

Aplicando Mineração de Dados para Apoiar a Tomada de Decisão na Segurança Pública do Estado de Alagoas

Lucas M. Braz¹, Rafael Ferreira¹, Diego Dermeval¹,
Douglas Vêras¹, Marcilio Lima¹, Willy Tiengo¹

¹Instituto de Computação – Universidade Federal de Alagoas (UFAL)
Campus A. C. Simões, BR 104 - Norte, Km 97, C. Universitária, Maceió/AL
GrOW - Grupo de Otimização da Web
<http://www.grow.ic.ufal.br>

{lmonteirobraz, rafaelflmello, diegodermeval, douglassverass}@gmail.com

Abstract. *Data mining is becoming increasingly common in both the private and public sectors. In the area of public safety data mining can be used to: determine where the levels of crime are higher, define profiles victims and criminals and detect the days that occur the most number of crimes. The aim of this paper is to use data mining in the system SISGOP, a database that records the police reports of the occurrences of Maceió, in order to discover informations that aid in the strategical actions of the police department, based on the behaviour of the criminals and victims.*

Resumo. *Mineração de Dados ou Data Mining têm se tornado cada vez mais comum nos setores público e privado. Na área de segurança pública a Mineração de Dados pode ser utilizada para: determinar os locais com maior criminalidade, definir perfis de vítimas e criminosos, identificar a existência de quadrilhas e serial killers, detectar quais dias da semana em que ocorrem mais delitos e até mesmo as suas causas, entre tantas outras possibilidades. O objetivo deste artigo é utilizar técnicas de mineração sobre o sistema SISGOP, um banco de dados da Polícia Militar do Estado de Alagoas que armazena os boletins das ocorrências de algumas cidades de Alagoas, para descobrir informações que auxiliem ações estratégicas da polícia, baseadas no comportamento dos criminosos e vítimas.*

1. Introdução

Mineração de Dados ou *Data Mining* é a atividade de extrair informação de grandes bases de dados, cujo objetivo é a descoberta de fatos e padrões comportamentais, previamente ocultos. O uso da Mineração de Dados nos setores público e privado está se tornando cada vez mais comum. No setor privado, bancos, seguradoras, varejistas, entre outros, utilizam essa tecnologia com o intuito de reduzir custos e aumentar as vendas. Um caso de sucesso é o *Bank of America*, que utilizou técnicas de Mineração de Dados para selecionar dentre seus clientes aqueles com menor possibilidade de não honrar com empréstimos. Com base nessa informação, foi feito um trabalho de *marketing* na oferta de crédito aos correntistas selecionados. Em apenas três anos, o *Bank of America* aumentou sua lucratividade em cerca de trinta milhões de dólares [BankOfAmerica 2009]. No setor público, a mineração de dados é utilizada, principalmente, como forma de descobrir fraudes. Os governos

brasileiro e norte americano, por exemplo, monitoram transações financeiras procurando detectar padrões similares à lavagem de dinheiro do narcotráfico [Ministério 2006].

Na área de segurança pública a Mineração de Dados pode ser utilizada para: determinar os locais com maior criminalidade, definir perfis de vítimas e criminosos, identificar a existência de quadrilhas e *serial killers*, detectar quais dias da semana em que ocorrem mais delitos e até mesmo as suas causas, entre tantas outras possibilidades. Os benefícios em utilizar Mineração de Dados são muitos, por exemplo, descobrir os bairros que são mais perigosos pode induzir a uma melhor alocação dos policiais, bem como definir o perfil das vítimas pode ajudar a traçar campanhas preventivas focadas no público que se encaixa nesse perfil.

A Polícia Militar do Estado de Alagoas possui um banco de dados chamado SISGOP - Sistema de Gestão de Ocorrências Policiais, no qual ficam registrados os boletins das ocorrências policiais em algumas cidades de Alagoas. O objetivo deste artigo é utilizar esse banco de dados para obter informações privilegiadas a fim de auxiliar ações estratégicas da polícia baseadas no comportamento dos criminosos e vítimas, por meio da utilização de técnicas de Mineração de Dados.

Este artigo está organizado como segue. Na Seção 2, nós revisamos os trabalhos relacionados à aplicação de mineração de dados na segurança pública. Depois na Seção 3 são explicados alguns conceitos relacionados à este trabalho. A Seção 4 descreve o experimento realizado para obter padrões comportamentais no banco de dados do SISGOP-AI e na Seção 5 é mostrada as conclusões e os trabalhos futuros do artigo.

2. Trabalhos Relacionados

Atualmente, o uso de mineração de dados como apoio à tomada de decisão ocorre em diversas áreas e setores, tais como: governo, *marketing*, medicina, economia, engenharia, administração, etc [Carvalho 2005]. Tratando-se de governo, especificamente no setor de segurança pública, alguns trabalhos têm sido feitos ressaltando-se a utilização da mineração de dados como apoio ao planejamento estratégico adotado pela polícia. Um exemplo da utilização de técnicas do KDD (*Knowledge Discovery in Databases*) pelo departamento de polícia Richmond, como a mineração de dados, pode ser visto em [McCue 2006a].

Na monografia [Silva Filho 2007], mostraram-se quais tipos de crimes ocorriam em determinadas horas e locais, servindo assim como grande apoio ao planejamento estratégico policial, onde poderiam ser alocados efetivos naqueles locais entre as horas determinadas após o resultado da mineração. Entretanto, somente determinar a hora e o local de um crime muitas vezes é insuficiente para se tomar decisões estratégicas, sendo necessária a inclusão de outras variáveis (hora, local, entre outros) no processo de mineração, como pode ser visto no presente artigo.

Em [McCue 2006b], alguns desafios da mineração de dados na segurança pública foram levantados, tais como: dividir os dados temporais e espaciais em conjuntos, decidir quais atributos presentes em base de dados utilizar, etc. Mostra-se também um simples cenário ilustrativo para avaliar os resultados obtidos a partir de diferentes abordagens utilizando mineração de dados.

