



CHRISTIANE FALEIRO SIDNEY

**APLICAÇÃO DE MINERAÇÃO DE DADOS NO
BANCO DE DADOS DO ZONEAMENTO
ECOLÓGICO ECONÔMICO DE MINAS
GERAIS**

**LAVRAS - MG
2010**

CHRISTIANE FALEIRO SIDNEY

**APLICAÇÃO DE MINERAÇÃO DE DADOS NO BANCO DE DADOS
DO ZONEAMENTO ECOLÓGICO ECONÔMICO DE MINAS GERAIS**

Monografia de graduação apresentada ao Departamento de Ciência da Computação da Universidade Federal de Lavras como parte das exigências do curso de Sistemas de Informação para obtenção do título de Bacharel em Sistemas de Informação.

Orientador:

Dr. Ahmed Ali Abdallah Esmín

**LAVRAS - MG
2010**

CHRISTIANE FALEIRO SIDNEY

**APLICAÇÃO DE MINERAÇÃO DE DADOS NO BANCO DE DADOS
DO ZONEAMENTO ECOLÓGICO ECONÔMICO DE MINAS GERAIS**

Monografia de graduação apresentada ao Departamento de Ciência da Computação da Universidade Federal de Lavras como parte das exigências do curso de Sistemas de Informação para obtenção do título de Bacharel em Sistemas de Informação.

APROVADA em ____ de _____ de _____

Msc. Juliana Galvani Greghi UFLA

Dra. Marluce Rodrigues Pereira UFLA

Dr. Ahmed Ali Abdalla Esmín

Orientador

**LAVRAS - MG
2010**

RESUMO

Com os estudos e pesquisas relacionadas a questões ambientais e sócio-econômicas tem-se obtido grande volume de dados. Porém, essa grande quantidade de dados torna a análise humana onerosa e métodos tradicionais de recuperação de dados não são eficazes para obtenção de conhecimento ocultos em massas de dados. Desta forma, a mineração de dados surge como alternativa para auxiliar a descoberta automática de conhecimento. Este trabalho apresenta uma aplicação prática de mineração de dados na base de dados do Zoneamento Ecológico Econômico de Minas Gerais. Para isso, foi realizada toda a etapa de KDD (*Knowledge Discovery in Databases*), incluindo a implementação de um *software* que utiliza o algoritmo Apriori, para obtenção de regras de associação, da biblioteca WEKA (*Weikato Enviroment for Knowledge Analysis*). Os resultados obtidos poderão ser usados como uma ferramenta de apoio no planejamento e elaboração das políticas públicas para o governo e para a sociedade.

Palavras-chave: Mineração de dados. Regras de Associação. Banco de dados espacial. Zoneamento Ecológico Econômico do Estado de Minas Gerais.

ABSTRACT

Through studies and researches related to environmental and socio-economic has obtained large amounts of data. However, this large amount of data makes analysis and human costly traditional methods of data recovery are not effective for obtaining knowledge hidden in masses of data. Thus, data mining is an alternative to assist the automatic discovery of knowledge. This paper presents a practical application of data mining in the database of Ecological and Economic Zoning of Minas Gerais. For this, we performed the entire step of KDD (Knowledge Discovery in Databases), including the implementation of software that uses the Apriori algorithm to obtain association rules, the biblioteca WEKA (Weikato Environment for Knowledge Analysis). The results can be used as a support tool in planning and developing public policies for the government and society.

Keywords: Data Mining. Association Rules. Spatial Database. Ecological and Economic Zoning of Minas Gerais.

LISTA DE ILUSTRAÇÕES

Figura 1 Dado Espacial	15
Figura 2 Relacionamentos topológicos entre objetos espaciais	16
Figura 3 Relação entre KDD e Mineração de Dados	17
Figura 4 Processos de KDD.....	17
Figura 5 Intersecção de MD com outras disciplinas	18
Figura 6 Tarefa de Classificação.....	21
Figura 7 Site ZEE-MG	25
Figura 8 Ferramenta ZEE-MG.....	26
Figura 9 Mapa do Brasil com destaque na localização do estado de Minas Gerais	27
Figura 10 Mapa de Minas Gerais com os seus 853 municípios.....	27
Figura 11 Potencialidade Produtiva	29
Figura 12 Potencialidade Natural.....	29
Figura 13Potencialidade Humana	30
Figura 14 Potencialidade Institucional	30
Figura 15 ST_Centroid(Geometria A)	33
Figura 16 ST_Intersects(Geometria A, Geometria B).....	34
Figura 17 ST_Intersection(Geometria A, Geometria B)	34
Figura 18 Consulta SQL para classificação dos município em relação ao índice de malha ferroviária	35
Figura 19 Tabela união.....	36
Figura 20 Tela com as opções de indicadores para geração de regras de associação.....	37
Figura 21 Parâmetros de entrada para a associação	38
Figura 22 Tela com as regras geradas	39
Figura 23 Exemplo de regras geradas	40

Figura 24 Municípios enquadrados na regra 1.....	43
Figura 25 Municípios enquadrados na regra 2.....	44
Figura 26 Municípios enquadrados na regra 3.....	46
Figura 27 Municípios enquadrados na regra 4.....	47
Figura 28 Municípios enquadrados na regra 5.....	49
Figura 29 Municípios enquadrados na regra 6.....	51
Figura 30 Municípios enquadrados na regra 7.....	53
Figura 31 Municípios enquadrados na regra 8.....	54
Figura 32 Municípios enquadrados na regra 9.....	56

LISTA DE ABREVIATURAS

MD	Mineração de Dados
KDD	<i>Knowledge Discovery in Databases</i>
SGBD	Sistema Gerenciador de Banco de Dados
SQL	<i>Structured Query Language</i>
ZEE-MG	Zoneamento Ecológico Econômico do Estado de Minas Gerais

SUMÁRIO

1	INTRODUÇÃO.....	10
1.1	Contextualização e Motivação	10
1.2	Objetivos do Trabalho	11
1.2.1	Objetivo Geral	11
1.2.2	Objetivos Específicos	11
1.3	Estrutura do Trabalho	12
2	REFERENCIAL TEÓRICO.....	13
2.1	Banco de Dados.....	13
2.1.1	Banco de Dados Espaciais.....	13
2.2	Descoberta de Conhecimento Em Banco De Dados	16
2.3	Mineração de Dados.....	18
2.3.1	Técnicas de Mineração De Dados	19
2.3.2	Mineração de Dados Espaciais.....	23
2.4	Zoneamento Ecológico Econômico do Estado de Minas Gerais	23
3	MATERIAIS E MÉTODOS	28
3.1	Tipo de Pesquisa.....	28
3.2	Objeto do Estudo	28
3.3	Procedimentos Metodológicos	31
4	EXECUÇÃO DO TRABALHO.....	32
4.1.1	Tipo de Conhecimento	32
4.1.2	Seleção dos Dados.....	32
4.1.3	Limpeza e pré-processamento dos dados e Redução e projeção dos dados	32
4.1.4	Mineração de Dados.....	37
4.1.5	Interpretação dos Padrões Descobertos e Implantação do Conhecimento	40
5	RESULTADOS	42
6	CONCLUSÃO.....	57
7	REFERÊNCIAS BIBLIOGRÁFICAS	58

1 INTRODUÇÃO

1.1 Contextualização e Motivação

Nas sociedades organizadas sempre foi uma tarefa importante a coleta de informações relacionadas a distribuição geográfica de recursos minerais, animais e plantas. Com o advento da informática, o que era antes realizado apenas em documentos e mapas em papel, passou a ser feito em ambiente computacional (Câmara & Davis, 2001).

Com o grande incentivo no mundo à estudos e pesquisas relacionadas a questões ambientais e sócio-econômicas, motivados pela necessidade de informações para tomada de decisões, tem-se obtido grande volume de dados georreferenciados, que são aqueles que possuem localização espacial determinada por coordenadas geográficas e atributos descritivos.

Porém, essa grande quantidade de dados torna a análise humana onerosa e métodos tradicionais de recuperação de dados, mesmo que sejam sofisticados, não são eficazes para obtenção de conhecimentos ocultos em massas de dados. Portanto, o maior desafio é a busca de ferramentas e técnicas que auxiliem nesse processo de busca de informações novas e de valor.

Desta forma, a descoberta de conhecimento em bancos de dados surge como alternativa para auxiliar a descoberta automática de conhecimento através do processo completo de conversão de dados brutos em informações úteis (Tan, Steinbach & Kumar, 2009). A principal etapa da descoberta de conhecimento, onde são aplicados algoritmos inteligentes para extração de dados, é conhecida como mineração de dados.

Segundo Fayyad, Piatetsku-Shapiro e Smyth (1996), mineração de dados consiste na aplicação de análise de dados e algoritmos de descoberta de conhecimento que, sob as limitações aceitáveis de eficiência computacional,

produzem uma enumeração particular de padrões sobre os dados.

Mineração de dados espaciais é a descoberta de relações interessantes e características que podem existir implicitamente em bases de dados espaciais. Entretanto, o processo de descoberta de conhecimento em banco de dados espaciais é mais complexo que o de dados relacionais. Isso se aplica tanto a eficiência dos algoritmos, quanto a complexidade dos padrões possíveis que podem ser encontrados em um banco de dados espacial. A razão é que, em contraste com a mineração em bases de dados relacionais, algoritmos de mineração de dados espaciais tem que considerar a vizinhança entre os objetos, a fim de extrair conhecimento útil (Ester, Kriegel & Sander, 2001).

