



**Análise do Banco de Atendimentos da Defesa Civil do Estado de São Paulo**  
Analysis of São Paulo Civil Defense Response Database

Diego José Chagas<sup>1</sup>; Chou Sin Chan<sup>1</sup> & Alessandra Cristina Corsi<sup>2</sup>

<sup>1</sup> Centro de Previsão de Tempo e Estudos Climáticos /Instituto Nacional de Pesquisas Espaciais,  
Rodovia Presidente Dutra, km 39 - 12630-000 Cachoeira Paulista, SP, Brasil.

<sup>2</sup> Instituto de Pesquisas Tecnológicas, Av. Prof. Almeida Prado, 532, Cidade Universitária. 05508-901 São Paulo, SP, Brasil  
E-mails: [diego.chagas@cptec.inpe.br](mailto:diego.chagas@cptec.inpe.br); [chou.sinchan@cptec.inpe.br](mailto:chou.sinchan@cptec.inpe.br); [accorsi@ipt.br](mailto:accorsi@ipt.br)

Recebido em: 09/10/2009      Aprovado em: 08/05/2010

## Resumo

Nos últimos anos pode-se observar que a simples organização de dados já não representa um fator diferencial para as instituições, uma vez que, dependendo do seu volume, o método tradicional de análise e interpretação se torna extremamente lento e custoso. O uso de técnicas de mineração de dados surge como alternativa para tornar esse processo semi-automático. O presente trabalho tem por objetivo realizar um estudo de caso por meio de técnica de mineração de dados, baseado no programa WEKA aplicado em dados hidrometeorológicos e geomorfológicos coletados na região da Serra do Mar do Estado de São Paulo. Resultados obtidos a partir da aplicação da técnica de associação indicam que a presença de rocha e matacão em terrenos com cicatrizes e declividade alta são fatores relevantes para a ocorrência de deslizamentos.

**Palavras-chave:** Mineração de Dados; WEKA; Análise semi-automática

## Abstract

In recent years the simple data organization is no longer a differential factor for institutions, since, depending on their volume, the traditional method of analysis and interpretation is extremely slow and costly. The use of data mining techniques is an alternative to allow this process semi-automatic. The objective of this work is to carry out a case study of data mining technique based on the WEKA software applied to hydrometeorological and geomorphological data which were collected in the Serra do Mar region of São Paulo State. Results obtained from the application of the association technique indicate that the presence of rock and boulders at terrains with scars and high declivity are relevant factors for the landslide occurrence.

**Keywords:** Data Mining; WEKA; Semi-automatic analysis

## 1 Introdução

A região da Serra do Mar do Estado de São Paulo (Figura 1) é de grande importância para o Estado de São Paulo não só para desenvolvimento sustentável, por conter as porções remanescentes da Mata Atlântica, mas também pela participação no desenvolvimento econômico favorecido pelas ferrovias, dutovias e instalações industriais e portuárias. Porém, existe um problema recorrente na região que são os constantes deslizamentos de encostas que acarreta em grandes prejuízos e mortes (s.n. Projeto Serra do Mar, 2005).

O Plano Preventivo de Defesa Civil (PPDC) da Secretaria de Meio Ambiente do Estado de São Paulo foi formulado a partir de conhecimentos geológicos sobre os escorregamentos na região da Serra do Mar Paulista. O PPDC vem sendo operado continuamente no período chuvoso desde 1988/1989 nos municípios da Baixada Santista e Litoral Norte (Macedo *et al.*, 2005). O Plano utiliza como parâmetros operacionais: índice pluviométrico, previsões meteorológicas e vistoria de campo. O Banco de Dados contém os dados de vistoria de campo obtidos durante a operação do PPDC.