No livro [McCue 2007], descrevem-se não somente as diferentes abordagens de

extração de dados para auxiliar a formulação de estratégias, mas também fornece exemplos do mundo real que mostram como a mineração de dados identificou tendências criminais, antecipando crimes e refinando os dados utilizados como entrada.

Neste artigo, apresentamos a utilização da tarefa de mineração de agrupamento para procurar potencializar indicações de diretrizes para policiamento em termos de distribuição e das razões da ocorrência policial, encontradas a partir dos dados da vítima e do criminoso, presentes no boletim de ocorrência.

3. Embasamento teórico

Esta seção pretende definir o referencial teórico que embasou este trabalho. Por ser um assunto amplo e de domínio de várias áreas do conhecimento, tornou-se importante definir os conceitos abordados, em virtude de se ter uma melhor compreensão da abordagem utilizada.

3.1. Mineração de dados

Data mining ou mineração de dados é o processo de extração do conhecimento a partir de grandes volumes dados [Han and Kamber 2000]. Na realidade, a mineração de dados é apenas uma etapa de um processo geral que é o KDD, isto é, a descoberta de conhecimento em banco de dados. A Figura 1 mostra as etapas do KDD.

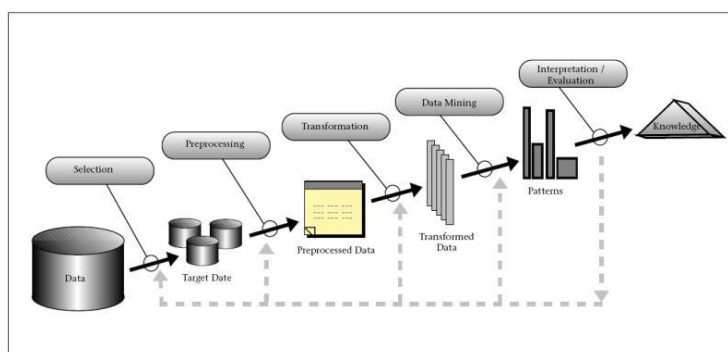


Figura 1. Etapas do KDD [Han and Kamber 2000]

Este processo mais abrangente possui uma metodologia própria para preparação e exploração de dados, interpretação de seus resultados e assimilação dos conhecimentos minerados. Mas a etapa de mineração de dados ficou mais conhecida, pois é nela que as técnicas de mineração são aplicadas.

Abaixo faz-se uma breve discussão a respeito de algumas tarefas de mineração de dados:

- **Classificação:** tarefa de mineração de dados supervisionada e preditiva, ou seja, a partir de dados rotulados se gera um classificador que pode ser uma árvore de decisão, essa fase é conhecida como treinamento. A partir desse classificador, em uma segunda etapa é possível classificar dados não rotulados para um dos rótulos se baseando na árvore de decisão gerada.
- **Agrupamento:** implementa algoritmos não supervisionados em que o rótulo da classe de cada amostra do treinamento não é conhecido, e o número ou conjunto

de classes a ser treinado pode não ser conhecido a priori. No agrupamento os objetos são agrupados de modo que a semelhança seja máxima dentro de cada cluster e mínima entre instâncias de clusters diferentes [Witten and Frank 1999].

- Regras de associação: assim como no agrupamento, possuem algoritmos não supervisionados, ou seja, a classe da tarefa de mineração não é determinada. Esse método busca regras sobre relações e co-ocorrências em bases de dados, por exemplo regras do tipo - se X ocorre na base de dados então Y também ocorre (com alguma relação à X) - esse método é muito usado para verificar associações em tabelas de transações. Para avaliar as regras precisamos de métricas: significância em uma associação, ela pode existir mas ser muito rara (Suporte) e confiança em uma associação, o antecedente pode acontecer várias vezes em uma base de dados mas nem sempre o conseqüente vai ser o mesmo (Confiança).

3.2. WEKA

O Weka [Han and Kamber 2000][Weka 2009] (Waikato Environment for Knowledge Analysis) - *software* de código aberto - é formado por um conjunto de implementações de algoritmos de diversas técnicas de Mineração de Dados. O Weka está implementado na linguagem Java, que fornece ao WEKA a característica de ser portátil, desta forma pode rodar nas mais variadas plataformas e aproveitando os benefícios de uma linguagem orientada a objetos como modularidade, polimorfismo, encapsulamento, reutilização de código dentre outros. Utilizamos a versão 3.4.11 em nossos experimentos. As principais interfaces desta ferramenta são:

- *Simple Client* - O usuário opera o WEKA com o uso de linhas de comando, sendo muito aproveitado por usuário avançados;
- *Explorer* - Visualização mais usada, que aborda o pré-processamento dos dados, mineração de dados (associação, classificação, agrupamento) e pós-processamento dos dados (apresentação de resultados);
- *Experimenter* - Ambiente para testes estatísticos entre algoritmos utilizados;
- *KnowledgeFlow* - Ferramenta para planejamento de ações (em fase de teste).

O Weka possui um formato próprio para o arquivo de entrada, o ARFF [Arff 2009]. Nele descrevemos os dados que são usados no processo de mineração. Antes de aplicar o algoritmo de mineração aos dados na ferramenta Weka, estes devem ser convertidos para o formato ARFF, esse processo pode ser feito de duas formas:

1. Criar um arquivo de texto composto de duas partes. A primeira contém uma lista de todos os atributos, onde devemos definir o tipo do atributo ou os valores que ele pode representar. A segunda parte consiste em instâncias, os registros a serem minerados.
2. Criar o arquivo ARFF implicitamente, onde cada atributo é definido via código e as instâncias são carregadas diretamente do banco de dados. Normalmente usado quando se está trabalhando com o *Simple Client*.