1.2 Objetivos do Trabalho

1.2.1 Objetivo Geral

O objetivo deste trabalho é a aplicação de mineração de dados no banco de dados geográficos do Zoneamento Ecológico Econômico de Minas Gerais para extração de padrões não triviais. O Zoneamento Ecológico Econômico (ZEE) é uma ferramenta que a sociedade e empreendedores têm para conhecimento das vulnerabilidades e potencialidades de cada região do estado de Minas Gerais.

O ZEE é de grande importância no planejamento e elaboração das políticas públicas e das ações em meio ambiente, orientando o governo e a sociedade civil na elaboração dos seus programas e em seus investimentos.

1.2.2 Objetivos Específicos

O trabalho ainda possui objetivos específicos, que são:

- Estudar e identificar técnicas de mineração de dados aplicáveis a descoberta de conhecimento na base de dados do Zoneamento Ecológico Econômico de Minas Gerais;
- Aplicar e adaptar a técnica de mineração de dados na base de dados para extração de características espaciais interessantes e;
- Analisar e relacionar o resultado do algoritmo.

1.3 Estrutura do Trabalho

Este trabalho está estruturado da seguinte maneira:

- O Capítulo 2 apresenta o Referencial Teórico, esclarecendo os principais conceitos utilizados no trabalho.
- O Capítulo 3 mostra a Metodologia do trabalho.
- O Capítulo 4 possui os Resultados deste trabalho.
- O Capítulo 5 mostra a Execução do trabalho.
- O Capítulo 6 finaliza o trabalho com a Conclusão.

2 REFERENCIAL TEÓRICO

2.1 Banco de Dados

Segundo Elmasri e Navathe (2005), um banco de dados é uma coleção de dados relacionados e possui as seguintes propriedades:

- Representa alguns aspectos do mundo real, sendo chamado, às vezes, de minimundo ou de universo de discurso. As mudanças no minimundo são refletidas em um banco de dados.
- É uma coleção lógica e coerente de dados com algum significado inerente.
- É projetado, construído e povoado por dados, atendendo a uma proposta específica.

De acordo com Date (2003), banco de dados é um sistema computadorizado cuja finalidade geral é armazenar informações e permitir que os usuários busquem e atualizem essas informações quando solicitar.

2.1.1 Banco de Dados Espaciais

“Banco de dados espaciais pode ser visto como uma extensão de um banco de dados padrão e tem como característica predominante ser composto por dados descritivos associados a uma representação geométrica que possui localização espacial. No caso de utilização de dados geográficos, o banco de dados espaciais pode ser nomeado como banco de dados geográficos” (Pivato, 2006).

Segundo Guting (1994), um sistema de banco de dados espacial possui características tais como: ser um sistema de banco de dados, oferecer tipos de dados espaciais em seu modelo de dados e linguagem de consulta, e suportar tipos de dados espaciais em sua implementação como indexação espacial e algoritmos para junção espacial.

Os objetos espaciais, de acordo com Câmara *et al* (2005), usam um ou mais pares de coordenadas para determinar sua localização no espaço. Desta forma, o objeto pode ter o tipo de dado espacial representado como ponto, reta, linha poligonal e polígono.

Ponto é um par ordenado (x, y) de coordenadas espaciais. O ponto pode ser utilizado para identificar localizações ou ocorrências no espaço.

Linha poligonal é um conjunto de pontos conectados. A linha é utilizada para guardar feições unidimensionais.

Polígono (área) é a região do plano limitada por uma ou mais linhas poligonais conectadas de tal forma que o último ponto de uma linha seja idêntico ao primeiro da próxima. A fronteira do polígono divide o plano em duas regiões: o interior e o exterior. Os polígonos são usados para representar unidades espaciais individuais como setores censitários, distritos, zonas de endereçamento postal e municípios. Para cada unidade, são associados dados oriundos de levantamento como censos e estatísticas de saúde.

A Figura 1 ilustra a composição do dado espacial, que possui suas informações descritivas e sua representação geométrica (ponto, reta, linha poligonal ou polígono).

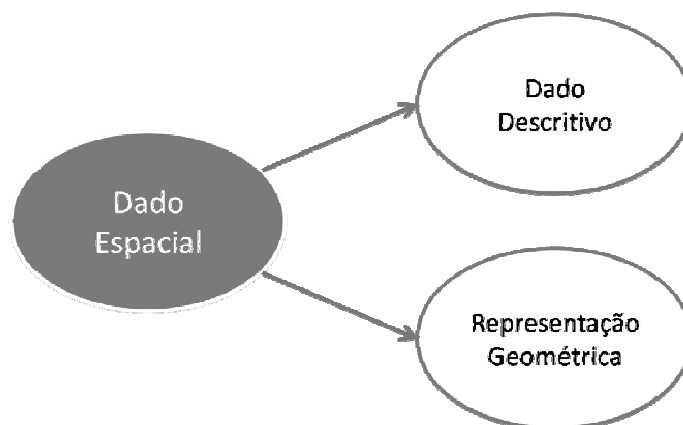


Figura 1 Dado Espacial

A principal diferença entre os bancos de dados espaciais e os não-espaciais está associada ao fato de que entidades geográficas endereçadas podem ser afetadas por características de entidades vizinhas. A influência mútua que duas entidades exercem entre si depende de fatores como a topologia, a distância e a direção entre elas (Borgony, 2003). As influências podem ser de ordem topológica, de direção ou distância.

Relação topológica: é a relação que define os objetos como adjacentes, dentro, disjuntos de outros objetos.

Relações de direção: define os objetos como acima, abaixo ou ao norte de, a sudoeste de outros objetos.

Relações de distância (métricas): define os objetos em relação a distância de outros objetos.

A Figura 2 ilustra possíveis relacionamentos topológicos entre objetos espaciais.

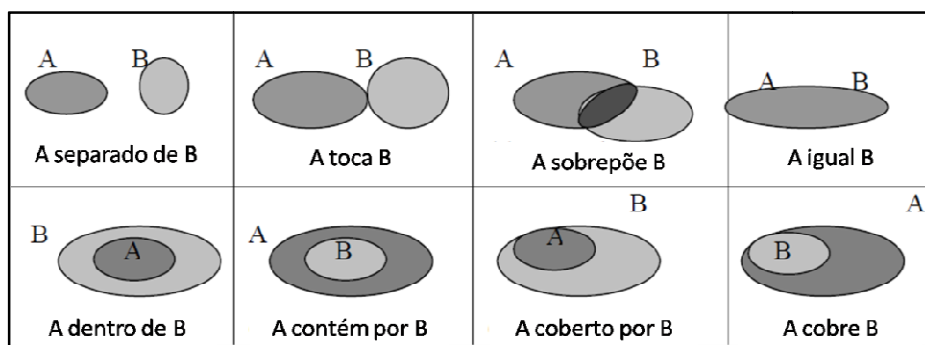


Figura 2 Relacionamentos topológicos entre objetos espaciais

2.2 Descoberta de Conhecimento em Banco de Dados

Segundo Fayyad, Piatetsky-Shapiro, Smyth e Uthurusamy (1996), nos últimos tempos, todo o mundo tem armazenado grandes quantidades de dados, superando consideravelmente as habilidades humanas de interpretação, gerando a necessidade de criação de técnicas e ferramentas que automatizem e analisem a base de dados de maneira inteligente.

O termo KDD – *Knowledge Discovery in Databases* foi criado em 1995 para designar o conjunto de processos, técnicas e abordagens que propiciam o contexto no qual a mineração de dados terá lugar (Braga, 2005). Assim, a relação existente entre descoberta de conhecimento e mineração de dados foi definida como na Figura 3.

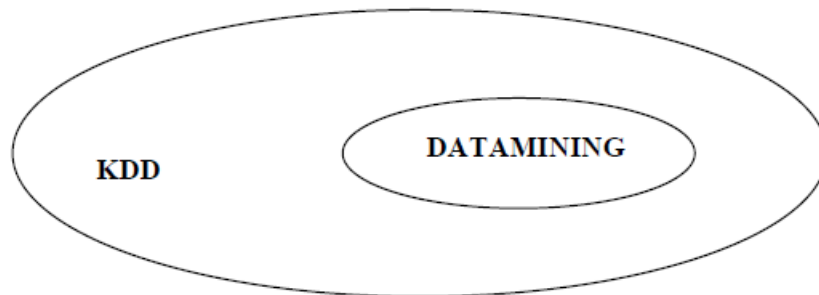


Figura 3 Relação entre KDD e Mineração de Dados

O processo global de descoberta de conhecimento, segundo Mitra e Acharya (2003) consiste na seqüência iterativa dos passos abaixo, representado na Figura 4.

- Definição do tipo de conhecimento que se deseja descobrir;
- Seleção de dados;
- Limpeza e pré-processamento dos dados;
- Redução e projeção dos dados;
- Mineração de dados;
- Interpretação dos padrões descobertos;
- Implantação do conhecimento.

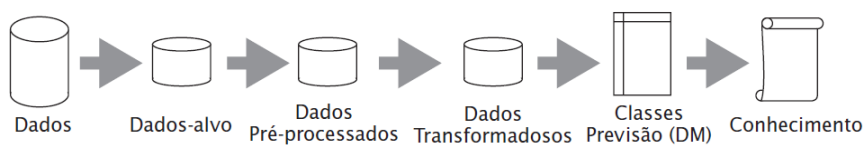


Figura 4 Processos de KDD

2.3 Mineração de Dados

A mineração de dados é uma área interdisciplinar que surgiu nos anos 80 por vários pesquisadores de áreas como inteligência artificial, estatística, banco de dados, computação paralela, guiados pela necessidade de obtenção de informações dentro de grandes bases de dados. A Figura 5 demonstra essa relação.