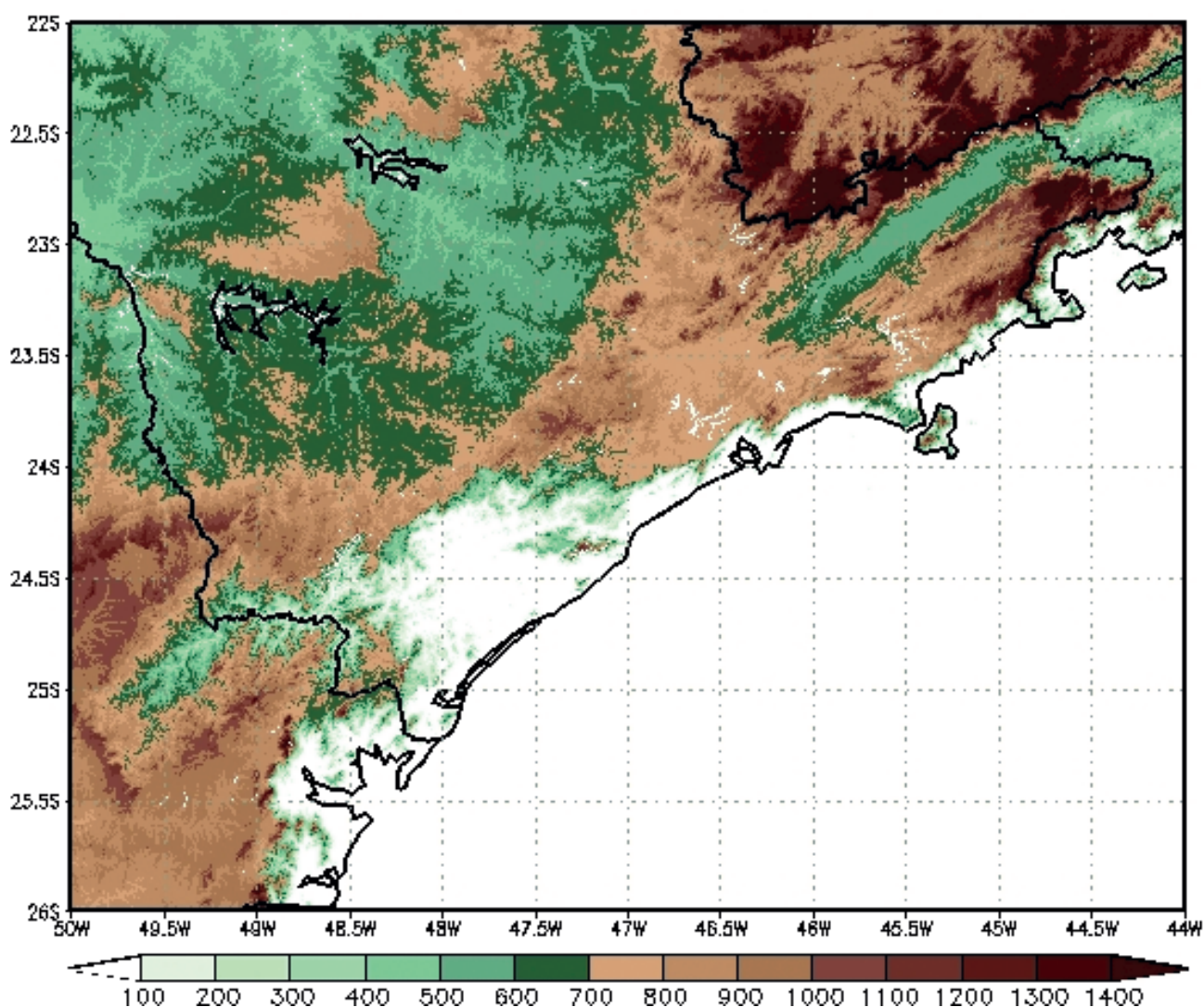


Figura 1 Topografia da Região da Serra do Mar do Estado de São Paulo.

O método tradicional de análise e interpretação dos dados pode ser uma tarefa artesanal. Por exemplo, a análise de um determinado evento de escorregamento de terra, gera relatórios com informações que podem ser utilizadas no gerenciamento do risco do desastre. Usualmente, esta forma de análise é lenta e subjetiva, o que dificulta os estudos em virtude do aumento da quantidade de informações a serem analisadas. À medida que o volume de dados aumenta, a capacidade de análise decresce. Assim, observa-se a necessidade de que o processo de análise seja parcialmente automatizado, ou seja, uma tarefa executada pelo usuário com o auxílio de ferramentas de análise de dados. Para isso, a mineração de dados apresenta-se como uma tecnologia que compõe etapas para extrair conhecimento de uma base de dados.

O presente trabalho tem por objetivo aplicar a técnica de mineração de dados baseado no aplicativo *Waikato Environment for Knowledge Analysis* (WEKA) sobre o banco de dados obtidos de operações do PPDC. Serão apresentadas as etapas do processo e a aplicação do algoritmo, a partir de

um estudo de caso com dados históricos coletados na região da Serra do Mar do Estado de São Paulo.

Este trabalho está organizado da seguinte maneira: a seção 2 contém a descrição da metodologia, a descrição da descoberta de conhecimento em bancos de dados e o algoritmo de mineração de dados empregado. Na seção 3 é apresentada a descrição do banco de dados de atendimentos, na seqüência, nas seções 4 e 5, os resultados e a conclusão respectivamente.

## 2 Metodologia

Os dados selecionados são oriundos do Banco de Dados de Atendimentos que contém informações geológico-geotécnicas e de uso e ocupação dos locais, gerado pelo Instituto de Pesquisas Tecnológicas (IPT) de São Paulo no período compreendido entre 1992 e 2007. Por isso, informações de diversas cidades do estado de São Paulo estão contidas neste Banco de Dados (BD) (Figura 2).

