a fundamentação teórica referente ao processo de atendimento à saúde da comunidade e as ferramentas para a coleta e análise de dados sobre as condições de saúde de um território, bem como os conceitos e ferramentas sobre o processo de descoberta de conhecimento em base de dados.

No capítulo 3 descreve-se a metodologia aplicada para atingir aos objetivos propostos, apresentando as etapas desenvolvidas durante o estudo.

No capítulo 4 são apresentados e discutidos os resultados obtidos pela aplicação das etapas descritas na metodologia.

O capítulo 5 apresenta as conclusões, contribuições e as perspectivas de trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta uma visão geral da relação entre o sistema de atenção à saúde no Brasil e território e a metodologia da estimativa rápida participativa como instrumento para coleta e análise de dados objetivando a construção do planejamento em saúde. Ainda, neste capítulo é abordado o processo de descoberta de conhecimento apresentando os principais conceitos referentes as suas etapas, sendo elas, o pré-processamento, mineração de dados e pós-processamento.

### 2.1 SISTEMA ÚNICO DE SAÚDE

Idealizado a partir de um processo histórico, o Sistema Único de Saúde (SUS) permitiu mudar o rumo do atendimento à saúde da população Brasileira de forma significativa. Com princípios como a universalidade, integralidade e equidade, a proposta do SUS foi melhorar a cobertura, qualidade e resolutividade da assistência prestada nos serviços públicos de saúde, sendo essas condições distribuídas de forma igualitária e como um direito a todos os brasileiros (BRASIL, 2001). A implantação do SUS também buscou reorganizar o modelo de assistência à saúde prestada, passando de um modelo de assistência curativista focado na doença, para um modelo de atenção integral a saúde focando o indivíduo, família e a comunidade (BRASIL, 2006).

Para reforçar os princípios do SUS, foi implantado, em 1994, o Programa de Saúde da Família (PSF), que tem por objetivo, trabalhar com o indivíduo e sua família por meio de ações preventivas, educacionais e de reabilitação. Este programa proporciona aos profissionais que compõem uma equipe de saúde, uma maior aproximação com a comunidade. Isto fortalece o vínculo entre a equipe e a comunidade, resgata as ações de prevenção e promoção à saúde, permitindo mudar o modelo assistencial curativista.

As equipes de saúde da família (ESF) que integram o PSF são compostas por médicos, enfermeiros, auxiliares de enfermagem, agentes comunitários de saúde, entre outros profissionais conforme as necessidades e as políticas públicas locais.

Cada ESF é responsável por uma população delimitada pelo número de habitantes, sendo que, segundo o MS (2002) “cada equipe se responsabiliza pelo acompanhamento de cerca de três mil a quatro mil e quinhentas pessoas ou mil famílias de determinada área, e estas passam a ter co-responsabilidade no cuidado à saúde”. Essa delimitação populacional forma um território, o qual pode ser considerado como ambiente de trabalho das equipes. Desta forma, reconhecê-lo facilita o entendimento das condições de saúde da população sobre sua responsabilidade melhorando a qualidade do atendimento prestado.

## 2.2 ATENÇÃO À SAÚDE E O TERRITÓRIO

Para atuar sob o enfoque do PSF, as equipes de saúde entendem que a situação de saúde do indivíduo é condicionada por fatores sociais, culturais, econômicos, espirituais e ambientais. Sobre esse último, a abordagem atual incorpora na paisagem natural as mudanças ocorridas pelas construções humanas e as conseqüências que essas podem ocasionar (CHIESA; KON, 2006). Esses fatores interagem com o indivíduo em um espaço denominado território, que por sua vez é entendido de diferentes formas.

O entendimento do conceito de território segue por duas grandes linhas de pensamento, uma que conceitua território de forma estática, considerando-o como uma área delimitada geograficamente por meio de critérios geopolíticos, denominando este espaço como território-solo. A outra linha defende o território como um espaço que possui limites geográficos, entretanto, é assimétrico e está em constante processo de transformação por conter uma dinâmica social. Esta última linha denomina o espaço em transformação como território-processo (UNGLERT, 1993).

Para conhecer a dinâmica de um território contemplando suas assimetrias e complexidades, Mendes (1993) sugere subdividir o território em unidades menores. Para isso uma subdivisão a ser considerada é o território-área, o qual corresponde à área de abrangência de uma unidade de saúde (US). Uma área de abrangência é limitada pela capacidade de atendimento de uma US e o fluxo e contra fluxo da população (MIRANDA, 2007). Assim, os limites geográficos do território área de abrangência, são definidos por barreiras que impedem o livre acesso da população quando procuram atendimento nas US. Podem ser consideradas

como barreiras de acesso, a distância entre o serviço de saúde e os domicílios, a inclinação do solo, a presença de rios, rodovias ou outros que dificultem chegar até a US (MENDES, 1993).

A partir da área de abrangência, o território é novamente subdividido em áreas menores denominadas de microáreas. A microárea é uma pequena área de formato assimétrico e condições econômicas e demográficas nem sempre homogêneas (SOUZA; MORAES; BERNARDES, 2005).

Algumas microáreas podem ser consideradas como microáreas de risco, isto por conterem características que podem interferir de forma direta ou indireta na saúde da população. Estas características provêm praticamente de duas condições, uma referente ao ambiente físico, o qual envolve as condições ambientais e de infra-estrutura como saneamento, abastecimento de água, arruamento, condições de habitação, entre outros. E a outra condição se associa ao perfil dos moradores, como condições sócio-econômicas, culturais e demográficas. Um conjunto de condições, quando deficientes ou precárias, aumenta a probabilidade dos moradores em desenvolver agravos à saúde. Assim, atualmente a classificação utilizada para as microáreas é dada como risco ou não risco.

A microárea por ser uma pequena divisão territorial, pode ser considerada como um espaço privilegiado para o combate de problemas de saúde. Isto devido à proximidade com que a equipe de saúde se encontra em relação à comunidade. Logo, ao reconhecer uma microárea de risco, a equipe de saúde tem a oportunidade de investir em ações mais direcionadas aos problemas encontrados (TEIXEIRA; PAIN; VILASBOAS, 2002).

Para finalizar a divisão do território, podem ser encontradas dentro das microáreas, pequenas unidades as quais são denominadas de território domicílio. Este território domicílio se refere a uma micro unidade social, ou seja, uma família.

Para ilustrar a divisão de um território, a Figura 1 representa uma área de abrangência dividida em microáreas sendo uma considerada de risco.

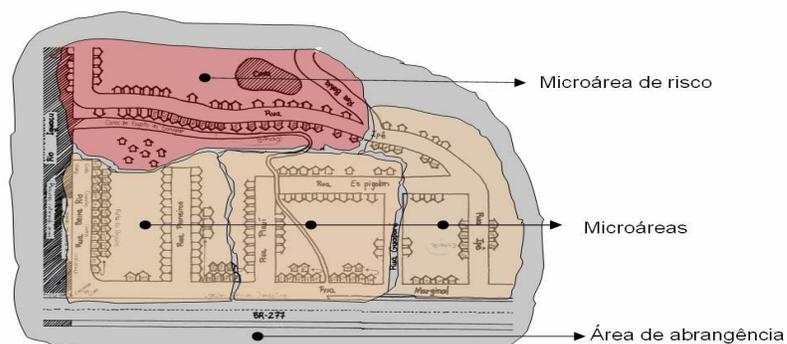


Figura 1 – Representação da divisão de um território em área de abrangência, microáreas e microáreas de risco

### **2.2.1 Reconhecendo o território**

Com base no conceito de território-processo, as equipes de saúde buscam reconhecer no território de sua responsabilidade, as condições e situações de saúde dos indivíduos identificando seus problemas, riscos e potencialidades como ponto de partida para gerar o planejamento e a programação de ações de saúde a nível local (MONKEN; BARCELLOS, 2005).

O reconhecimento dos riscos e potencialidades, contidas em um território, ocorre por um processo de coleta e análise de dados, o qual permite delinear os limites e identificar as condições das áreas de atuação das equipes de saúde. Ainda, é neste momento de reconhecimento do território, que se inicia o vínculo entre a equipe e a comunidade (TEIXEIRA; PAIN; VILASBOAS, 2002).

A importância de reconhecer os riscos e potencialidades em um território se justifica pelo fato que, a causalidade e a determinação de certas doenças se apresentam em diferentes proporções em um mesmo território, assim como, os recursos para combatê-los serem escassos. Portanto, torna-se fundamental identificar quais são as microáreas que contêm maiores riscos a fim de planejar as possíveis intervenções.

Para a identificação das microáreas de risco, podem ser aplicadas diversas metodologias, dentre elas a estimativa rápida participativa (ERP). A ERP se destaca de outras metodologias por apresentar vantagens como: proporcionar informações que representam condições específicas de uma dada população, ser de baixo custo e rapidez, envolver a população assim como alguns setores que compartilham os problemas locais (PIOVESAN et al., 2005).

### **2.2.2 Estimativa rápida participativa**

A ERP é uma metodologia para a coleta e análise de dados que proporciona a identificação das condições de vida e saúde de grupos distintos. Também permite, por meio da distinção de diferentes condições socioeconômicas, epidemiológicas e ambientais, identificar microáreas de risco em um território (BRASIL, 2000).

A ERP tem como objetivo fornecer informações para a equipe de saúde, por meio de uma coleta de dados sintética e que represente as condições locais e situações específicas (BRASIL, 2000). Para a aplicação da ERP é recomendado organizá-la em três etapas, sendo elas a organização dos detalhes, sistematização da área de trabalho e análise dos dados. Estas três etapas são descritas a seguir.

#### 2.2.2.1 Organização dos detalhes

Na etapa de organização dos detalhes, são coletados dados oriundos de bases de dados de diferentes sistemas de informação. As bases de dados consultadas podem ser de órgãos oficiais como o IBGE, secretarias estaduais e secretarias municipais. Outras bases de dados consultadas podem ser de órgãos não oficiais como os cartórios, instituições de ensino, empresas de pesquisas entre outros. Essas fontes são consideradas como fontes secundárias de dados (TAVARES; TAKEDA, 2004).

Para a organização dos dados, a coleta de dados referentes ao perfil socioeconômico, demográfico, epidemiológico e organizacional da comunidade, pode ser realizada por meio dos sistemas de informação disponíveis (SAVASSI; DIAS, 2007). Entretanto, os dados disponibilizados nos diferentes sistemas de informação de órgãos oficiais são agregados para representar grandes territórios, ou seja, representar características de estados, municípios e bairros. Consequentemente, esses dados representam um perfil genérico de um território, tornando-se insuficiente para permitir a identificação das diferenças contidas em pequenos espaços como às microáreas (CHIESA; KON, 2007).

Também é importante salientar que essas bases de dados nem sempre são atualizadas com a mesma rapidez que ocorrem as mudanças no espaço territorial. Isto torna sua utilização limitada pelas equipes, pois na maior parte do tempo, não refletem de forma atualizada a situação local. Ainda em decorrência das crescentes ocupações irregulares, a limitação dos sistemas de informação se agrava quando se pretende obter dados referentes às periferias das grandes cidades, pois nem sempre há informações atualizadas das mesmas.

### 2.2.2.2 Sistematização da área de trabalho

A segunda etapa se refere à sistematização da distribuição da área para o trabalho de campo, o qual é realizado com o auxílio de um mapa da região para a divisão empírica das áreas a fim de aperfeiçoar e facilitar o trabalho.

Subseqüente à coleta de dados de fontes secundárias é necessário que se percorra todo o território acompanhado de um mapa para facilitar a observação e delimitação empírica de possíveis microáreas de risco. Esta observação também orienta e distribui as equipes no território para a posterior aplicação de entrevistas para os informantes chave. Os informantes chave podem ser moradores antigos da região, presidentes de associação de moradores, comerciantes, professores, funcionários da rede de saúde entre outras pessoas que participam ativamente da vida da comunidade (SAVASSI; DIAS, 2007).

### 2.2.2.3 Análise dos dados

Nesta etapa, as equipes voltam ao território aplicando um olhar crítico sobre as condições da comunidade, e são aplicados questionários para os informantes chave, a fim de completar e desvendar condições não evidenciadas com os dados obtidos das fontes secundárias. Esses dados coletados diretamente com os moradores são considerados como fontes primárias de dados.

Os dados coletados por fontes primárias, secundárias e pelas observações realizadas, finalmente são agrupados em categorias e analisados. Essa análise busca desvelar os principais problemas, riscos e potencialidades que a comunidade apresenta, assim como identificar as microáreas que representam maior risco aos seus moradores. A análise dos dados deve receber significativa importância e ser realizada por pessoas capacitadas, pois, por meio dessa análise será possível identificar áreas de risco, que para o MS (2006):

A identificação dessas áreas, pessoas ou famílias que precisam de atenção especial é fundamental para que sejam programadas ações específicas e atividades de acompanhamento permanente àquela comunidade. A necessidade de atenção especial pode ser causada por uma situação de desequilíbrio já existente na área, ou porque tal comunidade está sujeita as situações de maior risco de adoecer ou morrer.

Apesar de fundamental, a análise dos dados é realizada de forma subjetiva pelos integrantes das equipes de saúde, pois, se considerar que esses possuem experiências e percepções diferentes, estes olhares podem influenciar tanto na identificação dos problemas, quanto na identificação das áreas consideradas de risco.

Perante esse contexto e dada à importância da análise dos dados, nota-se que este é um campo em que a área da computação pode oferecer suporte por meio de técnicas e ferramentas para a manipulação de dados. O suporte proveniente da área da computação pode otimizar e melhorar a análise dos dados. O processo de descoberta de conhecimento em base de dados é uma das áreas que pode ser utilizada para auxiliar na manipulação de dados para classificar as microáreas de risco e é apresentado na seção 2.3.

### 2.3 DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS

O avanço da informática permitiu um aumento rápido e significativo no armazenamento e recuperação de dados, no entanto esses grandes volumes de dados gerados podem algumas vezes ultrapassar a capacidade de análise pelos métodos tradicionais como a geração de relatórios, gráficos, planilhas entre outros. Impulsionada por esse volume crescente de dados e a necessidade de análise sobre eles, a ciência da computação tem se dedicado a desenvolver novas ferramentas e métodos de análise sobre as bases de dados. A área que vem se destacando neste contexto é o KDD (CARVALHO, 1999).

O KDD é um processo que busca extrair padrões, associações, modelos ou informações relevantes que permanecem ocultos em bases, repositórios e outras formas de armazenamento de dados. Segundo Navega (2002) esses padrões podem ser definidos como “unidades de informação que se repetem, ou então seqüências de informações que dispõem de uma estrutura que se repete”. Para Fayyad et al. (1996), o processo de KDD permite identificar padrões válidos, novos, potencialmente úteis e compreensíveis.

Para atingir o objetivo de extrair padrões novos, úteis e compreensíveis, o processo de KDD envolve diversas áreas como aprendizado de máquina, base de dados, estatística, reconhecimento de padrões, visualização e continua a envolver novas áreas. Devido a isto é considerado como uma área interdisciplinar (CARVALHO, 1999; FAYYAD et al, 1996).

Os padrões identificados são utilizados para formar novos conhecimentos que podem ser aplicados para o suporte à decisão. Atualmente, o KDD é aplicado em diversas áreas como a administração, análise de *marketing*, medicina, entre outros (HAN; KAMBER, 2001).

Ainda, o KDD é um processo que ocorre de forma interativa entre o especialista do domínio, o analista e o usuário final e também de forma iterativa visto que pode haver a necessidade de refazer alguma etapa do processo. O processo de KDD é dividido em etapas, as quais são apresentadas a seguir.

### 2.3.1 Etapas do processo de KDD

Segundo Rezende (2005), as etapas de KDD podem ser divididas em três, sendo elas: o pré-processamento, a extração de padrões e pós-processamento. Essas etapas são apresentadas na Figura 2 e comentadas a seguir.

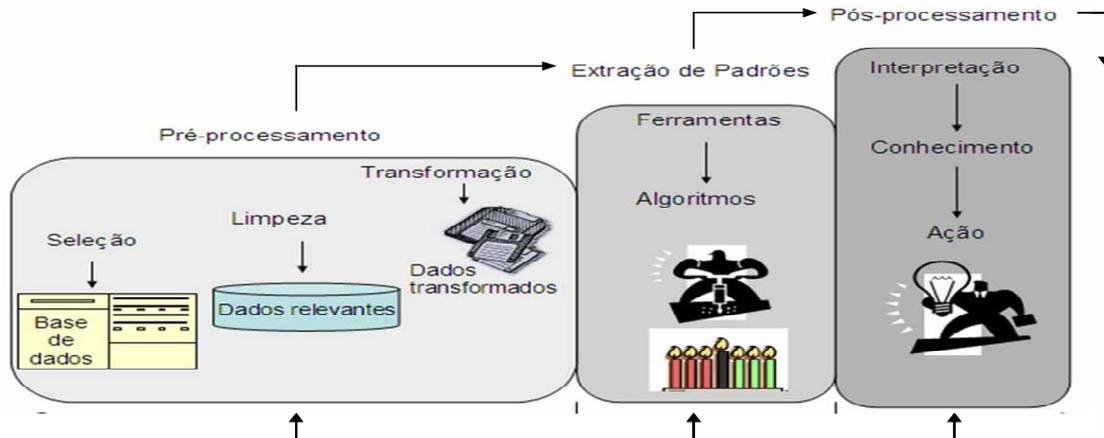


Figura 2 – Processo de descoberta de conhecimento  
Fonte: Adaptado de Fayyad, Piatetsky-Shapiro e Smyth (1996)

### 2.3.1.1 Pré-processamento

Nesta etapa inicia-se a definição do problema, a escolha dos atributos a serem utilizados e a organização e preparação da base de dados para a extração do conhecimento (REZENDE, 2005). A etapa de pré-processamento é considerada de grande importância devido às bases de dados não terem sido modeladas para a mineração de dados, pois, geralmente as bases disponibilizadas pelos sistemas de informação contêm falhas. Dentre as falhas encontradas estão os valores ausentes para muitos atributos, formatos incompatíveis com o formato requisitado pela ferramenta a ser utilizada na extração de padrões, entre outros.

Com o objetivo de adequar as bases para a extração de padrões, a etapa de pré-processamento recorre a técnicas como integração dos dados, limpeza e transformação dos dados, as quais são comentadas a seguir.

- a) **integração dos dados:** técnica utilizada quando há a necessidade de selecionar e integrar dados de outras bases para melhorar o conjunto de dados. Esta integração visa à unificação da fonte de dados (REZENDE, 2005);
  
- b) **transformação:** trata-se da adequação dos dados a serem utilizados pelos algoritmos na fase de extração de padrões. Essa adequação envolve métodos como a normalização de atributos, resumos por agrupamentos, entre outros. Um exemplo típico é a transformação de data de nascimento em idade ou mesmo estrato etário. Alguns algoritmos podem apresentar melhor desempenho quando os valores dos atributos são discretos, outros algoritmos processam apenas valores contínuos, assim a técnica utilizada para a adequação dos dados fica subordinada ao algoritmo a ser aplicado. A transformação também envolve a codificação dos dados para um formato que permita sua leitura pelo algoritmo e a ferramenta a ser utilizada. (GOLDSCHIMIDT; PASSOS, 2005);
  
- c) **limpeza:** os dados podem se encontrar incompletos ou ruidosos em decorrência da forma em que foram coletados ou até mesmo por erros de digitação no momento da entrada dos dados. Dados incompletos são aqueles onde há ausência de valores em um ou mais atributos e ruídos são valores que não condizem com o padrão normal

esperado. Assim, a limpeza tem como objetivo preencher as falhas. Esse preenchimento pode ocorrer manualmente pela eliminação de dados incompletos e ruidosos, pelo preenchimento das falhas com o auxílio de um especialista ou de forma automática pela predição de valores. Para bases com falhas nos dados, atualmente, há no mercado ferramentas que completam estas falhas com símbolos, auxiliando na etapa de pré-processamento, (GOLDSCHIMIDT; PASSOS, 2005)

### 2.3.1.2 Mineração de dados

Após a etapa de pré-processamento, inicia-se a fase de mineração de dados (MD) do inglês *Data mining (DM)*. Essa etapa é considerada como essencial na descoberta do conhecimento e envolve a escolha e aplicação da ferramenta e algoritmo a ser utilizado. Esse algoritmo é fundamentado em técnicas de inteligência artificial, dentre essas técnicas, estão a indução de regras, árvores de decisão, algoritmos genéticos entre outros (REZENDE, 2005).

Alguns autores utilizam o termo mineração de dados como uma alternativa para o termo extração de padrões, outros autores fazem uso do mesmo termo como sinônimo de KDD (CARVALHO, 1999). Neste estudo o termo mineração de dados será considerado como uma etapa do KDD, ou seja, a etapa referente à extração de padrões. A relação entre KDD e MD é ilustrada na Figura 3.