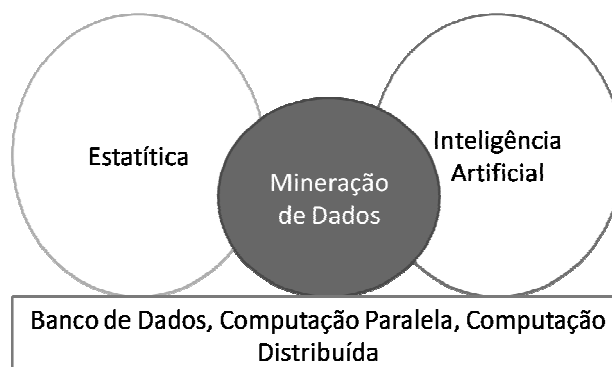


Figura 5 Intersecção de MD com outras disciplinas

A mineração de dados é definida como um processo de descoberta de padrões nos dados. O processo de mineração de dados deve ser automático ou semi-automático. Os padrões descobertos devem ser significantes na medida em que levam a alguma vantagem, normalmente vantagem econômica (Witten & Frank, 2005).

A mineração de dados é uma tecnologia que combina métodos tradicionais de análise de dados com algoritmos sofisticados para processar grandes volumes de dados (Tan, Steinbach & Kumar, 2009).

A mineração de dados provê um método automático para descobrir padrões em dados, sem a tendenciosidade e a limitação da análise baseada meramente na intuição humana (Braga, 2005).

“Procurar por padrões interessantes em uma forma particular de representação ou um conjunto de representações como: regras de classificação ou árvores, regressão, agrupamento e outros.” (Fayyad, Piatetsku-Shapiro & Smyth, 1996)

2.3.1 Técnicas de Mineração de Dados

As principais técnicas de mineração de dados são associação, classificação, agrupamento e detecção de anomalias,

2.3.1.1 Associação

“Associação é uma técnica de atividade descritiva de mineração de dados que extrai o conhecimento a partir de exemplos, construindo representações compreensíveis ao ser humano. A representação utilizada é uma regra indicando o quanto a presença de um conjunto de itens está relacionada com a presença de outro conjunto de itens distintos nos mesmos registros.” (Pivato, 2006)

Uma regra de associação é uma expressão de implicação no formato $X \rightarrow Y$, onde X e Y são conjuntos distintos de itens. A força de uma regra de associação pode ser medida em termos de seu suporte e confiança. O suporte determina a frequência na qual uma regra é aplicável a um determinado conjunto de dados, enquanto a confiança determina a frequência na qual os itens em Y aparecem em transações que contenham X (Tan, Steinbach & Kumar, 2009).

O suporte é utilizado para eliminar regras sem interesse porque uma regra que tenha baixo suporte pode acontecer simplesmente por coincidência. A confiabilidade de uma regra é obtida através da confiança, que observa o quão confiável é a inferência da regra. Através da confiança temos a estimativa da probabilidade de um item X aparecer em transações que possuam X (Tan, Steinbach & Kumar, 2009).

2.3.1.2 Classificação

Classificação, segundo Han e Kamber (2000), é o processo de encontrar um modelo ou função que descreve e distingue classes de dados e conceitos, com o propósito de usar este modelo para predizer objetos ainda não classificados.

Os dados de entrada da tarefa de classificação são um conjunto de registros. Os registros possuem um conjunto de atributos e um conjunto especial, designado como rótulo da classe. O papel da classificação é aprender uma função alvo que mapeie cada conjunto de atributos para um dos rótulos de classes pré-determinados. A função alvo é conhecida como modelo de classificação (Tan, Steinbach & Kumar, 2009). A Figura 6 mostra como é realizada a tarefa de classificação.



Figura 6 Tarefa de Classificação

A modelagem pode ser descritiva ou preditiva. A modelagem preditiva serve como ferramenta explicativa para se distinguir entre objetos e classes diferentes. Já a modelagem preditiva é usada para prever rótulos da classe de registros não conhecidos.

2.3.1.3 Agrupamento

“Classes, ou grupos conceitualmente significativos de objetos que compartilham características comuns, desempenham um papel importante em como as pessoas analisam e descrevem o mundo. De fato, seres humanos têm habilidade na divisão de objetos em grupos (agrupamento) e atribuir objetos particulares a esses grupos (classificação)” (Tan, Steinbach & Kumar, 2009).

Em (Berry & Linoff, 2004), agrupamento é a tarefa de segmentar uma população heterogênea dentro de um número de subgrupos homogêneos. No agrupamento não são predefinidas as classe, assim, os registros são agrupados baseados em suas próprias similaridades.

A intenção é que os objetos dentro de um grupo sejam semelhantes entre si e diferentes dos objetos dos outros grupos. Quanto maior a semelhança dentro de um grupo e maior a diferença entre grupos, melhor ou mais distinto será o agrupamento.

2.3.1.4 Detecção de Anomalias

Detecção de anomalias é a tarefa de encontrar objetos que sejam diferentes da maioria dos outros objetos. Objetos anômalos têm atributos que se desviam significativamente dos valores de atributos esperados ou típicos (Tan, Steinbach & Kumar, 2009).

O motivo da anomalia pode ser proveniente de dados de classes diferentes, variação natural e medidas de dados ou erros de coleta. Dados de classes diferentes ocorrem quando o objeto é de uma classe diferente. Segundo definição de Hawkins (1980), um elemento estranho é uma observação que difere tanto das outras que gera suspeita de que tenha sido gerada por um mecanismo diferente.

A variação natural retrata, numa distribuição normal, que a probabilidade de um objeto decresce à medida que ele se distancia do centro da distribuição. Assim, a maioria dos objetos está no centro, e a probabilidade de que um objeto seja diferente significativamente desse centro é pequena.

Já medidas de erro e erros de coleta são ocasionados por medições incorretas por erro humano, problemas com dispositivos e presença de ruído. Como exemplo, temos erros que podem ocorrer na inserção de dados oriundos de documentos em papel para planilhas e banco de dados.

Anomalias podem ser detectadas através de testes estatísticos que assumem uma distribuição de probabilidade ou modelo para os dados, ou utilizar

medidas de distância, onde os objetos que estão a uma considerável distância de qualquer outro são considerados anômalos. Pode-se também utilizar métodos baseados em desvio, analisando as diferenças das características principais de objetos em um grupo (Han & Kamber, 2000).

2.3.2 Mineração de Dados Espaciais

Mineração de dados espacial é uma sub-área de mineração de dados que lida com a extração de conhecimento implícito, relações espaciais, ou outros padrões interessantes não explicitamente armazenados em bases de dados espaciais (Han, Koperski & Stefanovic, 1997).

De acordo com Miller e Han (2001), padrões como classes, associações, regras e tendência possuem expressões espaciais bem definidas, desta maneira, existem técnicas de mineração de dados que conseguem recuperar esses padrões.

A classificação espacial mapeia objetos geográficos em categorias, considerando relacionamentos de distância, direção, conectividade e morfologia.

A associação espacial define regras de associação que utiliza predicados espaciais na cláusula.

O agrupamento espacial explora relações espaciais entre os objetos de dados para determinar agrupamentos inerentes aos dados de entrada.

E detecção de anomalias espaciais é definida como um objeto espacialmente referenciado, cujos atributos não espaciais parecem incoerentes com os outros objetos dentro de uma vizinhança.

2.4 Zoneamento Ecológico Econômico do Estado de Minas Gerais

O Zoneamento Ecológico Econômico do Estado de Minas Gerais (ZEE-MG) foi implementado com o objetivo de subsidiar o planejamento e orientações das políticas públicas e das ações em meio ambiente, por meio de um Macrodiagnóstico do Estado. Tem a coordenação da Secretaria de Estado de Meio Ambiente e Desenvolvimento Sustentável, participação de todas as Secretarias de Estado de Minas, de outras entidades e da sociedade civil.

O ZEE-MG consiste no diagnóstico dos meios geo-biofísico e sócio-econômico-jurídico-institucional, gerando respectivamente duas cartas, que são mapas geográficos ou topográficos, a carta de Vulnerabilidade Ambiental e a Carta de Potencialidade Social, que sobrepostas concebem áreas com características próprias, determinando o Zoneamento Ecológico-Econômico do Estado.

O ZEE-MG possui um site onde são apresentados, mais detalhadamente, os objetivos, a metodologia, a coordenação e as publicações. A Figura 7 apresenta o site do ZEE-MG, que esta disponível em http://www.zee.mg.gov.br/zee_externo.