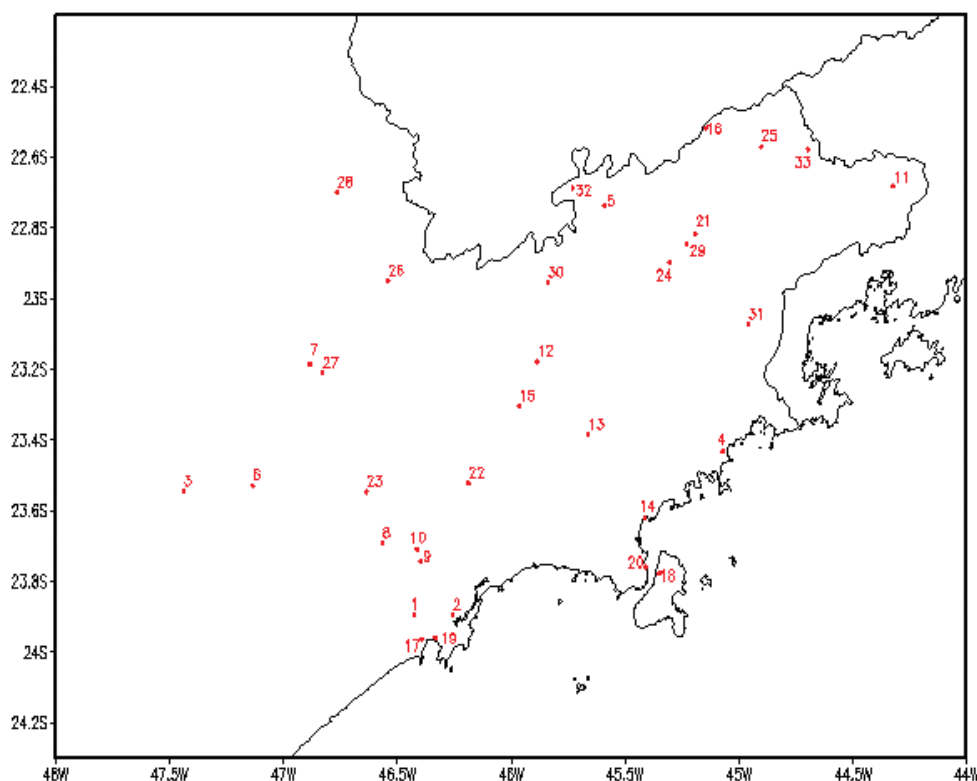


Figura 2 Localização das cidades estudadas neste trabalho. Os nomes das cidades são listados na Tabela 1.

Neste trabalho serão utilizados dados de algumas cidades que se encontram na região leste do estado. A Tabela 1 indica a posição destas cidades com as informações da latitude e longitude.

cód	cidade	latitude °S	longitude °W
1 -	Cubatão	-23.895	-46.425
2 -	Guarujá	-23.895	-46.256
3 -	Votorantim	-23.546	-47.437
4 -	Ubatuba	-23.433	-45.071
5 -	Campos do Jordão	-22.739	-45.591
6 -	São Roque	-23.529	-47.135
7 -	Jundiá	-23.186	-46.884
8 -	São Bernardo do Campo	-23.693	-46.565
9 -	Rio Grande da Serra	-23.744	-46.398
10 -	Ribeirão Pires	-23.710	-46.413
11 -	Bananal	-22.683	-44.323
12 -	São José dos Campos	-23.179	-45.886
13 -	Paraibuna	-23.386	-45.662
14 -	Caraguatatuba	-23.620	-45.413
15 -	Jacareí	-23.305	-45.965
16 -	Piquete	-22.520	-45.146
17 -	São Vicente	-23.963	-46.391
18 -	Ilhabela	-23.778	-45.358
19 -	Santos	-23.960	-46.333
20 -	São Sebastião	-23.760	-45.409
21 -	Guaratinguetá	-22.816	-45.192
22 -	Mogi das Cruzes	-23.522	-46.188
23 -	São Paulo	-23.547	-46.636
24 -	Roseira	-22.898	-45.305
25 -	Lavrinhas	-22.570	-44.902
26 -	Bragança Paulista	-22.951	-46.541
27 -	Várzea Paulista	-23.211	-46.828
28 -	Amparo	-22.701	-46.764
29 -	Aparecida	-22.846	-45.229
30 -	Monteiro Lobato	-22.956	-45.839
31 -	Cunha	-23.074	-44.959
32 -	São Bento do Sapucaí	-22.688	-45.730
33 -	Areias	-22.579	-44.696

Tabela 1 Latitude, longitude e código das cidades indicadas na Figura 1.

A seleção dos dados faz parte de uma etapa do processo de mineração de dados que está contemplada na técnica de Descoberta de Conhecimento em Banco de Dados. Todo o trabalho foi desenvolvido com base nesta técnica que será apresentada na próxima seção.