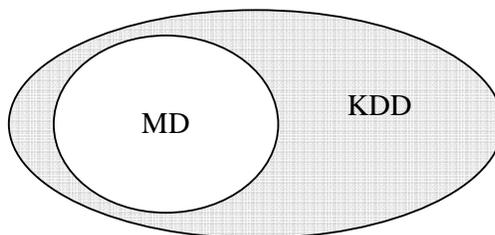


Figura 3 - Relação entre o processo de KDD e a etapa de MD  
Fonte: Carvalho (1999)

As aplicações da mineração de dados são classificadas em descritivas e preditivas. As aplicações descritivas consistem em identificar comportamentos intrínsecos dentro de conjuntos sem uma classe pré-definida. Para esse tipo de aplicação são utilizadas as tarefas de agrupamento, sumarização e regras de associação (REZENDE, 2005).

Em aplicações preditivas se busca demonstrar como certos atributos irão se comportar futuramente, isto ocorre por generalização de exemplos ou experiências passadas que possuem classes pré-definidas. Essa aplicação tem como intenção reconhecer a qual classe pertencerá um novo exemplo. As tarefas para este tipo de aplicação podem ser as classificações e as regressões, as quais se relacionam com o processo de apoio à decisão. Assim, a escolha da técnica a ser utilizada se fundamenta no objetivo e finalidade para qual o conhecimento descoberto será utilizado (REZENDE, 2005).

Neste estudo foi utilizada a atividade preditiva aplicando a tarefa de classificação, a qual será apresentada a seguir.

#### 2.3.1.2.1 Tarefa de classificação

A tarefa de classificação tem como objetivo encontrar modelos de relacionamento entre um atributo e uma classe pré-definida, denominada de classe alvo. Este relacionamento permite prever a qual classe um objeto pertence baseando-se em seus atributos (BOLOGNANI, 2005).

Para a tarefa de classificação, um exemplo simplificado para a área da saúde poderia ser classificar, por meio de sinais e sintomas apresentados, algumas pessoas como portadoras ou não de determinadas patologias previamente definidas (REZENDE, 2005).

O resultado da classificação pode ser representado na forma de árvores de decisão que, por sua vez, podem ser transformadas em regras. A seguir são abordados alguns conceitos referentes à construção de árvores de decisão e regras de classificação por terem sido utilizados neste estudo.

### 2.3.1.2.2 Árvore de decisão

A árvore de decisão é uma representação de um classificador pertencente à família dos algoritmos *Top-Down Induction of Decision Trees* (TDIDT). A árvore de decisão é construída por meio de uma estrutura composta de nós e relações que se formam partindo de uma raiz e seguindo em direção as folhas. A partir do nó raiz são formados os nós de decisão, sendo que estes últimos realizam testes sobre os atributos e para cada resultado emitem um ramo para uma subárvore. Em seguida, um ramo se liga aos nós folhas da árvore que representam as classes previamente definidas (REZENDE, 2005). Um exemplo de árvore de decisão é demonstrado pela Figura 4.

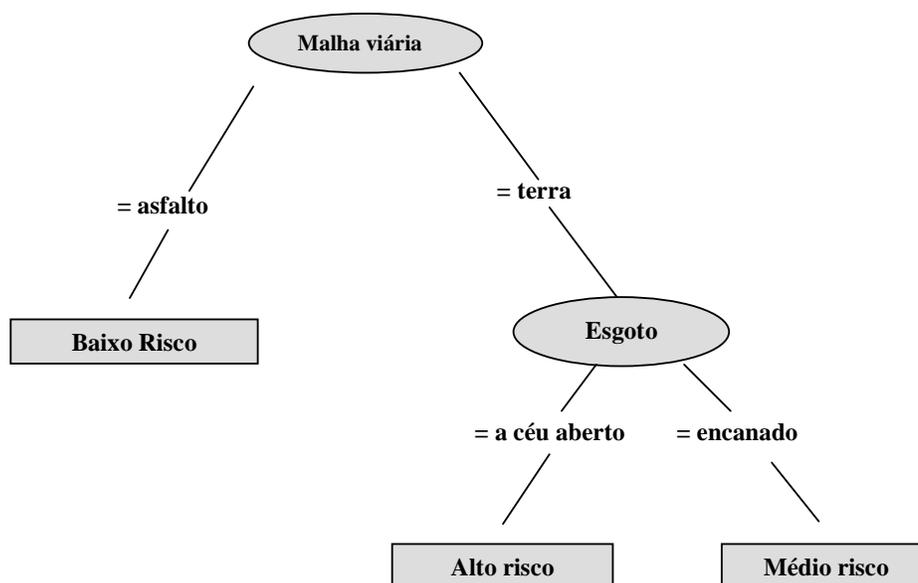


Figura 4 – Modelo de árvore de decisão

No exemplo apresentado na Figura 4, cada elipse representa um nó de decisão, neste nó é realizado um teste sobre um atributo pertencente ao conjunto e cada retângulo representa a classe pré-definida com o nível de risco contido no ambiente. Assim, para classificar um novo ambiente, segue-se do nó raiz em direção ao nó folha.

Uma árvore de decisão também pode ser representada por meio de um conjunto de regras que determinam o caminho a ser seguido. As regras são conjuntos de condições formadas por premissas e conclusões descritas no formato **SE <condição> ENTÃO <conclusão>**. As premissas podem ser ligadas pelos conectivos lógicos **NÃO**, **E** e **OU**. Quando a condição é satisfeita, a conclusão é considerada como verdadeira.

Dentre as vantagens na utilização de regras estão: explicitar o conhecimento garantindo maior legibilidade da base de conhecimento; permitem certa modularidade devido a cada regra ser considerada como uma peça do conhecimento; facilitam a edição pela modularidade permitindo modificar, inserir ou excluir regras conforme exista a necessidade (UNIVERSIDADE FEDERAL DO CEARÁ, 1995).

A representação por meio de regras para a árvore de decisão apresentada anteriormente na Figura 4 é:

**Regra 1**

**SE** malha viária = asfalto

**ENTÃO** : baixo risco

**Regra 2**

**SE** malha viária = terra

**E** esgoto= a céu aberto

**ENTÃO** : alto risco

**Regra 3**

**SE** malha viária = terra

**E** esgoto = encanado

**ENTÃO** : médio risco

Para Michalski (1983 *apud* REZENDE, 2005) uma regra prática é composta por menos de cinco condições em uma conjunção, poucas condições em uma disjunção e no máximo uma implicação. As regras por serem de fácil compreensão e intuitivas, se tornaram a forma mais conhecida para representar o conhecimento em Sistemas Especialistas.

Os sistemas especialistas constituem uma área de aplicação da inteligência artificial (IA) e apresentam raciocínio semelhante a um especialista. Esses sistemas têm como propósito auxiliar na tomada de decisões ou resoluções de problemas (MARQUES; MARIM, 2002).

### 2.3.1.3 Pós-processamento

Na etapa de pós-processamento, os resultados obtidos são analisados e interpretados. Collazos et al. (2002) referem que há três tipos de resultados obtidos pela mineração de dados, sendo eles: a confirmação do óbvio, um novo conhecimento, ou nenhuma relação significativa entre os dados. Ainda, na etapa de pós-processamento, são realizadas análises considerando desempenho e a qualidade do resultado obtido pela mineração de dados. Para avaliar o desempenho da mineração de dados, podem ser consideradas medidas como cobertura, erro e taxa de acerto.

Para avaliar os resultados, sob a perspectiva da qualidade, podem ser consideradas medidas como interessabilidade e compreensibilidade.

A interessabilidade, ou grau de interesse, reflete como especialistas julgam os padrões encontrados. As medidas de interessabilidade estão fortemente relacionada a utilidade que a regra apresenta para o usuário final, ou seja o quão interessante é o padrão descoberto para o usuário (REZENDE, 2005). Já a medida de compreensibilidade de uma regra pode ser avaliada de acordo com a facilidade de interpretação pelo ser humano. Os resultados obtidos na mineração de dados que apresentam menor quantidade de regras e regras com menor número de condições, tendem a serem mais bem compreendidos pelo usuário final (REZENDE, 2005).

A utilização destas e outras medidas possibilitam a avaliação de todas as etapas e podem indicar a necessidade de reiniciar o processo visando aprimorar as etapas em busca de melhores resultados (REZENDE, 2005).

### 3 METODOLOGIA

Considerando a importância em identificar, no ambiente físico, fatores que possam influenciar a saúde da coletividade, este estudo buscou identificar padrões para auxiliar a classificação de microáreas em relação aos riscos oriundos do ambiente físico. Para isso, as etapas de identificação dos atributos e aquisição dos dados, se constituíram em uma pesquisa exploratória metodológica. Nestas etapas foram aplicados instrumentos estruturados com afirmações fechadas para a coleta de dados (TOBAR; YALOUR, 2001).

Na etapa de aplicação do processo de KDD, a base de dados foi preparada e submetida à mineração de dados, sendo após avaliada e analisada concluindo a fase de pós-processamento. As três etapas que constituíram este estudo são apresentadas na Figura 5 e descritas a seguir.

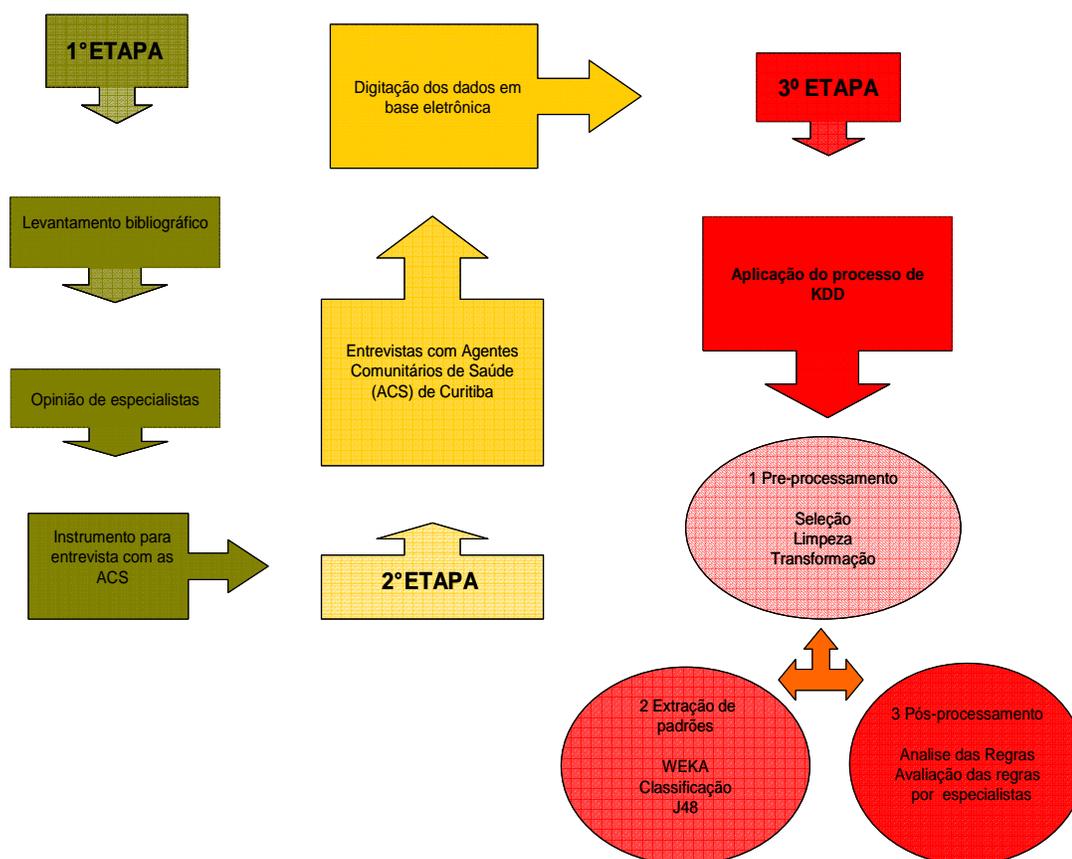


Figura 5 – Representação esquemática do percurso metodológico

### 3.1 CARACTERIZAÇÃO DO CAMPO DE ESTUDO

Este estudo foi desenvolvido na cidade de Curitiba, capital do estado do Paraná, fundada em 29 de março de 1693. Segundo o IBGE (2007), em 2007, sua população era de 1.797.408 habitantes distribuídos em 435,495 km<sup>2</sup>, sendo a sua extensão territorial dividida em 75 bairros, dos quais, aqueles que se encontram localizados próximos da área central possuem a maior renda por chefe de família e alta densidade populacional. Entretanto, alguns dos bairros localizados na periferia da cidade também apresentam alta densidade populacional, mas, renda mediana baixa. Considerando essa característica da cidade, a política de desenvolvimento prioriza seus investimentos em equipamentos sociais nos bairros de menor renda e alta densidade populacional, buscando melhorar a qualidade de vida de seus moradores (PREFEITURA MUNICIPAL DE CURITIBA, 2003).

Em relação aos indicadores de 2007, a cidade de Curitiba apresentou uma densidade demográfica de 4.127,28 hab/Km<sup>2</sup> e um índice de desenvolvimento humano (IDH) de 0,856 colocando a cidade na 16ª posição em relação às outras cidades Brasileiras (PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO, 2007).

Com relação as estatísticas apresentadas no IDH de 2003, a cidade possuía grau de urbanização de 100%. No ano de 2000, na área social, a cidade possuía 42.620 famílias em situação de pobreza<sup>iii</sup> totalizando 155.080 pessoas nesta situação (PREFEITURA MUNICIPAL DE CURITIBA, 2003).

Na área da saúde, as ações são desenvolvidas por equipes de saúde em parceria com a comunidade, sendo as equipes distribuídas em unidades de saúde de complexidades diferentes conforme a localização e necessidade da população.

Para desenvolver ações em epidemiologia, planejamento, informação em saúde e outras atividades passíveis de realização a nível regional, Curitiba está dividida em nove distritos sanitários, sendo elas Santa Felicidade, Boa Vista, Matriz, Portão, Pinheirinho, Cajuru, Boqueirão, Cidade Industrial de Curitiba e Bairro Novo (PREFEITURA MUNICIPAL DE CURITIBA, 2003).

Os nove Distritos sanitários abrigam 123 equipamentos de saúde, dos quais 47 são Unidades Básicas de Saúde, 48 Unidades Básicas com estratégia da Saúde da Família (totalizando 96 áreas de abrangência), 11 Unidades de Saúde Complexas, 7 Centros de

---

<sup>iii</sup> Situação de pobreza é a população calculada em função da renda familiar per capita de até meio salário mínimo.

Atendimento Psicossocial (CAPS), 8 Centros Municipais de Urgências Médicas (CMUM), um hospital com maternidade e um Laboratório de Análises Clínicas. Para viabilizar a estratégia de saúde da família a Secretaria Municipal de Saúde (SMS) disponibiliza 131 equipes e 1.149 ACS (PREFEITURA MUNICIPAL DE CURITIBA, 2007).

Além da estrutura física e dos recursos humanos disponibilizados pela prefeitura, esta rede ainda conta com instituições hospitalares e clínicas conveniadas para suporte ao atendimento. A distribuição espacial da rede municipal de saúde, assim como as divisões das regionais, podem ser observadas na Figura 6.

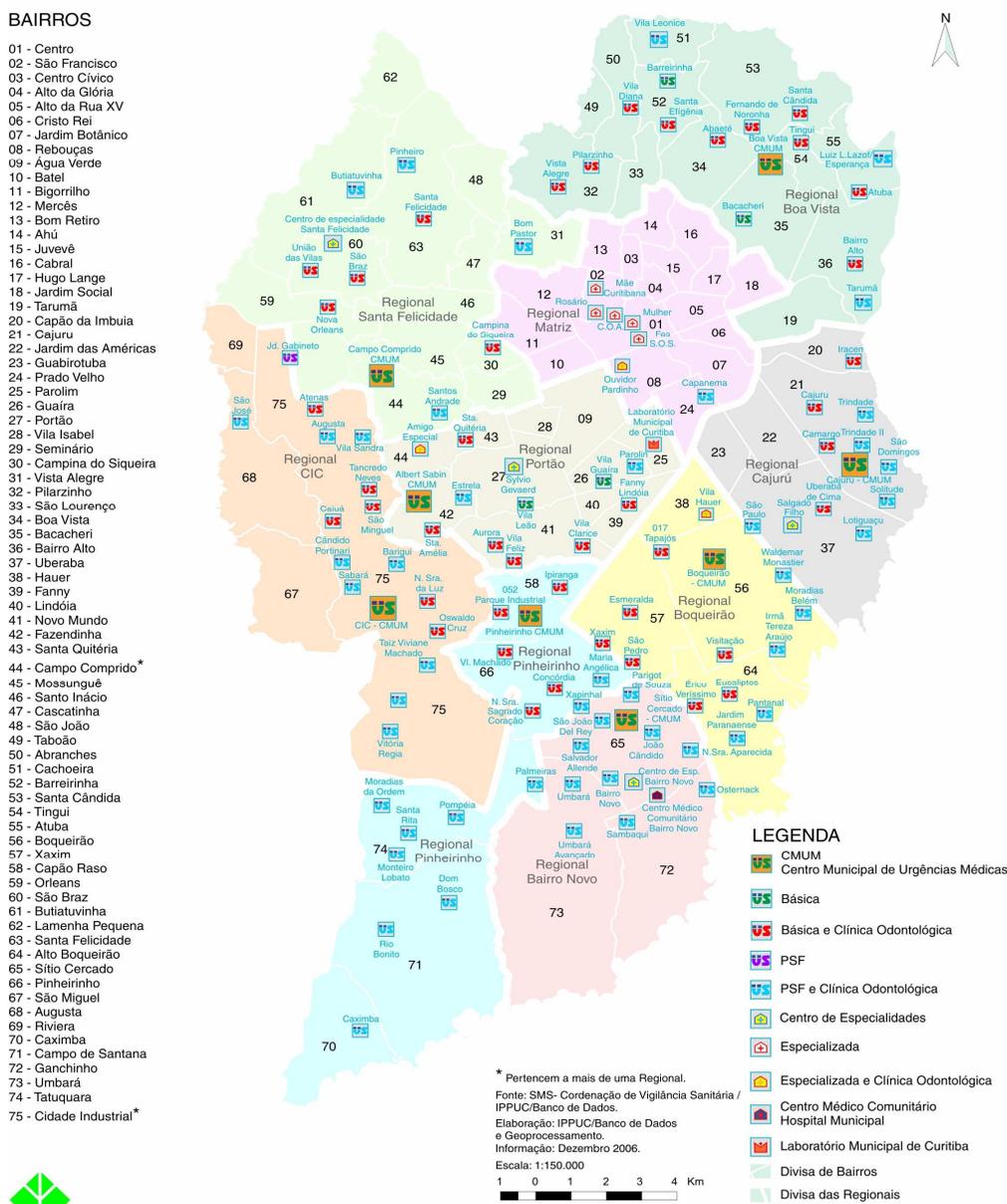


Figura 6 – Distribuição da rede municipal de saúde de Curitiba

Fonte: IPPUC (2008)

### 3.2 SUJEITOS ENVOLVIDOS NAS PESQUISAS

Para a etapa **identificação dos atributos** importantes para a classificação de microáreas de risco, foram envolvidos oito especialistas, sendo três médicos e cinco enfermeiros. A seleção dos especialistas ocorreu com base nos seguintes critérios de inclusão: profissionais com formação de nível superior nas áreas de enfermagem, medicina ou odontologia; com atuação na área da saúde coletiva por no mínimo dois anos; vinculados à área acadêmica e com titulação mínima de mestre.

Na etapa de **construção da base de dados**, foram considerados como sujeitos da pesquisa os ACS da cidade de Curitiba por serem moradores da região e ao mesmo tempo integrantes das equipes de saúde, portanto mais habilitados para preencherem o instrumento de coleta de dados, nesta etapa foram entrevistados 531 ACS. Como critérios de exclusão foram considerados aqueles que estavam de férias ou folga no período da coleta de dados e aqueles que estavam lotados no distrito sanitário Matriz, por este apresentar dinâmica diferenciada na assistência à saúde.

Para a etapa de **pós-processamento do processo de KDD**, o conjunto de regras foi avaliado por três especialistas da área de saúde coletiva. Os critérios de inclusão foram: com formação na área da saúde, atuantes na área de saúde coletiva há mais de dois anos e titulação mínima de especialista em saúde coletiva, pública ou da família.

### 3.3 CONSIDERAÇÕES ÉTICAS DO ESTUDO

Os preceitos éticos deste estudo foram resguardados pela aprovação do comitê de ética em pesquisa da Pontifícia Universidade Católica do Paraná (PUCPR) dos Projetos **Concepção de um sistema de apoio à decisão para caracterização de micro áreas de risco na comunidade (Reg. 1442)** (Anexo B) e **Desenvolvimento de Base de Dados Referente a Microáreas (Reg. 1771)** (Anexo C), sendo a viabilidade deste último, aprovada pelo Comitê de Ética em pesquisa da Secretaria Municipal de Saúde sob o registro 76/2007. Portanto ambos os projetos cumprem Resolução nº. 196/96 do Conselho Nacional de Saúde.

### 3.4 ETAPAS PARA O DESENVOLVIMENTO DO ESTUDO

A seguir são descritas as etapas realizadas para o desenvolvimento deste estudo, sendo elas a identificação de atributos para classificar as microáreas, a construção da base de dados e a aplicação do processo de KDD.