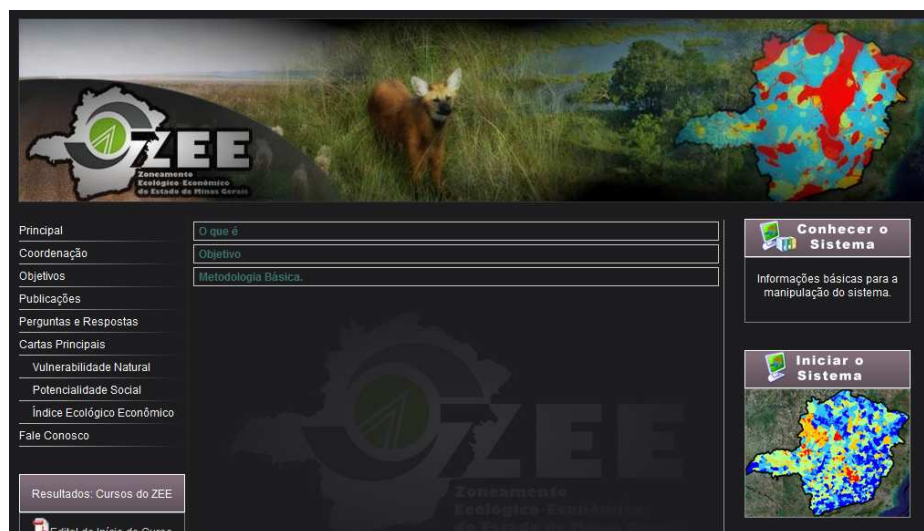


Figura 7 Site ZEE-MG

E a partir desse site é possível acessar a ferramenta do ZEE-MG, onde é possível realizar consultas por camadas, consultas por municípios, consultas espaciais, além de outras funções. A Figura 8 apresenta esta ferramenta, que está disponível em <http://www.zee.mg.gov.br/ferramenta.html>.

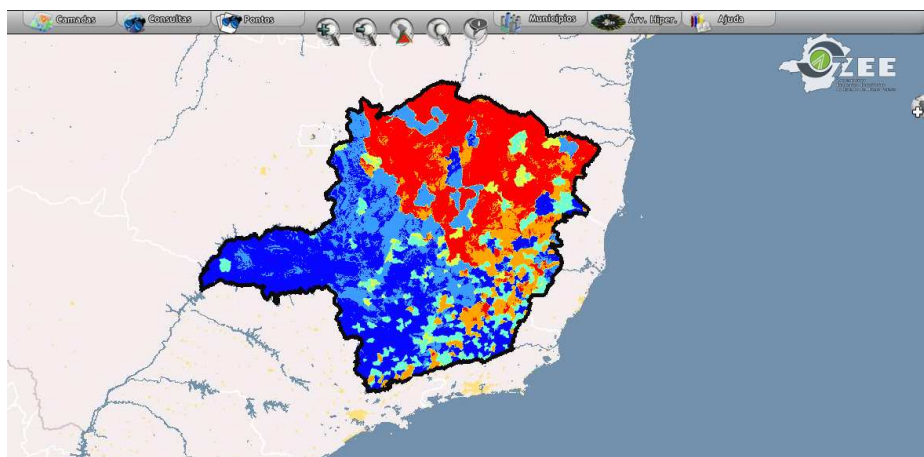


Figura 8 Ferramenta ZEE-MG

O estado de Minas Gerais tem a quarta maior extensão territorial do país, 586.528 km², e está localizado entre os paralelos de 14°13'58' ' e 22°54'00' ' de latitude sul e os meridianos de 39°51'32' ' e 51°02'35' ' a oeste de Greenwich. Está inserido no sudeste do Brasil, possuindo fronteiras com os estados de São Paulo, Mato Grosso do Sul, Goiás, Distrito Federal, Espírito Santo, Rio de Janeiro e Bahia, como mostra a Figura 9. Possui 853 municípios, representados na Figura 10, tendo a população estimada em mais de 20 milhões de habitantes e com a capital do estado em Belo Horizonte.



Figura 9 Mapa do Brasil com destaque na localização do estado de Minas Gerais

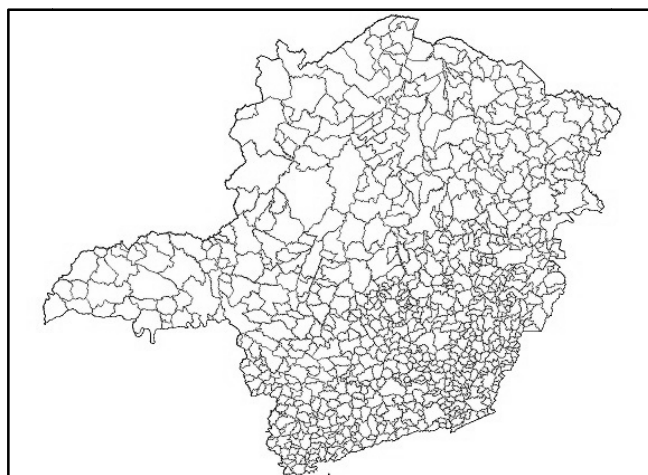


Figura 10 Mapa de Minas Gerais com os seus 853 municípios

3 MATERIAIS E MÉTODOS

Neste capítulo, será esclarecido o tipo de pesquisa utilizada para a fundamentação dos conceitos apresentados nos capítulos anteriores e a descrição de como o estudo foi efetuado.

3.1 Tipo de pesquisa

A pesquisa é um processo de construção do conhecimento que tem por objetivo gerar conhecimentos, corroborar ou refutar algum conhecimento pré-existente. Esta pesquisa pode ser classificada como pesquisa exploratória em relação ao seu objetivo, experimental de acordo com seus procedimentos e em laboratório por seu local de execução.

3.2 Objeto do estudo

Os dados utilizados no estudo são indicadores provenientes da Carta de Potencialidade Social do ZEE-MG, que é dividida nos componentes produtivo, natural, humano e institucional.

O componente produtivo é determinado pelas condições de infra-estrutura e sobre as atividades econômicas produtivas. Os indicadores do componente produtivo utilizados são os demonstrados na Figura 11.



Figura 11 Potencialidade Produtiva

O componente natural trata da utilização econômica dos recursos naturais como exploração mineral, intensidade do uso da terra e sua forma de ocupação e a preservação e conservação do meio ambiente. Os indicadores de componente natural utilizados são os demonstrados na Figura 12.



Figura 12 Potencialidade Natural

O componente humano abrange elementos como trabalho, população, renda, saúde, educação, habitação e segurança. Os indicadores desse componentes estão descritos na Figura 13.

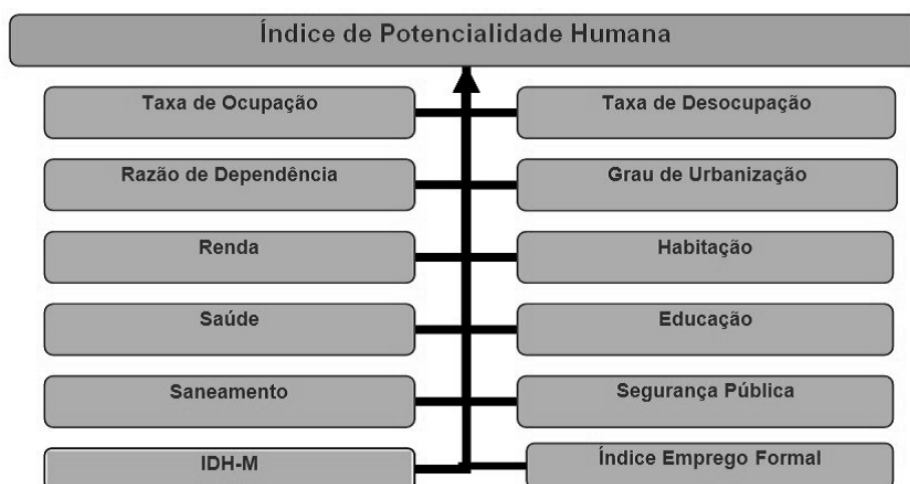


Figura 13 Potencialidade Humana

O componente institucional trata a capacidade do município em atender as demandas dos habitantes de acordo com o funcionamento adequado das instituições presentes. A Figura 14 mostra quais são os indicadores relacionados a esse componente.



Figura 14 Potencialidade Institucional

Cada índice é representado por uma tabela, que possui registros com a categorização para cada cidade do estado de Minas Gerais. De acordo com Scolforo *et al* (2008), a categorização de cada município foi obtida a partir da comparação entre este e todos os 853 municípios.

A categorização é representada por “A”, “B”, “C”, “D” e “E”, que significa respectivamente, “Ponto de Partida em Condições Muito Favoráveis”, “Ponto de Partida em Condições Favoráveis”, “Ponto de Partida em Condições Pouco Favoráveis”, “Ponto de Partida em Condições Precárias” e “Ponto de Partida em Condições Muito Precárias”.

Como exemplo, temos um município X que possui condições de saneamento muito precárias, condições de saúde muito precárias, e educação com condições muito precárias.

3.3 Procedimentos metodológicos

O presente trabalho foi desenvolvido de acordo com o processo de descoberta de conhecimento descrito por Mitra e Acharya (2003), que é dividido em: definição do tipo de conhecimento que se deseja descobrir; seleção de dados; limpeza e pré-processamento dos dados; redução e projeção dos dados; mineração de dados; interpretação dos padrões descobertos; e implantação do conhecimento.

4 EXECUÇÃO DO TRABALHO

Este capítulo aborda todos os passos realizados durante a execução do trabalho.

4.1.1 Tipo de conhecimento

Na fase de definição do tipo de conhecimento que se deseja descobrir foi escolhida a análise de associação.

Foi escolhido como relação topológica a relação contém, que foi usada para calcular qual classificação está contida, em relação a cada um dos indicadores, na área de cada município.

4.1.2 Seleção dos Dados

Foram escolhidos atributos da base de dados do Zoneamento Ecológico Econômico de Minas Gerais relacionados a Potencialidade Social.