3.3. SISGOP

As Polícias Militares do Brasil têm como missão [Constituição 1988] a execução do policiamento ostensivo preventivo e com repressão imediata. O Comando Geral da Polícia Militar determinou que seu Estado Maior elaborasse e implantasse, mediante a

autorização da Corregedoria Geral de Justiça, um banco de dados do SISGOP-Sistema de Gestão de Ocorrências Policiais da Polícia Militar, onde ficam registrados os boletins das ocorrências policiais atendidas pela Polícia Militar, no policiamento ostensivo, em nosso caso, em algumas cidades de Alagoas. O objetivo é dar agilidade e alcance ao atendimento à população, com isso a Polícia Militar passa a ter maior raio de ação e passa a atender a maior parte de ocorrências policiais simples, logo, desconcentra o procedimento de registrar ocorrência nas delegacias de Polícia Civil, dando maior fôlego a este órgão policial para dedicar-se a procedimentos mais complexos contidos em investigações criminais oriundas dos inquéritos policiais. Os dados volumosos cadastrados nos formulários das ocorrências policiais são uma fonte de conhecimento relativo às vítimas, aos criminosos e à natureza dos crimes, sendo armazenados, visando a recuperação de dados para análise estatística e de descoberta de conhecimento.

3.4. O modelo CRISP-DM

CRISP-DM[CRISP-DM 1996] (*Cross Industry Standard Process for Data Mining*) é uma metodologia para descoberta do conhecimento em base de dados, que impõe ao projeto um detalhado planejamento e avaliação do processo em suas fases, facilitando a organização, a compreensão e o controle dos eventos na coordenação do projeto. O modelo CRISP-DM foi idealizado objetivando ser uma metodologia padrão, as suas fases vão desde um planejamento na identificação dos objetivos sob a ótica da compreensão do negócio até a aplicação do conhecimento extraído. Dessa forma, o modelo CRISP-DM possui um formato de ampla iteratividade com as suas fases, as quais são flexíveis e sem ordenamento, facilitando retornos e avanços, com o encadeamento de passos dependendo do resultado satisfatório e a comunicação entre as fases. Portanto, este modelo incremental e iterativo com um controle rígido das fases, ajuda na identificação de riscos e gera um produto final mais cuidadoso, fornecendo mais oportunidades ao coordenador de garantir uma certa qualidade no projeto. A metodologia fornece um modelo constituído em seis fases descritas a seguir no ciclo das fases do processo da Figura 2.

Estas fases são detalhadas a seguir:

1. **Compreensão do negócio** (*Business understanding*): compreender os objetivos do negócio e a sua estrutura organizacional disponível, relacionando com os objetivos e metas da mineração de dados a serem alcançadas, e ainda, projetar e cogitar as tarefas, técnicas e ferramentas a serem utilizadas;
2. **Compreensão dos dados** (*Data understanding*): procura-se uma compreensão dos dados, sua localização, estrutura, formato, atributos chaves e tipos de valores; buscando itens mais significantes, qualitativos e quantitativos, que possam ter influência nos resultados e nos objetivos, e ainda, realiza agregações de informações e levanta-se possíveis problemas e soluções;
3. **Preparação dos dados** (*Data preparation*): comporta todas as atividades já descritas nas considerações iniciais desta seção pertinentes a operação de preparação de dados, como: selecionar, construir, codificar, enriquecer, normalizar, etc;
4. **Modelagem** (*Modeling*): é a fase de aplicação dos métodos de mineração de dados, com as técnicas propostas, estabelecendo seus respectivos parâmetros de funcionamento e resultados, de forma iterativa e incremental, avaliando o modelo de conhecimento extraído;

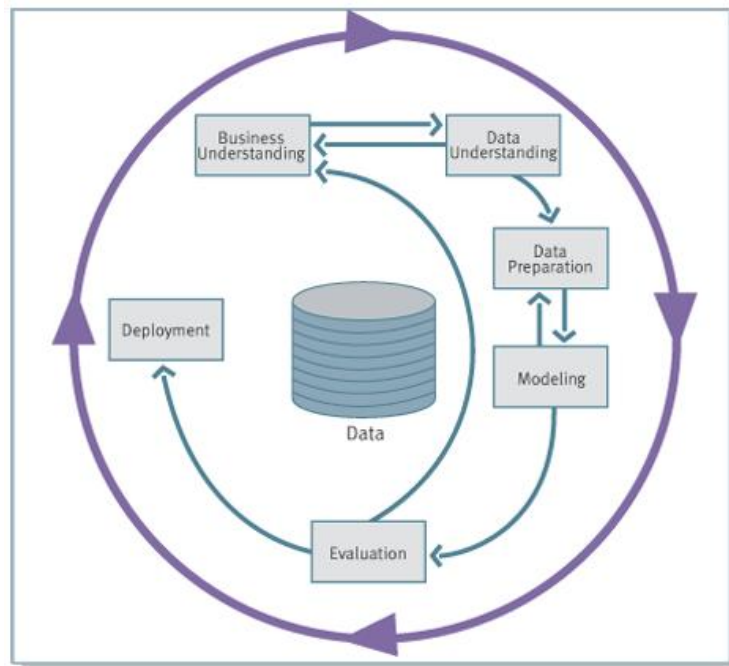


Figura 2. Ciclo das fases da metodologia CRISP-DM [CRISP-DM 1996]

5. **Avaliação** (*Evaluation*): é a fase de avaliar se as metas foram atingidas, revisar o projeto identificando informações despercebidas e caminhos equivocados que devem ser refeitos, relatar explicações e analisar o encerramento do projeto e sua passagem para próxima fase;
6. **Aplicação** (*Deployment*): planejar a aplicação estratégica dos resultados e o emprego rotineiro da mineração de dados, documentar e produz relatório final da implementação do projeto, e ainda, emitir um relatório geral de todo o processo identificando erros, aperfeiçoamentos e experiência adquirida.