#### 3.4.1 Etapa 1 – Identificação de atributos para classificar as microáreas

Como etapa inicial deste estudo, foi realizada uma pesquisa exploratória, por meio da aplicação de um questionário (Apêndice A) para oito especialistas, dentre eles cinco enfermeiros e três médicos, a fim de identificar atributos relevantes para avaliar o risco contido no ambiente físico de uma microárea.

A amostra desta etapa se limitou a oito participantes por ocorrer a partir deste número, a saturação dos dados<sup>iv</sup>, ou seja, as informações compartilhadas se tornaram repetitivas. O preenchimento do questionário pelos especialistas, ocorreu conforme a disponibilidade dos mesmos, sendo realizado entre os meses de fevereiro a abril de 2007.

Como modelo para a elaboração do questionário, foi utilizada a escala de Lickert, a qual é composta por uma lista de condições que podem ser afirmadas como essenciais, recomendáveis ou irrelevantes (WOOD; HABER, 2002).

O questionário foi elaborado a partir de um levantamento bibliográfico identificando possíveis atributos considerados como riscos ou potencialidades para a saúde de uma comunidade. Os atributos considerados riscos foram entendidos como aqueles que quando presentes aumentam a probabilidade dos indivíduos em sofrer agravos à saúde, dentre eles a falta de saneamento e energia elétrica, presença de fontes contaminadoras do solo ou ar, entre outros. Com base em Santos e Noronha (2001) e Tomasi et al. (2003), os atributos potencializadores para a saúde da comunidade foram considerados como aqueles que

---

<sup>iv</sup> Segundo Wood e Harber (2002) a saturação dos dados é entendida como “um ponto em que a coleta de dados pode cessar. Ocorre quando as informações que estão sendo compartilhadas com o pesquisador tornam-se repetitivas. As idéias transmitidas pelo participante foram compartilhadas antes por outros participantes; inclusão de participantes adicionais não resulta em novas idéias”.

constituem um meio disponível para melhorar a saúde da comunidade, como escolas, creches, acessibilidade entre outros.

Os atributos identificados na literatura foram classificados em categorias referentes às condições de infra-estrutura e ambiente da microárea e são descritos a seguir.

#### 3.4.1.1 Condições ambientais

- **Geografia:** essa categoria integra as questões geográficas da comunidade. Alguns fatores como a distância e inclinação do solo tornam-se relevantes em determinados casos devido à necessidade de locomoção dos usuários para a unidade de saúde, representando assim barreiras geográficas de acesso, ou onde se associam a presença de rios em terrenos baixos e planos representando a possibilidade de alagamento e a potencial veiculação de doenças (SOUZA; MORAES; BERNARDES,2005).
- **Vetores:** para este grupo foram selecionados os atributos referentes às zoonoses, ou seja, animais, roedores e outros que pela sua presença podem transmitir doenças através de suas excreções ou mordidas.
- **Poluição ambiental:** nesta categoria encontram-se atributos referentes a presença de plantações e indústrias poluentes pela possibilidade de contaminação do meio ambiente que estes podem causar. Esta contaminação pode ser por eliminação de gases ou produtos químicos contaminantes da hidrografia local ou do solo, o que tem ocorrido e atingido grandes populações influenciando diretamente na saúde das mesmas (CASTRO; GOUVEIA; ESCAMILA, 2003).

### 3.4.1.2 Condições de infra-estrutura

- **Equipamentos sociais:** esta categoria agrupou os atributos relacionados aos serviços disponibilizados à comunidade, sendo eles públicos ou de ordem religiosa.
- **Comércio:** nesta categoria foram incluídos estabelecimentos comerciais, pela sua importância em disponibilizar a venda de mantimentos para alimentação, higiene, entre outros.
- **Áreas de lazer:** nesta categoria foram agrupados locais destinados a recreação da comunidade. São locais de acesso público que permitem a realização de exercícios físicos como caminhada ou exercícios em grupos como vôlei ou futebol, além de ambientes que proporcionam momentos de relaxamento, tanto físicos como mental.
- **Urbanização:** para compor esta categoria foram selecionados os atributos relacionados ao processo de urbanização da comunidade, sua forma de organização, e a disponibilidade de recursos públicos em saneamento, esgoto e coleta de lixo.

Por meio do questionário elaborado, os especialistas qualificaram os atributos como: essenciais, recomendáveis ou irrelevantes para o estudo. Entre as afirmações essencial, recomendável ou irrelevante foi incluída uma escala contendo valores entre zero e um. Para esse valor foi considerado que quanto mais próximo de um, maior impacto de influência o atributo possui sobre a saúde da comunidade e assim maior representatividade para o estudo. A partir disto, para selecionar as variáveis mais representativas, foi definido que aquelas que mantiveram valores igual ou acima de 0.7, indicando ser essencial para compor o estudo, permaneceram pela sua relevância e os atributos que mantiveram valores abaixo de 0.7 foram eliminados.

### 3.4.2 Etapa 2 - Construção da base de dados

A base de dados utilizada neste estudo foi construída com dados adquiridos por meio de entrevistas com ACS de Curitiba. Para realizar as entrevistas foi utilizado um questionário com respostas fechadas (Apêndice B), elaborado com os atributos indicados como essenciais pelos especialistas na etapa anterior.

Os questionários foram aplicados aos ACS no período de setembro a dezembro de 2007, obtendo-se neste período 531 entrevistas representando uma amostra de 46,2% do total de ACS de Curitiba. Neste momento, é importante ressaltar que cada ACS entrevistada é responsável por uma microárea, logo, a amostra representa 46,2% das microáreas de Curitiba.

Na parte final do questionário havia um campo a ser preenchido pela opinião das ACS referente à como os ACS consideravam a microárea em que atuam. Como resposta fechada para este campo, o ACS poderia indicar se a microárea era considerada entre baixo, médio ou alto risco. Esta opinião foi utilizada como a classe alvo na etapa de mineração de dados.

Os questionários foram digitados em planilha Excel® contendo 41 colunas representando os atributos e seus possíveis valores e uma coluna representando a classe alvo para o estudo.

Devido ao fato de não ser possível identificar nos mapas disponibilizados os limites das microáreas, a Figura 7 aponta com um círculo as US responsáveis pelas microáreas envolvidas na pesquisa. Cabe ressaltar que cada área de abrangência pode ser responsável por uma ou mais microáreas, sendo que esta variação é dependente do tamanho da área de abrangência, quantidade de população e características locais.

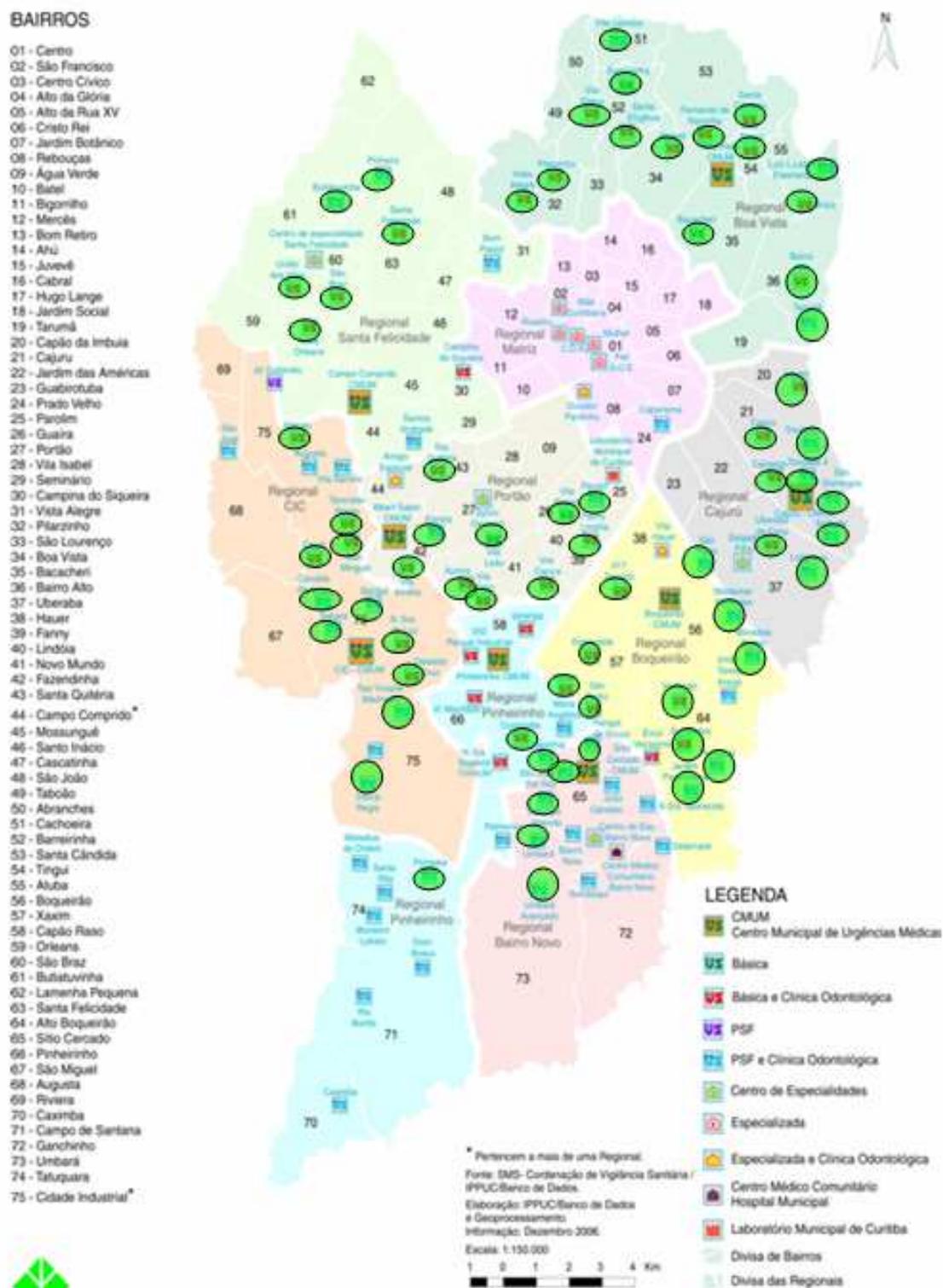


Figura 7 – Áreas de abrangência envolvidas no estudo  
 Fonte: Adaptado de IPPUC (2008).

### 3.4.3 Etapa 3 – Aplicação do processo de KDD

Nesta etapa foi aplicado o processo de KDD sobre a base de dados adquirida na etapa de aquisição e digitação dos dados. Esse processo seguiu as fases de pré-processamento, mineração de dados e pós-processamento conforme descritos a seguir.

#### 3.4.3.1 Pré-processamento

Na etapa de pré-processamento, a base foi submetida às técnicas de seleção, transformação e formatação dos dados, como são descritas a seguir.

- **Seleção dos dados**

Foi observado que durante a digitação das entrevistas, havia três atributos que mantiveram valor constante, esses atributos foram retirados da base por serem considerados redundantes e são apresentados no Quadro 1.

Atributo	Valores possíveis	Valores encontrados na base de dados
Unidade de Saúde	Presente / Ausente	Presente
Açude	Presente / Ausente	Ausente
Meio de transporte	Ônibus / Lotação	Ônibus

**Quadro 1 - Atributos com valores constantes retirados por serem considerados redundantes.**

Com a retirada dos atributos apresentados na Quadro 1 a base passou a conter 38 colunas representando os atributos e seus possíveis valores e uma coluna representando a classe alvo com os valores de baixo, médio ou alto risco.

- **Transformação dos dados**

Em alguns questionários foram assinalados múltiplos valores para um mesmo atributo. Para melhorar o desempenho do processo de mineração de dados, aqueles atributos com mais de um valor assinalado, foram transformados em novos atributos, porém, sem alterar o

significado original. O Quadro 2 apresenta os atributos modificados. A primeira coluna do Quadro 2 mostra a descrição do atributo original, na segunda os possíveis valores originais, a terceira mostra o atributos modificados e a quarta coluna os valores possíveis para os atributos modificados.

<b>Atributo original</b>	<b>Valor original</b>	<b>Atributo Modificado</b>	<b>Valores modificados</b>
Distribuição de energia elétrica	Regularizada e/ou clandestina	Distribuição de energia elétrica regularizada	Sim Não Parcialmente
Abastecimento de água	Clandestina, poço comum, bica, cisternas e poço artesiano.	Abastecimento de água por rede pública	Sim Não Parcialmente
Rede de esgoto	A céu aberto, fossa e encanado	Rede de esgoto	A céu aberto fossa encanado Misto
Destino do lixo	Em caçamba e coletado pela prefeitura, em terrenos vazios e coletados pela prefeitura	Destino do lixo coletado pela prefeitura	Sim Não Parcialmente
Distribuição das moradias	Organizadas, Aglomerado	Distribuição das moradias organizada	Sim Não Parcialmente
Barreiras de acesso	Rodovia, distante, rio, baía ou mar	Barreiras de acesso	Sim Não
Indústrias poluentes	Ar, solo, água, sonora	Indústrias poluentes	Presente Ausente
Área de construção das casas	Em processo de legalização e não legalizada	Área de construção das casas legalizada	Sim Não Parcialmente

**Quadro 2 - Relação dos atributos e valores modificados.**

- **Formatação da base de dados**

A formatação da base de dados foi realizada por meio da conversão do formato planilha *Excel*® para o formato *arff*. Esse último formato é próprio da ferramenta utilizada na etapa de mineração de dados e a conversão é realizada pela própria ferramenta no momento de importar dos dados.

### 3.4.3.2 Mineração de dados

Para a mineração de dados, foi utilizada a ferramenta operacional WEKA<sup>v</sup> (*Waikato Environment for Knowledge Analysis*) versão 3.5. WEKA é um *software* de domínio público, desenvolvido pelo curso de Ciência da Computação da Universidade de Waikato na Nova Zelândia. Esse *software* é composto por um conjunto de implementações de diversos tipos de algoritmos para tarefas de associação, classificação e clusterização (KIRKBY; FRANK; REUTEMANN, 2007).

Com o objetivo em descobrir, por meio de um conjunto de dados, classes diferentes de atributos para classes pré-definidas (classes alvo), a tarefa apropriada para esse problema foi à classificação mediante o algoritmo J48. É importante destacar, que o objetivo deste estudo foi descobrir padrões para serem utilizados na classificação de microáreas, assim, um resultado compreensível é de fundamental importância. Isto justifica a utilização da técnica de árvore de decisão para a tarefa de classificação, pois, a partir desta técnica é possível gerar regras de mais fácil compreensão e nas quais é possível identificar os atributos discriminantes.

A classe alvo utilizada na tarefa de classificação foi o atributo “**você considera a microárea**”, o qual contém três valores, sendo eles: **baixo, médio** e **alto risco**. Cabe lembrar que os valores deste atributo foram indicados conforme a percepção das ACS sobre sua microárea de atuação.

---

<sup>v</sup> <http://www.cs.waikato.ac.nz/ml/weka/>.

### 3.4.3.3 Pós-processamento

O arquivo de saída do algoritmo J48 foi apresentado nos formatos de árvore decisão e regras de produção no estilo **SE... ENTÃO...**. A fim de melhorar a compreensão das regras para os especialistas avaliadores, as regras foram separadas e após acrescentados os conectores lógicos **SE**, **E** e **ENTÃO** no início dos antecedentes e conseqüentes.

Nesta dissertação, a avaliação na etapa de pós-processamento foi realizada sob as perspectivas do desempenho computacional e da qualidade do conjunto de regras.

Para a avaliação do desempenho computacional foram consideradas as medidas de cobertura e acerto, sendo estas entendidas como:

- **Cobertura:** indica o número de exemplos cobertos pelas condições das regras e é indicada pelo primeiro número dentro dos parênteses no conseqüente da regra. Uma alta cobertura com alta taxa de acerto, pode indicar uma regra de senso comum.
- **Taxa de acerto:** apresenta a porcentagem de casos classificados corretamente em relação à cobertura, indicando a credibilidade da regra e foi calculada utilizando a seguinte expressão:  $(\text{cobertura} - \text{erro}) / (\text{cobertura}) * 100$ .

Para ilustrar a cobertura e o erro utilizados para calcular a taxa de acerto para cada regra, um exemplo é mostrado a seguir.

$$\text{ENTÃO MICROÁREA} = \text{médio risco } (\underline{42.0} / 6.0)$$

  
Cobertura

  
Erro

Para avaliação, sob a perspectiva da qualidade, foram consideradas as medidas de compreensibilidade e o quão surpreendente é a regra para o especialista. Para avaliar a compreensibilidade foi considerado o tamanho da regra, ou seja, o número de condições por regra, considerando o conceito de regra prática de Michalski. Cabe lembrar que para Michalski uma regra prática é composta por menos de cinco condições em uma conjunção.

Para avaliar o quão surpreendente foi a regra para o especialista, foi utilizado um instrumento (Apêndice C) no qual três especialistas da área de saúde coletiva avaliaram as regras indicando para cada uma, um entre três valores possíveis, sendo eles: <1> **Irrelevante**, quando avaliava a regra como incompatível com a realidade, <2> **Confirma seu conhecimento**, quando a regra apresentava padrões que confirmam seu conhecimento e <3> **Interessante**, para as regras que apresentavam padrões condizentes com a realidade, mas até então desconhecidos por ele.

## **4 RESULTADOS E DISCUSSÃO**

Considerando os objetivos definidos para este estudo, neste capítulo são apresentados e discutidos os resultados e para melhor compreensão este capítulo está dividido em três seções principais. A primeira é referente á etapa de identificação dos atributos, a segunda é dedicada aos resultados da etapa de construção da base de dados e, para finalizar, a terceira seção se refere aos resultados do processo de KDD.

### **4.1 RESULTADOS DA ETAPA DE IDENTIFICAÇÃO DE ATRIBUTOS**

Os 38 atributos relacionados no Quadro 3, considerados pelos especialistas como essenciais, podem colaborar para diversos fins, entre eles, na construção de um formulário para a orientação da coleta de dados no momento de realizar a estimativa rápida participativa em um território. Cabe lembrar que estes atributos são referentes apenas às condições do ambiente físico da microárea e envolvem as condições de infra-estrutura, hidrografia, solo, área de lazer, características da comunidade e vetores. Alguns destes atributos podem ser alterados por políticas públicas, outros, com a participação da comunidade. Assim, um sistema de informações que monitore as condições que estes atributos representam, pode colaborar para o planejamento de ações no nível local, assim como, pode servir para os gestores municipais realizarem planejamentos e distribuição de ações para a melhoria da qualidade de vida da população.

Equipamentos sociais		Infra-estrutura	
1	Escola pública	21	Distribuição de energia elétrica
2	Creche pública	22	Iluminação pública
3	Projetos públicos disponib. para crianças	23	Abastecimento de água
4	Igreja (qualquer religião)	24	Rede de esgoto
5	Policimento	25	Coleta de lixo
6	Unidade de saúde	26	Malha viária
Comércio		27	Barreiras de acesso da comunidade
7	Farmácia	28	Transportes
8	Mercado	Comunidade	
Áreas de lazer		29	Distribuição das moradias
9	Parques ou praças	30	Tipos de moradias
10	Cancha para esporte	31	Área da construção das casas
Hidrografia		32	Microárea considerada violenta
11	Rios	33	Existência de Indústrias poluentes
12	Cavas	34	Plantações
13	Córregos	35	Vazios Urbanos
14	Açudes	Vetores	
Solo		36	Gatos
15	Terreno inclinado	37	Cachorros
16	Terreno baixo	38	Pombos
17	Terreno úmido constantemente	39	Ratos
18	Terreno seco	40	Cavalos
19	Terreno plano	41	Vacas
20	Terreno área de aterro		

**Quadro 3 – Conjunto de atributos para classificar uma microárea em relação ao ambiente físico**

O Quadro 4 mostra os atributos considerados, pelos especialistas, como irrelevantes para a caracterização de microáreas sendo retirados do estudo e não constituindo a base de dados utilizada para o processo de KDD. Cabe lembrar que seguindo a metodologia proposta, estes atributos foram retirados do estudo por não atingirem valores igual ou acima de 0,7, logo, considerados pelos especialistas como irrelevantes.

	<b>Atributos</b>
1	Supermercado
2	Mercearia
3	Bar
4	Praças
5	Terreno ondulado
6	Clima
7	Terreno úmido ocasionalmente
8	Hospital

**Quadro 4 - Atributos considerados irrelevantes para caracterizar as microáreas**

No instrumento para avaliação dos atributos (Apêndice A), havia um campo para sugestões de outros atributos que fossem considerados relevantes para inclusão neste estudo, no entanto, não houve sugestões por parte dos especialistas.

#### 4.2 RESULTADOS DA CONSTRUÇÃO DA BASE DE DADOS

Como descrito na seção 3.4.2, após a identificação dos atributos essenciais e a elaboração de um instrumento de pesquisa, foram realizadas entrevistas com as ACS com o objetivo de adquirir dados referentes às microáreas.