4.1.3 Limpeza e pré-processamento dos dados e Redução e projeção dos dados

Nessa fase as tabelas dos indicadores foram importadas para o sistema gerenciador de banco de dados PostgreSQL 8.3, com extensão espacial PostGIS 1.5.

Como os dados das tabelas já possuíam categorização, não foi necessário discretizar os dados, mantendo-se assim, cinco classificações para cada indicador.

O cálculo para obtenção da classificação segundo o indicador foi realizado através das funções `ST_Centroid()`, `ST_Intersects()` e `ST_Intersection()` disponíveis no PostGis.

A função `ST_Centroid(Geometria A)` calcula o centro de massa da geometria A. A Figura 15 apresenta como o ponto central, o centro de massa da Geometria A.

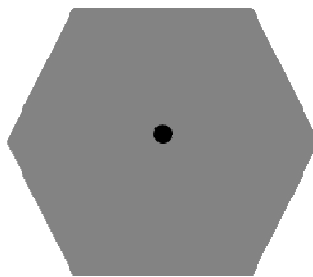


Figura 15 `ST_Centroid(Geometria A)`

A função `ST_Intersects(Geometria A, Geometria B)` verifica se a geometria A intersecciona a geometria B. Esta função retorna *true* se existe intersecção e *false* se não existe a intersecção. A Figura 16 apresenta a intersecção entre duas geometrias.

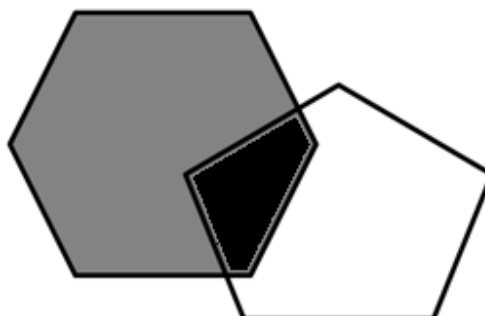


Figura 16 ST_Intersects(Geometria A, Geometria B)

A função `ST_Intersection(Geometria A, Geometria B)` retorna a intersecção entre as duas geometrias. A Figura apresenta a intersecção da Geometria A com Geometria B, obtendo assim a geometria resultante dessa função.

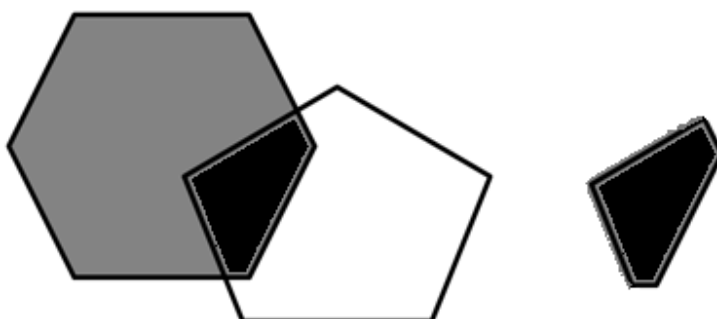


Figura 17 ST_Intersection(Geometria A, Geometria B)

Com o apoio dessas funções foi possível realizar a junção da tabela de municípios com cada tabela de indicadores.

A primeira consulta realizada tentou obter a classificação da geometria resultante da função `ST_Intersection(Geometria Município, Geometria`

Indicador) onde `ST_Intersects(Geometria Município, Geometria Indicador)` é *true*. Porém foram retornados, para cada município, mais de uma classificação, e isto aconteceu porque os polígonos dos índices não conferem exatamente com os polígonos de município. Desta maneira, os municípios estavam sendo classificados segundo os municípios vizinhos.

Para resolver esse problema foi utilizada a função `ST_Centroid()`. Com o auxílio dessa função foi calculado a classificação em relação a cada indicador utilizando o centro de massa de cada município, `ST_Centroid(Geometria Município)`. A Figura 18 exemplifica a consulta em SQL realizada em relação ao índice de malha ferroviária.

```
SELECT  m.gid,
        mf.gridcode,
        To_char(Max(ST_area(ST_Intersection(ST_Centroid(m.the_geom), mf.the_geom))) , '0.99999')
FROM    md.municipios m,
        md.indice_malha_feroviaria mf
WHERE   ST_Intersects(ST_centroid(m.the_geom), mf.the_geom)

GROUP BY m.gid,
         mf.gridcode
order by m.gid
```

Figura 18 Consulta SQL para classificação dos município em relação ao índice de malha ferroviária

Foi criada uma tabela denominada união que associa os 853 municípios a cada indicador, e com o uso da consulta SQL para cada indicador, foram classificados os municípios em relação aos indicadores. A Figura 19 mostra a tabela utilizada após a associação.

Edit Data - PostgreSQL 8.4 [localhost:5432] - md - md.uniao

Arquivo Editar Visualizar Ajuda

100 registros

gid	serial	indice_malha	indice_malha	indice_transp	indice_va_ag	indice_va_inc	indice_va_se	indice_expor	utilizacao_te	nivel_tecnolo	estrutura_fu	recursos_mir	indice_icms	ocupacao_ec	demografia	distribuicao	res
		smallint	smallint	smallint	smallint	smallint	smallint	smallint	smallint	smallint	smallint	smallint	smallint	smallint	smallint	smallint	smallint
1	1	1	2	1	3	1	4	1	3	3	2	4	5	4	2	2	4
2	2	1	4	5	4	5	4	5	3	3	1	1	5	3	5	3	5
3	3	1	5	1	3	5	5	5	3	3	4	1	5	4	4	3	5
4	4	1	1	1	3	4	2	1	4	3	2	1	5	3	2	1	3
5	5	1	3	1	2	5	5	4	3	3	3	1	5	2	4	2	4
6	6	1	1	1	2	2	2	1	2	3	2	1	1	5	3	2	4
7	7	1	2	1	2	2	2	1	3	4	3	1	5	5	2	1	3
8	8	1	4	5	2	5	2	4	3	3	4	4	4	4	4	4	5
9	9	1	1	1	3	2	1	1	4	4	2	1	1	3	3	1	3
10	10	1	4	1	2	5	5	5	4	3	2	1	4	4	5	4	5
11	11	1	1	1	1	2	3	1	2	3	1	5	1	3	2	3	3
12	12	1	1	1	1	4	1	1	4	3	1	1	1	1	2	2	2
13	13	1	5	1	3	3	1	1	3	4	1	4	1	5	3	1	4
14	14	1	1	1	2	2	2	1	1	3	4	3	1	1	5	3	2
15	673	1	1	1	3	1	4	4	1	1	1	1	1	1	1	2	1
16	674	1	1	1	1	2	4	4	1	1	1	1	1	3	1	1	1
17	675	1	2	1	2	3	5	1	1	1	3	5	1	1	2	3	1
18	676	1	2	1	4	2	4	1	1	1	1	1	1	5	1	1	1
19	677	1	1	1	3	2	2	1	1	1	1	1	5	2	1	1	1
20	678	1	2	1	3	1	2	1	1	1	1	1	1	1	1	1	1
21	679	1	5	1	1	4	4	1	1	1	3	1	1	1	3	5	2
22	680	1	3	5	5	1	4	1	2	1	1	5	4	2	1	2	1
23	681	1	1	1	2	3	4	1	2	1	1	1	1	2	1	1	1
24	682	1	3	1	2	2	5	1	2	1	1	1	1	3	1	2	1

Panel de rabiscos

100 registros.

Figura 19 Tabela união

4.1.4 Mineração de dados

Na fase de mineração de dados foi implementado um software em Java, que é uma linguagem extremamente portátil (Deitel & Deitel, 2003), denominado ZEE-MG Mining. Ele busca a tabela união, onde os dados já estão processados, e o usuário pode escolher os indicadores, que possuem uma descrição e também uma imagem de seu mapa, que serão utilizados na tarefa de associação. A Figura 20 mostra a tela onde o usuário visualiza e escolhe os indicadores.

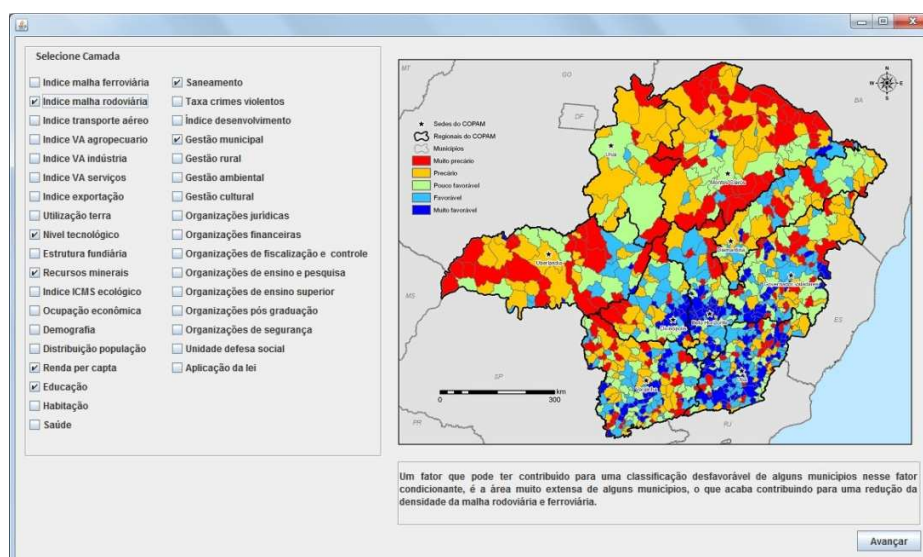
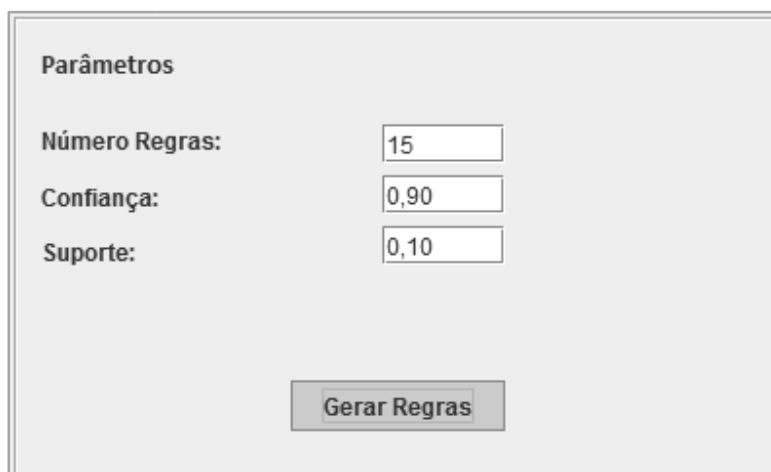