4. Experimento

O modelo CRISP-DM foi idealizado diante de situações práticas que expõe de forma objetiva cada fase e seus problemas, logo, favorece a coordenação do projeto pelo analista KDD e demonstração dos seus resultados. Neste contexto, tenta-se empregar as tarefas de agrupamento em nossa demonstração. Mostra-se nesta seção um experimento detalhado da aplicação de mineração de dados para descoberta de conhecimento na base da polícia militar, seguindo a metodologia CRISP-DM.

4.1. Compreensão do negócio

Nos formulários dos boletins de ocorrência cadastrados no banco de dados SISGOP, são registrados todos os dados relacionados ao crime (tipo, arma empregada, localidade, data, horário, etc), ao criminoso e a vítima (sexo, cor, idade, características físicas, estado civil, etc). Neste trabalho, busca-se fazer uma análise das condições de ocorrência do crime, da conduta do criminoso e da vítima com os dados específicos inerentes às ocorrências policiais. No entanto, existem limitações quanto a uma avaliação ampla de causas externas dos crimes em relação ao conteúdo do formulário de ocorrências citado, que poderiam relacionar informações, como: circulação de capital, operações de policiais, eventos de

grande porte, tráficos de drogas, iluminação, informações de boletins registrados pela Polícia Civil, entre outras. Porém, esta situação não prejudica o desenvolvimento da demonstração atual, apenas a complementar, pois as informações supracitadas visariam um alcance e profundidade maior ligado à atuação dos batalhões e a eventos da comunidade local. Neste caso, o processamento de descoberta do conhecimento será aplicado no agrupamento das condutas de comportamento de criminosos e vítimas para gerar perfis destas com as características que tornem uma pessoa em potencial vítima e informe como agem grupos criminosos em determinadas localidades. Desta maneira, tenta-se obter sugestões de ações de coordenação e distribuição de policiamento.

4.2. Compreensão dos dados

Os dados que são processados encontram-se no repositório SISGOP, na plataforma MySQL de banco de dados, com registros referentes as ocorrências de policiais atendidas pela Polícia Militar no ano de 2008, até o mês de julho. Os dados encontram-se armazenados em um conjunto de tabelas, conforme é mostrado na Figura 3. A estrutura de chaves das tabelas principais a serem trabalhadas são baseadas nos atributos e suas respectivas tabelas: bo, cod_bo; fato, cod_fato; envolvido, cod_envolvido. Os dados contidos em outras tabelas fornecem dados elementares de cadastro, como: graduação e posto do policial do atendimento da ocorrência, legislação aplicada, cidade, bairro, natureza do crime, unidades policiais militares cadastradas, etc.

Tabela	Ações	Registros	Tipo	Collation	Tamanho	Sobre Carga
ambiente		21	MyISAM	latin1_swedish_ci	2.5 KB	-
bairro		298	MyISAM	latin1_swedish_ci	12.2 KB	-
bo		3,583	MyISAM	latin1_swedish_ci	257.0 KB	-
bo_fato		0	MyISAM	latin1_swedish_ci	1.0 KB	-
cidade		104	MyISAM	latin1_swedish_ci	4.4 KB	-
envolvido		9,253	MyISAM	latin1_swedish_ci	2.3 MB	-
estabelecimento		170	MyISAM	latin1_swedish_ci	8.5 KB	-
fato		3,609	MyISAM	latin1_swedish_ci	1.3 MB	-
graduacao		15	MyISAM	latin1_swedish_ci	2.3 KB	-
lei		13	MyISAM	latin1_swedish_ci	2.3 KB	-
log		9,078	MyISAM	latin1_swedish_ci	679.7 KB	-
natureza		302	MyISAM	latin1_swedish_ci	59.7 KB	-
objeto		1,290	MyISAM	latin1_swedish_ci	80.7 KB	-
objeto_nome		43	MyISAM	latin1_swedish_ci	3.0 KB	-
objeto_tipo		5	MyISAM	latin1_swedish_ci	2.1 KB	-
profissoes		705	MyISAM	latin1_swedish_ci	25.5 KB	-
responsavel_bo		3,695	MyISAM	latin1_swedish_ci	503.2 KB	-
responsavel_tco		3,696	MyISAM	latin1_swedish_ci	123.3 KB	-
tco		3,703	MyISAM	latin1_swedish_ci	409.9 KB	-
unidade		24	MyISAM	latin1_swedish_ci	3.0 KB	-
usuarios		68	MyISAM	latin1_swedish_ci	8.9 KB	-
veiculo		556	MyISAM	latin1_swedish_ci	52.0 KB	-
vestimenta		27	MyISAM	latin1_swedish_ci	2.6 KB	-
23 tabela(s)	Soma	40,258	--	latin1_swedish_ci	5.8 MB	0 Bytes

Figura 3. Tabelas do banco de dados SISGOP

4.2.1. Preparação dos dados

Para facilitar a extração dos dados ocorreu uma junção das tabelas acima em uma única tabela que tivesse todas as informações relevantes a obtenção do produto de conheci-

mento. E assim, foi criada a tabela BO_fato_Envolvido, agregando as tabelas: BO, fato e envolvido. Evidentemente, foram feitas junções com outras tabelas, como: bairro, cidade, para obter informações complementares a respeito da ocorrência. Foi gerada uma instrução em SQL, com os atributos a serem utilizados no processo de descoberta do conhecimento em base de dados, para posteriormente tal consulta ser transformada em tabela. A seguir consta na Tabela 1 os atributos utilizados:

Tabela 1. Principais campos empregados no processo com a indicação de suas respectivas tabelas de dados

	Tabela	Atributos
01	BO	Opn, procedimento
02	Fato	data, hora, logradouro_fato, localidade, ponto_referencia, cod_cidade, cod_natureza, evasao, evasão, abordagem.
03	Envolvido	Tipo_envolvido, idade, sexo, grau_instrução, estado_civil, bairro, dia, hora, condições físicas.