Os resultados destas entrevistas foram utilizados para a construção da base de dados utilizada para a aplicação do processo de KDD. Os Gráficos de 1 a 8 ilustram a frequência absoluta dos dados na base, representando 531 microáreas distribuídas na cidade de Curitiba.

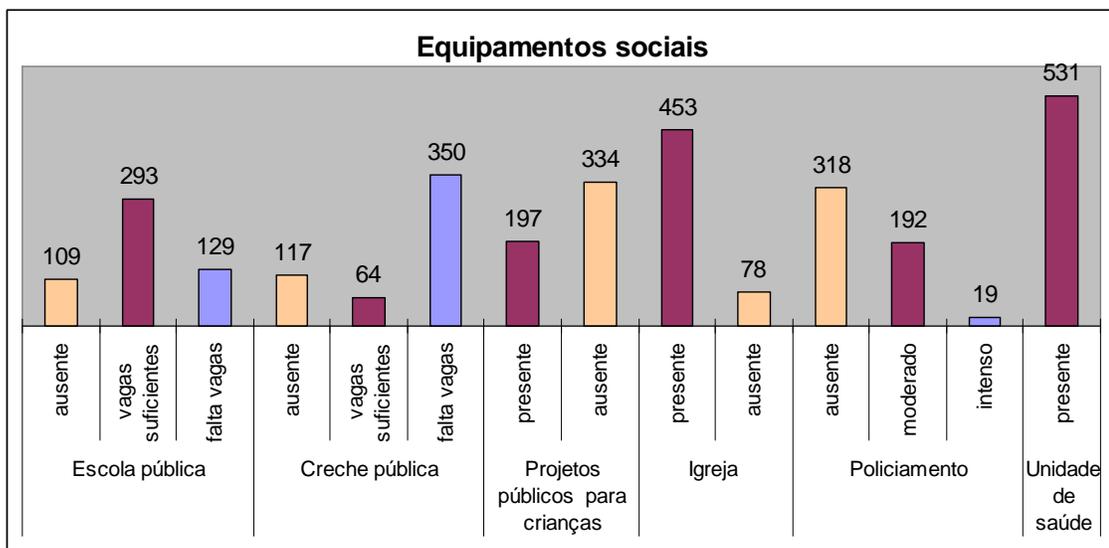


Gráfico 1 – Frequência absoluta dos valores dos atributos referentes aos equipamentos sociais na base de dados

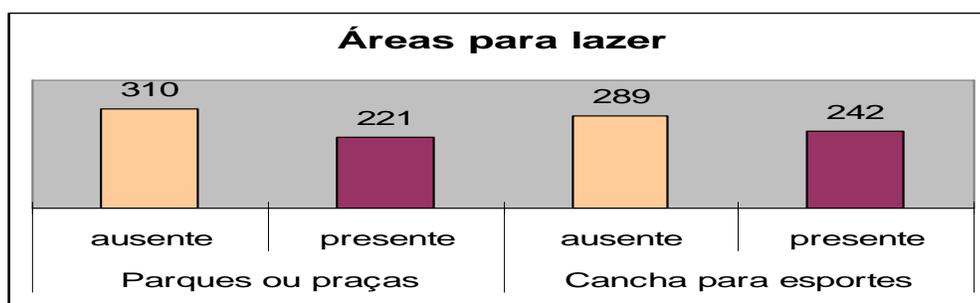


Gráfico 2 – Frequência absoluta dos valores dos atributos referentes às áreas de lazer na base de dados

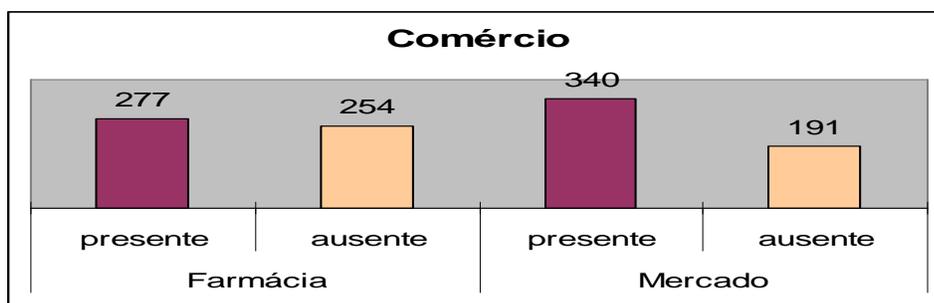


Gráfico 3 – Frequência absoluta dos valores dos atributos referentes ao comércio na base de dados

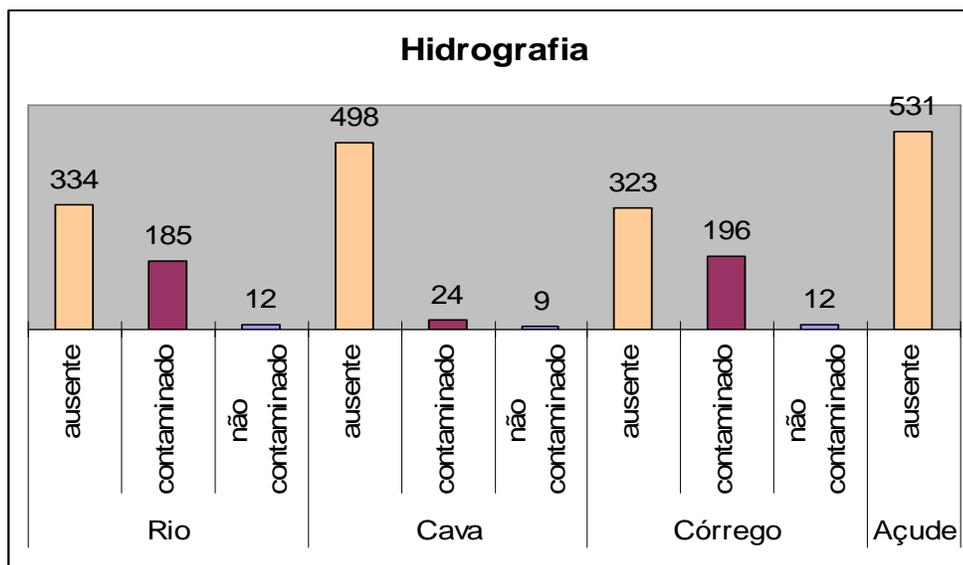


Gráfico 4 – Frequência absoluta dos valores dos atributos referentes à hidrografia na base de dados

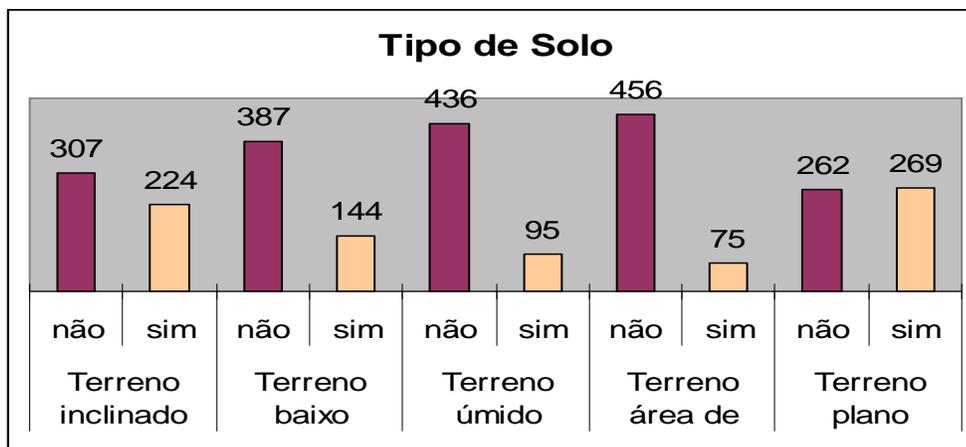


Gráfico 5 – Frequência absoluta dos valores dos atributos referentes ao tipo de solo na base de dados

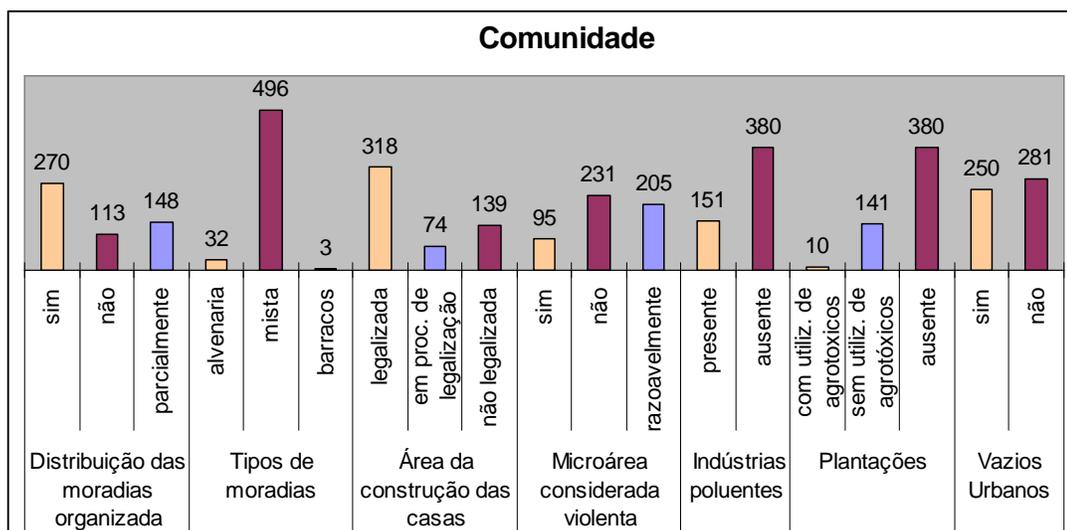


Gráfico 6 – Frequência absoluta dos valores dos atributos referentes às características da comunidade na base de dados

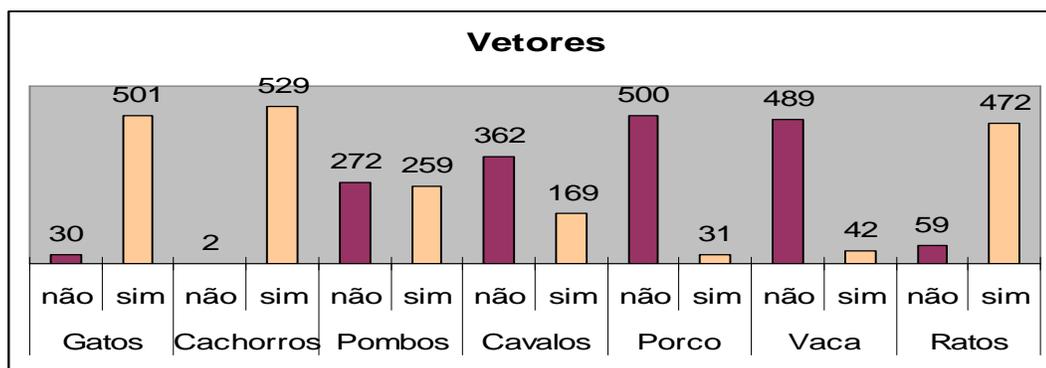


Gráfico 7 – Frequência absoluta dos valores dos atributos referentes aos vetores na base de dados

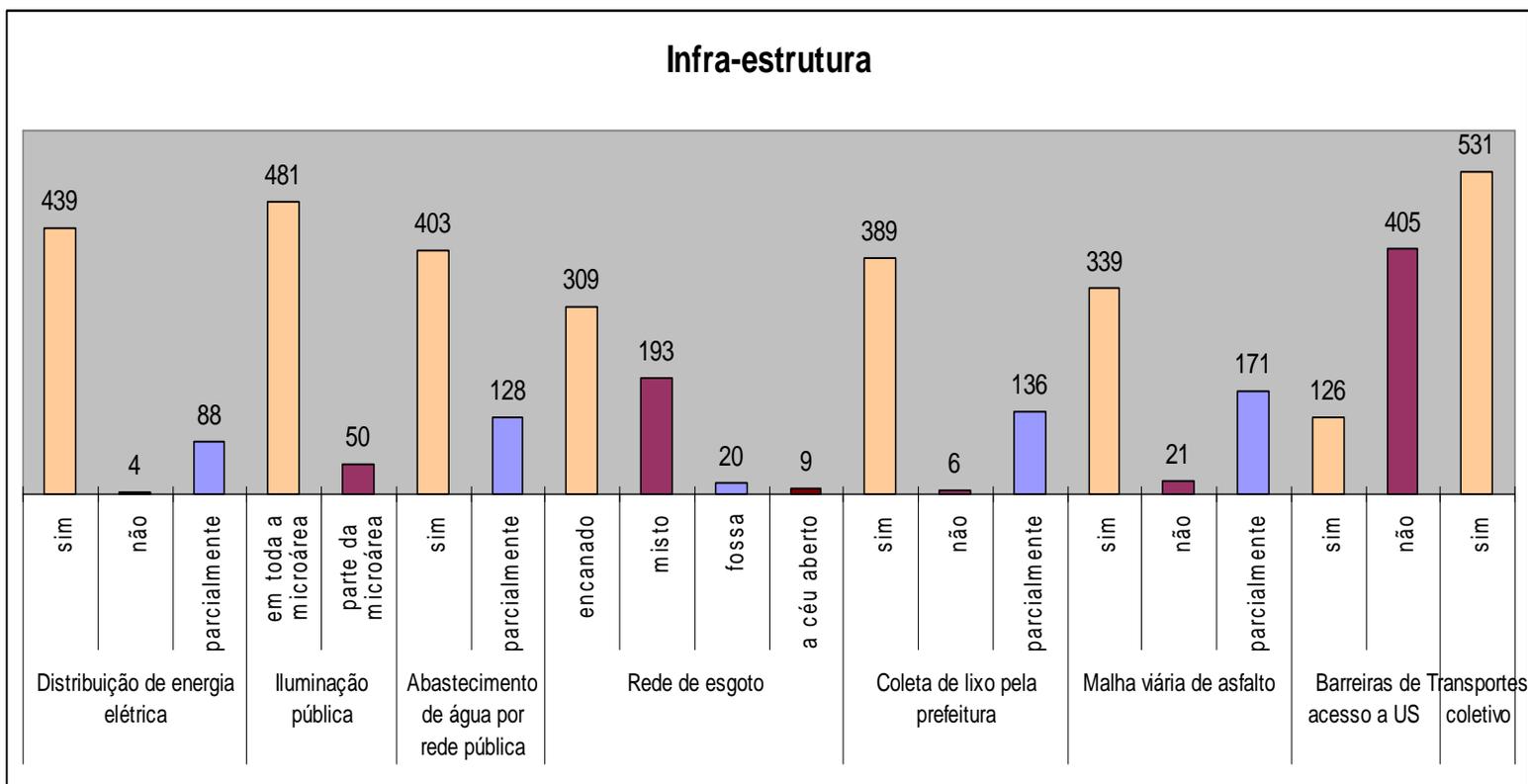


Gráfico 8 - Frequência dos valores dos atributos referentes às características de infra-estrutura na base de dados

Os atributos apresentados nos Gráficos de 1 a 8 subsidiaram as ACS a realizarem uma avaliação quanto ao risco do ambiente físico da microárea em que atuam. Este risco foi apontado, como descrito na seção 3.4.2, em baixo, médio e alto risco e como resultado desta questão, o Gráfico 9, representa a frequência com que o valor do atributo microárea, utilizado como classe alvo, se distribui na base de dados.

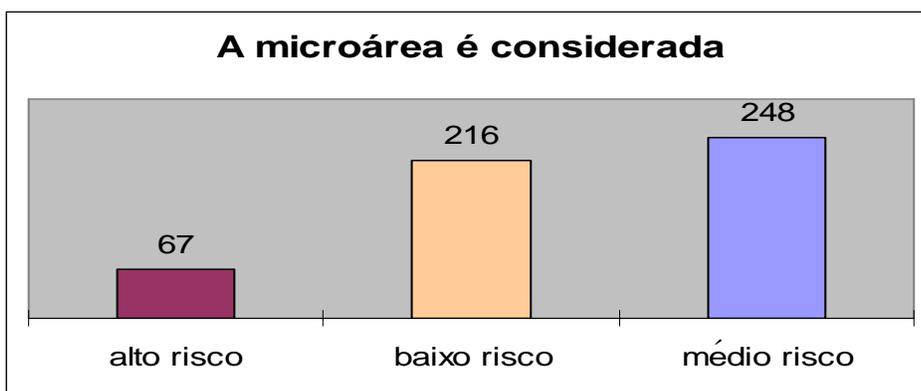


Gráfico 9 – Distribuição da frequência absoluta do valor do atributo utilizado como classe alvo

É possível notar no Gráfico 9, o desbalanceamento da frequência entre as classes, ou seja, a frequência de cada classe não é proporcional, sendo que esta condição pode interferir no resultado da etapa de mineração de dados. Entretanto, esta questão não foi trabalhada neste estudo. Observa-se também, no Gráfico 9, que 12,7% das microáreas foram consideradas pelas ACS como de alto risco, o que indica que nestas áreas existem mais condições precárias, logo, oferece maior risco a população ali residente. É importante ressaltar que ao considerar uma microárea como de alto risco, as ACS demonstraram que o conceito de saúde não está limitado simplesmente na ausência da doença, como era definido antigamente, mas envolve também questões sociais, econômicas e ambientais. Isto mostra um avanço na área da saúde, uma vez que as ACS exercem papel fundamental nas US e são responsáveis tanto por ajudar a classificar microáreas de risco, como a contribuir com a comunidade expondo seus problemas e potencialidades para as autoridades responsáveis pela saúde da comunidade.

### 4.3 RESULTADOS DA ETAPA DE KDD

Os resultados apresentados nesta seção estão divididos de acordo com as etapas do processo de KDD desenvolvidas no âmbito deste trabalho, ou seja, pré-processamento, mineração de dados e pós-processamento.

#### 4.3.1 Resultados da etapa de Pré-processamento

A partir de testes realizados na fase de pré-processamento, foi possível notar que o tratamento de transformação dos valores permitiu aumentar a taxa de acerto do classificador e diminuir o número de regras geradas, conforme mostra a Tabela 1. Os valores referentes a taxa de acerto e erro do classificador, assim como o tempo de processamento, são calculados pela própria ferramenta WEKA e podem ser visualizados no arquivo de saída. Ainda, cabe ressaltar que o computador utilizado para processar as bases (original e pré-processada) foi sempre o mesmo e no momento do processamento apenas a ferramenta WEKA estava sendo executada.

**Tabela 1 – Comparação dos resultados obtidos entre as bases original e pré-processada.**

	Taxa de acerto	Taxa de erro	Tempo de processamento ( em segundos)	Nº. de regras geradas	Regras com cobertura zero
<b>Base original</b>	87,54%	12,45%	0,14	130	57
<b>Base pré-processada</b>	88,70%	11,29%	0,17	79	10

Na tabela 1 ainda é possível notar que, após realizar a transformação dos dados conforme descrito na metodologia, foi obtido um conjunto com menor número de regras com cobertura igual a zero significando uma melhora no desempenho da mineração de dados. Ainda, com a transformação dos valores, as regras ficaram de mais fácil compreensão, isto, devido ao valor do atributo no antecedente da regra ficar mais objetivo. O Quadro 5 apresenta um exemplo comparativo de regras geradas a partir da base original e da base pré-processada.

Base original	Base pré-processada
Regras de difícil compreensão  SE microárea é considerada violenta = não E Área da construção das casas = legalizada E Distribuição das moradias = <b>irregular_e_aglomerada_e_organizada</b>  ENTÃO = médio_risco	Regras mais compreensíveis  SE microárea é considerada violenta = não E Área da construção das casas = legalizada E Distribuição das moradias organizadas = <b>parcialmente</b>  ENTÃO = médio_risco

**Quadro 5 – Comparação das regras obtidas no processamento entre a base original e a base pré-processada**

#### 4.3.2 Resultados da etapa de Mineração de dados

Como resultado da tarefa de classificação na etapa mineração de dados foi obtida uma árvore de decisão para classificar uma microárea em relação ao ambiente físico. Esta árvore foi convertida em 79 regras com o objetivo de melhorar a leitura e a compreensão pelos especialistas avaliadores. As regras foram expressas no seguinte formato:

**Antecedente:** SE<condição>

**Conseqüente:** ENTÃO <conclusão>

Cada uma das 79 regras possui um valor de cobertura e de erro associados. A cobertura corresponde ao número de exemplos cobertos pela regra, sendo assim, foram eliminadas aquelas cujo valor de cobertura foi igual à zero, significando regras não cobertas por exemplos na base de dados, **permanecendo 68 regras para serem avaliadas.**

A Tabela 2 apresenta os resultados estatísticos referentes as medidas de desempenho e qualidade consideradas para avaliar o conjunto de 68 regras. Cabe lembrar que as medidas para avaliar o desempenho foram a cobertura e a taxa de acerto, e para a qualidade a quantidade de condições por regra conforme descrito na metodologia. Na Tabela 2, a primeira coluna identifica as medidas para a avaliação; a segunda, a média (*m*) seguida de um valor identificado pelo símbolo “±”, representando o desvio padrão (DP); a terceira e quarta colunas

apresentam respectivamente os valores mínimo (*Min*) e máximo (*Max*) da amostragem e a quarta coluna o valor do intervalo da amostragem (*Int*).