## 2.1 Descoberta de Conhecimento em Bancos de Dados

A técnica de Descoberta de Conhecimento em Bancos de Dados, popularmente conhecida como *Knowledge Discovery in Databases (KDD)*, se refere a um conjunto de etapas que tenta extrair conhecimento a partir de uma base de dados. Segundo Fayyad *et al.* (1996 a), *KDD* é o processo não trivial de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis.

Um processo de *KDD* consiste nas seguintes etapas: seleção, pré-processamento, transformação, mineração de dados e análise (Fayyad *et al.*, 1996 a, b) consecutivamente (Figura 3).

Na etapa de seleção dos dados são identificadas as informações, dentre os conjuntos de dados existentes, que devem ser consideradas durante o processo de *KDD*. Para uma tarefa mais objetiva, é necessário selecionar variáveis de interesse. As variáveis podem ser do tipo qualitativa ou quantitativa. Variáveis qualitativas ou categóricas são finitas, podem ser nominais (tipo de solo, tipo de vegetação, ocorrência de deslizamento, etc.) ou ordinais, neste caso existe uma ordenação entre as categorias (temperatura baixa, média, alta, chuva

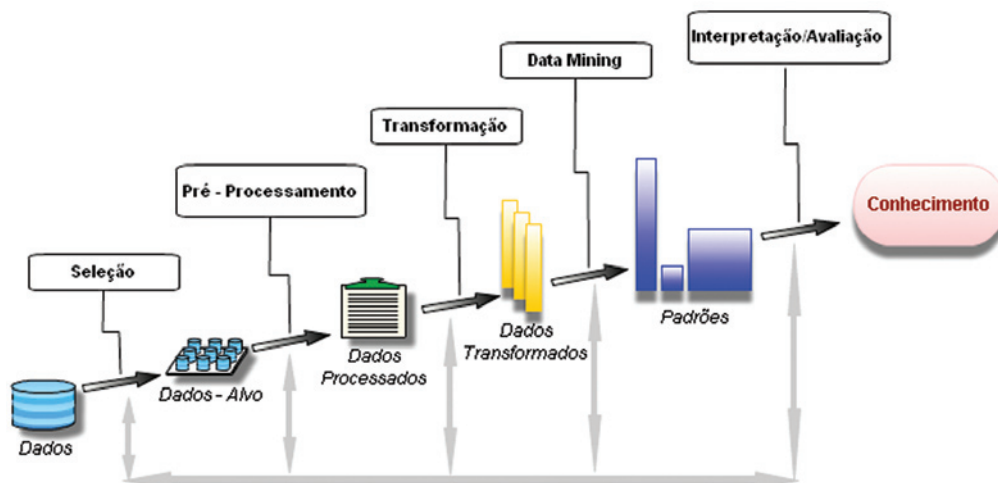


Figura 3  
Visão geral das etapas do processo de KDD.

fraca, intensa, moderada ou ordem de classificação de acordo com um requisito pré-estabelecido). Variáveis quantitativas assumem valores numéricos e podem ser do tipo contínua, onde os valores obedecem a uma escala contínua (tais como temperatura, pressão atmosférica e umidade) ou discreta, que pode assumir um número finito, como o número de ocorrências de chuva, quantidade de precipitação, durante um período em uma cidade, por exemplo.

Depois de selecionados, os dados precisam passar por um pré-processamento para tornarem-se compreensíveis e limpos, ou seja, dados completos, sem repetição de registros e ruídos. Essa tarefa é um requisito essencial em mineração de dados e visa qualificar os dados. O objetivo principal da etapa de pré-processamento é preparar os dados para os algoritmos de mineração que serão aplicados. Os problemas mais comuns encontrados e que devem receber tratamento são: dados ausentes e/ou incompletos, registros repetidos, ruídos, dentre outros fatores que podem interferir na integridade e qualidade dos dados.

Depois de pré-processados os dados estão qualificados, porém uma aplicação de algoritmos de mineração de dados exige um formato padrão. Este formato geralmente varia de acordo com o software adotado e o algoritmo que será aplicado. Por esse motivo os dados pré-processados passam pela etapa de transformação para um formato específico que será exigido pela etapa de mineração de dados.

Posteriormente à etapa de transformação dos dados, é realizada a etapa de mineração de dados, que é definida como o processo de descoberta de padrões em dados. O processo deve ser automático ou, mais usualmente, semi-automático. Os padrões descobertos devem ser significantes, oferecendo alguma vantagem, usualmente uma vantagem econômica (Witten & Frank, 2000).

Por fim deve haver uma interpretação dos resultados para gerar conhecimento através do padrão encontrado. Se na etapa de interpretação dos resultados o analista concluir que o padrão encontrado não possui significado ou não é satisfatório, deve se retornar a uma das etapas anteriores para refazer o processo de KDD (Fayyad *et al.*, 1996 a).

Finalmente, o produto encontrado deve ser o conhecimento útil, que pode ser aplicado

na organização a fim de aperfeiçoar processos ou serviços. Além disso, aconselha-se fazer uma documentação sobre os resultados que podem ser usados futuramente para reanálise (Fayyad *et al.*, 1996 a).