Foram detectados vários atributos com ausência de valores, desta forma em razão do alto grau de ausência de dados o atributo foi descartado para não gerar dados tendenciosos ou inconsistentes.

Os dados que forem detectados a ausência de valores seriam preenchidos com os valores medianos respectivos, através do item de limpeza de valores ausentes e seu referido subitem de preenchimento de informações ausentes por medidas estatísticas. No entanto, em razão do alto grau de ausência de dados poderá o atributo ser descartado para não gerar dados tendenciosos.

A Figura 4, a seguir, ilustra a tela do WEKA de organização, avaliação e estatística de dados a serem tratados. Esta é a tela principal de pré-processamento, a qual encontra-se no lado esquerdo os atributos utilizados (ver mais detalhadamente na Figura 5) e no lado direito, acima, a distribuição descritiva e distribuição dos dados do atributo (mais detalhes na Figura 6).

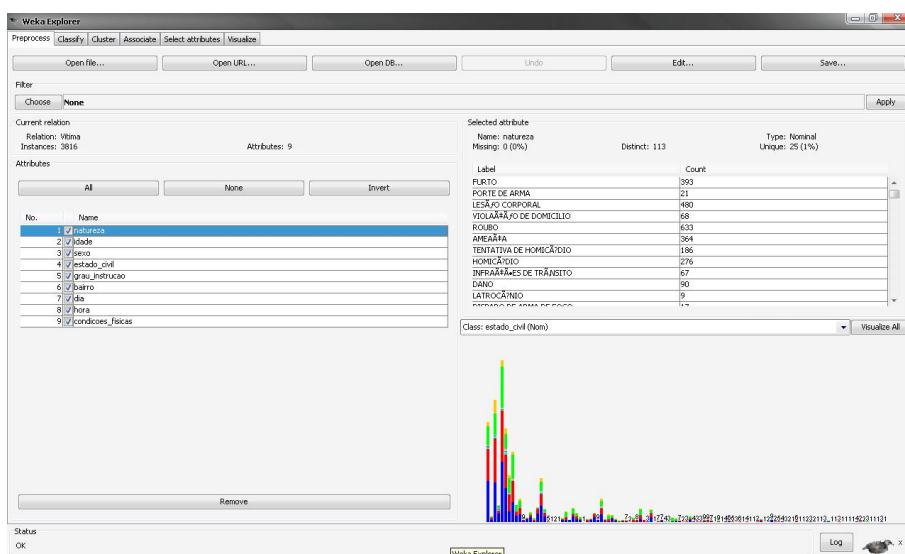


Figura 4. Tela de pré-processamento - WEKA

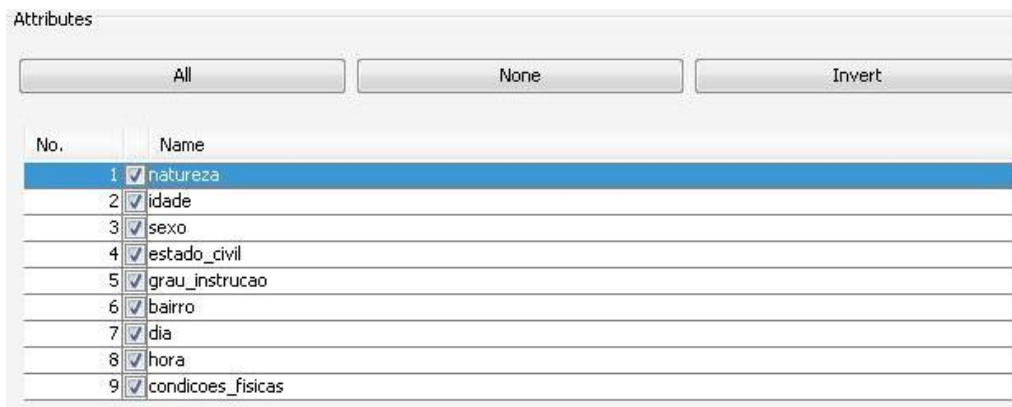


Figura 5. Atributos utilizados - WEKA

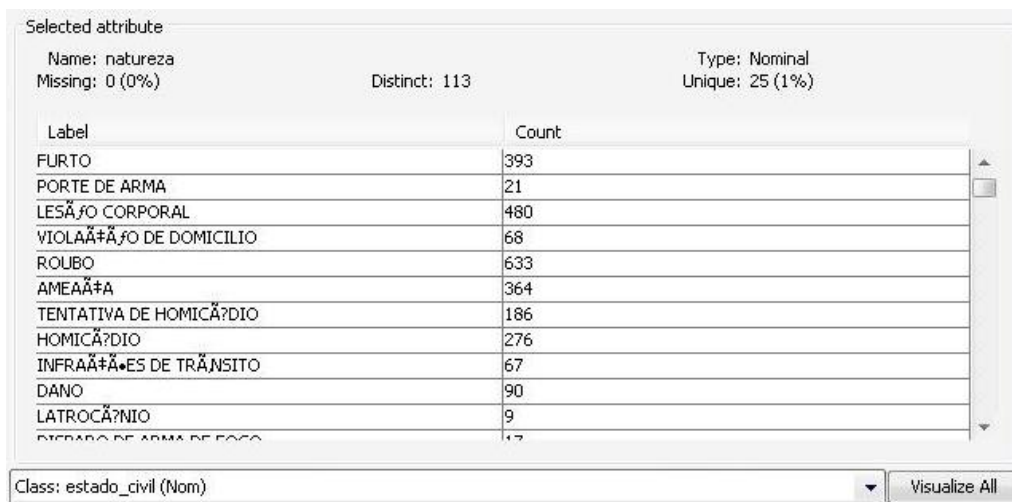


Figura 6. Distribuição descritiva - WEKA

4.3. Modelagem

A demonstração realizada tem como principal objetivo mostrar o potencial de aplicação do processo de descoberta de conhecimento em base de dados, pois pelo fato de que os formulários de atendimento de ocorrências policiais, chamados Boletins de Ocorrência, utilizados pela Polícia Militar não foram projetados para o desenvolvimento de tal aplicação e nem passaram por uma atualização ou reavaliação neste sentido. Portanto, nossa maior preocupação é evidenciar indícios sutis de que é possível estabelecer algumas diretrizes para o planejamento do policiamento. Com isso, após o pré-processamento, onde os dados foram tratados, as informações foram processadas na aplicação, conforme as Figuras 7 e 8, ilustram a aplicação da ferramenta, em termos de agrupamento, nas condutas dos criminosos e vítimas.