**Tabela 2 – Medidas para avaliação do desempenho do conjunto de regras obtido na etapa de mineração de dados.**

Medidas para a avaliação do desempenho	Média ( <i>m</i> ) com Desv. Padrão ( $\pm$ )	Mínimo ( <i>Min</i> )	Máximo ( <i>Max</i> )	Intervalo ( <i>Int</i> )
Quantidade de condições por regra	5,74 $\pm$ 2,11	2	12	10
Cobertura	7,84 $\pm$ 15,00	1	111	110
Taxa de acerto	91,55% $\pm$ 14,00	33,33%	100%	66,67%

Para o conjunto das 68 regras, a média ( $m = 5,74 \pm 2,11$ ) da quantidade de condições por regra manteve-se próxima ao conceito de regra prática onde 73,5% das regras mantiveram entre 4 e 7 condições. Em relação à taxa de acerto, a média para o conjunto de regras foi de 91,55% com desvio padrão de  $\pm 14,00$ , demonstrando um resultado satisfatório para o conjunto de regras obtido.

Considerando que o processo de KDD foi aplicado a uma base instanciada com dados obtidos de entrevistas com ACS e que estas expressaram seu conhecimento ao indicar o atributo utilizado como classe alvo, pode-se considerar que quanto maior a cobertura, mais próximo a regra está do senso comum.

#### 4.3.3 Resultados do Pós-processamento

Para estimar o quão surpreendente foi a regra para o especialista, todas as 68 regras foram submetidas à avaliação por três especialistas para indicarem, sob a perspectiva da prática, quais as regras que são consideradas inconsistentes, úteis ou surpreendentes. Cabe lembrar que, para realizar esta avaliação foi utilizado um questionário fechado contendo três valores, onde <1> representava uma regra inconsistente, <2> uma regra que confirmava o conhecimento do especialista e <3> uma regra contendo um novo conhecimento.

Com base no valor atribuído na avaliação realizada pelos especialistas foi calculada a mediana do valor de quão surpreendente é a regra, ressaltando que quanto maior a mediana,

mais surpreendente é a regra. O Quadro 6 apresenta a mediana de quão surpreendente foi cada regra, assim como uma visão geral dos valores de quantidade de condições por regras, cobertura e taxa de acerto para cada regra.

<b>Número da regra</b>	<b>Mediana do quão surpreendente foram as regras para os especialistas</b>	<b>Quantidade de condições por regra</b>	<b>Cobertura</b>	<b>Taxa de acerto %</b>	<b>Consequente</b>
1	2	8	7	71,4	baixo risco
2	2	8	9	88,9	médio risco
3	2	7	15	93,3	baixo risco
4	2	6	3	66,7	baixo risco
5	2	6	111	84,7	baixo risco
6	1	5	10	100	baixo risco
7	2	5	6	66,7	médio risco
8	1	5	2	100	baixo risco
9	2	5	5	80,0	baixo risco
10	2	5	11	81,8	médio risco
11	2	4	5	100	baixo risco
12	2	4	3	100	baixo risco
13	1	4	4	100	médio risco
14	2	6	8	100	baixo risco
15	2	6	3	66,7	médio risco
16	1	6	8	100	médio risco
17	1	6	4	100	baixo risco
18	2	5	4	100	médio risco
19	1	5	6	100	baixo risco
20	1	5	2	100	médio risco
21	1	5	2	100	baixo risco
22	1	5	1	100	baixo risco
23	2	5	44	90,9	médio risco
24	1	5	1	100	baixo risco
25	2	7	21	71,4	médio risco
26	1	12	3	66,7	médio risco
27	1	12	2	100	baixo risco
28	1	11	4	100	médio risco
29	2	10	4	100	baixo risco
30	1	9	7	100	baixo risco
31	1	9	2	100	médio risco
32	1	8	3	100	médio risco
					Continua

					Conclui
33	1	6	5	80,0	médio risco
34	2	4	3	66,7	médio risco
35	1	5	4	100	alto risco
36	2	5	46	87,0	médio risco
37	1	6	3	66,7	alto risco
38	1	7	9	77,8	baixo risco
39	2	6	14	78,6	médio risco
40	2	5	4	100	médio risco
41	1	3	3	33,3	médio risco
42	1	4	4	100	médio risco
43	1	5	2	100	baixo risco
44	2	5	4	100	alto risco
45	2	6	2	100	alto risco
46	1	4	2	100	baixo risco
47	1	2	7	85,7	médio risco
48	1	2	6	100	baixo risco
49	2	6	21	95,2	médio risco
50	2	7	2	100	médio risco
51	2	7	2	100	alto risco
52	1	5	12	91,7	médio risco
53	1	8	2	100	médio risco
54	1	8	2	100	baixo risco
55	1	6	2	100	baixo risco
56	1	5	2	100	médio risco
57	1	5	3	100	alto risco
58	1	3	1	100	baixo risco
59	1	5	5	100	médio risco
60	1	6	2	100	médio risco
61	2	7	3	100	alto risco
62	1	7	3	66,7	médio risco
63	2	6	8	100	alto risco
64	2	4	10	100	alto risco
65	1	4	2	100	médio risco
66	2	3	3	66,7	alto risco
67	2	2	1	100	médio risco
68	2	2	9	100	alto risco

**Quadro 6 – Mediana do quão surpreendente foram as regras para os especialistas e visão geral das médias de avaliação para cada regra**

A partir da mediana do quão surpreendente foi a regra, o conjunto de regras foi dividido em dois subconjuntos denominados de **Conjunto A** e **Conjunto B**. No **Conjunto A** estão as regras consideradas como inconsistentes pelos especialistas e no **Conjunto B** estão as regras que confirmam o conhecimento dos especialistas.

Nota-se, pelo valor da mediana apresentada no Quadro 6, que não foram identificadas regras consideradas como um novo conhecimento. Este fato pode ser justificado pelo conjunto de atributos para a mineração de dados ter sido determinado por especialistas na etapa de identificação de atributos. Assim, pretende-se em trabalhos futuros, ampliar o conjunto de atributos e o número de registros buscando obter novos resultados e regras mais surpreendentes.

#### 4.3.3.1 Avaliação do Conjunto A

O **conjunto A**, contendo as regras consideradas pelos especialistas como inconsistentes, é formado pelas 37 regras apresentadas no Quadro 7.

Identificação	Descrição do antecedente	Conseqüente
<b>REGRA 6</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = legalizada, E Terreno úmido constantemente = não, E Abastecimento de água por rede pública = parcialmente:	baixo_risco (10.0)
<b>REGRA 8</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = legalizada, E Terreno úmido constantemente = sim, E Córrego = ausente:	baixo_risco (2.0)
<b>REGRA 13</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = em_processo_de_legalização, E Terreno inclinado = sim:	médio_risco (4.0)
<b>REGRA 16</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = parcialmente, E barreira de acesso à unidade de saúde = não, E Distribuição de energia elétrica regularizada = sim, E Distribuição das moradias organizadas = parcialmente, E Terreno área de aterro = não:	médio_risco (8.0)
<b>REGRA17</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = parcialmente, E barreira de acesso à unidade de saúde = não, E Distribuição de energia elétrica regularizada = sim, E Distribuição das moradias organizadas = parcialmente, E Terreno área de aterro = sim:	baixo_risco (4.0)
<b>REGRA 19</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = parcialmente, E barreira de acesso à unidade de saúde = não, E Distribuição de energia elétrica regularizada = sim, E Distribuição de energia elétrica regularizada = parcialmente:	baixo_risco (6.0)
<b>REGRA 20</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = parcialmente, E barreira de acesso à unidade de saúde = sim, E Terreno baixo = sim, E Plantações = ausente:	médio_risco (2.0)
Continua		

		Continuação
<b>REGRA 21</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = parcialmente, E barreira de acesso à unidade de saúde = sim, E Terreno baixo = sim, E Plantações = sem_utilização_de_agrotóxicos:	baixo_risco (2.0)
<b>REGRA 22</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = parcialmente, E barreira de acesso à unidade de saúde = sim, E Terreno baixo = sim, E Plantações = com_utilização_de_agrotóxicos:	baixo_risco (1.0)
<b>REGRA 24</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = sim, E Córrego = não contaminado:	baixo_risco (1.0)
<b>REGRA 26</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = sim, E Córrego = ausente, E Igreja (qualquer religião) = presente, E Cancha para esporte = ausente, E Destino do lixo coletado pela prefeitura = sim, E Terreno plano = sim, E Vazios urbanos com presença de lixo = não, E Farmácia = presente, E A escola pública oferece vaga suficiente para crianças da microárea = sim:	médio_risco (3.0/1.0)
<b>REGRA 27</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = sim, E Córrego = ausente, E Igreja (qualquer religião) = presente, E Cancha para esporte = ausente, E Destino do lixo coletado pela prefeitura = sim, E Terreno plano = sim, E Vazios urbanos com presença de lixo = não, E Farmácia = presente, E A escola pública oferece vaga suficiente para as crianças da microárea = não:	baixo_risco (2.0)
<b>REGRA 28</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = sim, E Córrego = ausente, E Igreja (qualquer religião) = presente, E Cancha para esporte = ausente, E Destino do lixo coletado pela prefeitura = sim, E Terreno plano = sim, E Vazios urbanos com presença de lixo = não, E Farmácia = ausente:	médio_risco (4.0)
<b>REGRA 30</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = sim, E Córrego = ausente, E Igreja (qualquer religião) = presente, E Cancha para esporte = ausente E Destino do lixo coletado pela prefeitura = sim, E Terreno plano = não:	baixo_risco (7.0)
<b>REGRA 31</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = sim, E Córrego = ausente, E Igreja (qualquer religião) = presente, E Cancha para esporte = ausente, E Destino do lixo coletado pela prefeitura = sim, E Vaca = sim:	médio_risco (2.0)
<b>REGRA 32</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = sim, E Córrego = ausente, E Igreja (qualquer religião) = presente, E Cancha para esporte = ausente, E Destino do lixo coletado pela prefeitura = parcialmente:	médio_risco (3.0)
<b>REGRA 33</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = sim, E Córrego = ausente, E Igreja (qualquer religião) = ausente:	médio_risco (5.0/1.0)
<b>REGRA 35</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = não, E cavalos = sim:	alto_risco (4.0)
<b>REGRA 37</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = parcialmente, E Plantações = sem_utilização_de_agrotóxicos, E Vazios urbanos com presença de lixo = não	alto_risco (3.0/1.0)
		Continua

		Conclui
<b>REGRA 38</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = parcialmente, E Plantações = sem_utilização_de_agrotóxicos, E Vazios urbanos com presença de lixo = não, E cavalos = sim:	baixo_risco (9.0/2.0)
<b>REGRA 41</b>	SE microárea é considerada violenta = razoavel, E Ratos = sim, E Cava = não_contaminado:	médio_risco (3.0/2.0)
<b>REGRA 42</b>	SE microárea é considerada violenta = razoável,, E Ratos = sim, E Cava = contaminado, E Farmácia = presente:	médio_risco (4.0)
<b>REGRA 43</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = contaminado, E Farmácia = ausente, E Distribuição das moradias organizadas = sim:	baixo_risco (2.0)
<b>REGRA 46</b>	SE microárea é considerada violenta = razoável, E Ratos = não, E Rede de esgoto = encanado, E Indústrias poluentes = presente:	baixo_risco (2.0)
<b>REGRA 47</b>	SE microárea é considerada violenta = razoável, E Rede de esgoto = encanado	médio_risco (7.0/1.0)
<b>REGRA 48</b>	SE microárea é considerada violenta = razoável, E Rede de esgoto = misto	baixo_risco (6.0)
<b>REGRA 52</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = encanado, E Terreno área de aterro = não , E Igreja (qualquer religião) = presente:	médio_risco (12.0/1.0)
<b>REGRA 53</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = encanado, E Terreno área de aterro = não , E Rio = ausente, E Terreno úmido constantemente = não, E Igreja (qualquer religião) = ausente, E Projetos públicos para crianças = presente:	médio_risco (2.0)
<b>REGRA 54</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = encanado, E Terreno área de aterro = não , E Rio = ausente, E Terreno úmido constantemente = não, E Igreja (qualquer religião) = ausente, E Projetos públicos para crianças = ausente:	baixo_risco (2.0)
<b>REGRA 55</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = encanado, E Terreno área de aterro = não , E Rio = ausente, E Terreno úmido constantemente = sim:	baixo_risco (2.0)
<b>REGRA 56</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = encanado, E Terreno área de aterro = sim, E Cancha para esporte = presente:	médio_risco (2.0)
<b>REGRA 57</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = encanado, E Terreno área de aterro = sim, E Cancha para esporte = ausente:	alto_risco (3.0)
<b>REGRA 58</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = fossa:	baixo_risco (1.0)
<b>REGRA 59</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = misto, E Terreno área de aterro = não, E Parque ou praça = presente:	médio_risco (5.0)
<b>REGRA 60</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = misto, E Terreno área de aterro = não, E Parque ou praça = ausente, E Área da construção das casas = legalizada:	médio_risco (2.0)
<b>REGRA 62</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = misto, E Terreno área de aterro = não, E Parque ou praça = ausente, E Área da construção das casas = não_legalizada, E Destino do lixo coletado pela prefeitura = parcialmente:	médio_risco (3.0/1.0)
<b>REGRA 65</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = misto, E porco = sim:	médio_risco (2.0)

**Quadro 7 – Regras consideradas inconsistentes pelos especialistas**

A Tabela 3 apresenta os resultados referentes ao conjunto A, assim como mostra as medidas utilizadas para avaliar o conjunto de regras e segue a mesma descrição da Tabela 2.

**Tabela 3 – Medidas para avaliação do Conjunto A**

Medidas para avaliação	Média ( <i>m</i> ) com Desv. Padrão ( $\pm$ )	Mínimo ( <i>Min</i> )	Máximo ( <i>Max</i> )	Intervalo ( <i>Int</i> )
Quantidade de condições por regra	5,89 $\pm$ 2,40	2	12	10
Cobertura	3,84 $\pm$ 2,65	1	12	11
Taxa de acerto	93,74% $\pm$ 14,43	33,3%	100%	66,7%

Apesar do número máximo da amostragem para a quantidade de condições por regras ser igual a 12, o **Conjunto A** possui 67,7% de regras próximas a média de  $5,89 \pm 2,4$ , fato este que induz a descartar a hipótese que foi por dificuldades de interpretação que os especialistas julgaram as regras deste conjunto como inconsistentes.

Apesar do desvio padrão se encontrar alto para a média da taxa de acerto ( $m = 93,74\%$ ,  $\pm 14,43$ ), 78,37% das regras possui taxa de acerto igual a 100%, ou seja, acima da média do conjunto total e apenas 10,81% se encontram abaixo do limite inferior do desvio padrão.

É possível observar na Tabela 3 que **Conjunto A** possui média de cobertura baixa ( $m = 3,84$ ,  $\pm 2,65$ ,  $Max = 12$ ). Regras com baixa cobertura podem conter um padrão surpreendente, no entanto, os especialistas podem ter considerado estes padrões como inconsistentes, devido aos padrões divergirem dos seus conhecimentos prévios, fato este descrito na literatura.

Contudo, apesar dos especialistas considerarem as regras do **Conjunto A** como “inconsistentes”, as medidas estatísticas para a cobertura e a taxa de acerto mostram um bom desempenho computacional. Assim, este conjunto pode conter algumas regras surpreendentes, entretanto, ao avaliar as regras os especialistas podem estar sendo resistentes a aceitar novos padrões. Isto é descrito por Rezende (2005) como uma “tendência que o especialista possui a compreender melhor modelos que não contradizem seu conhecimento prévio”. Assim, o especialista passa a acreditar menos em regras que contrariam seus conhecimentos.

#### 4.3.3.2 Atributos apontados como a origem das inconsistências nas regras.

Para as regras consideradas inconsistentes, os atributos indicados por dois especialistas como causadores destas inconsistências são apresentados no Quadro 8.

<b>Conseqüente</b>	<b>Atributo</b>	<b>Valor</b>
Baixo risco	Abastecimento de água por rede pública	parcialmente
	Rede de esgoto	misto
	Microárea considerada violenta	sim
	Terreno úmido constantemente	sim
	Rede de esgoto	fossa
	Terreno úmido constantemente	sim
	Destino do lixo coletado pela prefeitura	parcialmente
	Terreno área de aterro	sim
	Distribuição das moradias organizadas	parcialmente
	Microárea considerada violenta	razoável
	Escola oferece vagas suficientes	não
	Ratos	sim
	Cavalos	sim
	Cava	contaminada
	Médio risco	Área da construção das casas
Destino do lixo coletado pela prefeitura		parcialmente
Distribuição das moradias organizadas		parcialmente
Barreira de acesso a unidade de saúde		sim
Microárea considerada violenta		sim
Microárea considerada violenta		razoável
Ratos		sim
Vaca		sim
Igreja		ausente
Rede de esgoto		misto
Alto Risco		Vazios urbanos com presença de lixo
	Rede de esgoto	encanado

**Quadro 8 - Atributos indicados como inconsistentes**

É possível observar no Quadro 8 que os atributos que originaram inconsistências nas regras com o conseqüente baixo risco, são relacionados a condições negativas na comunidade.

Esta evidência possivelmente reflete o modo de avaliar uma microárea pelos especialistas que participaram deste estudo. Ou seja, para considerar uma microárea como baixo risco, esta deve apresentar pouca ou nenhuma condição de precariedade.

## 4.3.3.3 Avaliação do Conjunto B

O **Conjunto B**, contendo as regras que confirmam o conhecimento dos especialistas, é formado pelas 31 regras descritas no Quadro 9.

Identificação	Descrição do antecedente	Consequente
<b>REGRA 1</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = legalizada, E Terreno úmido constantemente = não, E Abastecimento de água por rede pública = sim, E Córrego = contaminado, E Terreno plano = sim, E Projetos públicos para crianças = presente:	baixo_risco (7.0/2.0)
<b>REGRA 2</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = legalizada, E Terreno úmido constantemente = não, E Abastecimento de água por rede pública = sim, E Córrego = contaminado, E Terreno plano = sim, E Projetos públicos para crianças = ausente:	médio_risco (9.0/1.0)
<b>REGRA 3</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = legalizada, E Terreno úmido constantemente = não, E Abastecimento de água por rede pública = sim, E Córrego = contaminado, E Terreno plano = não:	baixo_risco (15.0/1.0)
<b>REGRA 4</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = legalizada, E Terreno úmido constantemente = não, E Abastecimento de água por rede pública = sim, E Córrego = não contaminado:	baixo_risco (3.0/1.0)
<b>REGRA 5</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = legalizada, E Terreno úmido constantemente = não, E Abastecimento de água por rede pública = sim, E Córrego = ausente:	baixo_risco (111.0/17.0)
<b>REGRA 7</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = legalizada, E Terreno úmido constantemente = sim, E Córrego = contaminado:	médio_risco (6.0/2.0)
<b>REGRA 9</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = não legalizada, E Terreno área de aterro = não, E Indústrias poluentes = presente:	baixo_risco (5.0/1.0)
<b>REGRA 10</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = não legalizada, E Terreno área de aterro = não, E Indústrias poluentes = ausente:	médio_risco (11.0/2.0)
<b>REGRA 11</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = não legalizada, E Terreno área de aterro = sim:	baixo_risco (5.0)
<b>REGRA 12</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = sim, E Área da construção das casas = em processo de legalização, E Terreno inclinado = não:	baixo_risco (3.0)
<b>REGRA 14</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = parcialmente, E barreira de acesso à unidade de saúde = não, E Distribuição de energia elétrica regularizada = sim, E Distribuição das moradias organizadas = sim, E Malha viária de asfalto = sim:	baixo_risco (8.0)
		Continua

		Conclui
<b>REGRA 15</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = parcialmente, E barreira de acesso à unidade de saúde = não, E Distribuição de energia elétrica regularizada = sim, E Distribuição das moradias organizadas = sim, E Malha viária de asfalto = parcialmente:	médio_risco (3.0/1.0)
<b>REGRA 18</b>	SE microárea é considerada violenta = não, E Destino do lixo coletado pela prefeitura = parcialmente, E barreira de acesso à unidade de saúde = não, E Distribuição de energia elétrica regularizada = sim, E Distribuição das moradias organizadas = não:	médio_risco (4.0)
<b>REGRA 23</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = sim, E Córrego = contaminado:	médio_risco (44.0/4.0)
<b>REGRA 25</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = sim, E Córrego = ausente, E Igreja (qualquer religião) = presente, E Cancha para esporte = presente:	médio_risco (21.0/6.0)
<b>REGRA 29</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = sim, E Córrego = ausente, E Igreja (qualquer religião) = presente, E Cancha para esporte = ausente, E Destino do lixo coletado pela prefeitura = sim, E Terreno plano = sim, E Vazios urbanos com presença de lixo = sim:	baixo_risco (4.0)
<b>REGRA 34</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = não	médio_risco (3.0/1.0)
<b>REGRA 36</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = parcialmente	médio_risco (42.0/6.0)
<b>REGRA 39</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = parcialmente, E Plantações = sem_utilização_de_agrotóxicos, E Vazios urbanos com presença de lixo = sim:	médio_risco (14.0/3.0)
<b>REGRA 40</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = ausente, E Malha viária de asfalto = parcialmente, E Plantações = com_utilização_de_agrotóxicos:	médio_risco (4.0)
<b>REGRA 44</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = contaminado, E Farmácia = ausente, E Distribuição das moradias organizadas = parcialmente:	alto_risco (4.0)
<b>REGRA 45</b>	SE microárea é considerada violenta = razoável, E Ratos = sim, E Cava = contaminado,, E Farmácia = ausente, E Distribuição das moradias organizadas = não:	alto_risco (2.0)
<b>Regra 49</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = encanado, E Terreno área de aterro = não, E Rio = contaminado, E barreira de acesso à unidade de saúde = não:	médio_risco (21.0/1.0)
<b>REGRA 50</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = encanado, E Terreno área de aterro = não, E Rio = contaminado, E barreira de acesso à unidade de saúde = sim, E Mercado = presente:	médio_risco (2.0)
<b>REGRA 51</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = encanado, E Terreno área de aterro = não, E Rio = contaminado, E barreira de acesso à unidade de saúde = sim, E Mercado = ausente:	alto_risco (2.0)
<b>REGRA 61</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = misto, E Terreno área de aterro = não, E Parque ou praça = ausente, E Área da construção das casas = não_legalizada, E Destino do lixo coletado pela prefeitura = sim:	alto_risco (3.0)
<b>REGRA 63</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = misto, E Terreno área de aterro = não, E Parque ou praça = ausente, E Área da construção das casas = em_processo_de_legalização:	alto_risco (8.0)
<b>REGRA 64</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = misto, E Terreno área de aterro = sim:	alto_risco (10.0)
<b>REGRA 66</b>	SE microárea é considerada violenta = sim, E Cava = ausente, E Rede de esgoto = a_ceu_aberto:	alto_risco (3.0/1.0)
<b>REGRA 67</b>	SE microárea é considerada violenta = sim, E Cava = não_contaminado:	médio_risco (1.0)
<b>REGRA 68</b>	SE microárea é considerada violenta = sim, E Cava = contaminado:	alto_risco (9.0)

**Quadro 9 – Regras que confirmam o conhecimento dos especialistas**

Considerando o **Conjunto B** como mais interessante, sob a perspectiva de quão surpreendente foi a regra para os especialistas, a Tabela 4 apresenta as medidas para avaliação do desempenho computacional deste conjunto.