Figura 20 Tela com as opções de indicadores para geração de regras de associação

O ZEE-MG Mining, após a escolha desses indicadores, gera um arquivo ARFF, que é o formato de entrada compatível com a biblioteca WEKA (Witten & Frank, 2005). O WEKA (*Waikato Environment for Knowledge Analysis*) é um pacote formado por um conjunto de implementações de algoritmos de diversas técnicas de Mineração de Dados. Este pacote foi utilizado na tarefa de mineração de dados para realizar a análise de associação.

O passo seguinte é a configuração dos parâmetros número máximo de regras, confiança e suporte. A Figura 21 mostra os campos para inserção desses valores.



Parâmetros

Número Regras:	<input type="text" value="15"/>
Confiança:	<input type="text" value="0,90"/>
Suporte:	<input type="text" value="0,10"/>

Figura 21 Parâmetros de entrada para a associação

Foi utilizado o algoritmo Apriori, que foi o primeiro algoritmo de associação que usou a poda baseada em suporte para controlar de forma sistemática o crescimento exponencial dos conjuntos de itens candidatos. Desta maneira, este algoritmo suporta grande quantidade de registros ou atributos de maneira satisfatória. Segundo Choa, Kimb e Kima (2002), o algoritmo Apriori possui as seguintes fases:

1. Todos os conjuntos de itens (*itemsets*) com suporte mínimo (frequentes) são gerados. Nesta fase é observado se um *itemset* de tamanho k é um *itemset* frequentes, então todos os *itemsets* com tamanho $k-1$ também devem ser *itemsets* frequente. Com esta propriedade, conjunto de itens candidatos de tamanho k são gerados a partir do conjunto de itens frequentes de tamanho $k-1$.
2. Geração de regras a partir do conjunto de todos os *itemsets* frequentes.

A Figura 22 mostra a tela onde as regras geradas pelo ZEE-MG Mining são apresentadas, onde ao lado direito são apresentadas as regras.

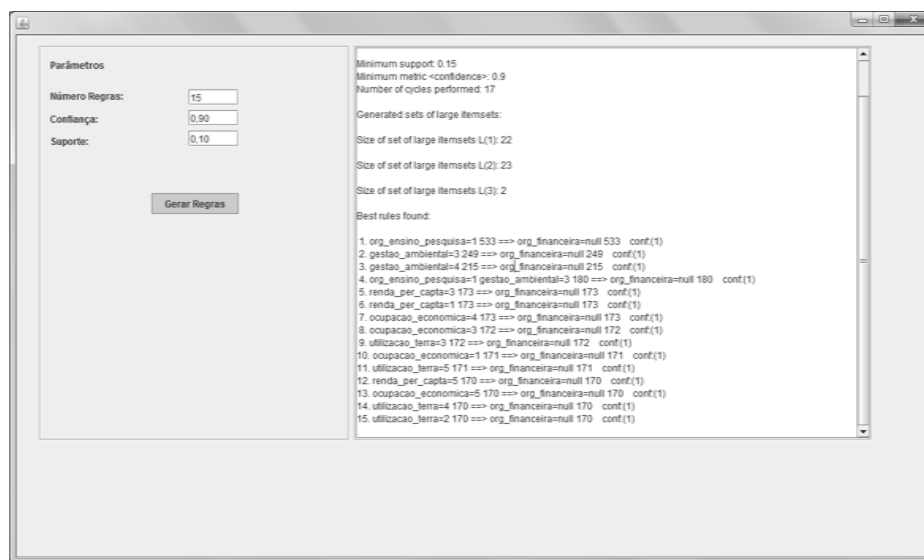


Figura 22 Tela com as regras geradas

A Figura 23 apresenta um exemplo de regras geradas.

```

Minimum support: 0.15
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 22
Size of set of large itemsets L(2): 23
Size of set of large itemsets L(3): 2

Best rules found:

1. org_ensino_pesquisa=1 533 ==> org_financeira=null 533  conf.(1)
2. gestao_ambiental=3 249 ==> org_financeira=null 249  conf.(1)
3. gestao_ambiental=4 215 ==> org_financeira=null 215  conf.(1)
4. org_ensino_pesquisa=1 gestao_ambiental=3 180 ==> org_financeira=null 180  conf.(1)
5. renda_per_capta=3 173 ==> org_financeira=null 173  conf.(1)
6. renda_per_capta=1 173 ==> org_financeira=null 173  conf.(1)
7. ocupacao_economica=4 173 ==> org_financeira=null 173  conf.(1)
8. ocupacao_economica=3 172 ==> org_financeira=null 172  conf.(1)
9. utilizacao_terra=3 172 ==> org_financeira=null 172  conf.(1)
10. ocupacao_economica=1 171 ==> org_financeira=null 171  conf.(1)
11. utilizacao_terra=5 171 ==> org_financeira=null 171  conf.(1)
12. renda_per_capta=5 170 ==> org_financeira=null 170  conf.(1)
13. ocupacao_economica=5 170 ==> org_financeira=null 170  conf.(1)
14. utilizacao_terra=4 170 ==> org_financeira=null 170  conf.(1)
15. utilizacao_terra=2 170 ==> org_financeira=null 170  conf.(1)

```

Figura 23 Exemplo de regras geradas

4.1.5 Interpretação dos Padrões Descobertos e Implantação do Conhecimento

O uso do algoritmo Apriori gera um grande número de regras. Por esse motivo, selecionar os padrões mais interessantes não é uma tarefa trivial. Portanto, foi necessário estabelecer um conjunto de critérios para avaliar a qualidade dos padrões da associação.

As medidas objetivas utilizadas para seleção de regras foram Suporte e Confiança, apresentadas na seção 2.3.1.1, que determinam, respectivamente, a frequência na qual uma regra é aplicável a um conjunto de dados e a frequência na qual os itens em Y aparecem em transações que contenham X.

Após a escolha das regras, foi entrevistado um especialista participante do desenvolvimento do ZEE-MG, onde cada regra foi classificada subjetivamente de acordo com a importância do conhecimento gerado:

- Conhecimento inesperado: Indica que aquele conhecimento é uma novidade para o especialista, algo que ele não havia pensado ou que contradiz um conhecimento prévio que ele tenha.
- Conhecimento útil: Indica que aquele conhecimento pode ser utilizado para tomar alguma decisão.
- Conhecimento prévio do especialista: Indica o conhecimento formado de acordo com experiências do especialista, podendo ser diferente para cada especialista.
- Conhecimento óbvio do domínio: Indica conhecimento bem consolidado no domínio. Assim, todo conhecimento óbvio do domínio é também um conhecimento prévio do especialista, porém o conhecimento óbvio, em geral, não é diferente para cada especialista.
- Conhecimento irrelevante: Indica que aquele conhecimento não é importante ou não é necessário de acordo com os objetivos definidos para o processo de mineração de dados durante a fase de identificação do problema.

Também será apresentado um mapa de Minas Gerais, onde os municípios em tom de cinza são os que se enquadram na regra gerada.

5 RESULTADOS

Este capítulo apresenta os testes e resultados obtidos durante o uso do software implementado. Foram obtidas 9 regras, sendo que cada regra possui diferentes índices envolvidos, assim como valor de suporte e confiança.

Regra 1:

Índices envolvidos:

- Carta dos recursos minerais das regiões estudadas, 2007;
- Carta das Exportações nas Regiões Estudadas, 2007.

Regra:

- recursos_minerais=1 501 ==> indice_exportacao=1 437 conf:(0.87)

Interpretação:

- Se recursos minerais = muito precário então índice de exportações = muito precário
- Suporte = 0,45 e Confiança = 0,87

Interpretação Geral:

- O município que possui condições muito precárias à exploração e extração dos recursos minerais, em 45% dos casos, são considerados muito precários em termos de exportação. E esta regra possui uma confiança alta, de 87%.

De acordo com classificação de especialista, esta regra possui conhecimento útil. E os municípios que se enquadram na regra são os indicados na Figura24.



Figura 24 Municípios enquadrados na regra 1

Regra 2:

Índices envolvidos:

- Carta do Indicador Gestão do Desenvolvimento Rural dos municípios do Estado de Minas Gerais, 2007;
- Carta do Saneamento para o Estado de Minas Gerais, ZEE-MG, 2007.