As Figuras 9 e 10 mostram algumas associações que foram geradas a partir do agrupamento realizado:

```

Clusterer output

estado_civil
grau_instrucao
bairro
dia
hora
condicoes_fisicas
Test mode: evaluate on training data

=== Model and evaluation on training set ===

kMeans
=====

Number of iterations: 3
Within cluster sum of squared errors: 18152.0

Cluster centroids:

Cluster 0
Mean/Mode: LESÃO CORPORAL Nao especificado MASCULINO Nao Informado(a) Nao especificado Jacintinho - Maceiã' Domingo 23
Std Devs: N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A

Cluster 1
Mean/Mode: AMEAÇA 25-30 FEMININO Casado(a) Nao especificado Jacintinho - Maceiã' Domingo 18 SEM LESÃO
Std Devs: N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A

Cluster 2
Mean/Mode: AMEAÇA 20-25 FEMININO Solteiro(a) Nao especificado Tabuleiro dos Martins - Maceiã' Quinta 20 SEM LESÃO
Std Devs: N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A

Cluster 3
Mean/Mode: HOMICÍDIO 40-45 MASCULINO Solteiro(a) Nao especificado Jacintinho - Maceiã' Domingo 20 SEM LESÃO
Std Devs: N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A

Cluster 4

```

Figura 7. Processamento do WEKA no módulo Explorer avaliando Agrupamento da conduta de vitimas do banco de dados do SISGOP-PMAL

```

Clusterer output

bigode
barba
costeleta
altura
peso
cor
Test mode: evaluate on training data

=== Model and evaluation on training set ===

kMeans
=====

Number of iterations: 4
Within cluster sum of squared errors: 10172.0

Cluster centroids:

Cluster 0
Mean/Mode: LESÃO CORPORAL MASCULINO Casado(a) Nao especificado 1º Grau Incompleto Centro - Palmeira dos Índios Quinta
Std Devs: N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A

Cluster 1
Mean/Mode: FURTO MASCULINO Solteiro(a) 15-20 1º Grau Completo Centro - Arapiraca Domingo 11 Nao Declarado Nao Declarado
Std Devs: N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A

Cluster 2
Mean/Mode: AMEAÇA MASCULINO Casado(a) 25-30 Nao especificado Jacintinho - Maceiã' Domingo 17 Nao Declarado Nao Declarad
Std Devs: N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A

Cluster 3
Mean/Mode: AMEAÇA MASCULINO Solteiro(a) 15-20 Nao especificado Centro - Arapiraca Domingo 21 Nao Declarado Nao Declarad
Std Devs: N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A

Cluster 4

```

Figura 8. Processamento do WEKA no módulo Explorer avaliando Cluster da conduta de criminosos do banco de dados do SISGOP-PMAL

01	<p>a) Atributos referentes à conduta dos criminosos: natureza; sexo; estado_civil; idade; grau_instrucao; bairro; dia; hora; aproximação; abordagem; evasão; bigode; barba; costeleta; altura; peso; cor.</p> <ol style="list-style-type: none"> 1. ROUBO MASCULINO Centro - Arapiraca Terça 11 A PE 2. AMEAÇA MASCULINO Solteiro(a) 15-20 1º Grau Completo Centro - Arapiraca Domingo 19h A PE 1,51 - 1,70m 61 - 80kg Pardo 3. LESÃO CORPORAL MASCULINO Casado(a) 30-35 1º Grau Completo Centro - Arapiraca Sabado 20h A PE 1,51 - 1,70m 61 - 80kg Pardo 4. LESÃO CORPORAL MASCULINO Solteiro(a) Vila Maria - Palmeira dos Índios Domingo 16h 5. LESÃO CORPORAL MASCULINO Solteiro(a) 15-20 Jacintinho - Maceió Quinta 12 1,51 - 1,70m 61 - 80kg Pardo 6. INFRAÇÕES DE TRÂNSITO MASCULINO Solteiro(a) 20-25 1º Grau Completo Pilar Domingo 17h 1,71 - 1,80m 61 - 80kg Branco 7. PORTE DE ARMA MASCULINO Centro - Arapiraca Sexta 14h 8. LESÃO CORPORAL FEMININO Solteiro(a) 15-20 Ponta Verde - Maceió Quinta 10h 9. PORTE DE ARMA MASCULINO Solteiro(a) 20-25 Tabuleiro dos Martins - Maceió Terça 23h 1,51 - 1,70m 61 - 80kg Branco 10. LESÃO CORPORAL MASCULINO Casado(a) 30-35 Tabuleiro dos Martins - Maceió Quinta 10h <p>b) Atributos referentes à conduta dos criminosos: natureza; sexo; estado_civil; idade; bairro; dia; hora.</p> <ol style="list-style-type: none"> 1. ROUBO MASCULINO Centro - Arapiraca Segunda 19h 2. PORTE DE ARMA MASCULINO Casado(a) Tabuleiro dos Martins - Maceió Quarta 3. LESÃO CORPORAL MASCULINO Casado(a) 30-35 Centro - Arapiraca Sabado 20h 4. LESÃO CORPORAL MASCULINO Solteiro(a) 15-20 Vila Maria - Palmeira dos Índios Domingo 16h 5. LESÃO CORPORAL MASCULINO Solteiro(a) 35-40 Farol - Maceió Quarta 12h 6. AMEAÇA MASCULINO Solteiro(a) 30-35 Jacintinho - Maceió Sexta 09h 7. PORTE DE ARMA MASCULINO 15-20 Benedito Bentes - Maceió Sexta 14h 8. LESÃO CORPORAL FEMININO Solteiro(a) 15-20 Ponta Verde - Maceió Quinta 10h 9. LESÃO CORPORAL MASCULINO Solteiro(a) 20-25 Centro - Arapiraca Terça 23h 10. TRÁFICO MASCULINO 30-35 Pajuçara - Maceió Quinta 10h
----	--