**Tabela 4 – Medidas para avaliação do conjunto B**

Medidas para avaliação	Média ( <i>m</i> ) com Desv. Padrão ( $\pm$ )	Mínimo ( <i>Min</i> )	Máximo ( <i>Max</i> )	Intervalo ( <i>Int</i> )
Quantidade de condições por regra	5,55 $\pm$ 1,71	2	10	8
Cobertura	12,61 $\pm$ 21,23	1	111	110
Taxa de acerto	88,92% $\pm$ 13,23	33,33%	100%	66,67%

Na Tabela 4 é possível notar que para o **Conjunto B** a quantidade de condições por regras ( $m = 5,55, \pm 1,71$ ) se manteve próxima a cinco, tornando as regras de mais fácil compreensão. Entretanto, a média para cobertura do **Conjunto B** ( $m = 12,61, \pm 21,23, Max III$ ) indica que há maior probabilidade de haver neste conjunto regras consideradas como senso comum.

Apesar de se tratar de um conjunto com regras que confirmam o conhecimento dos especialistas, a taxa de acerto ( $m = 88,92\%, \pm 13,23$ ) não superou a taxa de acerto do **Conjunto A** ( $m = 93,74\%, \pm 14,43$ ) e ainda, apenas 58,06% das regras obtiveram taxa de acerto acima da média geral para o **Conjunto B**.

#### 4.3.3.4 Comparação entre os Conjuntos A e B

A Tabela 5 apresenta uma comparação entre o **Conjunto A** (regras consideradas inconsistentes pelos especialistas) e o **Conjunto B** (regras que confirmam o conhecimento do especialista).

**Tabela 5 - Comparação entre o desempenho dos conjuntos A e B**

Conjunto	Quantidade de condições por regra		Cobertura		Taxa de acerto	
	A	B	A	B	A	B
Média ( <i>m</i> )	5,89	5,55	3,84	12,61	93,74%	88,92%
Desvio padrão	± 2,40	± 1,71	± 2,65	± 21,23	± 14,43	±13,23
Intervalo ( <i>Int</i> )	10	8	11	110	66,67%	33,33
Mínimo ( <i>Min</i> )	2	2	1	1	33,33%	66,67%
Máximo ( <i>Max</i> )	12	10	12	111	100%	100%

Na Tabela 5, nota-se que em média a quantidade de condições por regra no **Conjunto A** é discretamente maior em relação ao **Conjunto B** (**Conj. A**  $m = 5,89 \pm 2,40$  vs **Conj. B**  $m = 5,55 \pm 1,71$ ). Isto pode indicar que o tamanho da regra não interferiu no momento em que os especialistas avaliaram as regras.

Apesar da taxa de acerto ser superior para o **Conjunto A** (**Conj. A**  $m = 93,74\% \pm 13,23$  vs **Conj. B**  $m = 88,92\% \pm 13,23$ ), significando que computacionalmente as regras contidas no **Conjunto A** são melhores em relação ao **Conjunto B**, os especialistas julgaram as regras que formaram o **Conjunto B** como mais surpreendentes.

Observando a média de cobertura do **Conjunto B**, nota-se que esta obteve média muito superior em relação ao **Conjunto A**. Esta relação sugere que as regras contidas no **Conjunto B** tendem a representar o senso comum, o que foi confirmado pela opinião dos especialistas ao indicarem estas regras como a confirmação do seu conhecimento. Portanto as regras do **Conjunto B**, apesar de terem menor desempenho computacional, na avaliação dos especialistas são as que melhor classificam uma microárea em relação ao risco contido no ambiente físico.

#### 4.3.3.5 Resultados referentes aos atributos

Como descrito na metodologia, a base de dados utilizada na mineração de dados foi elaborada com todos os atributos considerados pelos especialistas como essenciais para classificar uma microárea de risco. Entretanto, alguns atributos podem ter um peso maior

nesta classificação. Assim, a partir da mineração de dados, as regras obtidas foram avaliadas buscando identificar os atributos que melhor diferenciam as microáreas em baixo, médio ou alto risco. Para isto foi estabelecido que, os atributos que apareceram nas cinco primeiras condições, são considerados como os mais importantes. Esta medida é justificada, uma vez que, uma condição de uma regra representa um nó, logo, as primeiras condições das regras, representam os nós mais próximos do nó raiz na árvore de decisão, sendo assim os mais importantes. Ainda, regras com número de condições próximo a cinco, estão mais próximas da representação do conhecimento humano.

O Quadro 10 apresenta os atributos que se mantiveram entre as cinco primeiras condições das regras, sendo assim apontados como os mais importantes. No Quadro 10 a primeira coluna se refere à identificação do atributo e a segunda coluna apresenta o número referente a posição da condição na regra.

<b>Atributo</b>	<b>Posição da condição na regra</b>
Microárea é considerada violenta	1
Cava	2
Destino do lixo coletado pela prefeitura	2
Ratos	2
Área da construção das casas	3
Barreira de acesso à unidade de saúde	3
Rede de esgoto	3
Distribuição de energia elétrica regularizada	4
Farmácia	4
Malha viária de asfalto	4
Terreno área de aterro	4
Terreno baixo	4
Terreno úmido constantemente	4
Indústrias poluentes	4
Porco	4
Terreno inclinado	4
Rio	5
Abastecimento de água por rede pública	5
Cancha para esporte	5
Cavalos	5
Córrego	5
Distribuição das moradias organizadas	5
Igreja (qualquer religião)	5
Parque ou praça	5
Plantações	5

**Quadro 10 – Relação entre o atributo e sua posição na regra**

Combinações entre os 25 atributos apresentados no Quadro 10, formaram 66,17% do total de regras geradas pela mineração de dados. Entretanto um fato que merece ser ressaltado é que, em algumas regras foram os atributos das condições mais distantes do início da regra que permitiram a distinção entre uma classificação de baixo, médio ou alto risco.

Este fato é contrário ao pré-suposto de que os atributos que compõem as condições mais próximas do início da regra são os que melhor classificam as microáreas. Contudo, isto não invalida a importância dos atributos do Quadro 10, mas sim deriva deste fato, a percepção de que foram os atributos mais distantes que refinaram a predição de uma classe para um conjunto de atributos. Os atributos que aparecem em **negrito** nas regras a seguir ilustram este fato.

#### REGRA 1

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = legalizada  
 E Terreno úmido constantemente = não  
 E Abastecimento de água por rede pública = sim  
 E Córrego = contaminado  
 E Terreno plano = sim  
**E Projetos públicos para crianças = presente:**

ENTÃO MICROÁREA = baixo\_risco

#### REGRA 50

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = encanado  
 E Terreno área de aterro = não  
 E Rio = contaminado  
 E barreira de acesso à unidade de saúde = sim  
**E Mercado = presente:**

ENTÃO MICROÁREA = médio\_risco

#### REGRA 2

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = legalizada  
 E Terreno úmido constantemente = não  
 E Abastecimento de água por rede pública = sim  
 E Córrego = contaminado  
 E Terreno plano = sim  
**E Projetos públicos para crianças = ausente:**

ENTÃO MICROÁREA = médio\_risco

#### REGRA 51

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = encanado  
 E Terreno área de aterro = não  
 E Rio = contaminado  
 E barreira de acesso à unidade de saúde = sim  
**E Mercado = ausente:**

ENTÃO MICROÁREA = alto\_risco

Pode-se observar que as regras 1 e 2 são semelhantes, porém, a classe a qual a microárea pertence é definido pela presença de projetos públicos para crianças. Para as regras 51 e 52 também semelhantes entre si, a definição da classe a qual pertence a microárea, é dada definida pela 7ª condição a qual é referente a presença de mercado.

O Quadro 11 apresenta outros atributos encontrados nas regras, que apesar de não serem encontrados nas primeiras cinco condições, auxiliam a prever a classe para um conjunto de atributos.

<b>Atributo</b>	<b>Posição na regra</b>
Mercado	7
Projetos públicos para crianças	8
Terreno plano	9
Vaca	9
Vazios urbanos com presença de lixo	06 e 10
A escola pública oferece vaga suficiente para crianças da microárea	12

**Quadro 11 – Atributos mais distantes do início das regras**

Outro comportamento identificado nos atributos durante o processo de mineração de dados, foi que, apesar de serem considerados como essenciais pelos especialistas na etapa de identificação dos atributos, os seis atributos apresentados no Quadro 12 não apareceram como condições nas regras, logo, podem ser considerados com menor peso na classificação das microáreas.

<b>Atributo</b>	<b>Posição na regra</b>
Policiamento	N.E.*
Gatos	N.E.*
Cachorros	N.E.*
Pombos	N.E.*
A creche pública oferece vagas suficiente. para as crianças das microáreas	N.E.*
Iluminação pública	N.E.*
Tipo de moradia	N.E.*

N.E.\* = Não encontrado

**Quadro 12 – Atributos não encontrados nas regras**

Os atributos apresentados no Quadro 12 apesar de serem considerados essenciais por especialistas, ao analisar o conjunto de regras, estes atributos não se apresentaram significativos, pois, não apareceram como condições nas regras. Portanto, considerando os resultados da mineração de dados, os atributos contidos nos quadros 10 e 11, formam um conjunto de atributos que melhor podem ser utilizados para caracterizar a microárea em baixo, médio ou alto risco.

## 5 CONCLUSÕES

Considerando que a classificação das microáreas de risco é uma importante ferramenta gerencial e assistencial, relacionada a distribuição dos recursos e cuidados para a população de um determinado território, realizá-la da melhor forma possível permite melhorar a qualidade nas ações desenvolvidas assim como o gerenciamento de recursos.

No entanto, por envolver o conhecimento e experiência de diferentes profissionais de saúde, a avaliação das microáreas é um processo subjetivo. Assim, para otimizar o processo de coleta e auxiliar na interpretação dos dados, este estudo oferece como principais contribuições um conjunto de atributos essenciais para a caracterização das microáreas quanto ao risco (Quadro 3) relacionado ao ambiente físico e um conjunto de regras (Quadro 9) que combinam estes atributos com o objetivo de classificar as microáreas quanto ao grau do risco (baixo, médio ou alto).

O conjunto de regras obtido pelo processo de KDD é capaz de prever a qual classe de risco uma determinada microárea pode pertencer, quando consideradas as características do ambiente físico. Entretanto, cabe ressaltar que o ambiente físico não se constitui no único fator para classificar uma microárea quanto ao risco. Pois, para complementar e tornar a classificação de risco mais efetiva, podem ser acrescentadas outras informações epidemiológicas, organizacionais e administrativas da comunidade.

Este estudo também oferece como contribuição uma base de dados referentes as características de 46,2% das microáreas de Curitiba. Esta base agrega dados referentes as condições físicas das microáreas fornecendo um perfil da situação atual da cidade em relação as condições de saneamento, infra-estrutura, equipamentos sociais e acesso.

Sobre o processo de KDD, este não identificou regras consideradas como surpreendentes, no entanto indicou regras que confirmaram o conhecimento de especialistas. A não identificação de regras surpreendentes ocorreu possivelmente devido ao conjunto de atributos utilizados no processo de KDD ter sido limitado aos considerados como essenciais pelos especialistas na primeira etapa.

Entretanto, se comparado o quão surpreendente foi cada regra para os especialistas, com os resultados das medidas de desempenho (cobertura e taxa de acerto), foi possível notar uma incoerência. Pois, considera-se como pressuposto, que as regras com melhores medidas

de desempenho são as regras mais interessantes para os especialistas, porém, neste estudo, o conjunto de regras consideradas como mais interessantes pelos especialistas obteve menor desempenho em relação ao conjunto de regras consideradas como inconsistentes.

Mesmo com esta contradição, foi possível atingir o objetivo proposto inicialmente, ou seja, extrair regras capazes de contribuir para a classificação de microáreas de risco. Estas regras proporcionam uma classificação melhor quanto ao risco, lembrando que atualmente as microáreas são consideradas como risco e não risco. Com a classificação em baixo, médio ou alto risco, será possível melhorar a priorização de recursos assim como as medidas de prevenção de doenças. Além disso, a utilização das regras propostas neste estudo permitem que a classificação de uma microárea possa ser realizada de forma mais rápida, menos subjetiva e mantendo um padrão para a classificação de microáreas entre as equipes de saúde.

## 5.1 TRABALHOS FUTUROS

Dentre as perspectivas de trabalhos futuros pode-se citar:

- elaborar um sistema de informação referente às condições do ambiente físico das microáreas, levando em consideração os atributos identificados neste estudo. Um sistema de informação contendo estes atributos poderia auxiliar no reconhecimento da situação e no planejamento local por auxiliar a desmascarar as desigualdades e as condições de cada região;
- utilizar as regras descobertas como base de conhecimento para um sistema especialista para classificar as microáreas em baixo, médio ou alto risco. Um sistema especialista com estas características pode auxiliar tanto profissionais quanto estudantes da área de saúde coletiva.
- identificar atributos referentes às áreas rurais e litorâneas, complementando assim o conjunto de atributos e ampliando a capacidade de utilização do conjunto de atributos.

## REFERÊNCIAS

BOLOGNANI, Silvia M. Z. **Grau de interessabilidade das regras:** estudo de caso de uma empresa no segmento de seguro. 2005. 96 f. Dissertação (Mestrado) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005.

BRASIL. Ministério da Saúde. Conselho Nacional de Saúde. **Resolução 196 de 10 de outubro de 1996.** Diretrizes e normas regulamentadoras da pesquisa envolvendo seres humanos. Brasília, 1996. Disponível em: <<http://conselho.saude.gov.br/comissao/conep/resolucao.htm>>. Acesso em: 10 ago. 2006.

BRASIL. Ministério da Saúde. **Programa Saúde da Família:** a implantação da unidade de saúde da família. Brasília, DF: Ministério da Saúde, 2000. Disponível em: <[http://dtr2004.saude.gov.br/dab/docs/publicacoes/cadernos\\_ab/abcad1.pdf](http://dtr2004.saude.gov.br/dab/docs/publicacoes/cadernos_ab/abcad1.pdf)>. Acesso em: 15 maio 2006. (Cadernos de atenção básica, 1).

BRASIL. Ministério da Saúde. **Gestão municipal de saúde:** textos básicos. Rio de Janeiro. Ministério da Saúde, 2001.

BRASIL. Ministério da Saúde. **Guia prático do programa da saúde da família.** Brasília, DF: Ministério da Saúde, 2002. v. 1. Disponível em: <[http://bvsmms.saude.gov.br/bvs/publicacoes/partes/guia\\_psf.pdf](http://bvsmms.saude.gov.br/bvs/publicacoes/partes/guia_psf.pdf)>. Acesso em: 13 jun. 2006.

BRASIL. Ministério da Saúde. **Manual do Programa de Saúde da Família - PSF.** Brasília, DF: Ministério da Saúde, 2006. Disponível em: <[www.saude.org.br](http://www.saude.org.br)>. Acesso em 15 maio 2006.

CARVALHO, Deborah R. **Data mining através de indução de regras e algoritmos genéticos.** 1999. 130 f. Dissertação (Mestrado) - Pontifícia Universidade Católica do Paraná, Curitiba, 1999.

CASTRO, Hermano A.; GOUVEIA, Nelson; ESCAMILA-CEJUDO, José A. Questões metodológicas para a investigação dos efeitos da poluição do ar na saúde. **Rev. bras. epidemiol.**, São Paulo, v. 6, n.2, 2003. Disponível em: <[http://www.scielo.org/scielo.php?script=sci\\_arttext&pid=S1415790X2003000200007&lng=en&nrm=iso](http://www.scielo.org/scielo.php?script=sci_arttext&pid=S1415790X2003000200007&lng=en&nrm=iso)>. Acesso em: 20 nov. 2006.

CHIESA, Anna Maria; KON, Rubens. Compreensão do território: instrumento de gestão em atenção primária à saúde. In: SANTOS, A.S. MIRANDA, S.M.R.C. (Org). **A enfermagem na gestão em atenção primária à saúde.** Barueri: Manole, 2006. p.309-22. (Série Enfermagem)

COLLAZOS, Kathya S. L.; BARRETO, Jorge M.; ROISENBERG, Mauro. Dificuldades na Aplicação de KDD em Medicina. In: II WORKSHOP DE INFORMÁTICA APLICADA À SAÚDE, 2. 2002, Itajaí. **Anais...** Itajaí: Univali, 2002.

FAYYAD, Usama; PIATESKY-SHAPIRO, Gregory; SMYTH, Padhraic. **From data mining to knowledge discovery in databases**. AI Magazine, 1996. p. 37-54.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Mineração de Dados: um guia prático**. Rio de Janeiro: Elsevier, 2005. p 255.

HAN, Jiawei; KAMBER, Micheline. **Data mining: concepts and techniques**. San Francisco: Morgan Kaufmann Publishers, 2001. 550 p.

KIRKBY, Richard; FRANK, Eibe; REUTEMANN, Peter. **WEKA Explorer User Guide for Version 3-5-6**. 2007. Hamilton, University of Waikato. Disponível em: <<http://64.233.169.104/search?q=cache:HBLz3jB9mLwJ:prdownloads.sourceforge.net/weka/ExplorerGuide-3.5.6.pdf%3Fdownload+WEKA+Explorer+User+Guide+for+Version+3-5-6&hl=pt-BR&ct=clnk&cd=1&gl=br>>. Acesso em: 01 jul.2007

MARQUES, Issac R.; MARIN, Heimar F. Sistemas de apoio à decisão em enfermagem. **Rev. Paul. Enf.** São Paulo, v.21, n.2, p.156-62, 2002.

MENDES, Eugênio V. Distritos Sanitários: conceitos-chave. In: MENDES, Eugênio V. (Org.). **Distritos sanitários: o processo social de mudança das práticas do Sistema Único de Saúde**. São Paulo: Hucitec 1993. p. 159-220.

MIRANDA, Sônia C. Gerenciamento da unidade básica de saúde: a experiência do enfermeiro. In: SANTOS, A.S. (Org); MIRANDA, S.M.R.C. (Org). **A enfermagem na gestão em atenção primária à saúde**. Barueri: Manole, 2007. p.81-110. (Série Enfermagem)

MONKEN, Maurício; BARCELLOS, Christovam. Vigilância em saúde e território utilizado: possibilidades teórica e metodológicas. **Cad. Saúde Pública**, Rio de Janeiro, v. 21, n. 3, p. 898-906, 2005.

NAVEGA, Sergio. Princípios Essenciais do Data Mining. In: INFOIMAGEM 2002, Cenadem, Novembro 2002. **Anais...** Disponível em: <<http://www.intelliwise.com/reports/i2002.htm>> Acessado em 28 jun. 2007.