Regra:

- $gestao_rural = 3\ 249 \implies saneamento = 1\ 229$ conf:(0.92)

Interpretação:

- gestão rural = pouco favorável então saneamento = muito precário
- Suporte = 0,25 e Confiança = 0,92

Interpretação Geral:

- O município que possui condições pouco favoráveis à gestão rural, em 25% dos casos, são considerados muito precários em saneamento. E esta regra possui uma confiança alta, de 92%.

De acordo com classificação de especialista, esta regra possui conhecimento útil. E os municípios que se enquadram na regra são os indicados na Figura 25.



Figura 25 Municípios enquadrados na regra 2

Regra 3:

Índices envolvidos:

- Carta das Exportações nas Regiões Estudadas, 2007;

- Carta do Saneamento para o Estado de Minas Gerais, ZEE-MG, 2007;
- Carta do Indicador Unidades de Defesa Social dos municípios do Estado de Minas Gerais, 2007.

Regra:

- $\text{indice_exportacao}=1 \text{ saneamento}=1 \text{ } 569 \implies \text{unidade_defesa_social}=1$
568 conf:(1)

Interpretação:

- Se índice de exportacao = muito precário e saneamento = muito precário então unidade de defesa social = muito precário
- Suporte = 0,60 e Confiança = 1

Interpretação Geral:

- O município que possui condições muito precárias em relação a exportação e muito precárias em relação a saneamento, em 60% dos casos, são considerados muito precários em relação à unidade de defesa social. Esta regra possui um alto grau de confiança, 100 % dos casos onde ocorrem condições muito precárias em relação a exportação e muito precárias para saneamento ocorre também defesa social muito precária .

De acordo com classificação de especialista, esta regra possui conhecimento inesperado e conhecimento útil. E os municípios que se enquadram na regra são os indicados na Figura26.

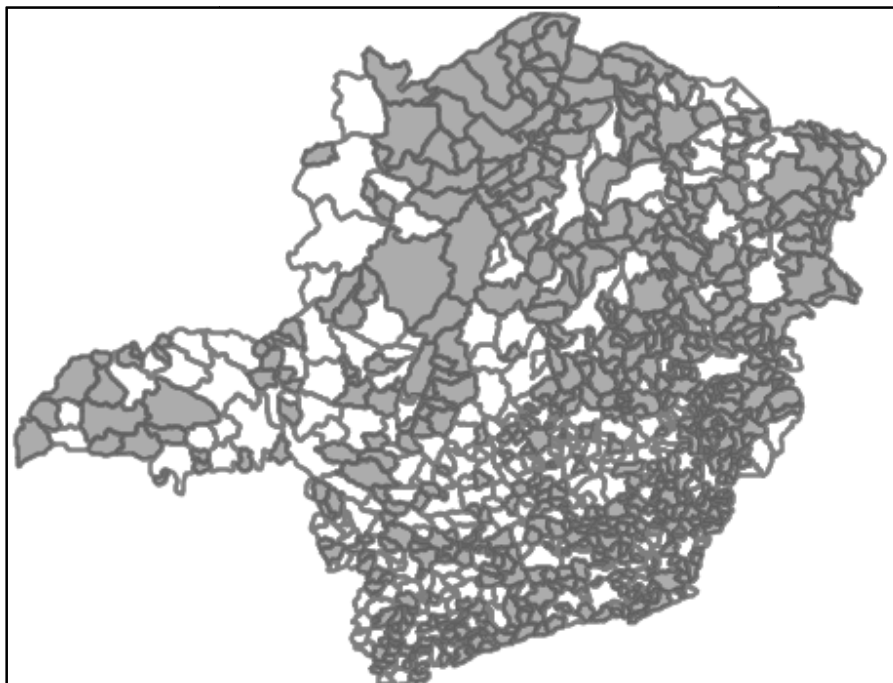


Figura 26 Municípios enquadrados na regra 3

Regra 4:

Índices envolvidos:

- Carta do Índice ICMS Ecológico das regiões estudadas, 2006;
- Carta do Indicador Unidades de Defesa Social dos municípios do Estado de Minas Gerais, 2007.

Regra:

- `indice_icms_ecologico=1 556 ==> unidade_defesa_social=1 539`
conf:(0.97)

Interpretação:

- Se índice ICMS ecológico = muito precário então unidade de defesa social = muito precário
- Suporte = 0,60 e Confiança = 0,97

Interpretação Geral:

- O município que possui condições muito precárias em relação ao índice de ICMS Ecológico, em 60% dos casos, são considerados muito precários em relação à unidade de defesa social. Isso com 97% de confiança.

De acordo com classificação de especialista, esta regra possui conhecimento útil. E os municípios que se enquadram na regra são os indicados na Figura 27.



Figura 27 Municípios enquadrados na regra 4

Regra 5:

Índices envolvidos:

- Carta da Renda per Capta para o Estado de Minas Gerais;
- Carta do Indicador Gestão Ambiental Municipal dos municípios do estado de Minas Gerais, 2007;
- Carta do Indicador Gestão Municipal dos municípios do estado de Minas Gerais, 2007.

Regra:

- $\text{renda_per_capta}=5 \text{ gestao_ambiental}=5 \text{ 74} \implies \text{gestao_municipal}=5 \text{ 59}$
conf:(0.8)

Interpretação:

- Se renda per capita = muito favorável e gestão ambiental = muito favorável então gestão municipal = muito favorável
- Suporte = 0,05 e Confiança = 0,80

Interpretação Geral;

- O município que possui condições muito favoráveis em renda per capita e muito favoráveis a gestão ambiental, em 5% dos casos, são considerados muito favoráveis em relação à gestão social. Isso com uma confiança de 80%.

De acordo com classificação de especialista, esta regra possui conhecimento prévio. E os municípios que se enquadram na regra são os indicados na Figura 28.

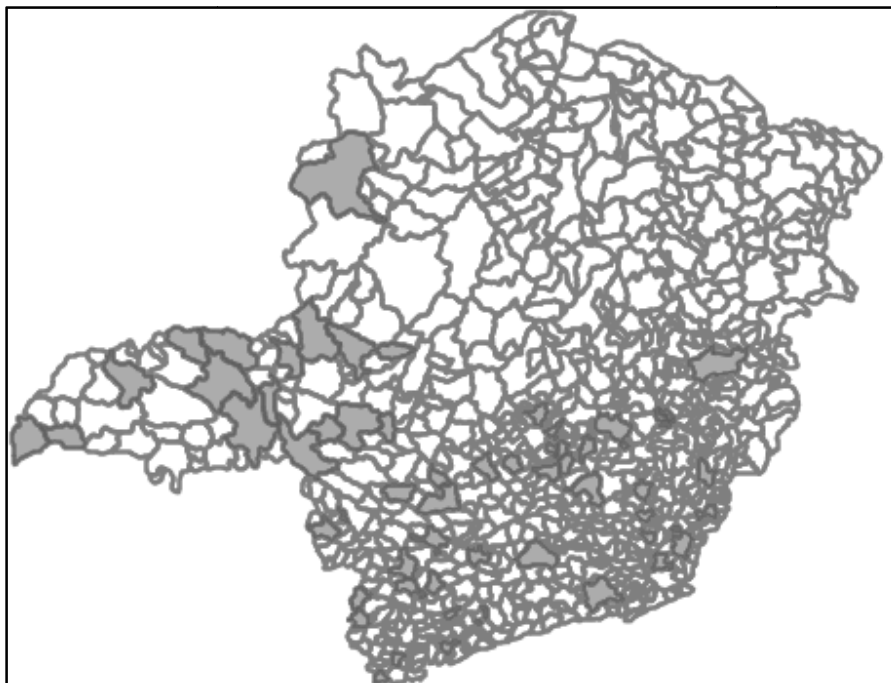


Figura 28 Municípios enquadrados na regra 5

Regra 6:

Índices envolvidos:

- Carta do Fator Condicionante Organizações de Segurança Pública dos municípios do estado de Minas Gerais, 2007;
- Carta do Fator Condicionante Organizações de Ensino e Pesquisa dos municípios do estado de Minas Gerais, 2007;
- Carta do Indicador Capacidade de Aplicação das Leis nos municípios do estado de Minas Gerais.

Regra:

- $org_seguranca=1 \quad org_ensino_pesquisa=1 \quad 146 \implies aplicacao_lei=1 \quad 146$
conf:(1)

Interpretação:

- Se organização de segurança pública = muito precária e organização de ensino e pesquisa = muito precária então aplicação da lei = muito precária
- Suporte = 0,15 e Confiança = 1

Interpretação Geral:

- O município que possui condições muito precárias a organização de segurança pública e muito precárias a organização de ensino e pesquisa, em 15% dos casos, são considerados muito precários em relação à aplicação da lei. Esta regra possui confiança de 100%, ou seja, todos os casos onde ocorrem condições muito precárias a organização de segurança pública e a organização de ensino e pesquisa, também ocorre condições precárias para pesquisa.

De acordo com classificação de especialista, esta regra possui conhecimento útil. E os municípios que se enquadram na regra são os indicados na Figura 29.