## 2.2 Algoritmo Apriori

O algoritmo *Apriori* foi aplicado com o intuito de analisar as informações referentes às características físicas dos locais onde ocorreram os deslizamentos. Este algoritmo permite identificar os itens nos dados que tem uma inter-relação, portanto, espera-se que possam ser identificados os tipos de características do terreno que têm correlação maior com deslizamento de terra. O *Apriori* permite que se escolha dentre outras informações, o número de regras que serão mostradas e o grau de confiabilidade mínima para a formação das regras. Neste caso, o algoritmo foi executado para gerar um conjunto formado por 5 regras de associação e com o grau de confiabilidade mínimo de 90%. O algoritmo inicia o processo de busca pelas associações identificando todas as regras que satisfazem o fator de suporte mínimo. Este fator é calculado da seguinte maneira: o fator de suporte de  $X$  é igual ao número de registros que contem  $X$ , dividido pelo número total de registros. Em seguida, identifica os itens mais frequentes gerando regras de associação que satisfazem o fator de confiança mínimo (Goldschmidt & Passos, 2005).

## 2.3 Waikato Environment For Knowledge Analysis (WEKA)

O Waikato Environment for Knowledge Analysis (WEKA) contempla um vasto conjunto de algoritmos de preparação de dados, aprendizagem de máquina e de validação de resultados (*J48*, *SimpleKmeans*, *Apriori*, *Naive Bayes*, dentre outros) (Witten & Frank, 2000). Nele são encontradas ferramentas para: pré-processamento dos dados, classificação, regressão, agrupamento, regras de associação e visualização.

O WEKA contempla diversas técnicas de mineração de dados e pode ser instalado em diferentes plataformas como Windows, Mac OS X e Linux. Desenvolvido na Universidade de Waikato (Nova Zelândia), o WEKA foi escrito em linguagem de programação Java, sendo distribuído livremente

por meio da Internet e pode ser modificado e redistribuído de acordo com os termos da GNU (GNU is Not Unix) General Public License. A confecção de aplicações com WEKA depende do ambiente de desenvolvimento Java (J2SE, Java 2 Standard Edition), instalado no computador onde se deseja executar a aplicação.

#### 2.4 Técnica de Mineração de Dados: Associação

A técnica de mineração por Associação é um processo que tenta expor as características e as tendências, ou seja, tenta determinar o relacionamento entre um conjunto de itens. A vantagem em relação às outras técnicas é simplicidade no uso.

Essa técnica utiliza dois fatores importantes: fator de suporte e fator de confiança. O fator de suporte indica a ocorrência relativa da regra de associação observada dentro de um conjunto de dados. O fator de confiança é o grau com o qual a regra é verdadeira entre os registros individuais (Diniz & Louzada-Neto, 2000). As características são apresentadas pelas afinidades entre elas e exibem, em uma estrutura de texto, os itens correlacionados.

### 3 Descrição do Banco de Atendimentos

Em casos de deslizamento, quando solicitado pela Defesa Civil Estadual, o IPT realiza vistoria na região onde ocorreu o evento detalhando as características do local (tipo de encosta, declividade, cortes no terreno, presença de vegetação, árvores ou postes inclinados, presença de blocos de rocha, lançamento de água, dentre outras). Além disso, verifica as informações dos pluviômetros que estão instalados nas regiões e combina com outros dados.

O Banco de Dados de Atendimentos é constituído por 3 tabelas: Atendimento Técnico, Informe Técnico e Vistoria Técnica. A relação (tabela) Atendimento Técnico é composta pelas seguintes informações (campos): código, número, município, data, referência, recomendações e equipe técnica. Estas informações dizem respeito às equipes técnicas que realizaram o atendimento indicando localidade, data e recomendações para as regiões de risco. Dentre as recomendações, se destacam as medidas preventivas.

A tabela Informe Técnico contém as informações: código, número, município, data,

referência, níveis de alerta (observação, atenção, alerta, alerta máximo), acumulado de chuva em 3 dias, Coeficiente de Ciclo Móvel (CCM), Coeficiente de Precipitação Crítica Efetiva (CPCef), Coeficiente de Precipitação Crítica Potencial (CPPot) (Tatizana *et al.*, 1987 a, b), previsão meteorológica, restauração dos sistemas de drenagem e restauração das vias de acesso, mudança de nível de “alerta”, de alerta máximo, de alerta, de atenção, de observação, para alerta máximo, para alerta, para atenção, para observação, outras recomendações e equipe técnica. Esses atributos informam os níveis de “alerta” para cada registro e indica a necessidade de alteração do mesmo em função da vistoria de campo e da previsão meteorológica.