PEREIRA, Martha Priscila B.; BARCELLOS, Christovam. O TERRITÓRIO NO PROGRAMA DE SAÚDE DA FAMÍLIA. **Hygeia Revista Brasileira de Geografia Médica e da Saúde**. Uberlândia, v.2, n.2, p.47-55, 2006. Disponível em: [www.hygeia.ig.ufu.br/](http://www.hygeia.ig.ufu.br/). Acesso em 13/01/2007.

PIOVESAN, Márcia F.; PADRÃO, Maria Valéria V.; DUMONT, Maria U.; GONDIN, Gracia M.; FLORES, Ouviaomar; PEDROSA, Jose I.; LIMA, Luiz F.M. Vigilância sanitária: uma proposta de análise dos contextos locais. **Rev. bras. epidemiol.**, São Paulo, v. 8, n. 1, p. 83-95, 2005.

PREFEITURA MUNICIPAL DE CURITIBA. Instituto de Pesquisa e Planejamento Urbano de Curitiba. Setor de Monitorização. **Qualidade de Vida caracterização**. Curitiba: PMC; 2003.

PREFEITURA MUNICIPAL DE CURITIBA. **Histórico da Secretaria Municipal de Saúde**. 2007. Disponível em: <<http://www.curitiba.pr.gov.br/saúde/sms/index.htm>> Acesso em: 14 jun 2007

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO. **Ranking do IDH-M dos municípios do Brasil**. Disponível em:<[http://www.pnud.org.br/atlas/ranking/IDHM%2091%2000%20Ranking%20decrecente%20\(pelos%20dados%20de%202000\).htm](http://www.pnud.org.br/atlas/ranking/IDHM%2091%2000%20Ranking%20decrecente%20(pelos%20dados%20de%202000).htm)> Acesso em 10 dez.2007

REZENDE, Solange O.; PUGLIESI, Jaqueline B.; MELNADA, Edson A.; DE PAULA, Marcos F. Mineração de dados In: REZENDE, S O (Org). **Sistemas inteligentes, fundamentos e aplicações**. Barueri: Manole, 2005. p. 307- 335

SANTOS, Simone M.; NORONHA, Cláudio P. Padrões espaciais de mortalidade e diferenciais sócio-econômicos na cidade do Rio de Janeiro. **Cad. Saúde Pública**, Rio de Janeiro, v. 17, n. 5, p. 1099-1110, 2001.

SAVASSI, Leonardo C.M.; DIAS, Ruth B. **Planejamento de ações na equipe**. Disponível em: <[www.smmfc.org.br/gesf](http://www.smmfc.org.br/gesf)>. Acessado em 20 jul.2007.

SOUZA, Cezarina M. N.; MORAES, Luiz R. S.; BERNARDES, Ricardo S. Doenças relacionadas à precariedade dos sistemas de drenagem de águas pluviais: proposta de classificação ambiental e modelos causais. **Cadernos de Saúde Coletiva**, Rio de Janeiro, v. 13, n. 1, p.157-68, 2005.

TAVARES, Mario R. G.; TAKEDA, Silvia M. P. Organização de serviços de atenção primária à saúde. In DUNCAN, B.B.; SCHMIDT, M.I.; GIUGLIANI, E.R.J.. **Medicina ambulatorial: condutas clínicas em atenção primária baseadas em evidências**. 3º edição. Porto Alegre (RS): Artmed; 2004.

TEIXEIRA, Carmen (Org.); PAIM, Jairnilson S. (Org.); VILASBÔAS, A. L. (Org.). **Promoção e Vigilância da Saúde**. 1. ed. Salvador: ISC-CEPS, 2002. v. 1. 128 p.

TOBAR, Frederico; YALOUR, Margot R. **Como fazer teses em saúde pública**: conselhos e idéias para formular projetos e redigir teses e informes de pesquisas. Rio de Janeiro: Fiocruz, 2001. p.168

TOMASI, Elaine. FACCHINI, Luiz A.; OSÓRIO, Alessander; FASSA, Anaclaudia G. Aplicativo para sistematizar informações no planejamento de ações de saúde pública. **Rev. Saúde Pública**, São Paulo, v. 37, n. 6, p.800-6, 2003. Disponível em: <[http://www.scielo.org/scielo.php?script=sci\\_arttext&pid=S003489102003000600017&lng=en&nrm=iso](http://www.scielo.org/scielo.php?script=sci_arttext&pid=S003489102003000600017&lng=en&nrm=iso)>. Acesso em: 20 set. 2006.

UNGLERT, Carmem V. S. **Territorialização em sistemas de saúde**. In: MENDES, EV. (Org.). **Distritos sanitários**: o processo social de mudança das práticas do Sistema Único de Saúde. São Paulo: Hucitec. 1993b. p. 221-295.

UNIVERSIDADE FEDERAL DO CEARÁ. LIA. **Expert Sinta**, Uma ferramenta visual para criação de sistemas especialistas. Manual do Usuário. 1995. Ceará. p. 47. Disponível em: <<http://64.233.169.104/search?q=cache:gmzLu49ULq0J:www.tce.pi.gov.br/adriana/manualdoSINTRA.doc+LIA.+Expert+Sinta,+Uma+ferramenta+visual+para+cria%C3%A7%C3%A3o+de+sistemas+especialistas.+Manual+do+Usu%C3%A1rio.&hl=pt-BR&ct=clnk&cd=4&gl=br.>> Acesso em: 12 nov.2006

WOOD, Geri L.; HABER, Judith. **Pesquisa em enfermagem**: métodos, avaliação crítica e utilização. Trad. de Ivone Evangelista Cabral. 4 ed. Rio de Janeiro: Guanabara Koogan, 2002. p.330.

**APÊNDICE A - INSTRUMENTO PARA VALIDAÇÃO DE ATRIBUTOS**

## Instrumento para validação de atributos

Participante Nº. \_\_\_\_\_

Formação: \_\_\_\_\_ ano: \_\_\_\_\_

Titulação: \_\_\_\_\_

Tempo de trabalho em saúde coletiva: \_\_\_\_\_

O conceito de **áreas de risco** para o Ministério da Saúde (MS) é considerado como “áreas que possuem fatores de riscos e/ou barreiras geográficas ou culturais (tudo aquilo que dificulta ou impede a chegada das equipes e o contato da comunidade com o serviço de saúde), ou ainda áreas com indicadores de saúde muito ruins” e o conceito de **fatores de risco** é considerado como “algo que favorece o surgimento ou o agravamento de uma doença, qualquer que seja ela: física, mental ou social”, **considerando estes conceitos e baseado em seu conhecimento, indique para os atributos relacionados abaixo, quais você considera essenciais, importantes ou irrelevantes para classificar uma microáreas em relação ao risco à saúde.**

1 - Equipamentos sociais													
		Essencial				importante			irrelevante				
		1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0	
	Escola												
	Creche												
	Projetos disponibilizados pela prefeitura para crianças												
	Igreja (qualquer religião)												
	Policimento												
	Unidades de saúde												
	Hospitais												
2 – Comércio													
		Essencial				importante			irrelevante				
		1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0	
	Farmácia												
	Supermercado												
	Mercearia												
	Bar												
3 - Áreas de Lazer													
		Essencial				importante			irrelevante				
		1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0	
	Parques ou praças												
	Praças												
	Cancha de esporte												
4- Hidrografia													
		Essencial				importante			irrelevante				





**APÊNDICE B – INSTRUMENTO PARA COLETA DE DADOS SOBRE AS  
CONDIÇÕES DAS MICROÁREAS**

## INSTRUMENTO PARA COLETA DE DADOS DAS MICROÁREAS

Nome da Unidade de abrangência: \_\_\_\_\_

Código da microárea: \_\_\_\_\_

Código ou nome da equipe de PSF: \_\_\_\_\_

### Na microárea:

A escola pública oferece vaga suficiente para as crianças da microárea	<input type="checkbox"/> sim	<input type="checkbox"/> não	<input type="checkbox"/> não existe escola
A creche pública oferece vaga suficiente para as crianças da microárea	<input type="checkbox"/> sim	<input type="checkbox"/> não	<input type="checkbox"/> não existe creche
Projetos públicos para crianças	<input type="checkbox"/> presente	<input type="checkbox"/> ausente	
Igreja (qualquer religião)	<input type="checkbox"/> presente	<input type="checkbox"/> ausente	
Policiamento	<input type="checkbox"/> moderado	<input type="checkbox"/> intenso	<input type="checkbox"/> ausente
Unidade de saúde	<input type="checkbox"/> presente	<input type="checkbox"/> ausente	
Farmácia	<input type="checkbox"/> presente	<input type="checkbox"/> ausente	
Mercado	<input type="checkbox"/> presente	<input type="checkbox"/> ausente	
Mercearia	<input type="checkbox"/> presente	<input type="checkbox"/> ausente	
Parque ou praça	<input type="checkbox"/> presente	<input type="checkbox"/> ausente	
Cancha para esporte	<input type="checkbox"/> presente	<input type="checkbox"/> ausente	
Rio	<input type="checkbox"/> presente <input type="checkbox"/> não existe	<input type="checkbox"/> contaminados <input type="checkbox"/> não contaminados	
Cava	<input type="checkbox"/> presente <input type="checkbox"/> não existe	<input type="checkbox"/> contaminadas <input type="checkbox"/> não contaminados	
Córrego	<input type="checkbox"/> presente <input type="checkbox"/> não existe	<input type="checkbox"/> contaminados <input type="checkbox"/> não contaminados	
Açude	<input type="checkbox"/> presente	<input type="checkbox"/> contaminados <input type="checkbox"/> não contaminados	

	<input type="checkbox"/>	não existe				
Terreno	<input type="checkbox"/>	inclinado	<input type="checkbox"/>	úmido constantemente		
	<input type="checkbox"/>	ondulado	<input type="checkbox"/>	seco		
	<input type="checkbox"/>	baixo	<input type="checkbox"/>	plano		
	<input type="checkbox"/>	área de aterro				
Distribuição de energia elétrica	<input type="checkbox"/>	regularizada	<input type="checkbox"/>	clandestina (gato)	<input type="checkbox"/>	não possui
Iluminação pública	<input type="checkbox"/>	em toda a microárea				
	<input type="checkbox"/>	apenas parte da microárea				
	<input type="checkbox"/>	não existe				
Abastecimento de água	<input type="checkbox"/>	rede pública	<input type="checkbox"/>	clandestina (gato)		
	<input type="checkbox"/>	poço comum	<input type="checkbox"/>	bica ou mina		
	<input type="checkbox"/>	poço artesiano	<input type="checkbox"/>	cisternas		
Rede de esgoto	<input type="checkbox"/>	encanado	<input type="checkbox"/>	a céu aberto	<input type="checkbox"/>	fossa
Destino do lixo	<input type="checkbox"/>	em terrenos vazios	<input type="checkbox"/>	realizada pela prefeitura	<input type="checkbox"/>	caçamba
Malha viária	<input type="checkbox"/>	asfalto	<input type="checkbox"/>	terra		
Indústrias poluentes	<input type="checkbox"/>	solo	<input type="checkbox"/>	ar	<input type="checkbox"/>	não existe
	<input type="checkbox"/>	água	<input type="checkbox"/>	sonora		
Plantações	<input type="checkbox"/>	com utilização de agrotóxicos	<input type="checkbox"/>	sem utilização de agrotóxicos	<input type="checkbox"/>	não existe
Vazios Urbanos	<input type="checkbox"/>	limpo	<input type="checkbox"/>	com presença de lixo	<input type="checkbox"/>	não existe
Existe barreira para chegar até a unidade de saúde como:	<input type="checkbox"/>	Rodovia	<input type="checkbox"/>	baía ou mar	<input type="checkbox"/>	não existe
	<input type="checkbox"/>	Rio	<input type="checkbox"/>	unidade de saúde distante		
Transporte coletivo	<input type="checkbox"/>	ônibus	<input type="checkbox"/>	lotação	<input type="checkbox"/>	barco
Área da comunidade	<input type="checkbox"/>	urbana	<input type="checkbox"/>	rural	<input type="checkbox"/>	litorânea
Distribuição das moradias	<input type="checkbox"/>	organizada	<input type="checkbox"/>	aglomerado	<input type="checkbox"/>	irregular
Moradias predominantes	<input type="checkbox"/>	alvenaria ou madeira			<input type="checkbox"/>	barracos

Área da construção das casas	<input type="checkbox"/> legalizada	<input type="checkbox"/> não legalizada	<input type="checkbox"/> em processo de legalização
Gatos	<input type="checkbox"/> sim	<input type="checkbox"/> não	
Cachorros	<input type="checkbox"/> sim	<input type="checkbox"/> não	
Pombos	<input type="checkbox"/> sim	<input type="checkbox"/> não	
Cavalos	<input type="checkbox"/> sim	<input type="checkbox"/> não	
Porco	<input type="checkbox"/> sim	<input type="checkbox"/> não	
Vaca	<input type="checkbox"/> sim	<input type="checkbox"/> não	
Ratos	<input type="checkbox"/> sim	<input type="checkbox"/> não	
A microárea é considerada violenta?	<input type="checkbox"/> sim	<input type="checkbox"/> não	<input type="checkbox"/> razoável

**Você considera esta microárea**  **de baixo risco**  **de médio risco**  **de alto risco**

**APÊNDICE C - INSTRUMENTO PARA OBTER O QUÃO SURPREENDENTE  
FORAM AS REGRAS PARA O ESPECIALISTA**

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2  
 Relation: Base\_de\_dados\_maio 2008.txt-weka.filters.unsupervised.attribute.  
 Instances: 531  
 Attributes: 38

A escola pública oferece vaga suficiente para as crianças da microárea  
 A creche pública oferece vaga suficiente para as crianças da microárea  
 Projetos públicos para crianças  
 Igreja (qualquer religião)  
 Policiamento  
 Farmácia  
 Mercado  
 Parque ou praça  
 Cancha para esporte  
 Rio  
 Cava  
 Córrego  
 Terreno inclinado  
 Terreno baixo  
 Terreno área de aterro  
 Terreno úmido constantemente  
 Terreno plano  
 Distribuição de energia elétrica regularizada  
 Iluminação pública  
 Abastecimento de água por rede pública  
 Rede de esgoto  
 Destino do lixo coletado pela prefeitura  
 Malha viária de asfalto  
 Indústrias poluentes  
 Plantações  
 Vazios urbanos com presença de lixo  
 Barreira de acesso à unidade de saúde  
 Distribuição das moradias organizadas  
 Área da construção das casas  
 Gatos  
 Cachorros  
 Pombos  
 cavalos  
 porco  
 Vaca  
 Ratos  
 Microárea é considerada violenta  
 Você considera esta microárea

Test mode: evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree

**REGRA 1**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = legalizada  
 E Terreno úmido constantemente = não  
 E Abastecimento de água por rede pública = sim  
 E Córrego = contaminado  
 E Terreno plano = sim  
 E Projetos públicos para crianças = presente:  
**ENTÃO MICROÁREA = baixo\_risco (7.0/2.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 2**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = legalizada  
 E Terreno úmido constantemente = não  
 E Abastecimento de água por rede pública = sim  
 E Córrego = contaminado  
 E Terreno plano = sim  
 E Projetos públicos para crianças = ausente:  
**ENTÃO MICROÁREA = médio\_risco (9.0/1.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 3**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = legalizada  
 E Terreno úmido constantemente = não  
 E Abastecimento de água por rede pública = sim  
 E Córrego = contaminado  
 E Terreno plano = não:  
**ENTÃO MICROÁREA = baixo\_risco (15.0/1.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 4**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = legalizada  
 E Terreno úmido constantemente = não  
 E Abastecimento de água por rede pública = sim  
 E Córrego = não\_contaminado:  
**ENTÃO MICROÁREA = baixo\_risco (3.0/1.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 5**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = legalizada  
 E Terreno úmido constantemente = não  
 E Abastecimento de água por rede pública = sim  
 E Córrego = ausente:  
**ENTÃO MICROÁREA = baixo\_risco (111.0/17.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 6**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim

E Área da construção das casas = legalizada  
 E Terreno úmido constantemente = não  
 E Abastecimento de água por rede pública = parcialmente:  
**ENTÃO MICROÁREA = baixo\_risco (10.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 7**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = legalizada  
 E Terreno úmido constantemente = sim  
 E Córrego = contaminado:  
**ENTÃO MICROÁREA = médio\_risco (6.0/2.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 8**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = legalizada  
 E Terreno úmido constantemente = sim  
 E Córrego = ausente:  
**ENTÃO MICROÁREA = baixo\_risco (2.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 9**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = não\_legalizada  
 E Terreno área de aterro = não  
 E Indústrias poluentes = presente:  
**ENTÃO MICROÁREA = baixo\_risco (5.0/1.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 10**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = não\_legalizada  
 E Terreno área de aterro = não  
 E Indústrias poluentes = ausente:  
**ENTÃO MICROÁREA = médio\_risco (11.0/2.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 11**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = não\_legalizada  
 E Terreno área de aterro = sim:  
**ENTÃO MICROÁREA** = baixo\_risco (5.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 12**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = em\_processo\_de\_legalização  
 E Terreno inclinado = não:  
**ENTÃO MICROÁREA** = baixo\_risco (3.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 13**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = sim  
 E Área da construção das casas = em\_processo\_de\_legalização  
 E Terreno inclinado = sim:  
**ENTÃO MICROÁREA** = médio\_risco (4.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 14**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = parcialmente  
 E barreira de acesso à unidade de saúde = não  
 E Distribuição de energia elétrica regularizada = sim  
 E Distribuição das moradias organizadas = sim  
 E Malha viária de asfalto = sim:  
**ENTÃO MICROÁREA** = baixo\_risco (8.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 15**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = parcialmente  
 E barreira de acesso à unidade de saúde = não  
 E Distribuição de energia elétrica regularizada = sim  
 E Distribuição das moradias organizadas = sim  
 E Malha viária de asfalto = parcialmente:  
**ENTÃO MICROÁREA** = médio\_risco (3.0/1.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)

3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 16**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = parcialmente  
 E barreira de acesso à unidade de saúde = não  
 E Distribuição de energia elétrica regularizada = sim  
 E Distribuição das moradias organizadas = parcialmente  
 E Terreno área de aterro = não:  
**ENTÃO MICROÁREA** = médio\_risco (8.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 17**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = parcialmente  
 E barreira de acesso à unidade de saúde = não  
 E Distribuição de energia elétrica regularizada = sim  
 E Distribuição das moradias organizadas = parcialmente  
 E Terreno área de aterro = sim:  
**ENTÃO MICROÁREA** = baixo\_risco (4.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 18**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = parcialmente  
 E barreira de acesso à unidade de saúde = não  
 E Distribuição de energia elétrica regularizada = sim  
 E Distribuição das moradias organizadas = não:  
**ENTÃO MICROÁREA** = médio\_risco (4.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 19**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = parcialmente  
 E barreira de acesso à unidade de saúde = não  
 E Distribuição de energia elétrica regularizada = sim  
 E Distribuição de energia elétrica regularizada = parcialmente:  
**ENTÃO MICROÁREA** = baixo\_risco (6.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 20**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = parcialmente  
 E barreira de acesso à unidade de saúde = sim

E Terreno baixo = sim  
 E Plantações = ausente:  
**ENTÃO MICROÁREA = médio\_risco (2.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 21**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = parcialmente  
 E barreira de acesso à unidade de saúde = sim  
 E Terreno baixo = sim  
 E Plantações = sem\_utilização\_de\_agrotóxicos:  
**ENTÃO MICROÁREA = baixo\_risco (2.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 22**

SE microárea é considerada violenta = não  
 E Destino do lixo coletado pela prefeitura = parcialmente  
 E barreira de acesso à unidade de saúde = sim  
 E Terreno baixo = sim  
 E Plantações = com\_utilização\_de\_agrotóxicos:  
**ENTÃO MICROÁREA = baixo\_risco (1.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 23**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = ausente  
 E Malha viária de asfalto = sim  
 E Córrego = contaminado:  
**ENTÃO MICROÁREA = médio\_risco (44.0/4.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 24**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = ausente  
 E Malha viária de asfalto = sim  
 E Córrego = não\_contaminado:  
**ENTÃO MICROÁREA = baixo\_risco (1.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 25**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = ausente  
 E Malha viária de asfalto = sim  
 E Córrego = ausente  
 E Igreja (qualquer religião) = presente  
 E Cancha para esporte = presente:  
**ENTÃO MICROÁREA = médio\_risco (21.0/6.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 26**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = ausente  
 E Malha viária de asfalto = sim  
 E Córrego = ausente  
 E Igreja (qualquer religião) = presente  
 E Cancha para esporte = ausente  
 E Destino do lixo coletado pela prefeitura = sim  
 E Terreno plano = sim  
 E Vazios urbanos com presença de lixo = não  
 E Farmácia = presente  
 E A escola pública oferece vaga suficiente para crianças da microárea = sim:  
**ENTÃO MICROÁREA = médio\_risco (3.0/1.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 27**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = ausente  
 E Malha viária de asfalto = sim  
 E Córrego = ausente  
 E Igreja (qualquer religião) = presente  
 E Cancha para esporte = ausente  
 E Destino do lixo coletado pela prefeitura = sim  
 E Terreno plano = sim  
 E Vazios urbanos com presença de lixo = não  
 E Farmácia = presente  
 E A escola pública oferece vaga suficiente para as crianças da microárea = não:  
**ENTÃO MICROÁREA = baixo\_risco (2.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 28**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = ausente  
 E Malha viária de asfalto = sim  
 E Córrego = ausente  
 E Igreja (qualquer religião) = presente  
 E Cancha para esporte = ausente  
 E Destino do lixo coletado pela prefeitura = sim