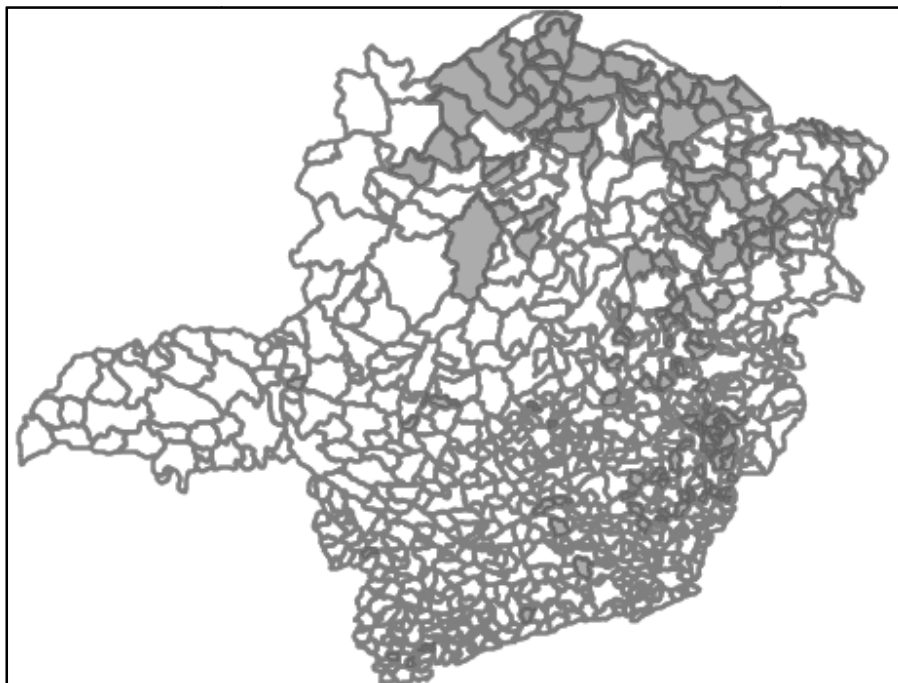


Figura 29 Municípios enquadrados na regra 6

Regra 7:

Índices envolvidos:

- Carta dos recursos minerais das regiões estudadas, 2007;
- Carta do Índice ICMS Ecológico das regiões estudadas, 2006.

Regra:

- recursos_minerais=1 501 ==> indice_icms_ecologico=1 359
conf:(0.72)

Interpretação:

- Se recursos minerais = muito precário então índice ICMS ecológico = muito precário
- Suporte = 0,40 e Confiança = 0,72

Interpretação Geral:

- O município que possui condições muito precárias a exploração e extração dos recursos minerais, em 40% dos casos, são considerados muito precários ao índice de ICMS Ecológico. Esta regra apresenta uma confiança de 72%.

De acordo com classificação de especialista, esta possui conhecimento inesperado. E os municípios que se enquadram na regra são os indicados na Figura30.



Figura 30 Municípios enquadrados na regra 7

Regra 8:

Índices envolvidos:

- Carta da Habitação para o Estado de Minas Gerais, ZEE-MG, 2007;
- Carta dos Recursos Minerais das regiões estudadas, 2007.

Regra:

- $habitacao=1\ 172 \implies recursos_minerai=1\ 142$ conf:(0.83)

Interpretação:

- Se habitação = muito precária então recursos minerais = muito precário
- Suporte = 0,15 e Confiança = 0,83

Interpretação Geral:

- O município que possui condições muito precárias para habitação, em 15% dos casos, são considerados muito precários a exploração e extração dos recursos minerais. Esta regra possui uma confiança de 83%.

De acordo com classificação de especialista, esta regra possui conhecimento inesperado. E os municípios que se enquadram na regra são os indicados na Figura31.



Figura 31 Municípios enquadados na regra 8

Regra 9:

Índices envolvidos:

- Carta do Fator Condicionante Organizações de Ensino e Pesquisa dos municípios do estado de Minas Gerais;
- Carta do Indicador de Pós-Graduação e Pesquisa dos municípios do estado de Minas Gerais, 2007.

Regra:

- 1. org_ensino_pesquisa=1 533 ==> org_pos_graduacao=1 533
conf:(1)

Interpretação:

- Se organização de ensino e pesquisa = muito precário então organização de pós-graduação = muito precário
- Suporte = 0,50 e Confiança = 1

Interpretação Geral:

- O município que possui condições muito precárias em relação a organizações de ensino e pesquisa, em 50% dos casos, são considerados muito precários para organização de pós-graduação. E esta regra possui uma grande confiança, de 100%, que indica que todos os casos onde ocorreram ensino e pesquisa muito precárias, ocorreram também pós graduação muito precária.

De acordo com classificação de especialista, esta regra possui conhecimento óbvio. E os municípios que se enquadram na regra são os indicados na Figura32.



Figura 32 Municípios enquadrados na regra 9

6 CONCLUSÃO

Com a finalização do trabalho e observando os objetivos propostos e os resultados alcançados, pode se concluir que a aplicação de mineração de dados no banco de dados do Zoneamento Ecológico Econômico de Minas Gerais é uma boa solução para extração de padrões não triviais.

O objetivo de estudar e identificar técnicas de mineração de dados aplicáveis a descoberta de conhecimento na base de dados utilizada, foi obtido através da técnica de análise de associação. Esta técnica foi aplicada e adaptada de acordo com a base de dados, então as regras geradas foram analisadas através de medidas objetivas e subjetivas.

Para obtenção dos resultados todas as etapas de KDD devem ser realizadas com devida importância, sendo que a etapa de Limpeza e Pré-processamento, para este trabalho, teve grande influência no resultado final pelo pré-processamento realizado através da extensão PostGIS do sistema gerenciador de banco de dados PostgreSQL.

Portanto, é possível evidenciar que técnicas de mineração de dados, que associam dados socioeconômicos a dados espaciais podem contribuir no apoio à tomada de decisão no planejamento e elaboração de programas sociais e ambientais favorecendo, assim a melhoria da qualidade de vida da população.

Trabalho futuros poderão utilizar outras técnicas de mineração de dados no banco de dados do Zoneamento Ecológico Econômico de Minas Gerais para extrair outros tipos de informações relevantes. Esta metodologia empregada no trabalho também poderá ser implementada através de plataforma WEB para maior disponibilidade desta ferramenta.

7 REFERÊNCIAS BIBLIOGRÁFICAS

BERRY, M. J. A; LINOFF, G. S. **Data Mining Techniques**. Wiley, Rio de Janeiro, 2004.

BORGONY, V. **Algoritmos e Ferramentas de Descoberta de Conhecimento em Bancos de Dados Geográficos**. PPGC da UFRGS, 2003.

BRAGA, L. P. V. **Introdução à Mineração de Dados**. E-papers, Rio de Janeiro, 2005.

CÂMARA, G.; CASANOVA, M.; DAVIS, C.; VINHAS, L.; QUEIROZ, G. R. **Bancos de Dados Geográficos**. MundoGEO, 2005.

CÂMARA, G.; DAVIS, C.; MONTEIRO, A. M. V. **Introdução à Ciência da GeoInformação**. INPE/DPI, São José dos Campos, 2001.

CHOA, Y. H.; KIMB, J. K.; KIMA, S. H.. **Personalized Recommender System Based on Web Usage Mining and Decision Tree Induction**. Expert Systems with Applications. 2002.

DATE, C. J. **Introdução à Sistemas de Banco de Dados**, Elsevier, 2003.

DEITEL, H. M.; DEITEL, P. J. **Java: Como programar**. 4. ed. [S.l.]: Bookman, 2003.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de Banco de Dados**. Pearson Education, 2005.

ESTER, M.; KRIEGEL, H. P.; SANDER, J. Algorithms and Applications for Spatial Data Mining. In: MILLER, H. J.; HAN, J. **Geographic Data Mining and Knowledge Discovery**, 2001.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: **Proceeding of The Second Int. Conference on Knowledge Discovery and Data Mining**, p. 82-88, 1996.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH P.; UTHURUSAMY, R. **Advances in knowledge Discovery & Data Mining**. AAAI/MIT, 1996.

GUTING, R. H. An Introduction to Spatial Database Systems. **The International Journal on Very Large Data Bases**, [S.1], V.3, N.4, P.357 – 399, Oct. 1994.

HAN, J.; KOPERSKI, K.; STEFANOVIC, N. GeoMiner: A System Prototype for Spatial Data Minig. In: **ACM-SIGMOD Int'l Conf. on Management of Data**, 1997.

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. Morgan Kaufmann, 2000.

HAWKINS, D. **Identification of Outliers**. Chapman and Hall, 1980.

MITRA, S.; ACHARYA, T. **Data Mining: Multimedia, Soft Computing and Bioinformatics**. John Wiley & Sons, 2003.

MILLER, H. J.; HAN, J. **Geographic data mining and knowledge discovery: An overview**. Taylor and Francis, 2001.

NG, R. T.; HAN, J. CLARANS: A Method for Clustering Objects for Spatial Data Mining, **IEEE Transactions on Knowledge and Data Engineering**, v. 14, n. 5, Set/Out, 2002.

PIVATO, M. A. **Mineração de regras de associação em dados georreferenciados**. Universidade de São Paulo - São Carlos, 2006.

SCOLFORO, J. R. S.; OLIVEIRA, A. D.; CARVALHO, L. M. T. **Zoneamento ecológico-econômico do Estado de Minas Gerais: componente socioeconômico**. Editora UFLA, 2008.

TAN, P., STEINBACH, M., KUMAR, V. **Introdução ao Data Mining**. Ciência Moderna, 2009.

WITTEN, I. H., FRANK, E. **Data Mining: practical machine learning tools and techniques**. Elsevier, 2005.