Na descrição acima, o Coeficiente de Ciclo Móvel (CCM) se refere a um índice que avalia a chuva por um período de tempo com o objetivo de prever condições que podem favorecer o escorregamento do terreno. O CCM é calculado através do acumulado de chuva de junho até o dia 1º de maio do ano subsequente dividido pelo acumulado normal de chuva no mesmo período. Em seguida, o Coeficiente de Precipitação Crítica Efetiva (CPCef) é calculado a partir de índices de chuva registrados nos postos pluviométricos e o Coeficiente de Precipitação Crítica Potencial (CPCPot) é calculado a partir de índice de chuva estimado com base na previsão meteorológica e índices de chuva registrados nos postos pluviométricos. Estes índices são utilizados para monitorar as áreas de riscos de corridas de detritos. Na tabela Informe Técnico os valores destes índices estavam incompletos em grande parte do conjunto de dados, por isso, apesar de serem importantes, não foram considerados. Estes índices são aplicados a setores específicos no município de Cubatão, na área do Pólo Industrial.

A relação Vistoria Técnica contém os campos código, número, data, localização, nome do morador, condições de acesso a área, Características do local (talude natural, talude de corte, distância da moradia a base do talude, aterro compactado, aterro lançado, distância da moradia ao topo do aterro, alta declividade, média declividade, baixa declividade, estrutura em solo favoráveis a segurança, estrutura em solo desfavoráveis a segurança, presença de blocos de rocha e matacão, presença de paredões rochosos, presença de lixo), Água (concentração de água de chuva em superfície, lançamento de água servida em superfície, presença de fossas, presença

de rede de esgoto, presença de rede de água, surgências de água), Evidências de movimentação (trincas no terreno, trincas na moradia, inclinação de árvores/postes/muros, degraus de abatimento, cicatrizes de escorregamentos, feições erosivas, muros embarrigados), Caracterização em caso de escorregamento (volume mobilizado m<sup>3</sup>, alcance a partir da base, data e horário da ocorrência), Vegetação no talude ou proximidade (presença de árvores, vegetação rasteira, área desmatada, área de cultivo), Tipo de moradia (alvenaria, madeira, misto), Tipo de movimentação ocorrida ou esperada (escorregamento natural, escorregamento de corte, escorregamento de aterro, escorregamento superficial, queda de blocos, rolamento de matacão, deslocamento rochoso, corridas, solo/rocha, inundação), Grau de risco (iminente/providência imediata, não iminente/manter o local em observação), número de moradias em risco, estimativa do número de pessoas para remoção e equipe técnica.

Como se pode observar, a relação Vistoria Técnica é a mais detalhada do BD pois contém informações sobre as características do terreno onde ocorreu o escorregamento. As informações apresentam uma caracterização geológico-geotécnica da área vistoriada.

## 4 Resultados

### 4.1 Seleção dos Dados do Banco de Atendimentos

Para a realização do estudo de caso foram extraídas inicialmente as variáveis que podem ser úteis para descrever as características de escorregamentos na região da Serra do Mar.

Para tanto foram selecionados os seguintes campos da tabela Vistoria Técnica: aterro compactado, aterro lançado, alta declividade, média declividade, baixa declividade, presença de blocos de rocha e matacão, presença de paredões rochosos, presença de lixo, surgências de água, trincas no terreno, cicatrizes de escorregamentos e vegetação rasteira, totalizando 306 registros.

### 4.2 Pré-processamento dos Dados

Neste estudo de caso, a etapa de pré-processamento correspondeu à exportação de dados, originalmente, em tabelas do Access para o editor de

planilhas Excel. Após este procedimento, os arquivos foram editados para eliminar as colunas que não faziam parte do conjunto de dados selecionados para o domínio da aplicação.

O problema mais comum encontrado nas tabelas foram registros com dados incompletos que tiveram que ser eliminados. Na tabela Vistoria Técnica os dados selecionados tinham valores do tipo *booleans* (verdadeiro ou falso) que foram substituídos por sim e não respectivamente, indicando ou não a presença de certos tipos de características no terreno onde ocorreu o escorregamento.

Nos campos da relação Vistoria Técnica, os campos com valores correspondentes a “não” foram substituídos pelo ponto de interrogação “?”, que representa valores desconhecidos. Essa medida foi tomada para que o algoritmo interprete somente os valores correspondentes a “sim”, que indica a presença de determinadas características físicas nos locais onde ocorreram os escorregamentos. Desta maneira, na execução do algoritmo o processo buscará os relacionamentos mais fortes que existem entre as variáveis considerando somente o valor “sim”. Isso implicará que, quando o algoritmo encontrar uma característica específica (Trincas no Terreno, por exemplo) ele buscará quais outros atributos estão relacionados àquela característica.

### 4.3 Transformação dos Dados

Depois de pré-processados, os dados passaram pela etapa de transformação. No caso do software WEKA é a transformação dos dados para o formato *Attribute-Relation File Format* (ARFF). Os dados foram armazenados em um arquivo texto composto por três partes. A primeira parte é chamada de Relação, que é a linha inicial do arquivo deve iniciar com a expressão @relation seguida de uma palavra chave que identifique a relação.