E Terreno plano = sim  
 E Vazios urbanos com presença de lixo = não  
 E Farmácia = ausente:  
**ENTÃO MICROÁREA = médio\_risco (4.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

### **REGRA 29**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = ausente  
 E Malha viária de asfalto = sim  
 E Córrego = ausente  
 E Igreja (qualquer religião) = presente  
 E Cancha para esporte = ausente  
 E Destino do lixo coletado pela prefeitura = sim  
 E Terreno plano = sim  
 E Vazios urbanos com presença de lixo = sim:  
**ENTÃO MICROÁREA = baixo\_risco (4.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

### **REGRA 30**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = ausente  
 E Malha viária de asfalto = sim  
 E Córrego = ausente  
 E Igreja (qualquer religião) = presente  
 E Cancha para esporte = ausente  
 E Destino do lixo coletado pela prefeitura = sim  
 E Terreno plano = não:  
**ENTÃO MICROÁREA = baixo\_risco (7.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

### **REGRA 31**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = ausente  
 E Malha viária de asfalto = sim  
 E Córrego = ausente  
 E Igreja (qualquer religião) = presente  
 E Cancha para esporte = ausente  
 E Destino do lixo coletado pela prefeitura = sim  
 E Vaca = sim:  
**ENTÃO MICROÁREA = médio\_risco (2.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 32**

SE microárea é considerada violenta = razoavel

E Ratos = sim

E Cava = ausente

E Malha viária de asfalto = sim

E Córrego = ausente

E Igreja (qualquer religião) = presente

E Cancha para esporte = ausente

E Destino do lixo coletado pela prefeitura = parcialmente:

**ENTÃO MICROÁREA** = médio\_risco (3.0)

Para você esta regra é:

1 - ( ) Irrelevante (incompatível com a realidade)

2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)

3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 33**

SE microárea é considerada violenta = razoavel

E Ratos = sim

E Cava = ausente

E Malha viária de asfalto = sim

E Córrego = ausente

E Igreja (qualquer religião) = ausente:

**ENTÃO MICROÁREA** = médio\_risco (5.0/1.0)

Para você esta regra é:

1 - ( ) Irrelevante (incompatível com a realidade)

2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)

3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 34**

SE microárea é considerada violenta = razoavel

E Ratos = sim

E Cava = ausente

E Malha viária de asfalto = não

**ENTÃO MICROÁREA** = médio\_risco (3.0/1.0)

Para você esta regra é:

1 - ( ) Irrelevante (incompatível com a realidade)

2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)

3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 35**

SE microárea é considerada violenta = razoavel

E Ratos = sim

E Cava = ausente

E Malha viária de asfalto = não

E cavalos = sim: alto\_risco (4.0)

Para você esta regra é:

1 - ( ) Irrelevante (incompatível com a realidade)

2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)

3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

**REGRA 36**

SE microárea é considerada violenta = razoavel

E Ratos = sim

E Cava = ausente

E Malha viária de asfalto = parcialmente

**ENTÃO MICROÁREA** = médio\_risco (42.0/6.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

### **REGRA 37**

SE microárea é considerada violenta = razoavel

E Ratos = sim

E Cava = ausente

E Malha viária de asfalto = parcialmente

E Plantações = sem\_utilização\_de\_agrotóxicos

E Vazios urbanos com presença de lixo = não

**ENTÃO MICROÁREA = alto\_risco (3.0/1.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

### **REGRA 38**

SE microárea é considerada violenta = razoavel

E Ratos = sim

E Cava = ausente

E Malha viária de asfalto = parcialmente

E Plantações = sem\_utilização\_de\_agrotóxicos

E Vazios urbanos com presença de lixo = não

E cavalos = sim:

**ENTÃO MICROÁREA = baixo\_risco (9.0/2.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

### **REGRA 39**

SE microárea é considerada violenta = razoavel

E Ratos = sim

E Cava = ausente

E Malha viária de asfalto = parcialmente

E Plantações = sem\_utilização\_de\_agrotóxicos

E Vazios urbanos com presença de lixo = sim:

**ENTÃO MICROÁREA = médio\_risco (14.0/3.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

### **REGRA 40**

SE microárea é considerada violenta = razoavel

E Ratos = sim

E Cava = ausente

E Malha viária de asfalto = parcialmente

E Plantações = com\_utilização\_de\_agrotóxicos:

**ENTÃO MICROÁREA = médio\_risco (4.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

### **REGRA 41**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = não\_contaminado:  
**ENTÃO MICROÁREA = médio\_risco (3.0/2.0)**

Para você esta regra é:  
 1 - ( ) Irrelevante (incompatível com a realidade)  
 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)  
 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 42**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = contaminado  
 E Farmácia = presente:  
**ENTÃO MICROÁREA = médio\_risco (4.0)**

Para você esta regra é:  
 1 - ( ) Irrelevante (incompatível com a realidade)  
 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)  
 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 43**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = contaminado  
 E Farmácia = ausente  
 E Distribuição das moradias organizadas = sim:  
**ENTÃO MICROÁREA = baixo\_risco (2.0)**

Para você esta regra é:  
 1 - ( ) Irrelevante (incompatível com a realidade)  
 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)  
 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 44**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = contaminado  
 E Farmácia = ausente  
 E Distribuição das moradias organizadas = parcialmente:  
**ENTÃO MICROÁREA = alto\_risco (4.0)**

Para você esta regra é:  
 1 - ( ) Irrelevante (incompatível com a realidade)  
 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)  
 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 45**

SE microárea é considerada violenta = razoavel  
 E Ratos = sim  
 E Cava = contaminado  
 E Farmácia = ausente  
 E Distribuição das moradias organizadas = não:  
**ENTÃO MICROÁREA = alto\_risco (2.0)**

Para você esta regra é:  
 1 - ( ) Irrelevante (incompatível com a realidade)  
 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)  
 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 46**

SE microárea é considerada violenta = razoavel

E Ratos = não  
 E Rede de esgoto = encanado  
 E Indústrias poluentes = presente:  
**ENTÃO MICROÁREA** = baixo\_risco (2.0)

Para você esta regra é:  
 1 - ( ) Irrelevante (incompatível com a realidade)  
 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)  
 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 47**

SE microárea é considerada violenta = razoavel  
 E Rede de esgoto = encanado  
**ENTÃO MICROÁREA** = médio\_risco (7.0/1.0)

Para você esta regra é:  
 1 - ( ) Irrelevante (incompatível com a realidade)  
 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)  
 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 48**

SE microárea é considerada violenta = razoavel  
 E Rede de esgoto = misto: ( encanado e fossa ou a céu aberto)  
**ENTÃO MICROÁREA** = baixo\_risco (6.0)

Para você esta regra é:  
 1 - ( ) Irrelevante (incompatível com a realidade)  
 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)  
 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **Regra 49**

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = encanado  
 E Terreno área de aterro = não  
 E Rio = contaminado  
 E barreira de acesso à unidade de saúde = não:  
**ENTÃO MICROÁREA** = médio\_risco (21.0/1.0)

Para você esta regra é:  
 1 - ( ) Irrelevante (incompatível com a realidade)  
 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)  
 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 50**

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = encanado  
 E Terreno área de aterro = não  
 E Rio = contaminado  
 E barreira de acesso à unidade de saúde = sim  
 E Mercado = presente:  
**ENTÃO MICROÁREA** = médio\_risco (2.0)

Para você esta regra é:  
 1 - ( ) Irrelevante (incompatível com a realidade)  
 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)  
 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 51**

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = encanado

E Terreno área de aterro = não  
 E Rio = contaminado  
 E barreira de acesso à unidade de saúde = sim  
 E Mercado = ausente:  
**ENTÃO MICROÁREA** = alto\_risco (2.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 52**

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = encanado  
 E Terreno área de aterro = não  
 E Igreja (qualquer religião) = presente:  
**ENTÃO MICROÁREA** = médio\_risco (12.0/1.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 53**

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = encanado  
 E Terreno área de aterro = não  
 E Rio = ausente  
 E Terreno úmido constantemente = não  
 E Igreja (qualquer religião) = ausente  
 E Projetos públicos para crianças = presente:  
**ENTÃO MICROÁREA** = médio\_risco (2.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 54**

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = encanado  
 E Terreno área de aterro = não  
 E Rio = ausente  
 E Terreno úmido constantemente = não  
 E Igreja (qualquer religião) = ausente  
 E Projetos públicos para crianças = ausente:  
**ENTÃO MICROÁREA** = baixo\_risco (2.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 55**

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = encanado  
 E Terreno área de aterro = não  
 E Rio = ausente

E Terreno úmido constantemente = sim:  
**ENTÃO MICROÁREA** = baixo\_risco (2.0)

#### **REGRA 56**

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = encanado  
 E Terreno área de aterro = sim  
 E Cancha para esporte = presente:  
**ENTÃO MICROÁREA** = médio\_risco (2.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 57**

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = encanado  
 E Terreno área de aterro = sim  
 E Cancha para esporte = ausente:  
**ENTÃO MICROÁREA** = alto\_risco (3.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 58**

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = fossa:  
**ENTÃO MICROÁREA** = baixo\_risco (1.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 59**

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = misto  
 E Terreno área de aterro = não  
 E Parque ou praça = presente:  
**ENTÃO MICROÁREA** = médio\_risco (5.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 60**

SE microárea é considerada violenta = sim  
 E Cava = ausente  
 E Rede de esgoto = misto  
 E Terreno área de aterro = não  
 E Parque ou praça = ausente  
 E Área da construção das casas = legalizada:  
**ENTÃO MICROÁREA** = médio\_risco (2.0)

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 61**

SE microárea é considerada violenta = sim

- E Cava = ausente
- E Rede de esgoto = misto
- E Terreno área de aterro = não
- E Parque ou praça = ausente
- E Área da construção das casas = não\_legalizada
- E Destino do lixo coletado pela prefeitura = sim:

**ENTÃO MICROÁREA = alto\_risco (3.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 62**

SE microárea é considerada violenta = sim

- E Cava = ausente
- E Rede de esgoto = misto
- E Terreno área de aterro = não
- E Parque ou praça = ausente
- E Área da construção das casas = não\_legalizada
- E Destino do lixo coletado pela prefeitura = parcialmente:

**ENTÃO MICROÁREA = médio\_risco (3.0/1.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 63**

SE microárea é considerada violenta = sim

- E Cava = ausente
- E Rede de esgoto = misto
- E Terreno área de aterro = não
- E Parque ou praça = ausente
- E Área da construção das casas = em\_processo\_de\_legalização:

**ENTÃO MICROÁREA = alto\_risco (8.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 64**

SE microárea é considerada violenta = sim

- E Cava = ausente
  - E Rede de esgoto = misto
  - E Terreno área de aterro = sim:
- ENTÃO MICROÁREA = alto\_risco (10.0)**

Para você esta regra é:

- 1 - ( ) Irrelevante (incompatível com a realidade)
- 2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)
- 3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 65**

SE microárea é considerada violenta = sim

E Cava = ausente

E Rede de esgoto = misto

E porco = sim:

**ENTÃO MICROÁREA** = médio\_risco (2.0)

Para você esta regra é:

1 - ( ) Irrelevante (incompatível com a realidade)

2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)

3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 66**

SE microárea é considerada violenta = sim

E Cava = ausente

E Rede de esgoto = a\_ceu\_aberto:

**ENTÃO MICROÁREA** = alto\_risco (3.0/1.0)

Para você esta regra é:

1 - ( ) Irrelevante (incompatível com a realidade)

2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)

3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 67**

SE microárea é considerada violenta = sim

E Cava = não\_contaminado:

**ENTÃO MICROÁREA** = médio\_risco (1.0)

Para você esta regra é:

1 - ( ) Irrelevante (incompatível com a realidade)

2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)

3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

#### **REGRA 68**

SE microárea é considerada violenta = sim

E Cava = contaminado:

**ENTÃO MICROÁREA** = alto\_risco (9.0)

Para você esta regra é:

1 - ( ) Irrelevante (incompatível com a realidade)

2 - ( ) Confirma seu conhecimento (apresenta padrões que confirmam seu conhecimento)

3 - ( ) Interessante (apresenta padrões condizentes com a realidade, mas até então desconhecidos para você)

Number of Leaves : 79  
 Size of the tree : 133

Time taken to build model: 0.17 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	471	88.7006 %
Incorrectly Classified Instances	60	11.2994 %
Kappa statistic	0.8091	
Mean absolute error	0.1222	
Root mean squared error	0.2471	
Relative absolute error	30.4952 %	
Root relative squared error	55.2395 %	
Total Number of Instances	531	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Área	Class
0.903	0.12	0.868	0.903	0.885	0.944	médio_risco
0.907	0.076	0.891	0.907	0.899	0.961	baixo_risco
0.761	0.004	0.962	0.761	0.85	0.968	alto_risco

=== Confusion Matrix ===

a	b	c	<-- classified as
224	22	2	a = médio_risco
20	196	0	b = baixo_risco
14	2	51	c = alto_risco

**ANEXO A – TERMOS DE CONSENTIMENTO LIVRE E ESCLARECIDO**

## TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Eu \_\_\_\_\_ Rg n° \_\_\_\_\_

Estou sendo convidado a participar de um estudo denominado:  
**DESENVOLVIMENTO DE BASE DE DADOS PARA MICROÁREAS**

Cujo objetivo é: **construir um base de dados contendo dados reais referentes às características de: equipamentos sociais, comércio, áreas de lazer, hidrografia, solo, infra-estrutura, vetores, atividades da unidade de saúde e comunidade de microáreas da cidade de Curitiba.**

Este estudo resultara em um sistema de informação com dados atuais que ira ajudar as equipes de saúde da família no planejamento das ações locais de saúde.

Por ser o questionário sobre as condições sócio-ambientais das microáreas, este estudo não oferece risco à saúde dos participantes envolvidos.

Sei que para o avanço da pesquisa a participação de voluntários é de fundamental importância. Caso aceite participar desta pesquisa eu responderei a um questionário elaborado pelo pesquisador, que consta de questões fechadas e objetivas, referentes a variáveis que influenciam na saúde da comunidade.

Estou ciente que minha privacidade será respeitada, ou seja, meu nome, ou qualquer outro dado confidencial, será mantido em sigilo. A elaboração final dos dados será feita de maneira codificada, respeitando a ética da confidencialidade.

Estou ciente que posso me recusar a participar do estudo, ou retirar meu consentimento a qualquer momento, sem precisar me justificar, nem sofrer qualquer dano.

Estão garantidas todas as informações que eu queira saber antes, durante e depois do estudo.

Li este termo, fui orientado quanto ao conteúdo desta pesquisa e compreendi a natureza e o objetivo do estudo do qual fui convidado a participar.

O pesquisador envolvido com o referido projeto é ALTAIR VON STEIN JUNIOR (ENFERMEIRO MESTRANDO DO PROGRAMA DE PÓS GRADUAÇÃO EM TECNOLOGIA EM SAÚDE DA PUCPR), com quem poderei manter contato pelos telefones: (41) 9176-5239 (41) 3666-4528.

Concordo, voluntariamente em participar desta pesquisa, sabendo que não receberei nem pagarei nenhum valor econômico por minha participação.

\_\_\_\_\_  
Assinatura do sujeito de pesquisa\_\_\_\_\_  
Assinatura do pesquisador

Curitiba \_\_\_\_\_ de \_\_\_\_\_ de 2007

## TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Eu \_\_\_\_\_ Rg n° \_\_\_\_\_

ESTOU SENDO CONVIDADO A PARTICIPAR DE UM ESTUDO DENOMINADO “SISTEMA DE APOIO À DECISÃO NA CLASSIFICAÇÃO DE MICRO ÁREAS PARA O PLANEJAMENTO DE AÇÕES LOCAIS DA EQUIPE DE SAÚDE DA FAMÍLIA” CUJO OBJETIVO É DESENVOLVER UM SISTEMA DE APOIO À DECISÃO PARA A CLASSIFICAÇÃO DE MICRO ÁREAS, AUXILIANDO O PROCESSO DE DECISÃO NO PLANEJAMENTO DAS AÇÕES LOCAIS PELA EQUIPE DO PROGRAMA DE SAÚDE DA FAMÍLIA .

Sei que para o avanço da pesquisa a participação de voluntários é de fundamental importância. Caso aceite participar desta pesquisa eu responderei a uma escala de validação de variáveis elaborado pelos pesquisadores, que consta de questões fechadas e objetivas, referentes a variáveis que influenciam no processo saúde doença de uma comunidade.

Estou ciente que minha privacidade será respeitada, ou seja, meu nome, ou qualquer outro dado confidencial, será mantido em sigilo. **A elaboração final dos dados será feita de maneira codificada, respeitando o imperativo ético da confidencialidade.**

**Estou ciente que posso me recusar a participar do estudo, ou retirar meu consentimento a qualquer momento, sem precisar me justificar, nem sofrer qualquer dano.**

Os pesquisadores envolvidos com o referido projeto são ALTAIR VON STEIN JUNIOR (ENFERMEIRO MESTRANDO DO PROGRAMA DE PÓS GRADUAÇÃO EM TECNOLOGIA EM SAÚDE DA PUCPR), LAUDELINO CORDEIRO BASTOS ( DOUTOR EM SISTEMAS DE INFORMAÇÃO EM SAÚDE) E ANDREIA MALUCELLI ( MESTRE EM VISÃO COMPUTACIONAL), com quem poderei manter contato pelos telefones: (41) 9176-5239, (41) 3666-4528.

Estão garantidas todas as informações que eu queira saber antes, durante e depois do estudo. Li, portanto, este termo, fui orientado quanto ao teor da pesquisa acima mencionada e compreendi a natureza e o objetivo do estudo do qual fui convidado a participar. Concordo, voluntariamente em participar desta pesquisa, sabendo que não receberei nem pagarei nenhum valor econômico por minha participação.

\_\_\_\_\_  
Assinatura do sujeito de pesquisa

\_\_\_\_\_  
Assinatura dos pesquisadores

Curitiba \_\_\_\_\_ de \_\_\_\_\_ de 2007

**ANEXO B - APROVAÇÃO DO COMITÊ DE ÉTICA EM PESQUISA (PESQ.1)**

**Ref. “ Concepção de um sistema de apoio à decisão para caracterização de microáreas de risco na comunidade”**



Pontifícia Universidade Católica do Paraná  
Pró-Reitoria Acadêmica e de Pesquisa  
Diretoria de Pesquisa e Programas Stricto Sensu

Curitiba, 23 de novembro de 2006.

Of. 709/06/CEP-PUCPR

Ref. "Concepção de um sistema de apoio à decisão para caracterização  
de micro áreas de risco na comunidade"

---

Prezado (a) Pesquisador (es),

Venho por meio deste informar a Vossa Senhoria que o Comitê de Ética em Pesquisa da PUCPR, no dia 22 de novembro do corrente ano aprovou o Projeto Intitulado "Concepção de um sistema de apoio à decisão para caracterização de micro áreas de risco na comunidade", pertencente ao Grupo III, sob o registro no CEP nº 1442, e será encaminhado a CONEP para o devido cadastro. Lembro ao senhor (a) pesquisador (a) que é obrigatório encaminhar relatório anual parcial e relatório final a este CEP.

Atenciosamente,

Profª M. Sc Ana Cristina Miguez Ribeiro  
Coordenadora do Comitê de Ética em Pesquisa - PUCPR

Ilmo Sr  
Altair Von Stein Junior

**ANEXO C - APROVAÇÃO DO COMITÊ DE ÉTICA EM PESQUISA (PESQ.2)**

**Ref. “Desenvolvimento de base de dados referente as características de microáreas”**



PUCPR

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ**NÚCLEO DE BIOÉTICA  
COMITÊ DE ÉTICA EM PESQUISA

Curitiba, 04 de setembro de 2007.

Parecer 517/CEP-PUCPR

Ref. **“Desenvolvimento de base de dados referente as características de micro áreas”**

Prezado (a) Pesquisador (es),

Venho por meio deste informar a Vossa Senhoria que o Comitê de Ética em Pesquisa da PUCPR, no **dia 29 de julho** do corrente ano aprovou o Projeto Intitulado **“Desenvolvimento de base de dados referente as características de micro áreas I”**, pertencente ao Grupo III, sob o registro no **CEP n° 1771**, e será encaminhado a CONEP para o devido cadastro. Lembro ao senhor (a) pesquisador (a) que é obrigatório encaminhar relatório anual parcial e relatório final a este CEP.

Atenciosamente,

A handwritten signature in black ink, appearing to be 'S. Surugi de Siqueira', written over a horizontal line.

Prof. Sergio Surugi de Siqueira

Coordenador do Comitê de Ética em Pesquisa - PUCPR



Ilma Sr.

**Altair Von Stein Júnior**