Figura 9. Agrupamento - Criminoso

02	<p>a) Atributos referentes à conduta dos vitima: natureza; idade; sexo; estado_civil; grau_instrucao ; bairro; dia; hora; condicoes_fisicas.</p> <ol style="list-style-type: none"> 1) ROUBO-MASCULINO-Casado(a)- Centro - Arapiraca Domingo- 19h-SEM LESÃO 2) HOMICÍDIO- 25-30- MASCULINO Solteiro(a) -Jacintinho - Maceió- Terça 21h 3) FURTO- 30-35- MASCULINO- Casado(a)- Centro - Arapiraca- Terça -10h - SEM LESÃO 4) LESÃO CORPORAL 35-40 MASCULINO Casado(a) -Jacintinho - Maceió- Sexta 22h 5) ROUBO 20-25 MASCULINO Solteiro(a)- Farol - Maceió-Quarta 11h SEM LESÃO 6) AMEAÇA 35-40 FEMININO Solteiro(a)- Jacintinho - Maceió Domingo 14h SEM LESÃO 7) ROUBO- FEMININO-Jacintinho - Maceió-Domingo 17h SEM LESÃO 8) ROUBO -MASCULINO Casado(a) -Tabuleiro dos Martins - Maceió- Sabado 14h SEM LESÃO 9) HOMICÍDIO MASCULINO -Jacintinho - Maceió Sabado 20h OBITO 10) FURTO 45-50 MASCULINO Casado(a) Centro - Maceió Sexta 18h SEM LESÃO <p>b) Atributos referentes à conduta dos vitima: natureza; idade; sexo; estado_civil; bairro; dia; hora; condicoes_fisicas.</p> <ol style="list-style-type: none"> 1) ROUBO MASCULINO Casado(a) Centro - Arapiraca Domingo 19h SEM LESÃO 2) HOMICÍDIO 25-30 MASCULINO Solteiro(a) Jacintinho - Maceió Terça 21h OBITO 3) FURTO 30-35 MASCULINO Casado(a) Centro - Arapiraca Terça 20h SEM LESÃO 4) FURTO 45-50 MASCULINO Casado(a) Centro - Maceió Sexta 06h SEM LESÃO 5) LESÃO CORPORAL 35-40 MASCULINO Casado(a) Jacintinho - Maceió Sexta 22h COM LESÃO 6) ROUBO 20-25 MASCULINO Solteiro(a) Farol - Maceió Quarta 11h SEM LESÃO 7) AMEAÇA 35-40 FEMININO Solteiro(a) Jacintinho - Maceió Domingo 14 SEM LESÃO 8) ROUBO FEMININO Jacintinho - Maceió Domingo 20h SEM LESÃO 9) ROUBO MASCULINO Casado(a) Jacintinho - Maceió Sábado 14h SEM LESÃO 10) HOMICÍDIO MASCULINO Adjacências - Atalaia Sábado 20h OBITO
----	---

Figura 10. Agrupamento - Vítima

4.3.1. Avaliação

Os resultados dos agrupamentos fornecidos pelo WEKA não significam o número objetivo de crimes, e sim, que criminosos agem com conduta semelhantes a ponto de

formarem grupos característicos de horário, dia da semana, bairro, cidade, natureza de crimes, estado civil, etc. Portanto, servem como diretrizes que norteiam policiamento. Diante disto, temos que os resultados foram satisfatórios, pois podemos citar, por exemplo, a regra 7 e 8, da Figura 10, item "b", a qual reza a conduta das vítimas: *mulheres são vitimas de roubo e ameaça no jacintinho entre as 14hs e 20hs no dia de domingo*, logo, pode-se distribuir policiamento neste sentido, orientar vítimas em sua conduta de risco e tentar descobrir a razão da formação deste grupo, tudo com o objetivo de se evitar o delito. Há vários exemplos de regras geradas informando a conduta de criminosos, mas pode-se destacar que na cidade de Arapiraca várias condutas criminosas foram encontradas, basta observar as regras de associação de grupo expostas na Figura 9, no item "a", de numerações 1, 2, 3 e 7. Além disso, no mesmo local, item "b", pode-se citar: o crime de Porte de Arma, nos bairro do Tabuleiro e Jacintinho. Assim, fornece indícios que auxiliam na distribuição de policiamento, a busca por criminosos que atuam com estas características e motiva a investigação de saber o porquê de criminosos atuarem daquela forma com aquelas características, no sentido de obter vertentes de combate a criminalidade. No entanto, vale o esclarecimento de que os grupos (linhas 3 e 4) de todas as alíneas "b" possuem maior credibilidade do que as alíneas "a", pelo motivo de terem sido retirados a nível de pré-processamento atributos que possuíam grande números de instâncias com ausência de dados.