A segunda parte é chamada de Atributos, formada por um conjunto de linhas que devem iniciar com @attribute seguida do nome de identificação do atributo e seu tipo. Os tipos podem ser nominais (as alternativas devem aparecer separadas por vírgulas e delimitadas por chaves {}), numéricos (onde o nome do atributo deve ser seguido pela palavra-chave REAL) ou string (uma longa lista arbitrária de caracteres).

Por fim a terceira parte dos Dados, inicia-se o conjunto com @data onde cada linha corresponde a uma instância e deve ter valores separados por vírgula, a declaração de cada dado deve corresponder aos atributos anteriormente especificados.

#### 4.4 Escolha de algoritmos de mineração de dados

Para escolher o melhor algoritmo, foram feitos alguns testes em amostras do conjunto de dados (28 registros da tabela), a partir da execução de diferentes algoritmos de associação de dados, que é uma técnica de mineração de dados. Esta técnica procura uma relação entre os dados, por isso é a mais indicada para este estudo de caso. Para este tipo de tarefa o software WEKA dispõe de 3 algoritmos: Apriori, Predictive Apriori e Tertius. Dentre eles o *Apriori* se destaca, pois além de ser amplamente utilizado, percebeu-se que sua saída é simplificada e por isso mais fácil de ser interpretado.

#### 4.5 Análise dos resultados gerados pelo algoritmo Apriori

Após a execução de todos os passos anteriores, é feita a análise e interpretação dos resultados obtidos pela aplicação dos algoritmos. Nesta análise os atributos da tabela Vistoria Técnica utilizados foram: Aterro Compactado, Aterro Lançado, Alta Declividade, Média Declividade, Baixa Declividade, Rocha/Matacão, Paredões Rochosos, Lixo/Entulho, Surgências de Água, Trincas no Terreno, Cicatrizes de escorregamento e Vegetação Rasteira.

Regra	Formação da Regra	Nº de casos	Implica em /Nºcasos	Confiança
1	Rocha/Matacão + Cicatriz de Escorregamento	64	Alta Declividade/62	0.97
2	Aterro Lançado + Trincas no Terreno	54	Alta Declividade/50	0.93
3	Rocha/Matacão + Vegetação Rasteira	54	Alta Declividade/50	0.93
4	Aterro Lançado + Lixo/Entulho	51	Alta Declividade/47	0.92
5	Aterro Lançado + Cicatriz de Escorregamento	92	Alta Declividade/84	0.91

Tabela 2 Análise dos Dados da Vistoria Técnica. Esta tabela exhibe em ordem crescente o número da regra, os atributos que se relacionam entre si, o número de casos que essa relação foi encontrada, número de casos que implica em declividade alta e por fim o fator de confiança da regra.

A Tabela 2 mostra o resultado da execução do algoritmo *apriori* nos dados. Neste caso serão consideradas para análises todas as regras que incluem o índice de confiabilidade e número de instâncias associadas a cada uma delas. As regras são apresentadas em ordem de acordo com o fator de confiança. A Tabela 2 apresenta o resultado da aplicação do algoritmo nos dados.

Todas as regras implicam em declividade alta, essa característica é predominante nos dados de origem. Dos 306 registros, 247 são classificados como declividade alta, 7 registros como declividade média e 12 registros como declividade baixa. Isso explica porque todas as regras implicaram em declividade alta, sendo assim é um fator de grande contribuição para a ocorrência de deslizamentos de encostas. Abaixo é descrita uma análise detalhada para cada regra.

A primeira regra apresentou 97% de confiança e a associação feita pelo algoritmo indica que em 64 casos existe uma relação entre a presença de rocha e matacão com cicatrizes de escorregamento.

Nestes locais foram contabilizados 62 registros com declividade alta.

Analisando a segunda regra verifica-se que ela apresentou 93% de confiança e a associação feita pelo algoritmo indica que em 54 casos existe uma relação entre a presença de aterro lançado com trincas no terreno e cicatrizes de escorregamento. Nestes locais foram contabilizados 50 registros com declividade alta.

Assim como a segunda, a terceira regra apresentou 93% de confiança e a associação feita pelo algoritmo indica que em 54 casos existe uma relação entre a presença de rocha e matacão com vegetação rasteira. Nestes locais foram contabilizados 50 registros com declividade alta.

A quarta regra apresentou 92% de confiança e a associação feita pelo algoritmo indica que em 51 casos existe uma relação entre a presença de aterro lançado juntamente com lixo e entulho. Nestes locais foram contabilizados 47 registros com declividade alta.

A última regra encontrada possui 91% de confiança e a associação feita pelo algoritmo indica que em 92 casos existe uma relação entre a presença



de aterro lançado com cicatrizes de escorregamento. Nestes locais foram contabilizados 84 registros com declividade alta.

A base de dados fornecida para objeto de estudo do algoritmo é constituída de 12 atributos, porém as regras foram construídas com 7 atributos: rocha e matacão, cicatrizes de escorregamento, alta declividade, aterro lançado, trincas no terreno, vegetação rasteira e lixo/entulho. Sendo assim, algumas características não participaram da construção das regras que são: aterro compactado, media declividade, baixa declividade, paredões rochosos e surgências de água.

## 5 Conclusões

Existe uma diversidade de técnicas de mineração de dados, por isso muitos problemas podem ser tratados através de soluções que buscam encontrar padrões em dados. Neste trabalho foi aplicada a técnica de associação, representada pelo algoritmo *Apriori*. A partir da aplicação deste algoritmo foi possível estudar diferentes situações nos dados em estudo.

As técnicas de mineração de dados foram exploradas com o auxílio do aplicativo WEKA. Observou-se que o aplicativo apresenta interface amigável e intuitiva, bem como uma vasta documentação com explicações para o uso dos diferentes pacotes (filtros, algoritmos, dentre outros) disponíveis.

O estudo de caso foi realizado com a finalidade de organizar o Banco de Atendimentos do IPT. Os resultados obtidos com a aplicação do algoritmo *Apriori* permitiram concluir que existem diversas características que estão relacionadas com a ocorrência de escorregamentos de encostas. Analisando todos os atributos que participaram da formação das regras verifica-se que os atributos rocha e matacão, cicatrizes de escorregamento e aterro lançado influenciara consideravelmente nos casos de escorregamento, pois são freqüentemente encontrados na formação das principais regras. As regras 2, 4, e 5 mostram a influência da ação antropogênica nesse locais onde são encontrados a presença de aterro lançado e lixo/entulho. Já os atributos aterro compactado, mdia declividade, baixa declividade, paredões rochosos e surgências de água que não foram representados em nenhuma regra

sugerem que eles não são fatores determinantes na maioria dos casos de escorregamento.

A partir desse trabalho, pretende-se estender a utilização do WEKA incluindo atributos meteorológicos ao banco de dados.

## 6 Referências

- Diniz, C. A. R. & Louzada-Neto, F. 2000. *Data Mining: Uma introdução*. São Paulo. Associação Brasileira de Estatística. 134p.
- Fayyad, U.M.; Piatetsky-shapiro, G.; Smyth, P. & Uthurusamy, R. 1996a. *Advances in Knowledge Discovery and Data Mining*. Menlo Park: AAAI Press. 625p.
- Fayyad, U.M.; Piatetsky-shapiro, G. & Smyth, P. 1996b. *From Data Mining to Knowledge Discovery in Databases*. AI Magazine, Disponível em: [www.daedalus.es/fileadmin/daedalus/doc/MineriaDeDatos/fayyad96.pdf](http://www.daedalus.es/fileadmin/daedalus/doc/MineriaDeDatos/fayyad96.pdf).
- Goldschmidt, R. R. & Passos, E. L. 2005. *Data Mining: um guia prático*. Rio de Janeiro: Editora Campus – Elsevier. 255p.
- Macedo, E. S.; Santoro, J. & Araujo, R. E. 2004. Plano Preventivo de Defesa Civil (PPDC) para deslizamentos. In: SIMPÓSIO BRASILEIRO DE DESASTRES NATURAIS (SIBRADEN), Florianópolis, Santa Catarina.
- Projeto Serra do Mar. 2005. *Estudos da Previsibilidade de Eventos Meteorológicos Extremos na Serra do Mar*. Agência Financiadora: Fundação de Amparo à Pesquisa do Estado de São Paulo. Coordenadora do Projeto: Dra. Chou Sin Chan. Centro de Previsão de Tempo e Estudos Climáticos. Disponível em: <http://www.cptec.inpe.br/serradomar/objetivo.shtml>.
- Tatizana, C.; Ogura, A. T.; Cerri, L.E.S. & Rocha, M.C.M. 1987a. *Análise de Correlação entre Chuvas e Deslizamentos – Serra do Mar – Município de Cubatão*. In: CONGRESSO BRASILEIRO DE GEOLOGIA E ENGENHARIA, 5, *Anais*, São Paulo, v. 2 , p. 225-236.
- Tatizana, C.; Ogura, A.T.; Cerri, L.E.S. & Rocha, M.C.M. 1987b. *Modelamento Numérico da Análise de Correlação entre Chuvas e Deslizamentos aplicados à Encosta da Serra do Mar*. In: CONGRESSO BRASILEIRO DE GEOLOGIA E ENGENHARIA, 5, *Anais*, São Paulo, v. 2 , p. 237-248.
- Witten, I. H. & Frank E. 2000. *Data Mining: Practical Machine Learning Tools and Techniques with Java implementations*. San Diego: Academic Press. 371p.