4.4. Aplicação

A título de aplicação, temos que para usar outros métodos de mineração, como associação, é necessário um alto índice de preenchimento e de forma confiável do boletim de ocorrência policial. Portanto, é necessário que se faça um estudo com o objetivo de facilitar o preenchimento do formulário e sua atualização em termos de projeto para que se possa categorizar ou discretizar melhor um maior número de atributos, por exemplo, características físicas, abordagens, evasão, local de entrada, entre outros. Para se ter uma idéia do índice de não preenchimento podemos citar o atributo grau_instrução do criminoso, onde cerca de 61% dos dados não estão preenchidos. Outro fato que não podemos deixar de ressaltar é que se os boletins de ocorrências policiais fossem unificados entre a Polícia Civil e Polícia Militar existiria um universo maior de dados e com melhor credibilidade. Com isso, seriam facilitadas as ações estratégicas de ambas as instituições baseadas no conhecimento gerado pela aplicação KDD. Embora tenhamos usado apenas a técnica de agrupamento, temos que levar em consideração de utilizar outras, como por exemplo a classificação. A tarefa de classificação poderia ser empregada para a Polícia Militar na tomada de decisões de ocorrências de gerenciamento de crise com reféns, utilizando as características da conduta do seqüestrado e da ocorrência. Já para Polícia Civil, a tarefa de classificação serviria para detectarmos o perfil do criminoso ou do grupo delituoso pelas características da ocorrência. Contudo, seria necessário ter um cadastro do estudo de casos destes tipos de ocorrências complexas. Assim, a ferramenta WEKA, aplicada com as regras da técnica de agrupamento, utilizando o método de algoritmo *SimpleKmeans*, gerou um resultado que atingiram os objetivos do nosso trabalho, o qual era de demonstrar a possibilidade em termos do potencial das aplicações de obtenção do conhecimento em base de dados, proporcionando elementos para formar diretrizes de melhor distribuição de policiamento, orientações de vítimas, evidências que possibilitem a prevenção e repressão imediata de crimes e redução da violência, e ainda, fornecer indícios da causa e de onde há ocorrências. Embora, seja necessário uma reavaliação do formulário do boletim de

ocorrência com o objetivo de fornecer maior utilidade das informações na aplicação das ferramentas de mineração de dados.

5. Conclusão e Trabalhos Futuros

Neste trabalho foi apresentada uma abordagem para descoberta do conhecimento a partir dos dados das ocorrências atendidas pela Polícia Militar de Alagoas e que estão registradas no banco de dados SISGOP (Sistema de Gestão de Ocorrência Policiais da Polícia Militar). A proposta apresentada teve o intuito de demonstrar a aplicação de tarefas de mineração de dados na Segurança Pública, visando justamente se beneficiar dessas informações para conhecer o comportamento dos criminosos e das vítimas em geral, como também mapear (geograficamente) os crimes e suas possíveis causas.

A aplicação teve como base o emprego da tarefa de agrupamento, utilizando o algoritmo *SimpleKmeans*, em relação as condutas dos criminosos e das vítimas, procurando potencializar indicações de diretrizes para policiamento em termos de distribuição e das razões da ocorrência policial. Foram obtidos resultados relevantes, gerando justamente, a grupos de características comuns de condutas criminosas e de situações de risco das vítimas, conforme foi demonstrado.

O experimento realizado neste artigo utilizou uma base de dados incompleta, na qual vários atributos, que eram essenciais para a avaliação dos resultados esperados, estavam ausentes. Por exemplo, devido ao alto índice de ausência do atributo "bigode" na base de dados, foram geradas regras que agregam pouco valor ao objetivo deste trabalho. Diante disso, como trabalho futuro há a necessidade de obter uma base mais completa para garantir melhores resultados.

Por conseguinte, fica evidente o potencial existente na descoberta de conhecimento em favor da Segurança Pública, pois os resultados gerados podem servir de informação privilegiada para elaborar ações estratégicas, que visem reduzir a violência e os índices de criminalidade no Estado, principalmente, de maneira pró-ativa, baseadas nas causas de ocorrências.

Referências

- Arff (2009). Attribute-relation file format. Disponível em <http://www.cs.waikato.ac.nz/ml/weka/arff.html>. Acessado em 18 de maio de 2009.
- BankOfAmerica (2009). Disponível em <http://www.bankofamerica.com/index.cfm?page=about>. Acessado em 18 de maio de 2009.
- Carvalho, L. A. V. (2005). *A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração*. 1 edition.
- Constituição (1988). Constituição da república federativa do brasil. Disponível em http://www.planalto.gov.br/ccivil_03/Constituicao/Constitui%C3%A7ao.htm. Acessado em 06 de Abril de 2009.
- CRISP-DM (1996). Cross industry standard process for data mining. Disponível em <http://www.crisp-dm.org/>. Acessado em 06 de Abril de 2009.
- Han, J. and Kamber, M. (2000). *Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann.

- McCue, C. (2006a). Data mining and crime analysis in the richmond police department.
- McCue, C. (2006b). Data mining and predictive analytics in public safety and security. *IT Professional*, 8(4):12–18.
- McCue, C. (2007). *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis*. Butterworth-Heinemann, Newton, MA, USA.
- Ministério (2006). Programa nacional de capacitação e treinamento para o combate à lavagem de dinheiro (pnld). Disponível em [http://bvc.cgu.gov.br/bitstream/123456789/1536/1/Programa+Nacional+de+CapacitaD\).pdf](http://bvc.cgu.gov.br/bitstream/123456789/1536/1/Programa+Nacional+de+CapacitaD).pdf). Acessado em 09 de maio de 2009.
- Silva Filho, L. A; Santos, F. H. M. (2007). A utilização de sistemas de informação no apoio à tomada de decisão na segurança pública do estado do pará. *Trabalho de Conclusão de Curso. Belém (UFPA)*.
- Weka (2009). Disponível em <http://www.cs.waikato.ac.nz/ml/weka/>. Acessado em 09 de Março de 2009.
- Witten, I. H. and Frank, E. (1999). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann.