



UNIVERSIDADE ESTADUAL DO NORTE DO PARANÁ
CAMPUS LUIZ MENEGHEL



ALESSANDRO ROBERTO LUZ

**APLICANDO MINERAÇÃO DE DADOS NO SISTEMA
ESTADUAL DE REGISTRO DE OCORRÊNCIAS
POLICIAIS BOLETIM DE OCORRÊNCIA UNIFICADO**

Bandeirantes

2012

ALESSANDRO ROBERTO LUZ

**APLICANDO MINERAÇÃO DE DADOS NO SISTEMA
ESTADUAL DE REGISTRO DE OCORRÊNCIAS
POLICIAIS BOLETIM DE OCORRÊNCIA UNIFICADO**

Trabalho de Conclusão de Curso submetido ao Campus Luiz Meneghel da Universidade Estadual do Norte do Paraná, como requisito parcial para a obtenção do grau de Bacharel em Sistemas de Informação.

Orientador: Prof. Me. Glauco Carlos Silva

Bandeirantes

2012

ALESSANDRO ROBERTO LUZ

**APLICANDO MINERAÇÃO DE DADOS NO SISTEMA
ESTADUAL DE REGISTRO DE OCORRÊNCIAS
POLICIAIS BOLETIM DE OCORRÊNCIA UNIFICADO**

Trabalho de Conclusão de Curso
submetido ao Campus Luiz Meneghel da
Universidade Estadual do Norte do
Paraná, como requisito parcial para a
obtenção do grau de Bacharel em
Sistemas de Informação.

COMISSÃO EXAMINADORA

Prof. Me. Glauco Carlos Silva
Campus Luiz Meneghel

Prof. Me. André Luis Andrade Menolli
Campus Luiz Meneghel

Prof. Me. Ricardo Gonçalves Coelho
Campus Luiz Meneghel

Bandeirantes, ___ de _____ de 2012.

A minha esposa Kelly, pela compreensão e paciência.

Ao meu filho Henrique, razão do meu viver.

AGRADECIMENTOS

Primeiramente a Deus, pela oportunidade de estudo e crescimento.

Ao Prof. Me. Orientador Glauco Carlos Silva, pelo apoio e incentivo.

Aos companheiros Jacob Rockembach e Eduardo Marcelo Castella, respectivamente, Delegados de Polícia da Agência de Inteligência e Coordenação de Informática da Polícia Civil.

Aos profissionais Jair, Valter e Rubens, analistas da Celepar.

Aos professores pela dedicação e difusão do conhecimento.

Aos amigos e colegas de Curso, companheiros de jornada.

Enfim, a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

*"Ama-se mais o que se
conquista com esforço."*

Benjamin Disraeli

RESUMO

O uso da computação está presente em praticamente todos os setores, sobretudo auxiliando o homem na solução de problemas que seriam humanamente inviáveis de se resolver. Um dos ramos do conhecimento computacional é a Inteligência Artificial, área de pesquisa cujo objetivo é, basicamente, fazer os computadores realizarem tarefas inteligentes como o ser humano. A Inteligência Artificial propõe desenvolver soluções que simulem a capacidade humana de perceber, raciocinar, decidir e resolver problemas, analogamente à inteligência humana. Neste contexto, a Mineração de Dados é uma das áreas de Inteligência Artificial empregada na exploração de grandes quantidades de dados à procura de padrões consistentes que permitam extrair um conhecimento implícito e até então desconhecido, mas potencialmente útil ao homem. A extração de regras de associação é uma das tarefas empregadas no processo de descoberta do conhecimento. A proposta do trabalho é a aplicação desta técnica de mineração de dados no âmbito da atividade policial, extraindo conhecimento do sistema informatizado de registro de ocorrências do Estado do Paraná, a aplicação denominada Boletim de Ocorrência Unificado.

Palavras-chave: mineração de dados; extração do conhecimento; Inteligência Artificial; regras de associação; polícia; boletim de ocorrência.

ABSTRACT

The use of computers is present in almost all sectors, especially helping to solve human problems that would be humanly to solve. One branch of the computational expertise is Artificial Intelligence, research area whose goal is basically to make computers perform tasks such as intelligent human being. Artificial Intelligence proposes to develop solutions that simulate the human ability to perceive, think, decide and solve problems, analogous to human intelligence. In this context, the Data Mining is one of the Artificial Intelligence techniques employed in the exploration of large amounts of data looking for consistent patterns that allow one to extract implicit knowledge, unknown, but potentially useful to man. The extraction of association rules is one of the tasks employed in the process of knowledge discovery. The proposed work is the application of this technique of data mining within the police activity, extracting knowledge from the computerized system of registration of occurrences of Paraná State, the application called Unified Police Reports.

Keywords: data mining, knowledge extraction, Artificial Intelligence, association rules, police, police report.

LISTA DE ABREVIATURAS

ABIN	Agência Brasileira de Inteligência
AC	Análise Criminal
BO	Boletim de Ocorrência
BOU	Boletim de Ocorrência Unificado
CELEPAR	Companhia de Informática do Paraná
CIN	Coordenação de Informática
DEPEN	Departamento Penitenciário do Paraná
DNISP	Doutrina Nacional de Inteligência de Segurança Pública
EDA	<i>Exploratory Data Analysis</i>
IA	Inteligência Artificial
ISP	Inteligência da Segurança Pública
JDBC	<i>Java Database Connectivity</i>
KDD	<i>Knowledge Discovery in Database</i>
MD	Mineração de Dados
ML	<i>Machine Learning</i>
RAM	<i>Random Access Memory</i>
SESP	Secretaria de Estado da Segurança Pública
SGBD	Sistema Gerenciador de Banco de Dados
SISBIN	Sistema Brasileiro de Inteligência
SISP	Subsistema de Inteligência de Segurança Pública
SQL	<i>Structured Query Language</i>
TI	Tecnologia da Informação
XML	<i>Extensible Markup Language</i>

LISTA DE FIGURAS

Figura 1: Esquema simplificado da mineração de dados.	7
Figura 2: Reunião ocorrida na CELEPAR	14
Figura 3: <i>Screenshot</i> do software <i>RapidMiner</i>	15
Figura 4: Diagrama esquemático do processo de mineração de dados.....	18
Figura 5: Tarefas da mineração de dados.	22
Figura 6: Tela inicial do registro de ocorrência.....	40
Figura 7: Continuando no registro de ocorrência.	40
Figura 8: Diminuindo o número de conexões simultâneas no banco de dados.....	46
Figura 9: Configurando a quantidade de memória disponível para o <i>PostgreSQL</i> ...	46
Figura 10: Mensagem de alerta para criação do repositório de trabalho.	47
Figura 11: Criando um novo repositório.	48
Figura 12: Criando um repositório e finalizando a configuração.....	48
Figura 13: <i>Script</i> SQL de criação do banco de dados no ambiente <i>Windows</i>	49
Figura 14: <i>Script</i> SQL de criação do banco de dados no ambiente <i>Linux</i>	50
Figura 15: Resultado da consulta no editor SQL mostrando as linguagens procedurais instaladas.	51
Figura 16: Modelo lógico do banco de dados após exclusão de tabelas e seleção de atributos.....	55
Figura 17: Trecho do <i>Script</i> SQL para mapeamento dos valores dos atributos.	61
Figura 18: Discretizando o atributo idade estimada com faixa de valores.....	62
Figura 19: Representação do algoritmo da <i>FP – Tree</i>	26
Figura 20: Modelo de mineração de dados no <i>RapidMiner</i>	64
Figura 21: Representação textual das regras de associação extraídas.	65
Figura 22: Representação gráfica das associações extraídas.	65
Figura 23: Associações geradas.	66

SUMÁRIO

1	INTRODUÇÃO	7
1.1	Contextualização do problema	8
1.2	Motivação e justificativa	9
1.3	Objetivo.....	11
1.3.1	Objetivos específicos	11
1.4	Organização do trabalho.....	12
1.5	Metodologia	13
1.6	Atividades preliminares.....	13
1.7	Materiais e métodos.....	15
2	CONCEITOS GERAIS SOBRE MINERAÇÃO DE DADOS.....	16
2.1	Descoberta de conhecimento em banco de dados	16
2.2	Etapas da mineração de dados	17
2.2.1	Identificação do problema	18
2.2.2	Pré-Processamento.....	18
2.2.3	Extração de padrões	19
2.2.4	Pós-processamento	19
2.3	Tarefas da mineração de dados	20
2.3.1	Tarefas de previsão.....	20
2.3.2	Tarefas descritivas	21
2.4	Mineração por regras de associação.....	23
2.5	O algoritmo <i>FP-Growth</i>	25
3	MINERAÇÃO DE DADOS NA SEGURANÇA PÚBLICA	27
3.1	Polícia: noções gerais e estrutura policial brasileira	27
3.2	Polícia e a Tecnologia da Informação.....	29
3.3	Inteligência policial.....	30
3.4	Inteligência e Segurança Pública.....	32
3.5	Análise criminal e mineração de dados	34
3.6	O boletim de ocorrência.....	37
3.7	A informatização do registro de ocorrência.....	38
3.7.1	Funcionalidades do BOU	41
4	PREPARANDO O AMBIENTE DE TRABALHO	43
4.1	Preparando o sistema operacional	43
4.2	Instalação do sistema gerenciado de banco de dados PostgreSQL.....	44

4.3	Instalação do <i>RapidMiner</i>	46
4.4	Efetuação da restauração do banco de dados.	49
5	PRÉ-PROCESSAMENTO	52
5.1	Entendimento dos Dados.....	52
5.2	Segmentação do banco de dados	53
5.3	Seleção de atributos	54
5.3.1	Justificativas para a escolha de atributos.....	55
5.4	Redução da quantidade de dados	58
5.5	Mapeamento de valores	60
5.6	Discretização	61
6	MINERANDO OS DADOS.....	63
6.1	Mineração perfil do envolvido	63
6.2	Mineração arma x idade do envolvido	65
6.3	Mineração droga x arma	66
6.4	Mineração meio empregado x idade do envolvido.....	66
6.5	Mineração situação do envolvido x droga.....	67
6.6	Mineração título penal x sexo do envolvido	67
7	PÓS-PROCESSAMENTO.....	69
7.1	Analisando os padrões descobertos	69
8	CONCLUSÃO.....	72
9	REFERÊNCIAS.....	74
	APÊNDICE	78

1 INTRODUÇÃO

O grande avanço percebido em diversas áreas da computação nos últimos anos tem permitido o armazenamento de grandes volumes de dados. É comum muitas aplicações empresariais chegarem na ordem de *gigabytes*, *terabytes* e até *petabytes* de informações (CARVALHO, 2005). Porém, nem sempre os sistemas e metodologias tradicionais de recuperação, pesquisa e análise de dados são suficientes para atender a demanda de informações das organizações, motivo pelo qual a necessidade de ferramentas computacionais apropriadas que viabilizem a análise dessa grande quantidade de informações. Como o volume de dados tem aumentado drasticamente, a análise manual das informações tem se tornado impraticável pelo ser humano (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 2012). Mais do que isso, é preciso que o resultado dessa análise seja compreensível pelo ser humano.

Nesse cenário surgem as técnicas de mineração de dados (MD) ou *data mining*, consistente na aplicação de algoritmos apropriados que procuram descobrir em bases de dados novos padrões válidos, inteligíveis ao ser humano e potencialmente úteis. É a extração de um conhecimento implícito, até então desconhecido, que poderá ser útil ao homem (Figura 1).



Figura 1: Esquema simplificado da mineração de dados (Adaptado de Quoniam *et al.* (2001, p. 21)).

A aplicação prática da MD nos diversos ramos do conhecimento tem crescido ultimamente. Ferramenta de apoio para planos de vendas e *marketing* (RUD, 2001), aplicações científicas, estudos de doenças no campo da medicina são apenas

algumas das situações que podem ser beneficiadas por essa área de Inteligência Artificial.

Tendo em vista essa versatilidade do processo de extração do conhecimento e motivado pela experiência profissional como Delegado de Polícia do Estado do Paraná, o presente trabalho propõe a aplicação das técnicas de mineração de dados no âmbito policial, como instrumento de auxílio à tomada de decisão no campo da segurança pública. Para tanto, o trabalho foi realizado na principal base de dados do sistema de registro de ocorrências denominado Boletim de Ocorrência Unificado.

Com estudo e aprofundamento nas técnicas que compõem a tecnologia em questão, objetiva-se a extração de um conhecimento implícito que possa ser útil e significativo.

O desafio é grande, partindo-se da dificuldade burocrática em se obter a base pretendida objeto da pesquisa, na formulação das hipóteses, na identificação e aplicação dos algoritmos apropriados e, ao final, na interpretação dos resultados alcançados.

1.1 Contextualização do problema

Técnicas de mineração de dados estão cada vez mais sendo utilizadas em diversas áreas. Mediante a coleta e análise de dados relativos a clientes é possível identificar hábitos de consumo e elaborar perfis de compra; da mesma forma é possível estabelecer correlações entre diversos produtos adquiridos pelo cliente visando elaborar ofertas e vendas cruzadas; no campo da biologia, a MD também pode ser utilizada para estudo e classificação de espécies da flora e fauna; na ciência e medicina o estudo do genoma e fatores genéticos podem ser auxiliados com algoritmos de agrupamentos (PONCE; KARAHOCA, 2009) ou na implementação de grande bases de dados contendo sintomas de doenças e dados clínicos que poderão ser pesquisados de forma inteligente por algoritmos apropriados (FERREIRA *et al.*, 2012). Outrossim, órgãos governamentais estão utilizando técnicas de mineração de dados para auxiliar na apuração de crimes ou descoberta de fraudes. Carvalho (2005, p. 5) cita como exemplo o trabalho do

governo dos EUA que utiliza técnicas de MD para identificar padrões de transferência de fundos internacionais que possam indicar lavagem de dinheiro de dinheiro do narcotráfico.

Seguindo essa linha de raciocínio, o contexto do presente projeto situa-se na aplicação de mineração de dados no campo policial. A pesquisa baseia-se no conhecimento e experiência do pesquisador como policial, exercendo o cargo de Delegado de Polícia. Valendo-se do conhecimento especialista no domínio do problema como elemento norteador para as pesquisas, o trabalho propõe enriquecer os métodos de inteligência policial já disponíveis para a prevenção e combate ao crime.

1.2 Motivação e justificativa

A curiosidade é inerente ao ser humano, o qual está sempre buscando conhecimento, extraindo informações, realizando interpretações. Em contrapartida, a evolução dos meios computacionais permitiu que a produção e propagação do conhecimento chegassem a um patamar nunca antes visto. Mas antes de prosseguir, é forçoso tecer breves considerações sobre o domínio do problema alvo de pesquisa.

Essencialmente, a função da Polícia Civil resume-se na investigação criminal, ou seja, na apuração das infrações penais, apurando se houve ou não a existência do crime e sua respectiva autoria. O trabalho de investigação é uma atividade complexa e demanda a realização de diversos atos. Mas há um fator em comum em toda a atividade investigativa: a informação. A informação é a matéria prima fundamental no processo de apuração criminal. É por intermédio da informação que as correlações entre fatos são realizadas, que indícios são constatados ou que decisões são tomadas. Somente assim o trabalho investigativo vai tomando forma em busca do seu objetivo que é a elucidação do crime.

Por sua vez, a informação pode chegar ao conhecimento da autoridade policial de diversas maneiras. É possível tomar conhecimento de determinado fato por meios informais como conversas, denúncias anônimas, boatos, constatação

pessoal ou por documentos elaborados para esse fim dentre os quais o que se destaca é o boletim de ocorrência.

Tendo em mente o âmbito estadual, não é difícil imaginar o quanto de informação é diariamente produzida somente com os dados inseridos nos boletins de ocorrência ou, por outro modo, o quanto de volume de dados é passível de análise para eventuais providências. As centenas de Delegacias de Polícia espalhadas pelo Estado produzem os boletins de ocorrência que retratam o panorama da criminalidade do Paraná.

Muito embora essa massa de dados represente uma rica fonte para análise e estudo, como autoridade policial é possível intuir que as informações coletadas no contexto do registro de ocorrência deveriam ter um tratamento qualitativo. Para tanto, a principal fonte de conhecimento criminal é consubstanciada pela base de dados dos registros de ocorrência.

Lamentavelmente, praticamente não há aplicação de ferramentas inteligentes de análise desses dados, o que representa um grande desperdício de potencial de conhecimento.

Seria de grande valia, se houvesse o emprego de técnicas que pudessem extrair um significado implícito nestas bases de dados, revelando informações valiosas que poderiam servir de suporte a tomada de decisão. Vários questionamentos poderiam ser respondidos ou algumas verdades reveladas, como por exemplo, a correlação entre os vários tipos de delitos, a influência de aspectos temporais na prática delituosa, análise de motivação delituosa ou até mesmo a influência geográfica no mapa da criminalidade são algumas questões que poderiam ser respondidas com um trabalho computacional apropriado. E é possível ir ainda mais longe se fosse considerado que a confiabilidade de tais resultados poderia influenciar no planejamento estratégico de prevenção e combate à criminalidade no contexto da segurança pública e em âmbito estadual.

Nesse cenário é que se propõe a aplicação de técnicas de mineração de dados visando a descoberta de um novo conhecimento implícito na base de dados do sistema Boletim de Ocorrência Unificado.

Trata-se de um novo território a ser explorado e um grande desafio a ser vencido. O trabalho poderá servir de semente para projetos mais ambiciosos, que igualmente possam contribuir para a melhoria da segurança pública do Paraná.

1.3 Objetivo

Objetiva-se com este trabalho promover a aplicação de técnicas de mineração de dados na base de dados do Boletim de Ocorrência Unificado (BOU) visando a extração de conhecimento implícito e útil no contexto da segurança pública, assim como avaliar o resultado alcançado mediante análise comparativa com dados estatísticos previamente disponíveis da área de segurança pública.

As informações extraídas e analisadas com qualidade suficiente poderão servir de suporte à tomada de macro decisões em patamar estratégico e gerencial da Polícia Civil, servindo de motivação para trabalhos futuros mais aprofundados.

Outras hipóteses passíveis de análise computacional poderão ser levantadas no decorrer do trabalho, enriquecendo o processo de extração do conhecimento.

1.3.1 Objetivos específicos

No decorrer da pesquisa, alguns objetivos foram atingidos para aprendizagem e implementação do trabalho:

- Identificar o problema, formulando-se hipóteses e questionamentos passíveis de solução por intermédio da descoberta do conhecimento pelas técnicas de mineração de dados;
- Extrair e analisar uma base de dados objeto de estudo e de MD;
- Aplicar técnicas de pré-processamento visando preparar a base de dados para a extração do conhecimento mediante MD;
- Saber identificar a viabilidade de utilização dos diversos algoritmos utilizados no processo de mineração de dados tendo em vista os objetivos a serem alcançados com a descoberta do conhecimento;
- Aplicar algoritmos para a extração do conhecimento;
- Utilizar técnicas de testes e validação para garantir a confiabilidade do resultado alcançado;

- Fazer uso de recursos visuais para que os resultados sejam compreensíveis ao ser humano;
- Interpretar e avaliar o resultado apresentado como novo conhecimento descoberto em consonância com as regras que compõem o domínio do problema;

1.4 Organização do trabalho

Definidos o escopo de trabalho e a delimitação do tema é possível antever as etapas de pesquisa e os tópicos a serem abordados na elaboração do texto. Sem ter a pretensão, por ora, de esgotar os pontos que serão discutidos é possível organizar a monografia seguinte forma:

- ✓ Capítulo 1: introdução sobre o trabalho e o que se pretende com a realização da pesquisa;
- ✓ Capítulo 2: abordagem sobre os conceitos envolvendo inteligência artificial e mineração de dados;
- ✓ Capítulo 3: situa-se o leitor no domínio do problema, apresentando noções do trabalho policial investigativo, contextualização da mineração de dados na segurança pública e considerações sobre o sistema e o escopo da base de dados objeto de pesquisa - o Boletim de Ocorrência Unificado;
- ✓ Capítulo 4: comentários sobre a configuração do ambiente de trabalho;
- ✓ Capítulo 5: detalhes sobre a etapa de pré-processamento com a preparação dos dados alvos de mineração;
- ✓ Capítulo 6: descrição do trabalho prático realizado na mineração de dados propriamente dita, com métodos adotados objetivando a descoberta do conhecimento;
- ✓ Capítulo 7: descrição sobre a etapa de pós-processamento com considerações sobre os resultados apresentados;
- ✓ Capítulo 8: conclusão.

1.5 Metodologia

O presente trabalho foi fundamentalmente baseado na pesquisa bibliográfica, com revisão de literatura, e experimentos consubstanciados pela exploração da base de dados disponível, utilização de ferramentas e aplicação de técnicas pertinentes à mineração de dados.

Em linhas gerais, o trabalho consistiu nas seguintes etapas:

- Revisão bibliográfica: estudo das teorias, conceitos e técnicas relativas à mineração de dados e o domínio do problema;
- Obtenção da base de dados: obter nos *datacenters* da CELEPAR (Companhia de Informática do Paraná) a réplica do banco de dados referente ao sistema Boletim de Ocorrência Unificado;
- Mineração e validação das regras obtidas: extração das regras obtidas pelos experimentos;
- Visualização e interpretação dos resultados: análise, interpretação e avaliação das informações extraídas das regras.

1.6 Atividades preliminares

Inicialmente, a proposta do trabalho foi colocada em discussão em contato mantido com o Delegado Chefe da Agência de Inteligência da Polícia Civil, ocasião em que foram aventados alguns projetos no âmbito da Tecnologia de Informação que poderiam ser desenvolvidos no intuito de enriquecer a inteligência policial e auxiliar no suporte a tomada de decisões. A princípio vislumbrou-se na possibilidade de se realizar uma exploração na principal base de dados utilizada pela Polícia Civil: o Boletim de Ocorrência Unificado. De fato, este foi o domínio definido para a realização do projeto.

Concluiu-se após discussões com até então pretense orientador de pesquisa, Professor e Mestre Glauco Carlos Silva, que a implementação de técnicas de inteligência artificial e mineração de dados seriam perfeitamente viáveis e adequadas para a consecução do objetivo proposto. Aceita a viabilidade da pesquisa acadêmica, a partir daí iniciaram-se as tratativas visando à obtenção da base de dados pretendida.

Vale ressaltar que a base de dados citada contém dados sigilosos por natureza – registro de todas as ocorrências policiais do Estado – fruto do trabalho policial diário e encontra-se armazenada nos *datacenters* da CELEPAR, empresa responsável pelo desenvolvimento dos softwares utilizados nos órgãos públicos do Paraná. Desse modo, considerando a relevância das questões envolvidas, foi necessário vencer alguns entraves burocráticos.

Sendo assim, o Delegado Chefe da Coordenação de Informática da Polícia Civil (CIN) foi cientificado acerca da proposta do trabalho e mostrou-se favorável à ideia. Em contrapartida, foi realizada reunião com a diretoria da CELEPAR (Figura 2) que, igualmente, sinalizou de forma positiva na autorização ao acesso às bases de dados e disponibilização de *backups* para fins de pesquisa. Por conseguinte, foi disponibilizada uma réplica utilizada em desenvolvimento do banco de dados do sistema Boletim de Ocorrência Unificado contendo dados referentes aos anos de 2005 a 2011.



Figura 2: Reunião ocorrida na CELEPAR ocasião em que foram acertados os detalhes para a disponibilização da base de dados. Da esquerda para direita: analistas Rubens e Valter, Dr. Castella (Delegado Chefe da CIN), Prof. Orientador Glauco e acadêmico Alessandro

1.7 Materiais e métodos

O trabalho foi desenvolvido com a utilização de algumas ferramentas e tecnologias. De antemão foi averiguado que a base de dados objeto de estudo foi modelada no sistema gerenciador de banco de dados relacional *PostgreSQL*. Trata-se de um dos SGBD's *open source* mais utilizados no mercado.

Todavia, as principais atividades do processo de extração do conhecimento foram realizadas com o auxílio do software de código aberto *RapidMiner*. Cuida-se de uma solução inovadora desenvolvida pela empresa *Rapid-i* que agrega uma gama completa de recursos destinados à mineração, integração e transformação de dados, bem como métodos de modelagem. O software tem tido grande aceitação na comunidade. Para o desenvolvimento do trabalho foi utilizada uma versão *community* da ferramenta, sem custos de licença (Figura 3).

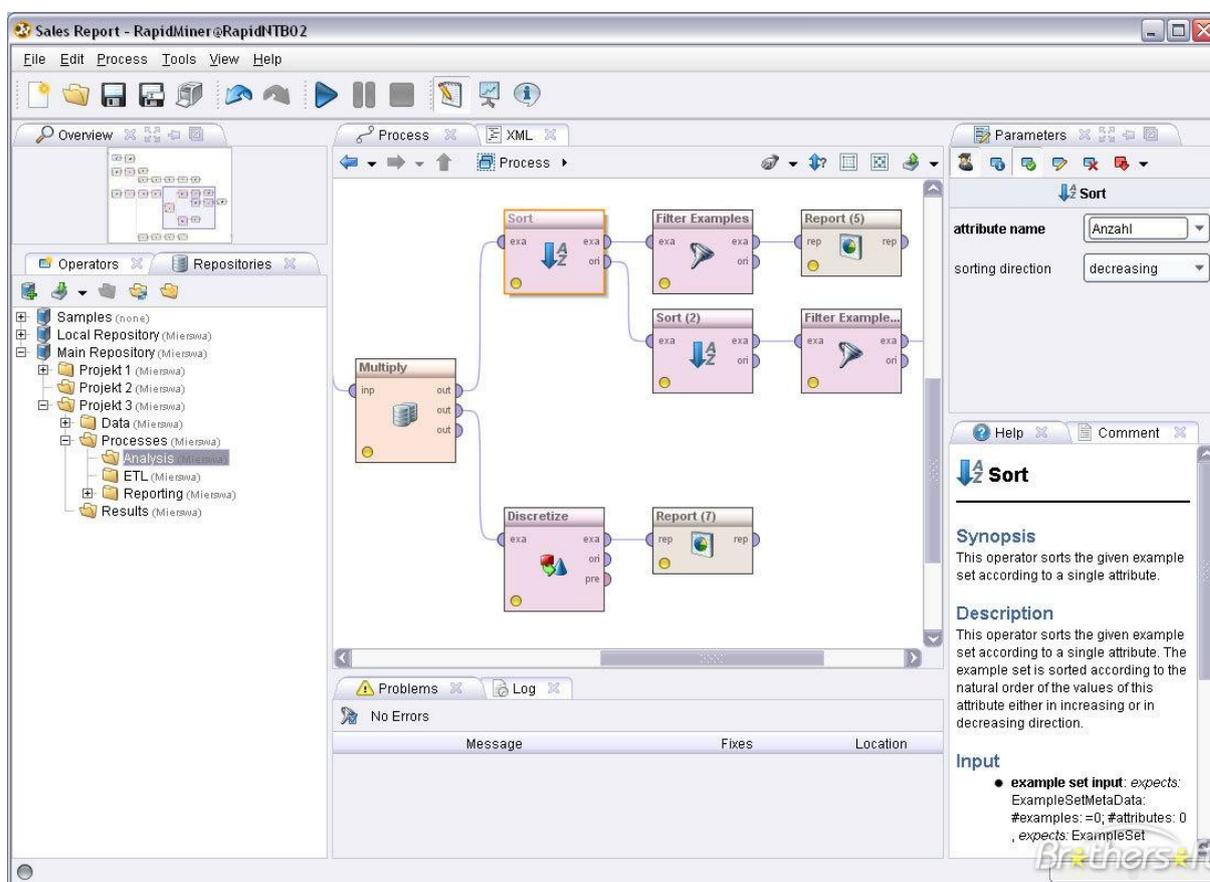


Figura 3: Screenshot do software *RapidMiner* (Fonte: www.rapid-i.com)

2 CONCEITOS GERAIS SOBRE MINERAÇÃO DE DADOS

2.1 Descoberta de conhecimento em banco de dados

Conforme já foi dito, o avanço na área de Tecnologia da Informação têm viabilizado o armazenamento de grande quantidade de dados. É comum em aplicações corporativas de grande porte a presença de bases de dados da ordem de centenas de *terabytes*.

Partindo da premissa que a análise de grandes quantidades de dados é inviável ao ser humano e que as organizações demandam por respostas que não são eficazmente respondidas pelos métodos tradicionais de consulta de dados, surgem nesse cenário as técnicas de mineração de dados que promovem a descoberta de um conhecimento implícito que possa ser útil ao homem.

Segundo Tan, Steinbach e Kumar (2009, p.3),

“Mineração de dados é o processo de descoberta automática de informações úteis em grandes depósitos de dados. As técnicas de mineração de dados são organizadas para agir sobre grandes bancos de dados com o intuito de descobrir padrões úteis e recentes que poderiam, de outra forma, permanecer ignorados.”

No mesmo sentido, Rezende (2005, p.309) leciona que “Extração do Conhecimento de Base de Dados é o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados.”

O termo mineração de dados está inserido numa área do conhecimento computacional denominada Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Database – KDD*), presente no contexto de inteligência artificial.

Segundo Gorunescu (2011, p. 2) esse processo de descoberta do conhecimento está fundamentado em três pilares:

- **Estatística:** o pilar mais velho sem a qual talvez não existisse a mineração de dados. Métodos estatísticos de análise exploratória de dados (*Exploratory*

Data Analysis - EDA) permitem identificar relações entre diferentes variáveis. Tais métodos se utilizam de técnicas que são aplicáveis na mineração de dados, tais como: métodos computacionais – estatística descritiva, correlação, tabelas de frequência, séries temporais, dentre outros; e visualização de dados, ou seja, artifícios que possibilitam representar as informações de forma visual para melhor entendimento (gráficos, por exemplo);

- **Inteligência Artificial (IA):** ramo da computação que contribui com técnicas de processamento de informação baseadas no modelo de raciocínio humano. Outrossim, métodos de aprendizagem de máquina (*Machine Learning – ML*) são amplamente utilizados na mineração de dados com a aplicação de técnicas que permitam o computador aprender com “treinamento”. São utilizadas estruturas formais (máquinas) para fazer inferência (CLARKE; FOKOUÉ; ZHANG, 2009, p. 2);
- **Banco de dados:** é considerado o terceiro pilar da mineração de dados fornecendo informações a serem extraídas com a aplicação dos métodos adequados.

2.2 Etapas da mineração de dados

O processo de mineração de dados é dividido em etapas. Embora não haja completo consenso na literatura quanto à divisão das etapas que compõem o processo, é possível dividi-las da seguinte forma (Figura 4):

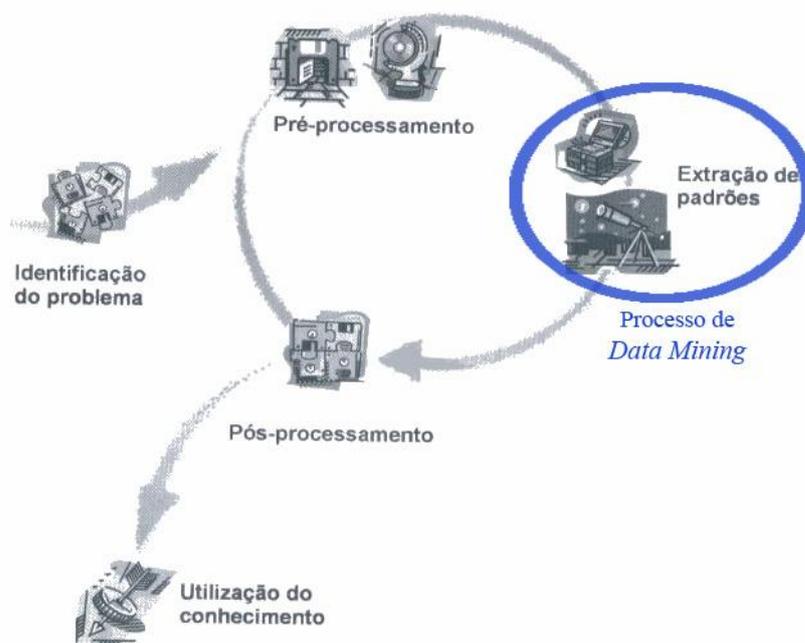


Figura 4: Diagrama esquemático do processo de mineração de dados. (REZENDE, 2005, p. 312)

2.2.1 Identificação do problema

Para o sucesso do processo de extração do conhecimento, é fundamental o “estudo do domínio da aplicação e a definição de objetivos e metas a serem alcançadas no processo de Mineração de Dados”. (REZENDE, 2005, p. 312). É importante a participação de especialistas no problema a ser proposto, fornecendo informações, definindo metas, objetivos e restrições que viabilizarão o êxito do processo.

2.2.2 Pré-Processamento

Nem sempre os dados disponíveis para análise estão adequados ao processo de extração do conhecimento. Sendo assim, a “etapa de pré-processamento compreende as funções relacionadas à captação, à organização, ao tratamento e à preparação dos dados para a etapa da Mineração de Dados.” (GOLDSCHMIDT e

PASSOS, 2005, p. 26). Trata-se de métodos de tratamento, limpeza e redução de dados que devem ser aplicados antes do processo de extração do conhecimento. Esta etapa engloba uma série de atividades que podem ser ou não aplicadas cumulativamente, dentre as quais é possível citar: limpeza, transformação, seleção e redução de dados, redução de valores, dentre outros.

2.2.3 Extração de padrões

É a extração do conhecimento propriamente dita visando alcançar os objetivos propostos na fase de identificação do problema. Trata-se de processo iterativo, que pode ser executado diversas vezes o quanto for necessário para o sucesso da extração. Esta etapa compreende a escolha da tarefa de mineração a ser empregada, escolha de algoritmos e o processo de extração em si. As tarefas do processo de extração do conhecimento são divididas em dois grandes grupos: tarefas de previsão e descritivas, as quais serão explicadas mais adiante.

2.2.4 Pós-processamento

Esta etapa engloba a visualização, a análise e a interpretação do modelo de conhecimento gerado pela etapa de mineração de dados. Nesta fase, o especialista do domínio da aplicação avalia os resultados alcançados, podendo definir novas alternativas de investigação dos dados. “Em geral, é nesta etapa que o especialista em KDD e o especialista no domínio da aplicação avaliam os resultados obtidos e definem novas alternativas de investigação dos dados” (GOLDSCHMIDT e PASSOS, 2005, p. 55).

Além disso, o conhecimento extraído deve ser compreensível ao ser humano. Após a análise, caso o conhecimento não seja aproveitável ou de interesse do usuário, o processo poderá se repetir com novos parâmetros de extração objetivando a solução do problema.

Domingues e Rezende (2005, p. 28) sugerem alguns métodos e procedimentos que poderão ser aplicados nesta fase, tais como:

- **Avaliação:** significa avaliar o conhecimento extraído da base de dados aplicando critérios como precisão, compreensibilidade, interessabilidade, dentre outros;
- **Interpretação e explicação:** visualizar, documentar, comparar, modificar o conhecimento extraído de forma a torna-lo compreensível ao usuário final;
- **Filtragem:** selecionar ou filtrar o conhecimento extraído do conjunto de dados mediante a aplicação de vários mecanismos ou técnicas.

Após a análise do conhecimento no pós-processamento, as informações extraídas poderão alimentar sistemas inteligentes ou servir de suporte à tomada de decisão.

Outra abordagem muito utilizada para pós-processar o conhecimento descoberto consiste em separar as técnicas de análise em duas categorias: métodos subjetivos (ou *user-driven*) e objetivos (ou *data-driven*) (CARVALHO *et al.*, 2003, p. 135). O método subjetivo utiliza as crenças ou conhecimento prévios do usuário, a partir do qual o sistema irá minerar o conjunto de padrões mediante a aplicação do algoritmo de MD apropriado. Em contrapartida, o método objetivo não necessita de um conhecimento previamente estabelecido.

2.3 Tarefas da mineração de dados

Basicamente, as tarefas inerentes à mineração de dados estão divididas em dois grandes grupos (TAN; STEINBACH; KUMAR, 2009, p. 9): tarefas de previsão e tarefas descritivas.

2.3.1 Tarefas de previsão

Também chamadas de tarefas de predição ou mineração de dados preditiva. Tem como objetivo prever o valor de determinado atributo com base nos valores de outros atributos. Resume-se na “avaliação do valor futuro de algum índice baseando-se em dados do comportamento passado deste índice” (CARVALHO, 2005, p. 21). Neste grupo estão inseridas as tarefas de classificação e regressão.

A classificação é uma das tarefas mais utilizadas no processo de mineração de dados, pois se assemelha às atividades cognitivas do ser humano no auxílio à compreensão do ambiente. O indivíduo está sempre classificando o que percebe ao seu redor, criando classes de entidades e dando a cada classe tratamento diferenciado. Conforme Rezende (2005, p. 323) a “tarefa de classificação é uma função de aprendizado que mapeia dados de entrada, ou conjuntos de dados de entrada, em um número finito de categorias.” Cada exemplo pertence a uma classe, num conjunto de classes pré-definidas.

A tarefa de regressão é semelhante à classificação, porém, o atributo a ser previsto é contínuo (numérico) e em vez de discreto.

2.3.2 Tarefas descritivas

Geralmente de cunho exploratório, objetivam derivar padrões, isto é, correlações, tendências, grupos, trajetórias e anomalias que resumam os relacionamentos subjacentes nos dados. Silva (2004, p. 15) ressalta que os “dados são peneirados à procura de padrões de ocorrências frequentes, tendências e generalizações sobre os dados sem intervenção ou direção do usuário.” Dentre os métodos empregados neste grupo cite-se a extração de regras de associação, sumarização e *clustering*.

Uma “regra de associação caracteriza o quanto a presença de um conjunto de atributos nos registros de uma Base de Dados implica a presença de algum outro conjunto distinto de atributos nos mesmos registros” (AGRAWAL e SRIKANT, 1994 apud REZENDE, 2005, p. 323).

A tarefa de clusterização ou agrupamento particiona os registros de uma base em subconjuntos de forma que os elementos de um subconjunto ou *cluster*

compartilhem propriedades comuns que os diferenciem dos elementos de outros clusters (GOLDSCHMIDT e PASSOS, 2005, p. 74).

Por sua vez, a sumarização ou descrição de conceitos estabelece métodos para descrever concisamente um subconjunto de dados, isto é, permite identificar e apresentar de forma compreensível as principais características contidas em um conjunto de dados.

A Figura 5 sintetiza as tarefas mencionadas.

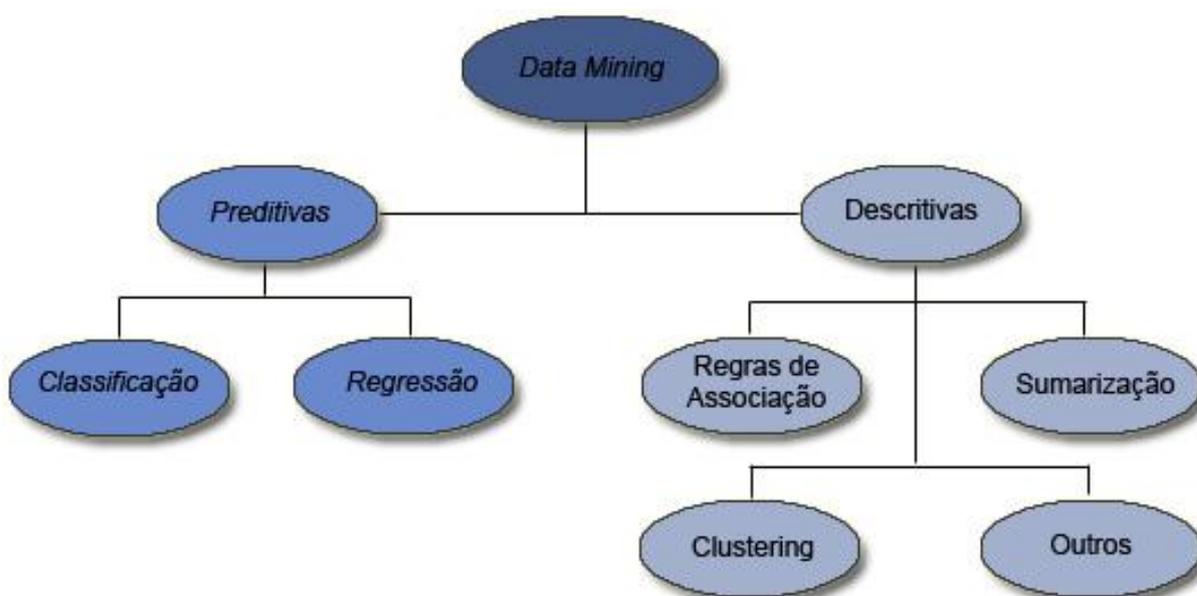


Figura 5: Tarefas da mineração de dados (REZENDE, 2005 p. 318).

Escolhida a tarefa a ser empregada, o próximo passo é escolher os algoritmos apropriados à extração do conhecimento.

O presente trabalho priorizou as atividades descritivas, especialmente a extração e regras de associação. Questões como perfil dos envolvidos em uma ocorrência policial ou averiguação de relacionamentos entre crimes, aspectos geográficos e outros poderão ser respondidas com a utilização de técnicas descritivas.

2.4 Mineração por regras de associação

Com o avanço da tecnologia, sobretudo dos meios de armazenamento de dados, tem sido cada vez mais comum o acúmulo de grande quantidade de informações. Situação corrente é do registro dos dados referentes a compras de clientes em balcões de mercadorias, cujo objetivo é registrar os itens comprados pelos clientes.

A ideia é útil, haja vista que pode ser interessante para o varejista analisar os dados para averiguar o comportamento de compra dos seus clientes. Tais informações podem ser valiosas para servir de suporte a outros setores relacionados ao negócio, como *marketing*, vendas, estoque, dentre outros.

O contexto é apropriado para a extração do conhecimento por regras de associação, aplicada para descobrir relacionamentos interessantes implícitos em grandes quantidades de dados. Os relacionamentos descobertos são representados na forma de itens frequentes ou regras de associação. Nesta etapa, o objetivo é descobrir regras de associação do tipo $X \rightarrow Y$ (leia-se SE (X) ENTÃO (Y)), onde X e Y são conjuntos de itens, $X \cap Y = \emptyset$. O significado da regra é que os conjuntos de itens X e Y frequentemente ocorrem juntos em uma mesma transação (CALIL *et al.*, 2008, p. 210).

O Quadro 2 exemplifica itens adquiridos em diversas transações de compra e venda:

QUADRO 2 – TRANSAÇÕES DE COMPRA E VENDA

	Itens
1	{Pão, Leite}
2	{Pão, Fraldas, Cerveja, Ovos}
3	{Leite, Fraldas, Cerveja, Cola}
4	{Pão, Leite, Fraldas, Cerveja}
5	{Pão, Leite, Fraldas, Cola}

(TAN; STEINBACH; KUMAR, 2009, p. 390).

{Fraldas \rightarrow Cerveja}

Observa-se um forte relacionamento entre os itens cerveja e fralda de modo a formar uma regra. É o clássico exemplo de associação descoberta por uma rede de lojas varejista no qual o cliente que comprava fraldas também adquiria cerveja. Esse tipo de conhecimento descoberto pode ser útil para identificar novas oportunidades de vendas.

A ideia também se aplica a outros domínios, tais como pesquisas científicas, médicas, análise geográficas, análise criminais, dentre outros. Portanto, é possível intuir que a opção por extração de conhecimento por regras de associação neste trabalho não foi aleatória. Regras de associação são úteis para descobrir padrões, itens frequentes em uma base de dados e correlação entre atributos. Em um repositório de dados sobre ocorrências policiais, a tarefa descritiva poderia revelar relacionamentos entre perfil do envolvido e ocorrências; meios empregados na prática delituosa e situação do envolvido; emprego de armamento com relação a ocorrência envolvendo entorpecentes; relação entre o sexo do envolvido e delitos, dentre outros.

Geralmente, os algoritmos de mineração de regras de associação abordam o problema em duas etapas:

1. **Geração de conjuntos de itens frequentes:** busca encontrar todos os conjuntos de itens (*itemsets*) que satisfaçam o número mínimo de transações que contenham um determinado conjunto de itens (*minsup*).
2. **Geração de regras:** visa extrair as regras dos conjuntos de itens frequentes conforme o grau de confiança estipulado.

Conceitos comuns na mineração de dados por regras de associação são o suporte e confiança. O suporte diz respeito a frequência na qual uma regra é aplicada a determinado conjunto de dados. A confiança se refere à frequência na qual os itens em Y aparecem em transações que contenham X (TAN; STEINBACH; KUMAR, 2009, p. 392).

Então, dado um conjunto de transações, descobrem-se todas as regras possíveis quando o suporte e a confiança são iguais ou superiores aos limites pré-definidos (GORONESCU, 2011, p. 251).

A medida de suporte é importante porque um baixo suporte pode indicar que a regra aconteceu por coincidência ou que ela não tenha tanta relevância para o

negócio.

Por sua vez, a confiança mede a confiabilidade da inferência feita por uma regra.

Tais índices auxiliam na tarefa de interpretação das regras geradas. A interpretação deve ser feita com cautela, tendo em mente as regras de negócio e a causalidade de surgimento das regras.

2.5 O algoritmo *FP-Growth*

O algoritmo *FP-Growth* foi desenvolvido com a intenção de ser uma nova abordagem para a extração de regras de associação, buscando suprir as deficiências do conhecido algoritmo *Apriori*, visando melhor desempenho computacional.

A proposta do algoritmo *FP-Growth* é extrair os *itemsets* frequentes de um modo mais eficiente, prescindindo da etapa de geração de candidatos. Para tanto, o algoritmo utiliza uma estrutura em memória denominada *FP-Tree*, que representa a base de dados. Desse modo, é possível obter os *itemsets* frequentes mediante recursão, utilizando-se estratégias de busca de padrões em árvores. A *FP-Tree* otimiza a utilização de memória, reduzindo o número de varreduras na base de dados, necessitando de apenas dois acessos a disco para a construção da árvore (OYAMA, 2010, p. 114).

Segundo, Hunyadi (2011, p. 378), a maior vantagem do algoritmo *FP-Growth* perante os demais algoritmos de mesmo propósito é que ele realiza apenas duas varreduras na base de dados e pode ser utilizado em grandes conjuntos de dados.

Os nós da *FP-Tree* armazenam informações referentes aos *1-itemsets* frequentes da base de dados, tais como a frequência de ocorrência de cada *itemset*, que será necessária para a extração dos padrões frequentes. O algoritmo da *FP-Tree* pode ser expresso da seguinte forma (Figura 19):

```

Entradas: Base de dados D, minsup
Saída: FP-Tree T

F := 1-itemsets_frequentes(D, minsup);
L = ordena_decrescente(F);
criaNoh(T, null);
para cada transação Trans em D {
    [p|P] = ordena (Trans, L);
    Insere_tree([p|P, T);
}
retorne T;

```

Figura 6: Representação do algoritmo da *FP-Tree* (HAN, PEI, YIN, 2000 *apud* OYAMA, 2010, p.115)

1. Realiza-se uma varredura na base de dados obtendo-se os *1-itemsets* frequentes;
2. O conjunto de *1-itemsets* frequentes é ordenado de forma decrescente pelo valor de suporte;
3. Adição do nó raiz da árvore;
4. Nova varredura da base de dados realizada em duas etapas: ordenação dos itens e inserção do nó na árvore;

Após a inserção e ordenação dos nós, inicia-se a etapa de mineração propriamente dita da *FP-Tree* com o objetivo de descobrir os padrões frequentes. De modo geral, o processo consiste em gerar os padrões condicionais para cada *1-itemset* existente.

Em razão da otimização de espaço em disco e memória, aliado à capacidade de lidar com grande volume de dados, o algoritmo *FP-Growth* revelou ser a escolha apropriada para a realização do trabalho.

3 MINERAÇÃO DE DADOS NA SEGURANÇA PÚBLICA

3.1 Polícia: noções gerais e estrutura policial brasileira

O homem é um ser essencialmente social. Porém, convivendo na comunidade com seus pares é natural que surjam conflitos. Partindo do princípio que é vedado ao particular agir com a própria força na solução das desavenças, o Estado chamou para si a tarefa da pacificação das lides sociais através do Poder Judiciário.

Por sua vez, a liberdade de agir do indivíduo no meio em que vive não é absoluta, mas limitada pela moral e por mecanismos de controle social como as normas jurídicas e o poder de polícia estatal. A convivência social é regida pela regra de que a liberdade de agir do indivíduo vai até onde começar o direito do outro.

Para Pietro (1997, p. 94), o poder de polícia é a “atividade do Estado consistente em limitar o exercício dos direitos individuais em benefício do interesse público”. É o poder conferido à Administração de impor limites ao exercício de direitos e de atividades individuais em função do interesse público.

O tratamento jurídico da atividade policial inicia-se no artigo 144, da Constituição Federal, de 1988, inserido no Título V, Seção III, Capítulo III – Da Segurança Pública:

Art. 144. A segurança pública, dever do Estado, direito e responsabilidade de todos, é exercida para a preservação da ordem pública e da incolumidade das pessoas e do patrimônio, através dos seguintes órgãos:

I - polícia federal;

II - polícia rodoviária federal;

III - polícia ferroviária federal;

IV - polícias civis;

V - polícias militares e corpos de bombeiros militares.

Da norma é possível extrair alguns significados importantes: a obrigatoriedade do Estado em assegurar ao cidadão segurança pública e gratuita; todos tem sua parcela de responsabilidade na consecução da segurança; o objetivo fundamental da segurança pública é a preservação da ordem pública e da incolumidade das pessoas e do patrimônio; e que a segurança pública se realiza através da Polícia, organizada nos órgãos policiais constitucionalmente enumerados.

Segundo Mirabete (2003, p.87), a “Polícia é uma instituição de direito público, destinada a manter a paz pública e a segurança individual”. A atividade policial divide-se em duas grandes áreas: administrativa e judiciária. A polícia administrativa, preventiva ou ostensiva atua preventivamente na garantia da ordem pública e impedindo o cometimento de crimes, ou seja, de fatos que ponham em risco bens individuais ou coletivos. Em contrapartida, a polícia judiciária ou polícia de investigação tem caráter repressivo, atuando após a prática de uma infração penal e recolhendo elementos para provar a existência do crime e sua respectiva autoria. Conforme a Constituição Federal, às polícias civis, dirigidas por Delegados de Polícia de carreira, incumbem, ressalvada a competência da União, as “funções de polícia judiciária e a apuração das infrações penais, exceto as militares” (art. 144, § 4º).

Resumidamente, a estrutura policial brasileira está organizada da seguinte forma:

- Polícia Federal: destinada a apurar infrações penais contra a ordem política e social ou em detrimento de bens, serviços e interesses da União ou de suas entidades autárquicas e empresas públicas, bem como outras infrações cuja prática tenha repercussão interestadual ou internacional e exija repressão uniforme; prevenir e reprimir o tráfico ilícito de entorpecentes e drogas afins, o contrabando e o descaminho; exercer as funções de polícia marítima, aeroportuária e de fronteira; exercer, com exclusividade, as funções de polícia judiciária da União;
- Polícia Rodoviária Federal: destina-se ao patrulhamento ostensivo das rodovias federais;
- Polícia Ferroviária Federal: realiza o patrulhamento ostensivo das ferrovias federais;

- Polícias Civis: organizadas em cada Estado da federação, são dirigidas por Delegados de Polícia de carreira e incumbidas das funções de polícia judiciária e apuração das infrações penais, com exceção das infrações militares e ressalvada a competência da União;
- Polícias Militares: polícias estaduais de caráter ostensivo para a preservação da ordem pública;
- Corpos de Bombeiros Militares: incumbidos da execução de atividades de defesa civil;

3.2 Polícia e a Tecnologia da Informação

É evidente o progresso que a humanidade tem passado ao longo dos anos. A evolução foi percebida em várias áreas do conhecimento, bem como em fatores socioeconômicos. Infelizmente, a elevação dos índices de criminalidade representa o lado negativo desta evolução. Todos os dias, os veículos de imprensa noticiam tragédias e fatos criminosos que chocam a sociedade. Como se não bastasse, novas modalidades de crimes estão cada vez mais em destaque: crimes do “colarinho branco”, delitos praticados por meios eletrônicos (internet, por exemplo), fraudes, e outros. Ninguém está totalmente a salvo, qualquer um é uma vítima em potencial. Enfim, segurança pública é o assunto sempre em evidência e ponto de preocupação para os governantes.

Partindo-se da premissa que o combate ao crime também deve evoluir e que os organismos policiais de prevenção e repressão devem exercer suas atividades utilizando-se somente de meios legais, é preciso pensar em novas formas de se promover segurança pública. Mais do que isso, é fundamental lançar mão de recursos de inteligência que possam auxiliar no trabalho policial.

A repressão isolada ao crime, desprovida de análise técnica não é mais suficiente. É imperioso estar à frente dos criminosos, definindo alvos, identificando riscos, estabelecendo metas, analisando fatores circunstanciais do crime.

Na esteira desse raciocínio, Castella (2005, p. 130/1) preconiza que

No âmbito de uma delegacia de polícia, o serviço de Inteligência é fundamental. Considerando-se que a atividade-fim da polícia civil é proceder a investigações para apurar a materialidade e a autoria de um delito, ao se fazer o trabalho de Inteligência, há como atuar, também de forma preventiva. Mapeando a incidência constante de determinadas infrações penais, dentro da área de atuação da Delegacia, e procedendo à avaliação da vida pregressa de marginais anteriormente identificados, que atuam praticando os mesmos ilícitos, cria-se a possibilidade de uma vigilância preventiva nesses setores.

Veja-se que o cenário é propício para a aplicação de técnicas de inteligência artificial, em especial a mineração de dados. A possibilidade de extração de um conhecimento implícito, até então desconhecido, em bases de dados que reflitam o panorama da criminalidade estadual favorece uma análise inteligente e estratégica da segurança pública. As informações obtidas poderão ser úteis na prevenção e combate ao crime ou auxiliar no processo de entendimento da criminalidade.

Ressalta-se que o campo jurídico/policial tem se mostrado promissor para o surgimento de pesquisas na área. Em trabalho análogo, Alberto Filho (2011) propõe a aplicação de técnicas de mineração de dados e de textos a partir de bases de dados da segurança pública, sobretudo boletins de ocorrência, para a extração de informações relevantes em auxílio a tomada de decisão. Castella (2005) relata sua experiência de pesquisa na aplicação de Gestão do Conhecimento com Inteligência Artificial na segurança pública. Cite-se também como exemplo o trabalho realizado por Benicasa e Paixão (2006) junto ao sistema de controle de inquéritos policiais do Ministério Público do Estado do Amapá. Aplicando-se o processo de descoberta do conhecimento, os pesquisadores conseguiram traçar um retrato significativo da criminalidade daquele Estado.

3.3 Inteligência policial

Segundo o Dicionário Houaiss (HOUAISS; VILLAR; FRANCO, 2001) o vocábulo inteligência, do latim *intelligentia*, significa: 1 capacidade de aprender. 2 astúcia. É a faculdade de entender, de compreender; destreza, habilidade para escolher os métodos e obter um bom resultado.

Pela análise do conceito é possível identificar três estágios: 1 - interpretar e

entender pessoas, coisas e fatos, mediante análise de situações, pessoas e eventos; 2 - resolução de problemas por meio do entendimento destes fatos; 3 - tomada de decisões através do raciocínio. Após a compreensão dos eventos e determinação de processos para resolução dos problemas, parte-se para o estabelecimento de ações a serem adotadas.

Esta pluralidade de abordagens propiciou o estudo da inteligência sob outros prismas como inteligência emocional, inteligência competitiva, inteligência financeira, inteligência de segurança pública, etc.

Transportando os conceitos para a realidade estatal, costuma-se aceitar que a inteligência pode ser estudada sob três aspectos (ROMÃO, 2012):

1. o conhecimento produzido, focando-se no produto. A atividade de inteligência trata de informações atinentes a segurança do Estado, motivo pelo qual as atividades desempenhadas visam obtê-las ou impedir que organizações não autorizadas as obtenham;
2. a estrutura funcional para obter os dados e as informações necessárias, com ênfase na organização. Tradicionalmente, as entidades responsáveis por realizar as atividades de inteligência são conhecidas por Serviços de Inteligência. São agências governamentais responsáveis pela coleta, análise e disseminação das informações relevantes para o processo de tomada de decisões e para a implementação de políticas públicas na área de segurança, defesa nacional, política externa e ordem pública;
3. os meios utilizados (processos) para a obtenção e produção do conhecimento. Nesse aspecto, o conhecimento é tratado em três frentes principais: a) produção e análise das informações que serão difundidas visando o assessoramento final do usuário final do conhecimento criado; b) contra inteligência: salvaguarda dos dados, conhecimento, instalações e pessoal das instituições; c) planejamento e implementação de operações visando a obtenção de dados negados ou não disponíveis.

No Brasil, a Lei nº 9.883, de 7 de dezembro de 1.999 instituiu o Sistema

Brasileiro de Inteligência (SISBIN) e a Agência Brasileira de Inteligência (ABIN). O SISBIN foi regulamentado pelo Decreto nº 4.376, de 13 de setembro de 2002, que dispõe sobre sua organização e seu funcionamento. Além de estabelecer a estrutura basilar dos serviços de inteligência no país, a lei também se encarregou de estabelecer a definição legal de inteligência. Dispõe o §2º, do artigo 1º da lei:

§ 2º Para os efeitos de aplicação desta Lei, entende-se como inteligência a atividade que objetiva a obtenção, análise e disseminação de conhecimentos dentro e fora do território nacional sobre fatos e situações de imediata ou potencial influência sobre o processo decisório e a ação governamental e sobre a salvaguarda e a segurança da sociedade e do Estado.

O Decreto nº 4.376/02 acompanhou a conceituação e praticamente repetiu a redação anterior, senão vejamos:

Art. 2º Para os efeitos deste Decreto, entende-se como inteligência a atividade de obtenção e análise de dados e informações e de produção e difusão de conhecimentos, dentro e fora do território nacional, relativos a fatos e situações de imediata ou potencial influência sobre o processo decisório, a ação governamental, a salvaguarda e a segurança da sociedade e do Estado.

Analisando o texto é possível constatar que a atividade de inteligência está voltada para o assessoramento do governo, sobretudo no que se refere à segurança pública e soberania do Estado. Trata-se, pois, de importante instrumento de suporte para a tomada de decisões.

3.4 Inteligência e Segurança Pública

Ficou claro que a atividade de inteligência serve como ferramenta de auxílio e assessoramento ao governo, fornecendo informações que poderão ser úteis no processo de tomada de decisão. Por conseguinte, é evidente que a inteligência pode ser de grande valia no âmbito da Segurança Pública.

Na esfera federal o Decreto nº 3.695, de 21 de dezembro de 2.000 ficou encarregado de criar o Subsistema de Inteligência de Segurança Pública. Segundo o artigo 1º:

Art. 1º Fica criado, no âmbito do Sistema Brasileiro de Inteligência, instituído pela Lei no 9.883, de 7 de dezembro de 1999, o Subsistema de Inteligência de Segurança Pública, com a finalidade de coordenar e integrar as atividades de inteligência de segurança pública em todo o País, bem como suprir os governos federal e estaduais de informações que subsidiem a tomada de decisões neste campo.

Por sua vez, o SISP foi regulamentado pela Resolução nº 1, de 15 de julho de 2009, da Secretaria Nacional de Segurança Pública. Foram estabelecidas as bases do subsistema, sua finalidade e âmbito de atuação:

Art. 1º O Subsistema de Inteligência de Segurança Pública - SISP, que compõe o Sistema Brasileiro de Inteligência - SISBIN, constituído de rede própria e responsável pelo processo de coordenação e integração das atividades de inteligência de segurança pública no âmbito do território nacional, tem por objetivo fornecer subsídios informacionais aos respectivos governos para a tomada de decisões no campo da segurança pública, mediante a obtenção, análise e disseminação da informação útil, e salvaguarda da informação contra acessos não autorizados.

Além de estruturar a Inteligência da Segurança Pública (ISP) nacional, a resolução também foi útil em explicar expressamente em seu texto (incisos, do §4º, do art. 1º) alguns conceitos relacionados ao tema. Vale a pena transcrevê-los:

I - Inteligência: é a atividade que objetiva a obtenção, análise e disseminação de conhecimentos, dentro e fora do território nacional, sobre fatos e situações de imediata ou potencial influência sobre o processo decisório e a ação governamental e sobre a salvaguarda e a segurança da sociedade e do Estado;

III - Inteligência de Segurança Pública: é a atividade permanente e sistemática via ações especializadas que visa identificar, acompanhar e avaliar ameaças reais ou potenciais sobre a segurança pública e produzir conhecimentos e informações que subsidiem planejamento e execução de políticas de Segurança Pública, bem como ações para prevenir, neutralizar e reprimir atos criminosos de qualquer natureza, de forma integrada e em subsídio à investigação e à produção de conhecimentos;

IV - Inteligência Policial: é o conjunto de ações que empregam técnicas especiais de investigação, visando a confirmar evidências, indícios e a obter conhecimentos sobre a atuação criminosa dissimulada e complexa, bem como a identificação de redes e organizações que atuem no crime, de forma a proporcionar um perfeito entendimento sobre a maneira de agir e operar, ramificações, tendências e alcance de condutas criminosas;

X - Atividade de Inteligência de Segurança Pública: é a atividade técnico-especializada, permanentemente exercida e orientada para a produção e salvaguarda de conhecimentos de interesse da segurança pública que, por seu sentido velado e alcance estratégico, configurem segredos de interesse do Estado e das instituições, objetivando assessorar as respectivas chefias em qualquer nível hierárquico.

Não há dúvida acerca da importância que a ISP desempenha no papel de prevenção e combate da criminalidade, inclusive mediante ações integradas e de intercâmbio de informações entre os Estados da federação.

Todavia, para a consecução dos seus objetivos a ISP segue uma metodologia e padrões de ações previamente definidos, em alinhamento com os princípios constitucionais do Estado Democrático de Direito e com a ordem jurídica vigente. Tais preceitos foram sistematizados na Doutrina Nacional de Inteligência de Segurança Pública – DNISP (art. 1º, §3º, da citada Resolução nº 01/09), documento aprovado pela Portaria nº 22, de 22 de julho de 2009 da Secretaria Nacional de Segurança Pública. A DNISP consiste num conjunto de normas, métodos e princípios orientadores que, combinados aos pressupostos éticos, regem as atividades de inteligência de segurança pública, com objetivo de produção e proteção de conhecimentos relativos à Segurança Pública, tanto com o foco estratégico quanto tático (WENDT, 2012).

3.5 Análise criminal e mineração de dados

Não obstante, a ISP também pode valer-se de instrumentos auxiliares para a consecução dos seus objetivos. Nesse sentido, a Análise Criminal (AC) constitui valioso recurso que pode ser utilizado no contexto da Segurança Pública. Para se compreender o que é Análise Criminal, novamente é necessário auxílio da Resolução nº 01/09 que estabelece no art. 1º, §4º, inciso V a definição do conceito. Segundo o texto,

V - Análise Criminal: é um conjunto de processos sistemáticos direcionados para o provimento de informação oportuna e pertinente sobre os padrões do crime e suas correlações de tendências, de forma a apoiar a área operacional e administrativa no planejamento e distribuição de recursos para a prevenção e supressão de atividades criminosas;

Partindo da definição é possível, por interpretação, extrair o sentido e o significado da norma.

- “*Análise criminal é um conjunto de processos sistemáticos*”: pode-se definir processo como um conjunto de atos lógicos e sequencialmente ordenados visando um objetivo determinado. No processo há os elementos de entrada e a saída que é decorrente do processamento. Note-se que pelo conceito a Análise Criminal se vale de vários processos organizados (sistemáticos) e não apenas um.
- “*(...) direcionados para o provimento de informação oportuna e pertinente sobre os padrões do crime e suas correlações de tendências (...)*”: a finalidade dos processos é fornecer informação conveniente, isto é, útil e tempestiva (que seja utilizável no momento apropriado) relacionada a revelar padrões de crimes e suas tendências. Não basta analisar isoladamente os crimes. Para se extrair um padrão de comportamento ou de delito é necessária a análise de um conjunto de dados e que esses dados sejam ainda em quantidade suficiente para revelar correlações ou características comuns entre eles que possa caracterizar um padrão. Por sua vez, pela semelhança entre padrões é possível realizar um prognóstico da criminalidade, revelando tendências que poderão fundamentar, por exemplo, ações preventivas de segurança pública.
- “*(...) de forma a apoiar a área operacional e administrativa no planejamento e distribuição de recursos para a prevenção e supressão de atividades criminosas.*”: as informações produzidas pela Análise Criminal serão utilizadas em âmbito gerencial, servindo de apoio ao processo de tomada de decisão.

Enfim, foi possível constatar que a Análise Criminal é uma atividade complexa, realizada em várias etapas e quase sempre fazendo uso de uma grande massa de dados. Felizmente, atualmente existem diversas ferramentas computacionais que podem ser utilizadas na coleta, processamento, análise e gerenciamento dos dados e informações criminais. Há técnicas e aplicativos específicos que objetivam a realização de pesquisas em grandes bases de dados, averiguando correlações entre os dados e produzindo análises computadorizadas. (FERRO, 2006).

Em contrapartida, é improvável que um ser humano comum (no caso, um investigador ou operador de segurança pública) tenha capacidade cognitiva para

interpretar milhares de registros e informações no tempo hábil exigido, relacionando as informações analisadas com aspectos gerais de criminalidade ou até mesmo com o crime organizado.

Desse modo, a Análise Criminal aliada aos recursos de TI constitui importante recurso que favorece o desenvolvimento de novos meios de “processamento e identificação de delitos, em suas relações ou vínculos, transformando-se em um instrumento indiscutivelmente com o poder de produzir efeito real, efetivo, eficaz e eficiente no controle do fenômeno da criminalidade contemporânea” (JÚNIOR; DANTAS, 2012).

O cenário é propício à mineração de dados. Utilizando técnicas de descoberta do conhecimento em conjunto com conceitos e aplicações de MD, é possível atender aos objetivos da Análise Criminal como recurso para o entendimento da criminalidade. O terreno é fértil e há grande potencial na aplicação na área de segurança pública.

Seguindo a mesma linha de pensamento Júnior e Dantas (2012) discorrem que

A AC pode ser considerada uma técnica de mineração de dados na qual é possível estabelecer conexões entre registros com o propósito de desenvolver modelos baseados em padrões de relações. É mais aplicada nas investigações de comportamento humano, especialmente na área policial, quando determinadas "pistas" são ligadas entre si para solucionar crimes. HARRISON (apud GONÇALVES, 1999, p. 22)

Com uso da TI, os analistas criminais passaram a poder descobrir e interpretar relações ocultas entre informações contidas em grandes volumes de dados, inclusive com a utilização de diagramas intuitivos. Tais diagramas são representações gráficas capazes de indicar conexões de relevante significado para a investigação. Isso se traduz em um trabalho que não envolve apenas a TI. Também a "tecnologia do conhecimento" (e a cognição a ela associada) está envolvida com este processo, em face da necessidade de correlacionar uma imensurável quantidade de dados e informações referentes a uma atividade criminosa.

Portanto, é inegável a importância do emprego das ferramentas apropriadas de TI para o estudo da criminalidade. Nesse panorama, as técnicas de MD tem se mostrado eficaz na descoberta de conhecimentos até então ocultos em grandes volumes de dados, servindo de auxílio aos operadores de segurança pública.

3.6 O boletim de ocorrência

A notícia de um fato criminoso (*notitia criminis*) pode chegar ao conhecimento da autoridade policial por diversas maneiras, seja pelo contato direto com o fato, por meios verbais, ou por documentos elaborados especificamente para este fim. Comumente, o documento mais utilizado para noticiar um fato criminoso é o boletim de ocorrência. É nele em que é transcrita a comunicação oral do noticiante na unidade policial, consignando-se também as providências preliminares a serem adotadas.¹ Assim, grosso modo, o boletim de ocorrência é a transcrição da comunicação de um fato ocorrido e suas circunstâncias.

Apesar do boletim de ocorrência não estar expressamente previsto na legislação processual penal, possui validade jurídica e está enraizado na praxe policial como sendo o principal meio de comunicação de um delito. Geralmente, no âmbito da polícia judiciária, o boletim de ocorrência serve como ponto de partida para a instauração dos diversos procedimentos investigatórios, tais como inquérito policial, termo circunstanciado de infração penal e boletim de ocorrência circunstanciado, este último para apurar atos infracionais praticados por adolescentes infratores.

Seguindo a tramitação normal, depois de elaborado, o B.O. é encaminhado ao Delegado de Polícia que decidirá pela instauração imediata do procedimento

¹ A doutrina jurídica (QUEIROZ, 2000, p.33) classifica as formas de comunicação da *notitia criminis* em:

- **Cognição imediata ou informal:** ciência direta e verbal do fato ou pelo boletim de ocorrência. O boletim de ocorrência é considerado um meio informal de comunicação porque não está previsto no ordenamento jurídico;
- **Cognição mediata ou formal:** o Delegado de Polícia é explicitamente provocado a instaurar o competente procedimento investigatório mediante peças formais que lhe são encaminhadas descrevendo o fato criminoso. Nos termos do artigo 5º do Código de Processo Penal são a requisição da autoridade judiciária ou do Ministério Público, requerimento do ofendido ou requisição do Ministro da Justiça;
- **Cognição coercitiva:** mediante lavratura do auto de prisão em flagrante delito;
- **Cognição por delação ou informação referenciada** obtida em oitivas formais no curso do procedimento investigatório.

investigatório pertinente ou despachará o documento determinando a realização de outras providências preliminares.

Certamente, o B.O. é um dos documentos mais relevantes elaborados numa Delegacia de Polícia em razão do desdobramento de toda atividade policial posterior. Ademais, grande parte das decisões efetivadas na gestão de segurança pública fundamenta-se em dados estatísticos extraídos da base de boletins de ocorrência.

Antigamente, o boletim de ocorrência era elaborado manualmente, em arquivo de texto no computador ou até mesmo em máquinas de escrever. O processo era rudimentar, o documento não era padronizado e as estatísticas eram lacunosas. Mais do que nunca era preciso informatizar o processo, ou ainda, o boletim de ocorrência deveria servir de ferramenta de auxílio para gerar o retrato da criminalidade estadual.

3.7 A informatização do registro de ocorrência

A primeira iniciativa de informatização do registro de ocorrência deu origem ao sistema *BO On-line* que era uma funcionalidade de um sistema maior denominado *Polícia On-line*. O *Polícia On-line* integrava as diversas unidades da Polícia Civil mediante compartilhamento de um banco de dados comum que propiciava uma série de consultas úteis ao trabalho investigatório. Contudo, a solução ainda não era ideal, pois não era adotada por padrão nas demais corporações policiais, ocasionando fragmentariedade e falta de integração das informações.

Em janeiro de 2003 o Governo do Estado instituiu uma equipe multidisciplinar com a incumbência de se elaborar um projeto que viabilizasse o mapeamento criminal do Estado. O projeto foi denominado “Mapa do Crime – Planejamento e Segurança” e constatou a falta de padronização dos registros de ocorrências policiais, bem como a inexistência de integração entre as forças de segurança

pública, haja vista que havia quatro sistemas de registros de ocorrências.

Em setembro do mesmo ano, a Secretaria de Estado da Segurança Pública (SESP) iniciou estudos para o desenvolvimento de um novo sistema de registro de ocorrências policiais denominado Boletim de Ocorrência Unificado (BOU), que veio a se tornar o atual sistema de coleta de informações sobre ocorrências policiais do Estado do Paraná.

O BOU é composto por formulários que são utilizados pela Polícia Militar no atendimento ao público como também pela Polícia Civil no atendimento às vítimas de ilícitos. Ou seja, ambas as corporações policiais se utilizam de um mesmo sistema de registro de ocorrências compondo um único banco de dados. As Figuras 6 e 7 expõem algumas telas do sistema.

As informações constantes no sistema são utilizadas para aprimorar o planejamento policial e subsidiar com informações a tomada de decisão dos gestores da segurança pública.

O BOU foi lançado oficialmente em dezembro de 2005 e regulamentado pela Resolução nº 309, de 15 de dezembro de 2005, da Secretaria de Estado da Segurança Pública.

Mais tarde, em agosto de 2006, o BOU sofreu um processo de reengenharia com diversas atualizações técnicas, resultando na segunda versão do sistema.

Em janeiro de 2007 o BOU foi implantado em todo o Estado e é, atualmente, o único sistema informatizado de registro de ocorrências policiais no Paraná.

COMUNICAÇÃO DE OCORRÊNCIA (etapa 1 de 3)

(*) Campo de preenchimento obrigatório.

*Origem da Comunicação: 2 - DIRETAMENTE AO ÓRGÃO POLICIAL | *Data de Registro: 22/01/2007 | *Hora do Registro: 09:00

Envolvidos
 Noticiantes: 0 [Inserir](#) | Envolvidos (I.E.): 0 [Inserir/Alterar](#)

DADOS DA OCORRÊNCIA
 *Município: CURITIBA | *UF: PR | *Bairro:
 *Logradouro: | *Número: | Complemento:
 Proximidades:
 Coordenada X: | Coordenada Y: | Setor Censitário:

DESCRIÇÃO SUMÁRIA DA OCORRÊNCIA (para qualificação legal)

*Delegado de Plantão:

CONTINUAR CANCELAR

(*) Campo de preenchimento obrigatório.

Figura 7: Tela inicial do registro de ocorrência.

COMUNICAÇÃO DE OCORRÊNCIA (etapa 2 de 3)

(*) Campo de preenchimento obrigatório.

B.O. Nº: 2007/139 | Tipo de B.O.: Incompleto | Unidade Policial: 39

NATUREZA CONSTATADA
 Pesquisar:
 *Código: Tentada Consumada [Inserir](#)

Código	Natureza	Constatação	Excluir
Não há registros relacionados a este BO.			

TIPO DE AMBIENTE
 Pesquisar:
 *Código: [Inserir](#)

Código	Tipo Ambiente	Excluir
Não há registros relacionados a este BO.		

MEIO EMPREGADO
 Pesquisar:
 *Código: [Inserir](#)

Código	Meio empregado	Excluir
Não há registros relacionados a este BO.		

PROVIDÊNCIA DO POLICIAL ATENDENTE
 Pesquisar:
 *Código: [Inserir](#)

Código	Providência Policial	Excluir
Não há registros relacionados a este BO.		

DATA E HORA DO FATO
 *Início: Data: Hora: | Final: Data: Hora:

POLICIAL QUE ATENDEU A OCORRÊNCIA
 *RG (com dv): | *Nome: | *Função: AGENTE DE OPERAÇÕES POLICIAIS | Disparos Efetuados: | Nº Série Arma: [Inserir](#)

Nome	Função	RG	Disparos Efetuados	Nº Série Arma	Excluir
Não há registros relacionados a este BO.					

CONTINUAR VOLTAR CANCELAR

(*) Campo de preenchimento obrigatório.

Figura 8: Continuando no registro de ocorrência.

3.7.1 Funcionalidades do BOU

Dentre as funcionalidades do BOU, cite-se (NOTA... 2012):

- Utilização de um único banco de dados sobre ocorrências policiais, eliminando duplicidade de registros e melhorando a qualidade das informações;
- A *notitia criminis* pode ser comunicada em quaisquer das corporações policiais;
- Viabilização de análises estatísticas e mapeamento espacial das ocorrências (geoprocessamento);
- Possibilita monitorar as áreas de maior risco de criminalidade, otimizando as estratégias de segurança;
- Favorece a integração e cooperação entre as Polícias Civil e Militar do Estado.

O foco do trabalho consistiu na extração do conhecimento na base de dados representativa do Boletim de Ocorrência Unificado.

É no sistema do BOU que são realizados os registros das ocorrências atendidas pela Polícia Civil e Polícia Militar. O registro é unificado, ou seja, o registro da ocorrência pode ser feito tanto por policiais civis e militares, nas respectivas unidades policiais. Isso promoveu a padronização do preenchimento e do documento a ser elaborado, bem como possibilitou a estruturação de uma base de dados única referentes às ocorrências policiais do Estado.

No sistema são preenchidos campos como datas e horas do registro do documento e da ocorrência, natureza da ocorrência, descrição do fato, qualificação dos envolvidos, meio empregado no crime, endereço, providências adotadas e outros dados que procuram detalhar o fato ocorrido, fornecendo também elementos iniciais suficientes para o início do trabalho de investigação (nos casos necessários).

A documentação eletrônica do boletim de ocorrência viabiliza a aplicação de

políticas públicas de interesse da comunidade. “Implica dizer, distribuição de efetivos e equipamentos realizada conforme o volume e as características das ocorrências policiais, com ampla capacidade de remanejamento de pessoal, de acordo com a exigência dos índices de criminalidade.” (CASTELLA, 2005, p. 62)

Fica claro que esta base de dados representa uma rica fonte de pesquisa para técnicas de extração do conhecimento. Questões como relação entre crimes, indicativos acerca da localização, natureza das ocorrências mais praticadas ou um panorama geral da criminalidade podem ser respondidas mediante a aplicação, por exemplo, de técnicas descritivas de mineração de dados.

4 PREPARANDO O AMBIENTE DE TRABALHO

Para a realização do trabalho foi providenciado um *dump* (*backup*) da base de dados referente ao sistema Boletim de Ocorrência Unificado. Foi disponibilizada uma base de desenvolvimento que, embora não seja a base utilizada em produção, contém registros reais e consiste, basicamente, em uma réplica do banco utilizado pelos usuários do sistema.

O *backup* foi disponibilizado no arquivo *bkp_sa_boletim_sa_detencao.tar.gz*. O primeiro passo foi descompactar o arquivo, tarefa que gerou dois outros arquivos: *bkp_sa_boletim_011211.sql* e *bkp_sa_detencao_011211.sql*. Na verdade, tais arquivos representam dois esquemas de um banco de dados único, o banco *bounificado* criado no SGBD *PostgreSQL*. Neste trabalho só será utilizado o esquema do B.O. Unificado (*bkp_sa_boletim_011211.sql*).

Antes de tudo foi preciso efetuar uma análise preliminar do arquivo, abrindo-o em um editor de texto adequado, observando sua estrutura e alguns parâmetros de construção do esquema.

De início já foi possível constatar que o banco de dados foi implementado com a codificação *LATIN1* (ISSO-8859-1), que é o conjunto de caracteres adequado para representar o idioma português brasileiro. Outro ponto que chamou atenção foi o tamanho do arquivo: o arquivo *bkp_sa_boletim_011211.sql* possui 10.5GB retratando, assim, o enorme volume de dados armazenado.

4.1 Preparando o sistema operacional

Em seguida, foi providenciada a configuração do ambiente de trabalho. Optou-se pelo sistema operacional *Windows 7* que se mostrou bastante estável para a realização das tarefas.

Entretanto, o trabalho também poderia ser realizado no ambiente *Linux*, desde que alguns ajustes de configuração fossem realizados.

A principal diferença de se trabalhar nos dois ambientes é a forma como os

sistemas operacionais lidam com a codificação de caracteres.

O *Windows* não oferece suporte nativo à codificação *Latin1* (ISO-8859-1). A codificação similar desenvolvida pela *Microsoft* é a *WIN1252*, que é uma implementação mais completa da codificação *Latin1*. Para trabalhar com a codificação *WIN1252* é necessário efetuar a opção no *script* de criação do banco de dados, tarefa que será descrita posteriormente.

No *Linux*, para se conseguir trabalhar com banco de dados com suporte à codificação *Latin1* são necessárias algumas alterações nos arquivos de configuração do sistema, conforme os passos a seguir:

1. Editar o arquivo */etc/environment* alterando os campos *LANG* e *LANGUAGE* para que fiquem dessa forma:

```
LANG="pt_BR"  
LANGUAGE="pt_BR:pt:en"
```

2. Editar o arquivo */etc/default/locale*

```
LANG="pt_BR"  
LANGUAGE="pt_BR:pt:en"
```

3. Digitar o seguinte comando no *prompt*:

```
dpkg-reconfigure locales
```

Esse procedimento torna a codificação ISO-8859-1 disponível para o sistema operacional sem, no entanto, alterar a codificação *default* que é a UTF-8. Com isso será possível trabalhar com banco de dados na codificação *Latin1*.

4.2 Instalação do sistema gerenciado de banco de dados PostgreSQL

O passo seguinte é a instalação do Sistema Gerenciador de Banco de Dados *PostgreSQL*. O *PostgreSQL* é um gerenciador de banco de dados objeto-relacional de código aberto, oriundo do *Postgres* versão 4.2, desenvolvido na universidade da

Califórnia em *Berkeley*. Em 1994, a linguagem SQL foi incorporada no *Postgres* passando a se chamar *Postgres95*. Houve grande esforço no aprimoramento do produto, com reescrita de código, inclusão de novas funcionalidades, esforço focado na confiabilidade e portabilidade, bem como inclusão de pacotes *Linux*. Assim, a equipe de desenvolvimento percebeu que o gerenciador deveria se chamar *PostgreSQL*.

Atualmente, o *PostgreSQL* consiste numa ótima opção de SGBD devido a sua robustez, confiabilidade, quantidade de recursos, implementação de características de orientação a objetos e além do fato de ser de código aberto (GONZAGA, 2007 p. 4).

Também será instalado o software *pgAdmin*, que é uma ferramenta visual para a administração de bancos de dados no *PostgreSQL*.

A instalação do *PostgreSQL* no *Windows* é tranquila, bastando ao usuário apenas baixar e executar o instalador automático, seguindo os passos requeridos.

Já no *Linux* há mais opções de instalação. É possível instalar o *software* mediante os seguintes comandos no terminal:

```
~# sudo apt-get update (atualização das listas de pacotes)
~# sudo apt-get install postgresql (instala o pacote do postgresql)
~# sudo apt-get install pgadmin3 (instala o pgAdmin 3)
~# sudo apt-get autoremove (removendo pacotes não mais necessários)
```

Para saber se a instalação foi efetuada com sucesso, basta digitar o comando abaixo:

```
~# service postgresql-8.4 status
Running clusters: 8.4/main
```

A mensagem indica que o programa foi instalado corretamente e o servidor está no ar. Porém, recomenda-se fortemente que, tanto no ambiente *Windows* quanto *Linux*, a instalação seja efetuada pelo instalador automático disponibilizado no *site* oficial do produto. Além de ser um procedimento mais simples, este tipo de instalação possibilita a configuração de alguns parâmetros do banco de dados logo de início, sobretudo a escolha da codificação do banco de acordo com o sistema, evitando ajustes adicionais.

No *Linux*, antes de iniciar os trabalhos no *PostgreSQL* é preciso redefinir a senha do usuário *postgres*. A tarefa é executada com os comandos abaixo onde a senha foi redefinida, por exemplo, para “admin”:

```

~# su postgres
postgres@note: psql
postgres=# ALTER USER postgres WITH PASSWORD 'admin';

```

Para melhor desempenho do *PostgreSQL* e prevenção de problemas com falta de memória, é recomendável alterar algumas configurações no arquivo *postgresql.conf* localizado na pasta *PostgreSQL/9.1/data*:

- 1 Diminuição do número de conexões simultâneas no banco de dados (Figura 8).

<input checked="" type="checkbox"/>	<code>max_connections</code>	10	(change requires restart)
-------------------------------------	------------------------------	----	---------------------------

Figura 9: Diminuindo o número de conexões simultâneas no banco de dados.

- 2 Aumento da memória compartilhada, buffer temporário e memória de trabalho (Figura 9)

<input checked="" type="checkbox"/>	<code>shared_buffers</code>	1024MB	min 128kB
<input checked="" type="checkbox"/>	<code>temp_buffers</code>	512MB	min 800kB
<input checked="" type="checkbox"/>	<code>work_mem</code>	1024MB	min 64kB

Figura 10: Configurando a quantidade de memória disponível para o *PostgreSQL*

4.3 Instalação do *RapidMiner*

Prosseguindo na configuração do ambiente de trabalho, o próximo passo é a instalação do *software RapidMiner*. Trata-se de uma ferramenta *open source* desenvolvida na Alemanha e constitui uma solução completa para a implementação das tarefas de mineração de dados. Desenvolvido sob a plataforma Java, o *software* conta com diversos recursos avançados como uma interface gráfica amigável, inúmeros operadores, *scripts* baseados em XML, expansão das funcionalidades mediante instalação de *plugins* e possibilidade de integração direta com o banco de

dados via bibliotecas do *Java Database Connectivity* (JDBC). Além disso, a ferramenta incorpora todas as bibliotecas do *Weka*, conhecido *software* utilizado em mineração de dados.

A versão utilizada no trabalho foi a *Community Edition 5.2*, que é gratuita. Para instalar o programa no *Windows*, basta executar o instalador e seguir os passos, sendo que ao final os atalhos de acesso ao *software* já estarão disponíveis para o usuário.

A instalação no *Linux* também é simples e não requer esforço. O arquivo compactado *rapidminer-5.2.001.zip* pode ser baixado diretamente no site da empresa desenvolvedora no endereço eletrônico *www.rapid-i.com*.

Após o *download* do arquivo basta descompactá-lo em uma pasta da preferência do usuário. Para utilizar o programa é preciso executar o arquivo *rapidminer.jar* que encontra-se na pasta */lib* no diretório de instalação do *software*, lembrando que é necessário ter a *runtime* do Java previamente instalada no computador. O comando abaixo executa o programa:

```
~$java -jar rapidminer.jar
```

Na primeira execução o *software* alerta que o usuário ainda não possui um repositório local definido (Figura 10). O repositório é necessário para armazenar os dados e processos realizados na descoberta do conhecimento.

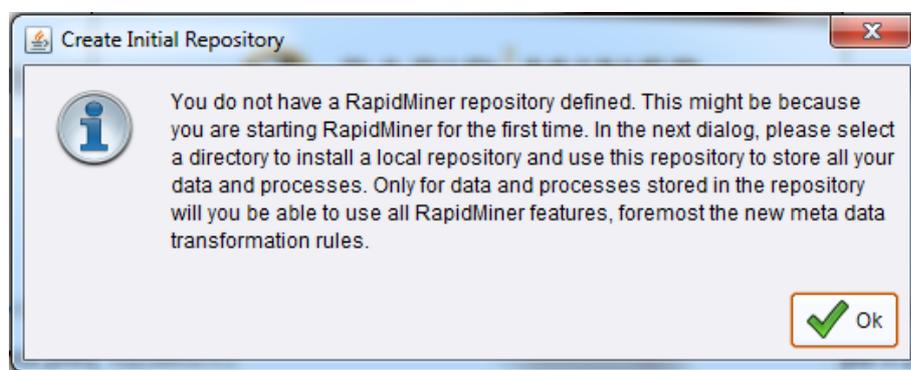


Figura 11: Mensagem de alerta para criação do repositório de trabalho do programa.

Na próxima tela é criado um repositório local com o nome de preferência do usuário, encerrando a configuração inicial clicando no botão "*Finish*". O ambiente de

trabalho está configurado e a ferramenta pronta para ser utilizada (Figuras 11 e 12).

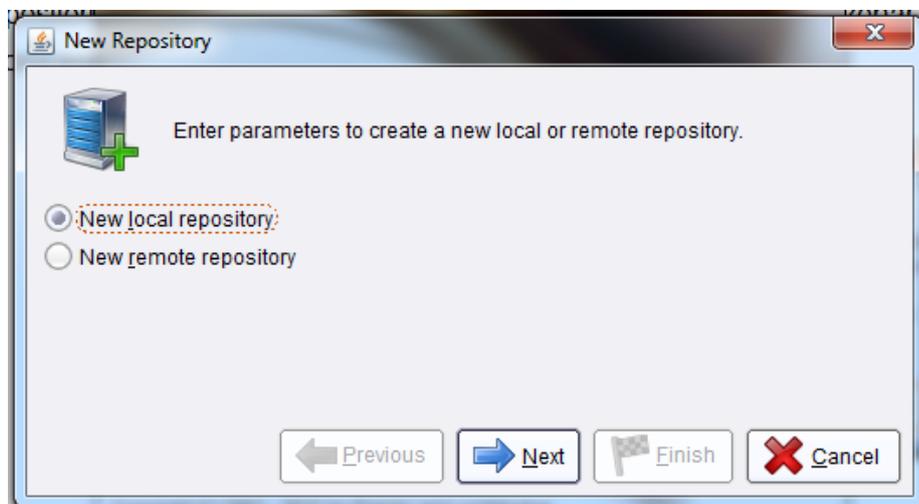


Figura 12: Criando um novo repositório.

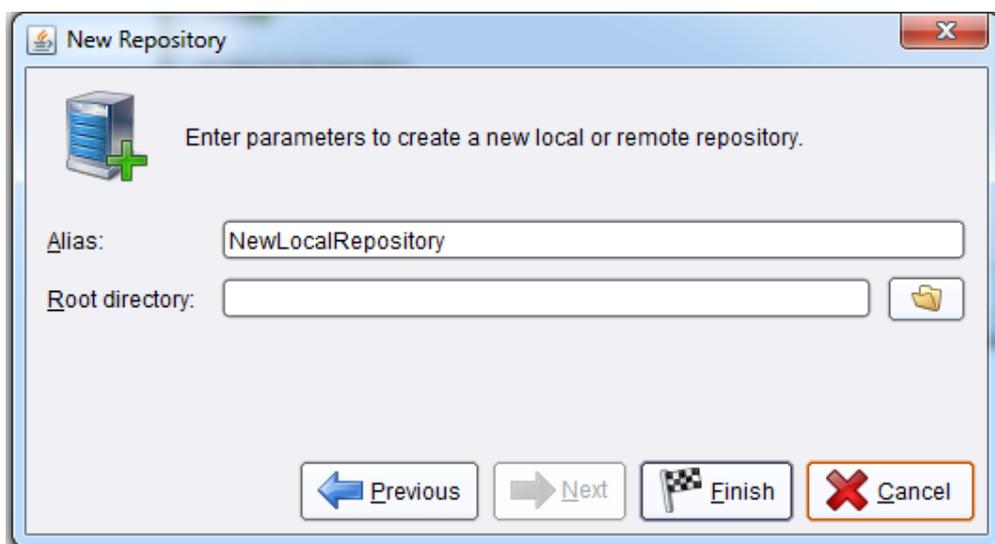


Figura 13: Criando um repositório e finalizando a configuração.

4.4 Efetuando a restauração do banco de dados.

Estando o ambiente de trabalho configurado, a etapa seguinte é a realização do *restore* do banco de dados, ou seja, a restauração do arquivo de *backup* da base de dados do sistema B.O. Unificado.

Conforme já dito, foi realizada uma análise preliminar dos requisitos necessários para o sucesso do *restore*. Inicialmente, é necessário criar o banco de dados denominado *bounificado* no *PostgreSQL*. A tarefa pode ser realizada no *pgAdmin* na janela do *SQL Editor* com a digitação do *script* abaixo, o qual foi fornecido com a documentação do banco. O *script* cria um banco de dados vazio, sem tabelas.

```
CREATE DATABASE bounificado
WITH OWNER = postgres
ENCODING = 'WIN1252'
TABLESPACE = pg_default
TEMPLATE = template0;
SET client_encoding = 'WIN1252';
SET check_function_bodies = false;
SET client_min_messages = warning;
COMMENT ON SCHEMA public IS 'Standard public schema';
CREATE SCHEMA sa_boletim;
ALTER SCHEMA sa_boletim OWNER TO postgres;
CREATE SCHEMA sa_historico;
ALTER SCHEMA sa_historico OWNER TO postgres;
```

Figura 14: *Script* SQL de criação do banco de dados no ambiente *Windows*.

No *Windows* deve-se prestar atenção na codificação a ser utilizada pelo banco de dados, optando-se pelo *encoding* *WIN1252* (Figura 13):

No *Linux* (Figura 14):

```

CREATE DATABASE bounificado
  WITH OWNER = postgres
       ENCODING = 'LATIN1'
       TABLESPACE = pg_default;
SET client_encoding = 'LATIN1';
SET check_function_bodies = false;
SET client_min_messages = warning;
COMMENT ON SCHEMA public IS 'Standard public
schema';
CREATE SCHEMA sa_boletim;
ALTER SCHEMA sa_boletim OWNER TO postgres;
CREATE SCHEMA sa_historico;
ALTER SCHEMA sa_historico OWNER TO postgres;

```

Figura 15: Script SQL de criação do banco de dados no ambiente *Linux*.

A principal diferença entre os *scripts* é a mudança de *encoding*. O *Windows* não trabalha nativamente com a codificação *Latin1* (ISSO-8859-1). Porém, a codificação *WIN1252* é a implementação similar da *Microsoft*, sendo inclusive mais ampla que a codificação *Latin1*. Por isso, a necessidade da modificação do *script*.

O banco de dados alvo da mineração implementa diversas funções de *triggers*. *Trigger* é um procedimento armazenado que é ativado por eventos no banco de dados. Esses eventos podem ser inserções (*INSERT*), atualizações (*UPDATE*) e exclusões (*DELETE*). Os *triggers* podem ser disparados antes da execução do evento (*BEFORE*) ou depois (*AFTER*).

As funções de *triggers* devem ser escritas em C ou em alguma linguagem procedural disponível no banco de dados, tais como *Ruby*, *Perl*, *Python*, entre outras. Atualmente existem quatro linguagens procedurais disponíveis na distribuição padrão do *PostgreSQL*: *PL/pgSQL*, *PL/Tcl*, *PL/Perl* e *PL/Python*. Mas é possível que o usuário defina outras linguagens. Para o trabalho foi utilizada a linguagem *PL/pgSQL*. Em alguns casos, pode ser necessário instalá-la no banco de dados a ser utilizado. A instalação foi efetuada com a digitação do seguinte comando no *SQL Editor* do *pgAdmin*:

```
CREATE LANGUAGE plpgsql;
```

Para constatar se a linguagem foi realmente instalada, basta realizar uma consulta simples que retornará as linguagens instaladas no banco de dados (Figura

15):

```
SELECT * FROM PG_LANGUAGE
```

	lanname name	lanowner oid	lanispl boolean	lanpltrusted boolean	lanplcallfoid oid	laninline oid	lanvalidator oid	lanacl aclitem[]
1	internal	10	f	f	0	0	2246	
2	c	10	f	f	0	0	2247	
3	sql	10	f	t	0	0	2248	
4	plpgsql	10	t	t	11640	11641	11642	

Figura 16: Resultado da consulta no editor SQL mostrando as linguagens procedurais instaladas.

Prosseguindo, a restauração do banco de dados foi executada, mediante a digitação do seguinte comando no *prompt*:

```
psql -h localhost -d bounificado -f bkp_sa_boletim_011211.sql -U postgres
```

Com todos os procedimentos realizados corretamente, a restauração foi realizada com sucesso com a criação das 79 tabelas que compõem o banco de dados.

5 PRÉ-PROCESSAMENTO

A etapa de pré-processamento é de fundamental importância no processo de descoberta de conhecimento, haja vista que a qualidade dos dados vai determinar a eficiência do algoritmo a ser aplicado.

Esta fase engloba desde o entendimento inicial dos dados para uma visão clara do domínio do problema até a aplicação de toda e qualquer operação necessária para a seleção, limpeza, transformação, redução e adequação visando otimizá-los para o processo de descoberta do conhecimento.

5.1 Entendimento dos Dados

Antes de se iniciar a aplicação das operações de pré-processamento propriamente ditas é relevante que se tenha entendimento dos dados a serem explorados como forma de orientar as decisões que poderão ser tomadas no processo, tanto na etapa de preparação quanto na mineração de dados. (NEVES, 2003, p. 32)

É preciso conhecer as tabelas, relacionamentos e atributos que compõem o banco de dados, bem como conhecer o teor das informações armazenadas. No caso em tela, o sistema é composto por 79 tabelas que armazenam dados que vão desde registros de *logs* até informações essenciais para o registro de uma ocorrência policial, tais como dados sobre o local, natureza do fato e informações sobre os envolvidos na ocorrência.

Da mesma forma, é relevante conhecer o volume de dados apresentado para que, caso necessário, seja possível traçar alguma estratégia de seleção de instâncias para viabilizar a mineração.

Nesta etapa preliminar seria ideal o acompanhamento do especialista do domínio ou que pelo menos o minerador tenha em mãos a documentação da base de dados para a correta compreensão da aplicação.

Felizmente, devido ao contato prático com o sistema de registro de

ocorrências somado ao conhecimento adquirido do domínio do problema foi possível elaborar uma visão bastante clara da base de dados, o que facilitou o direcionamento do trabalho de extração do conhecimento.

Com o entendimento formado, a seguir estão algumas operações de pré-processamento que foram realizadas.

5.2 Segmentação do banco de dados

Mineração de dados é um processo essencialmente iterativo, ou seja, sucessivas tentativas de extração do conhecimento podem ser realizadas até se obter um resultado satisfatório potencialmente útil ao objetivo proposto. A iteração também está presente na etapa de pré-processamento, desde a análise da base de dados até o tratamento final dos dados que serão minerados.

Nesse sentido, vale a pena relatar algumas dificuldades encontradas logo na fase de pré-processamento. As tentativas serviram para constatar o quanto é computacionalmente dispendioso manipular uma grande quantidade de dados. Para se ter uma ideia da dimensão do problema, há mais de 5 milhões de instâncias da tabela *tb_boletim* em todo o banco de dados. Somente os dados do ano de 2010 contem mais de 1 milhão de instâncias da mesma tabela. Mesmo trabalhando em um computador com configurações aparentemente robustas (processador *Intel i5* com 6 Gb de memória RAM), houve momentos críticos razão pela qual foi preciso traçar estratégias de seleção de dados para viabilizar a realização do trabalho. Uma simples instrução *select* de consulta poderia levar horas para ser processada. Nos inúmeros testes realizados foram constatados problemas como demora excessiva de processamento, estouro de memória, dentre outros. Veja-se, então, que a profundidade da extração do conhecimento é altamente influenciada pelas configurações de hardware disponíveis.

Diante disso, foi preciso iniciar a etapa de pré-processamento já diretamente no banco de dados, primeiramente segmentando a quantidade de dados que seriam alvo da pesquisa. Optou-se por trabalhar somente nos registros do ano de 2011, cuja tabela *tb_boletim* perfaz exatas 658941 instâncias.

5.3 Seleção de atributos

Definido o escopo geral da mineração, o passo seguinte é passarmos a seleção dos atributos que representarão o conhecimento a ser descoberto. Nem todas as informações contidas no banco de dados são relevantes ao objetivo proposto na mineração. Por exemplo, se o objetivo é descobrir padrões relacionados ao perfil dos envolvidos nas ocorrências policiais, nenhuma utilidade teriam dados relativos à *logs* do sistema. Desse modo, é preciso filtrar os atributos potencialmente úteis ao processo de mineração.

Consoante preceitua Neves (2003, p. 33), diante

“(...) da quantidade de dados disponível no contexto do domínio do problema é preciso selecionar as tabelas, atributos e instâncias das mesmas que estejam mais relacionadas aos objetivos do usuário a serem alcançados permitindo assim a geração de um novo conjunto de dados único e conciso que poderá ser submetido às demais subfases de pré-processamento bem como ao restante do processo de descoberta de conhecimento em banco de dados (DCBD)”.

Sendo assim, a providência inicial é a eliminação das tabelas claramente desnecessárias que, num primeiro momento, não guardam pertinência com os objetivos gerais da mineração a ser desenvolvida.

O segundo passo é a seleção dos atributos relevantes nas tabelas que restaram. Preservando-se os relacionamentos entre as diversas tabelas, foram selecionados atributos-chaves que armazenam as principais informações do sistema.

O diagrama da Figura 16 retrata a modelagem do sistema após a exclusão das tabelas desnecessárias. Os atributos selecionados estão em vermelho.

A escolha dos atributos deve ser fruto da reflexão sobre os objetivos gerais da mineração de dados. Mesmo que ainda não haja hipóteses bem definidas a serem exploradas, é importante que o minerador preveja quais informações poderão ser relevantes ao processo. Mais do que isso, no trabalho em tela, a escolha dos atributos levou em conta o melhor custo benefício da relação desempenho X eficácia. Isto é, o excesso de atributos poderia demandar um esforço computacional que possivelmente prejudicaria a viabilidade do trabalho levando em conta o hardware disponível. Em contrapartida, a escassez de atributos provavelmente dificultaria a extração de regras de associações significativas.

Desse modo, o Quadro 1 relaciona os atributos escolhidos com as respectivas justificativas. Também restaram consignadas as dimensões de cada atributo, ou seja, a quantidade de valores que cada um pode assumir no banco de dados.

Definidos os atributos, deve-se construir uma nova e única tabela com os atributos escolhidos, preservando-se os relacionamentos pré-existentes. Essa tabela é que será alvo da mineração de dados. O *script* SQL de criação da tabela encontra-se no Apêndice.

QUADRO 1 – ATRIBUTOS SELECIONADOS E JUSTIFICATIVAS

TABELA	ATRIBUTO	DIMENSÕES (quantidade de valores possíveis)
tb_boletim	cidade	4350
JUSTIFICATIVA: atributo que representa a cidade onde ocorreu o fato. É sempre válido analisar informações acerca da geografia do crime, averiguar associações que dizem respeito aos municípios em que os fatos são praticados. Vale ressaltar que tal atributo armazena qualquer localidade e não somente municípios paranaenses. A consistência da informação depende do correto preenchimento do usuário.		
tb_armaespecie	descrição	17
JUSTIFICATIVA: saber qual o tipo e letalidade das armas utilizadas nos crimes pode servir de base para equipar as forças policiais adequadamente. O atributo descreve armas como pistola, revolver, garrucha, escopeta, etc.		
tb_armacalibre	descrição	73
JUSTIFICATIVA: informações quanto ao calibre de armas podem ser úteis para relacionar quais calibres, permitidos ou restritos, são mais utilizados em práticas delituosas, inclusive sua correlação com determinados tipos de crimes. Além disso também pode justificar a utilização de equipamentos policiais adequados.		

QUADRO 1 – ATRIBUTOS SELECIONADOS E JUSTIFICATIVAS

continua

tb_entorpecente	descricao	14
JUSTIFICATIVA: o combate ao tráfico e uso de entorpecentes é fundamental para as forças de segurança. O atributo relaciona quais os tipos de drogas ilícitas conhecidas, de forma que é relevante saber qual tipo de droga tem maior incidência em aspectos geográficos ou em relação a pessoas envolvidas e até mesmo em termos quantitativos.		
tb_copelmunicipio	nome	401
JUSTIFICATIVA: informações relacionadas a aspectos geográficos são fundamentais para qualquer processo de mapeamento e análise criminal. Saber em quais regiões determinados crimes são praticados com mais frequência ou quais áreas são mais ou menos afetadas com a criminalidade pode ser relevante no processo de tomada de decisão.		
tb_tipolocal	descricao	22
JUSTIFICATIVA: o atributo “descrição” da tabela “tb_tipolocal” categoriza genericamente diversos tipos de locais em que pode ocorrer um fato como, por exemplo: residência, comércio, veículo, órgão público, dentre outros. Uma análise inteligente e associativa de tal atributo poderá revelar quais tipos de locais são mais “visados” ou propensos a práticas delituosas.		
tb_localespecifico	descricao	185
JUSTIFICATIVA: trata-se de uma especialização do atributo “descrição” da tabela “tb_localespecifico”, estabelecendo locais que são decorrentes daqueles anteriormente referenciados, tais como: teatro, shopping, lanchonete, via pública e outros.		
tb_natureza	tipo	4
JUSTIFICATIVA: o atributo descreve a natureza do fato ocorrido ou da chamada realizada estipulando, por exemplo, se o fato foi tentado ou consumado. Trata-se de informação interessante para o gestor da segurança que poderá analisar índices de incidência de crimes que efetivamente aconteceram e aqueles que nem chegaram a se consumir (tentados).		
tb_titulopenal	titulo	51
JUSTIFICATIVA: iniciando o processo de classificação delitiva este atributo enquadra o fato a determinadas categorias genéricas, tais como: crimes contra o patrimônio, contra a família, lei de tóxicos, dentre outros. É importante ter uma visão global dos fatos.		
tb_tipopenal	nomejuridico	953
JUSTIFICATIVA: o atributo informa o nome jurídico do fato delituoso, especializando a classificação da tabela acima referenciada. Dessa forma, é possível identificar com precisão os delitos noticiados no boletim de ocorrência, dado que poderá útil na análise criminal e extração do conhecimento. A título de exemplo, poderia ser de grande valia para o gestor correlacionar a prática de furtos com informações geográficas.		

QUADRO 1 – ATRIBUTOS SELECIONADOS E JUSTIFICATIVAS

continua

tb_meioempgrupo	descricao	13
JUSTIFICATIVA: existem diversas formas de se praticar um delito. O atributo categoriza genericamente os meios que podem ser empregados na prática delituosa como, por exemplo, arma de fogo, arma branca, disfarce, acesso, dentre outros. Ter em mente quais os meios mais comumente utilizados em determinadas práticas delituosas pode ser útil para aprimorar treinamentos policiais.		
tb_meioemptipo	descricao	177
JUSTIFICATIVA: o atributo especializa os meios descritos na tabela anterior, valendo a mesma justificativa de sua escolha.		
tb_envolvido	sexo	2
	situacaoenvolvido	16
	idadeestimada	contínuo
	grauinstrucao	8
	corcutis	5
	estadocivil	8
JUSTIFICATIVA: os atributos identificam as pessoas envolvidas em uma ocorrência. Vale lembrar que uma pessoa poderá figurar em um fato de diversas formas: como vítima, noticiante, testemunha, autor e outros. Sendo assim, será possível descobrir associações sobre o perfil dos envolvidos		
tb_provpolicialtipo	descricao	82
JUSTIFICATIVA: o atributo relaciona os tipos de providências policiais que poderão ser adotadas em face de uma ocorrência policial.		
TOTAL		6381 dimensões

5.4 Redução da quantidade de dados

Optou-se por trabalhar com os dados referentes aos registros de ocorrências realizadas no ano de 2011, totalizando 658.941 instâncias na tabela objeto da mineração de dados. Todavia, mesmo assim, o volume de dados ainda é expressivo, razão pela qual alguns problemas de desempenho computacional foram constatados, tais como lentidão no processamento e estouro de memória.

Diante disso, foi providenciada a redução na quantidade de dados mediante o emprego da técnica estatística de amostragem aleatória simples no universo populacional de registros. A amostragem aleatória simples estipula que cada um dos

elementos que compõem a população alvo tem igual probabilidade de ser escolhido para fazer parte de uma amostra.

A tabela alvo de MD foi criada com uma amostra de 300 mil registros aleatórios, utilizando a função SQL *RANDOM()*, conforme trecho do *script* abaixo:

```
(...)  
WHERE  
    tb_boletim.anoboletim = 2011 ORDER BY RANDOM() LIMIT 300000;
```

A redução da quantidade de dados a serem manipulados agilizou as consultas e facilitará as demais atividades da mineração de dados.

Contudo, a aplicação da técnica de amostragem não é suficiente para garantir a confiabilidade da amostra, ou seja, se ela é representativa do conjunto total de dados. É preciso calcular a taxa de erro amostral que é a diferença entre o valor estatístico e o verdadeiro valor do parâmetro que se deseja estimar.

Para amostragem aleatória simples, podemos calcular a taxa de erro amostral de acordo com as seguintes fórmulas (BARBETTA *apud* COSTA, 2011, p. 29):

$$n_0 = \frac{1}{E_0^2}$$

$$n = \frac{N * n_0}{N + n_0}$$

onde:

- N = tamanho da população;
- E_0 = erro amostral tolerável;
- n_0 = primeira aproximação do tamanho da amostra;
- n = tamanho da amostra.

Considerando que:

- N = 658.941 e $n = 300.000$, temos

$$n = \frac{N * \frac{1}{E_0^2}}{N + \frac{1}{E_0^2}}$$

$$300.000 = \frac{658.941 * \frac{1}{E_0^2}}{658.941 + \frac{1}{E_0^2}}$$

$$300.000 = \frac{\frac{658.941}{E_0^2}}{\frac{658.941 * E_0^2 + 1}{E_0^2}}$$

$$300.000 = \frac{658.941}{658.941 * E_0^2 + 1}$$

$$1976823 * 10^5 * E_0^2 + 300.000 = 658.941$$

$$E_0^2 = \frac{358.941}{1976823 * 10^5}$$

$$E_0 \cong 0,0013$$

Portanto a taxa de erro amostral é de aproximadamente apenas 0,13%, insignificante para o objetivo proposto, indicando que a mostra é significativa e reflete as características do universo populacional de registros da base de dados.

5.5 Mapeamento de valores

Analisando o banco de dados foi constatado que alguns atributos armazenam códigos numéricos inteiros para representar valores literais intrínsecos. Por exemplo, na tabela *tb_natureza* o número 3 significa natureza consumada. Os significados de todos os valores codificados estão presentes na documentação de desenvolvimento do sistema.

Porém, valores expressos somente em códigos numéricos podem dificultar a interpretação das regras de associação geradas. Sendo assim, foi necessária uma preparação especial dos dados, “traduzindo-se” os valores encontrados para os respectivos literais.

```
ALTER TABLE data_mining.tb_datamining_2011_def ADD COLUMN
natureza_descricao VARCHAR(15);
UPDATE data_mining.tb_datamining_2011_def SET natureza_descricao =
'CHAMADA' WHERE tiponatureza = '1';
UPDATE data_mining.tb_datamining_2011_def SET natureza_descricao =
'CONSUMADA' WHERE tiponatureza = '3';
UPDATE data_mining.tb_datamining_2011_def SET natureza_descricao =
'TENTADA' WHERE tiponatureza = '4';
ALTER TABLE data_mining.tb_datamining_2011_def DROP COLUMN tiponatureza;
```

Figura 18: Trecho do *Script* SQL para mapeamento dos valores dos atributos.

O tratamento foi realizado diretamente no banco de dados, mediante instruções *SQL*. Basicamente, a tarefa consistiu em adicionar uma nova coluna à tabela gerada a partir da seleção de atributos e preenche-la com os valores literais respectivos. Após, apagou-se a coluna original. A título de explicação foi transcrito um pequeno trecho do *script* para a melhor compreensão do processo (Figura 17):

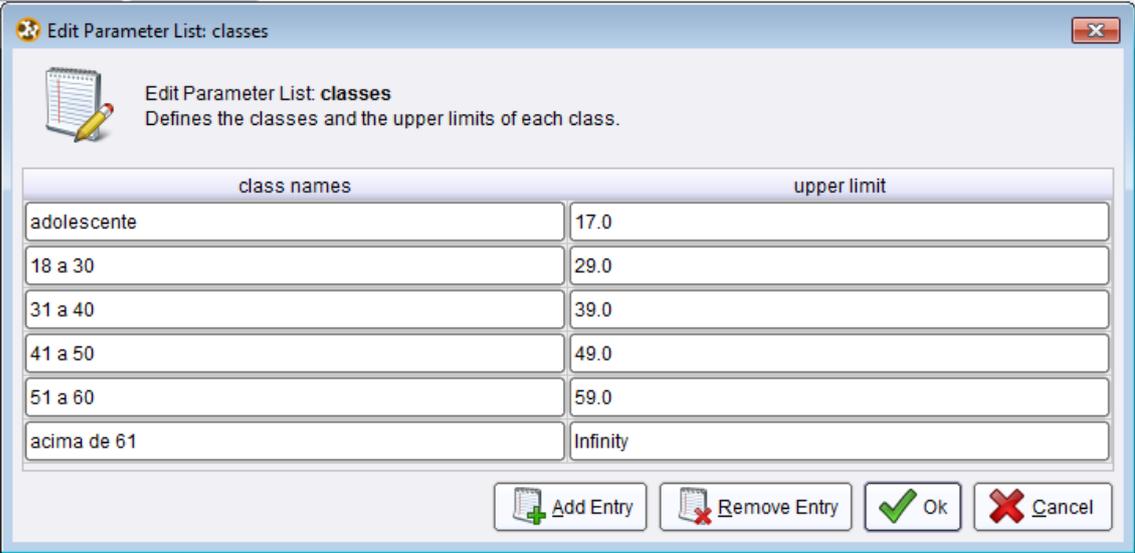
5.6 Discretização

Alguns algoritmos de mineração de dados requerem que os dados estejam na forma de atributos categorizados, especialmente algoritmos de associação. Dessa forma, é necessário transformar um atributo contínuo em um valor simbólico e categorizado (discretização). Atributos discretos são mais fáceis de compreender e de serem processados pelos algoritmos. Por sua vez, entende-se por atributos contínuos aqueles numéricos sem categoria definida.

Um exemplo de valor contínuo no presente trabalho é o campo “idade estimada” do envolvido no registro de ocorrência. O campo pode ser preenchido com qualquer valor inteiro (desde que seja uma idade humanamente possível) e não há categorias pré-definidas.

Para a aplicação do algoritmo de associação foi preciso discretizar o atributo. No *Rapidminer* a tarefa foi realizada mediante aplicação do operador *Discretize*

Com isso foi possível estabelecer faixas etárias, conforme os possíveis valores de idades. (Figura 18).



Edit Parameter List: classes

Edit Parameter List: **classes**
Defines the classes and the upper limits of each class.

class names	upper limit
adolescente	17.0
18 a 30	29.0
31 a 40	39.0
41 a 50	49.0
51 a 60	59.0
acima de 61	Infinity

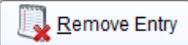
 Add Entry  Remove Entry  Ok  Cancel

Figura 19: Discretizando o atributo idade estimada com faixa de valores

6 MINERANDO OS DADOS

Neste capítulo são detalhas as minerações realizadas. Os parâmetros utilizados foram:

- Suporte mínimo variando de 0.001 (0,1%) a 0.1 (10%), equivalendo de 300 a 30.000 registros;
- Confiança mínima de 0.7 ou 70%;

6.1 Mineração perfil do envolvido

Descobrir o perfil dos envolvidos em ocorrências policiais pode ser útil para correlacionar aspectos sociais e econômicos da criminalidade. Com dados como grau de instrução, sexo, idade, estado civil, dentre outros é possível extrair conhecimento que possam trazer informações que vão além do aspecto penal do crime.

É importante ressaltar que esta mineração levou em consideração todo tipo de envolvidos em ocorrências, isto é, vítima, autor, testemunha dentre outros. A título de ilustração, a Figura 20 retrata o processo de mineração modelado no software *RapidMiner* e alguns operadores disponíveis utilizados, tais como:

- **Retrieve:** carrega a tabela de dados que será alvo da mineração;
- **Select Attributes:** dos atributos disponíveis, permite ao minerador selecionar aqueles que serão utilizados na seção de mineração;
- **Discretize:** operador responsável por realizar a discretização, com as configurações pertinentes;
- **Nominal to Binominal:** mapeia os valores de atributos nominais para valores binominais (somente dois valores);

- **FP-Growth:** este operador aplica o algoritmo *FP-Growth* com a construção da *FP-Tree* na geração de *itemsets*. Esse operador utiliza valores binominais como entrada para processamento;
- **Create Association Rules:** gera um conjunto de regras de associação a partir de um *itemset*.

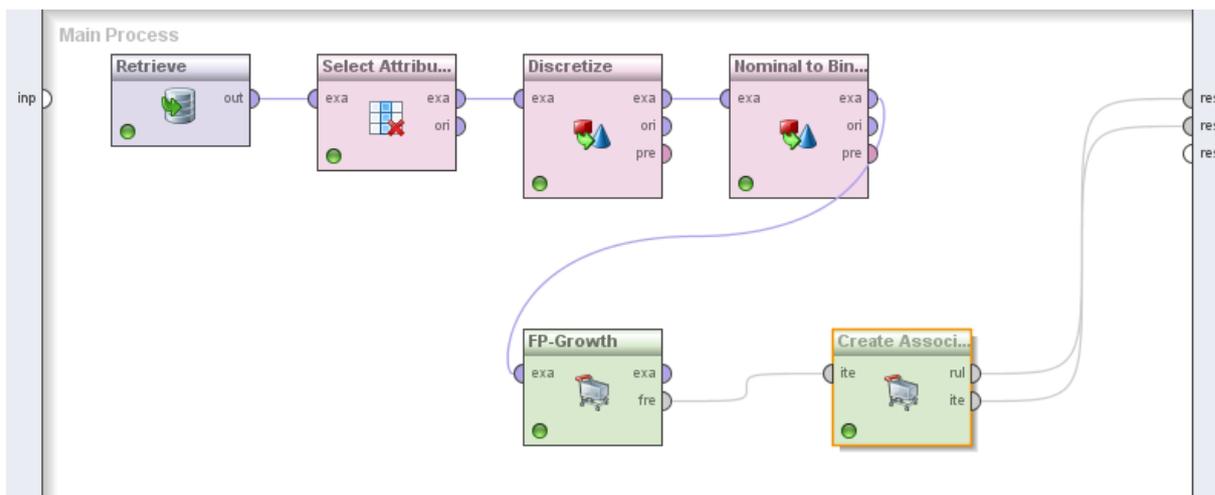


Figura 20: Modelo de mineração de dados no *RapidMiner*.

Logo de início foi possível confirmar a predominância de envolvidos do sexo masculino em ocorrências policiais. É uma informação facilmente constatada na prática.

Os resultados (suporte mínimo de 10%):

- Quando o envolvido for o autor do fato será do sexo masculino com 92% de confiança;
- Com 84,6% de confiança, quando o envolvido for solteiro, ele será do sexo masculino;
- Verifica-se grande incidência de envolvidos homens sem instrução escolar completa. Com 84,3% de certeza os envolvidos com o ensino fundamental incompleto são do sexo masculino;
- Foi constatada a predominância de envolvidos jovens. Com 83,3% de confiança, envolvidos com idade entre 18 e 30 anos são do sexo masculino;

- Com certeza de 76,4%, quando o envolvido for assistido na ocorrência, será também do sexo masculino;

A Figura 21 ilustra a representação textual das regras de associação extraídas, na Figura 22 a representação gráfica.

```

Association Rules
[situacao_envolvido = ASSISTIDO] --> [sexoenvolvido = M] (confidence: 0.764)
[idadeestimada = 18 a 30] --> [sexoenvolvido = M] (confidence: 0.833)
[grau_instrucao = 1o GRAU INCOMPLETO] --> [sexoenvolvido = M] (confidence: 0.843)
[estado_civil_envol = SOLTEIRO] --> [sexoenvolvido = M] (confidence: 0.846)
[situacao_envolvido = AUTOR] --> [sexoenvolvido = M] (confidence: 0.920)

```

Figura 21: Representação textual das regras de associação extraídas.

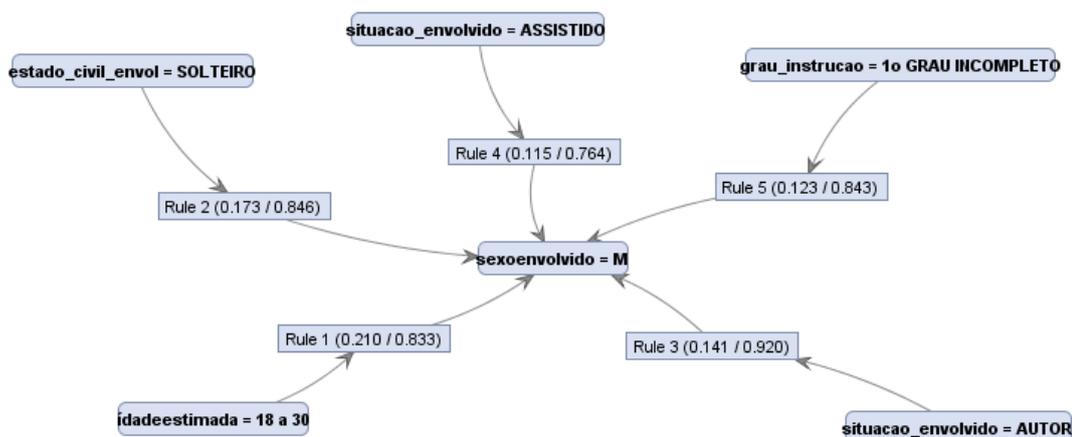


Figura 22: Representação gráfica das associações extraídas.

6.2 Mineração arma x idade do envolvido

Essa mineração buscou averiguar relacionamentos entre armas utilizadas nas ocorrências e a faixa etária dos envolvidos. Foi constatada a predominância do uso de arma de calibre 38 por jovens:

- Com 89,5% de confiança, revólver calibre 38 é utilizado por envolvidos de 18 a 30 anos (suporte mínimo de 1%).

6.3 Mineração droga x arma

Aqui o objetivo é correlacionar tipos de armas empregadas a praticas delituosas onde haja envolvimento de substâncias entorpecentes. Foi constatada a prevalência do uso de revólver calibre 38 em delitos com referência a drogas. O suporte mínimo foi de 0,001 (0,1%) (Figura 23):

- Com 92,1% de confiança há relação entre cocaína e revólver calibre 38;
- Assim como 91,5% e 89,6% de certeza, respectivamente, com “crack” e maconha.

```

Association Rules
[entorpecente = CRACK, calibrearma = 032,00] --> [especiearma = REVOLVER           ] (confidence: 0.728)
[entorpecente = MACONHA, especiearma = REVOLVER           ] --> [calibrearma = 038,00] (confidence: 0.739)
[especiearma = REVOLVER           , entorpecente = COCAINA] --> [calibrearma = 038,00] (confidence: 0.822)
[calibrearma = 028,00] --> [especiearma = ESPINGARDA       ] (confidence: 0.847)
[calibrearma = 036,00] --> [especiearma = ESPINGARDA       ] (confidence: 0.868)
[calibrearma = 038,00] --> [especiearma = REVOLVER         ] (confidence: 0.893)
[entorpecente = MACONHA, calibrearma = 038,00] --> [especiearma = REVOLVER           ] (confidence: 0.896)
[entorpecente = CRACK, calibrearma = 038,00] --> [especiearma = REVOLVER           ] (confidence: 0.915)
[calibrearma = 038,00, entorpecente = COCAINA] --> [especiearma = REVOLVER           ] (confidence: 0.921)

```

Figura 23: Associações geradas na mineração droga x arma.

6.4 Mineração meio empregado x idade do envolvido

Nesta mineração, objetivou-se relacionar os meios utilizados na prática do fato com a faixa etária do envolvido, bem como outras características individuais. O resultado trouxe algumas informações interessantes e outras triviais. Foi adotado suporte mínimo de 0,01:

- Quando há o emprego de droga ilícita, há participação de envolvido do sexo masculino com 78,3% de certeza;
- Homens de 41 a 50 anos aparecem com grande probabilidade de praticar fatos utilizando veículo automotor (confiança de 80,2%). Se a faixa etária cair para 31 a 40 anos, o grau de certeza é de 83,8% e de 18 a 30 com 87,1%;
- Da mesma forma, indivíduos do sexo masculino de 18 a 30 anos se envolvem em ocorrências com motocicletas numa margem de certeza de 91,3%;

6.5 Mineração situação do envolvido x droga

As pessoas podem figurar em ocorrências policiais sob diversas condições: autor do fato, vítima, testemunha, noticiante, dentre outros. Esta mineração procurou extrair conhecimento relacionado à situação do envolvido com substâncias entorpecentes. O suporte mínimo foi de 0,001:

- Com 84% de confiança, quando o envolvido tiver idade entre 18 a 30 anos e a ocorrência tiver relação com “crack”, ele estará figurando como autor do fato. Se for cocaína, a certeza é de 80,5%;

6.6 Mineração título penal x sexo do envolvido

De uma forma para a compreensão, o título penal relaciona os delitos de uma forma mais abrangente, mais geral. Aqui, procurou-se estabelecer relação entre títulos penais e sexo dos envolvidos. Com suporte mínimo de 0,001, a mineração confirmou alguns resultados anteriores e acrescentou:

- Com 93,1% de certeza, quando o título penal for Crimes contra a Família, o autor do fato será do sexo masculino. Essa informação confirma a realidade policial, onde os homens são autores predominantes em delitos contra a família;
- Com 92,9% de confiança, os crimes previstos no chamado Estatuto do Desarmamento são praticados por homem;
- Quando a situação do envolvido for autor do fato e o título penal for “Crimes contra o Patrimônio”, 90,7% de certeza do sexo do envolvido ser masculino;
- Da mesma forma, quando o título penal for “Crimes contra a Pessoa”, há 90,2% de certeza de o autor ser homem;
- Em crimes eleitorais também há grande envolvimento de homens. Com 88,3% de confiança, quando o envolvido for autor e o título for Código Eleitoral, o sexo do autor será masculino;

7 PÓS-PROCESSAMENTO

Vencidas as etapas de pré-processamento e extração de padrões, resta agora analisar os resultados alcançados, correlacionando o novo conhecimento descoberto com a realidade, tornando-o compreensível ao âmbito de aplicação da mineração.

7.1 Analisando os padrões descobertos

De modo geral, as minerações descobriram padrões aplicáveis no contexto social da criminalidade. Nesse sentido, o conhecimento descoberto enfatizou aspectos sociais, etários, relacionando-os também com informações atinentes a armas e substâncias entorpecentes.

Do ponto de vista subjetivo (*user-driven*), considerando a experiência profissional do pesquisador como especialista no domínio do problema, sem sombra de dúvida é possível afirmar que os resultados apresentados são plausíveis e factíveis na prática. Por exemplo, quando se verifica a esmagadora predominância de homens como envolvidos em ocorrências policiais ou que a faixa etária que se destaca em ocorrências é de indivíduos de 18 a 30 anos ou ainda a correlação entre armas de calibre 38 e substâncias entorpecentes, todas estas informações são comprovadas no dia a dia policial.

Todavia, para uma melhor validação do conhecimento gerado, muitas vezes não basta apenas a análise subjetiva do especialista. Sendo assim, revelou-se prudente a confrontação de alguns resultados apresentados com dados consistentes e confiáveis. Entretanto, é importante frisar que nem sempre essa abordagem é possível. Ora, se o conhecimento gerado é novo, implícito e até então desconhecido, é evidente que nem sempre será possível compará-lo com dados pré-existentes. Pensar o contrário seria inutilizar o objetivo da mineração de dados, pois o resultado do processo sempre dependeria de dados referenciais para sua validação. Por isso é que existe a etapa de pós-processamento, sendo que a participação do

especialista é de fundamental importância. É ele quem julgará a plausibilidade do conhecimento gerado, descartando os resultados absurdos e validando aqueles com consistência lógica.

Prosseguindo, embora algumas minerações realizadas não tenham focado especificamente em presos ou autores de crimes e sim em todos os envolvidos em ocorrências policiais (vítima, autor, testemunha, dentre outros), um parâmetro aceitável de avaliação é comparar os resultados apresentados com os dados relativos aos detentos do sistema carcerário estadual. Há explicação para a solução, isto é, se o indivíduo está preso é porque praticou alguma infração penal (foi autor) que teve origem em uma ocorrência policial que possivelmente foi registrada em um boletim de ocorrência. Além disso, tendo em vista o grande volume de dados minerados, a porcentagem do conhecimento extraído referente a envolvidos categorizados como autores de crimes é substancial.

Desse modo, utilizaram-se dados estatísticos do Departamento Penitenciário do Estado do Paraná – DEPEN, órgão gestor do sistema penitenciário estadual e unidade administrativa da Secretaria de Estado da Justiça e Cidadania. Foram analisados indicadores carcerários referentes ao mês de dezembro de 2011.² Conforme dito alhures, a ideia é comparar alguns resultados gerados pelo processo de mineração com os índices fornecidos pelo DEPEN, que representam dados consistentes e confiáveis.

Na mineração de dados realizada buscando traçar o perfil dos envolvidos em ocorrências policiais restou constatada a probabilidade de 92% dos envolvidos serem do sexo masculino. Pois bem, segundo o DEPEN, em dezembro/2011 o Estado do Paraná abrigava 33.586 presos (detentos em Delegacias e Penitenciárias). Deste total, 31.143 presos são do sexo masculino representando, portanto, aproximadamente 92,7% do universo da população carcerária.

Ainda conforme o relatório, dos 20.464 presos integrados ao Sistema Penitenciários (excluindo os detentos encarcerados em Delegacias), 9.570 são homens e possuem apenas o ensino fundamental incompleto (49,4%).

Continuando, segundo a pesquisa coordenada por Prá *et al.* (2012) com a

² Dados obtidos no *site* do Departamento Penitenciário do Paraná – DEPEN, no endereço eletrônico <http://www.depen.pr.gov.br> (FORMULÁRIO..., 2012).

finalidade de traçar o perfil dos presos reincidentes inseridos no sistema penitenciário estadual, verificou-se que 81% dos detentos entrevistados estão na faixa etária entre 21 e 40 anos. Analogamente, a mineração de dados realizada também ressaltou a predominância de jovens em ocorrências policiais: em 83,3% dos casos, envolvidos com idade entre 18 e 30 anos são do sexo masculino.

Obviamente, esse estudo comparativo não buscou realizar um confronto detalhado entre o conhecimento extraído mediante mineração e os dados estatísticos oficiais até porque a natureza de uma pesquisa estatística é essencialmente diferente do processo de descoberta do conhecimento por mineração de dados. Além disso, a descoberta de padrões resultou num conhecimento novo diferente dos resultados estatísticos. Como se não bastasse, a mineração propiciou um resultado mais profundo e minucioso, apresentando índices de probabilidade que expressam o grau de certeza dos padrões gerados.

Contudo, mesmo assim o método serviu para demonstrar a viabilidade do emprego de técnicas de mineração de dados no âmbito policial, bem como atestar a confiabilidade do trabalho realizado.

8 CONCLUSÃO

O âmbito de aplicação das técnicas de mineração de dados tem crescido nos últimos anos alcançando os mais diversos ramos do conhecimento, dada a utilidade prática do conhecimento descoberto por esse processo.

O presente trabalho procurou demonstrar a aplicabilidade da mineração de dados no âmbito policial, com o estudo da base de dados que armazena os registros de ocorrências policiais no âmbito estadual. Sem a pretensão de esgotar o assunto, foi descrito todo o ciclo de MD, abrangendo as etapas de pré-processamento, mineração de dados propriamente dita e pós-processamento.

No decorrer do trabalho, alguns obstáculos foram observados, tais como dificuldade de configuração de ambiente de trabalho para uma base de dados específica, requisitos de hardware insuficientes para viabilizar uma mineração computacionalmente custosa e definição de um escopo de trabalho bem definido pela seleção de atributos relevantes. Ademais, como em grande parte dos bancos de dados, a consistência das informações armazenadas pelo sistema depende da qualidade e nível de comprometimento do usuário operante. Por se tratar de uma base cujo preenchimento dos dados depende de certa valoração subjetiva, é previsível que algumas informações não correspondam totalmente à realidade. Para tentar contornar esse problema foi realizada a mineração numa quantidade considerável de dados, com uma amostra significativa de instâncias.

Em linhas gerais, as minerações efetivadas confirmaram alguns aspectos da realidade policial, como por exemplo, a constatação do grande envolvimento de jovens adultos em ocorrências, a relação entre entorpecentes e arma de fogo, dentre outros. Outros resultados obtidos, porém, são triviais como a predominância de indivíduos do sexo masculino no envolvimento com ocorrências policiais.

Também é importante ressaltar que não foram observados resultados absurdos, completamente discrepantes com a realidade, o que por si só, é um indicativo da validade da pesquisa e da consistência dos dados minerados. Todavia, para fins de suporte a tomada de decisão em âmbito estratégico-gerencial da segurança pública minerações mais específicas e direcionadas podem ser implementadas, conforme o problema a ser trabalhado pelo gestor da área.

De qualquer forma, antes de tudo o trabalho serviu para retratar alguns aspectos da criminalidade sob outro ângulo até então desconhecido (ou ao menos intuído) mediante a aplicação de técnicas e métodos consagrados de Inteligência Artificial, demonstrando que a utilização de ferramentas apropriadas pode servir de auxílio para o entendimento da criminalidade.

Atestada a viabilidade do emprego da técnica de mineração de dados, trabalhos futuros poderão aprofundar o tema seja realizando minerações em um nível de profundidade ainda maior, ou aplicando outros métodos de extração do conhecimento, como clusterização, por exemplo.

Portanto, o campo atuação desse ramo computacional da Inteligência Artificial é vasto e de relevante aplicabilidade prática. O presente trabalho poderá servir de inspiração para pesquisas mais profundas na área policial, extraindo conhecimento que poderão ser úteis ao processo de tomada de decisão.

9 REFERÊNCIAS

ALBERTO FILHO, Luiz. **Aplicação de Técnicas de Mineração de Dados e Textos no Apoio à Tomada de Decisão em Segurança Pública**. Disponível em: <http://www.ufpa.br/ppgcc/ppgcc/files/File/Seminario_Andamento/2008/SI/artigo_dissertacao_mestrado%20luiz.pdf>. Acesso em: 05 nov. 2011.

BENICAS, Alcides Xavier; PAIXÃO, Rodinei Silva. Mineração de Dados como Ferramenta para Descoberta do Conhecimento. In: SEMINFO - SEMINÁRIO DE INFORMÁTICA, 5., 2006, Torres. **Anais...**. Torres: Roland Teodorowitsch, 2006. p. 68 - 74.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília, DF, Senado, 1998. Disponível em: < http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm> Acesso em: 15 mar 2012.

BRASIL. Decreto nº 3.695, de 21 de dezembro de 2000. Cria o Subsistema de Inteligência de Segurança Pública, no âmbito do Sistema Brasileiro de Inteligência, e dá outras providências. Disponível em: < http://www.planalto.gov.br/ccivil_03/decreto/D3695.htm>. Acesso em: 15 mar 2012.

BRASIL. Decreto nº 4.376, de 13 de setembro de 2002. Dispõe sobre a organização e o funcionamento do Sistema Brasileiro de Inteligência, instituído pela Lei nº 9.883, de 7 de dezembro de 1999, e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/2002/d4376.htm> Acesso em: 15 mar 2012.

BRASIL. Lei nº 9.883, de 7 de dezembro de 1999. Institui o Sistema Brasileiro de Inteligência, cria a Agência Brasileira de Inteligência - ABIN, e dá outras providências. Disponível em: < http://www.planalto.gov.br/ccivil_03/Leis/L9883.htm> Acesso em: 15 mar 2012.

CALIL, Leonardo Aparecido de Almeida et al. Mineração de Dados e Pós-Processamento em Padrões Descobertos. **Publicatio Uepg**: Ciências Exatas e da Terra, Ciências Agrárias e Engenharias, Ponta Grossa, v. 14, n. 03, p.207-215, dez. 2008. Disponível em: <<http://www.revistas2.uepg.br/index.php/exatas/article/viewFile/946/782>>. Acesso em: 04 jun. 2012.

CARVALHO, Deborah Ribeiro et al. Ferramenta de Pré e Pós-processamento para Data Mining. In: SEMINCO - SEMINÁRIO DE COMPUTAÇÃO, 12., 2003, Blumenau. **Anais...**. Blumenau: Unversidade Regional de Blumenau, 2003. p. 131 - 140. Disponível em: <<http://www.inf.furb.br/seminco/2003/>>. Acesso em: 10 jun. 2012.

CARVALHO, Luís Algreto Vidal de. **Datamining - A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração**. Rio de Janeiro: Editora Ciência Moderna Ltda, 2005. 225 p.

CASTELLA, Eduardo Marcelo. **Investigação Criminal e Informática: Inteligência Artificial x Boletim de Ocorrência**. Curitiba: Juruá, 2005. 162 p.

CLARKE, Bertrand; FOKOUÉ, Ernest; ZHANG, Hao Helen. **Principles and Theory for Data Mining and Machine Learning**. New York: Springer, 2009. 781 p.

COSTA, Geraldo Vieira da. **Caderno de Estudos da Disciplina Metodologia da Pesquisa Aplicada a Projetos**. Disponível em: <http://pt.scribd.com/doc/59865232/6/CALCULO-DA-AMOSTRA-ALEATORIA-SIMPLES>. Acesso em: 18 abr. 2012.

DEPARTAMENTO PENITENCIÁRIO NACIONAL. **Formulário Categoria e Indicadores Preenchidos**. Sistema Integrado de Informações Penitenciárias - Infopen. Disponível em: <<http://www.depen.pr.gov.br/arquivos/File/Dezembro.pdf>>. Acesso em: 10 jun. 2012.

DOMINGUES, Marcos Aurélio; REZENDE, Solange Oliveira. Pós-processamento de Regras de Associação usando Taxonomias. **Infocomp: Journal of Computer Science**, Lavras, p. 26-37. jun. 2005. Disponível em: <<http://www.dcc.ufla.br/Infocomp/artigos/v4.2/art04.pdf>>. Acesso em: 10 jun. 2012.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. **From Data Mining to Knowledge Discovery in Databases**. Disponível em: <<http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>>. Acesso em: 14 mar. 2012.

FERREIRA, Carlos Alberto de Souza et al. **O Data Mining na Compreensão do Fenômeno da Dor: Uma Proposta de Aplicação**. Disponível em: <http://repositorium.sdum.uminho.pt/bitstream/1822/5577/1/Artigo_DM_Dor_VesaoFinal.pdf>. Acesso em: 17 jun. 2012.

FERRO JÚNIOR, Celso Moreira; DANTAS, George Felipe de Lima. **A descoberta e a análise de vínculos na complexidade da investigação criminal moderna**. Disponível em: <<http://jus.com.br/revista/texto/10002/a-descoberta-e-a-analise-de-vinculos-na-complexidade-da-investigacao-criminal-moderna>>. Acesso em: 13 mar. 2012.

FERRO, Alexandre Lima. Inteligência de Segurança Pública e Análise Criminal. **Revista Brasileira de Inteligência**, Brasília, v. 2, n. 2, p.77-92, abr. 2006. GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data Mining: um guia prático**. Rio de Janeiro: Elsevier, 2005. 261 p.

GONZAGA, Jorge Luiz. **Dominando o PostgreSQL**. Rio de Janeiro: Editora Ciência Moderna, 2007. 228 p.

GORUNESCU, Florin. **Data Mining: Concepts, Models and Techniques**. Berlin: Springer, 2011. 356 p.

HOUAISS, Antônio; VILLAR, Mauro de Salles; FRANCO, Francisco Manoel de Mello. **Míni Houaiss - Dicionário da Língua Portuguesa**. Rio de Janeiro: Objetiva,

2001.

HUNYADI, Daniel. **Performance comparison of apriori and fp-growth algorithms in generating association rules.** In: EUROPEAN COMPUTING CONFERENCE, 2011, Paris, France. Proceedings of the european computing conference. Paris, France: Wseas Press, 2011. p.376-381.

LENZA, Pedro. **Direito Constitucional Esquemático.** 8. ed. São Paulo: Editora Método, 2005. 573 p.

MIRABETE, Julio Fabbrini. **Código de Processo Penal Interpretado.** 11. ed. São Paulo: Atlas, 2003. 1902 p.

NEVES, Rita de Cássia David das. **Pré-Processamento no Processo de Descoberta do Conhecimento em Bases de Dados.** 2003. 137 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.

NOTA Técnica n. 001/08: Estatística Criminal da Segurança Pública do Estado do Paraná. elaborado por Secretaria de Estado da Segurança Pública - Coordenadoria de Análise e Planejamento Estratégico. Disponível em: <http://www.seguranca.pr.gov.br/arquivos/File/NOTA_TECNICA_001_08.pdf>. Acesso em: 02 abr. 2012.

OYAMA, Fernando Takeshi. **Mineração Multirrelacional de Regras de Associação em Grandes Bases de Dados.** 2010. 126 p. Dissertação (Mestrado) – Curso de Ciência da Computação, Universidade Estadual Paulista. São José do Rio Preto.

PIETRO, Maria Sylvia Zanella Di. **Direito Administrativo.** 8. ed. São Paulo: Atlas, 1997.

PONCE, Julio; KARAHOCA, Adem (Ed.). **Data Mining and Knowledge Discovery in Real Life Applications.** Croatia: In-tech, 2009.

PRÁ, Alcione et al. (Org.). **Pesquisa do Perfil Sócio Demográfico e Criminal dos Presos do Estado do Paraná sob o Ângulo da Reincidência.** Departamento Penitenciário do Estado. Disponível em: <http://www.depen.pr.gov.br/arquivos/File/perfil_presos.pdf>. Acesso em: 10 jun. 2012.

QUEIROZ, Carlos Alberto Marchi de (Org.). **Manual de Polícia Judiciária:** doutrina, modelos, legislação. São Paulo: Delegacia Geral de Polícia, 2000.

QUONIAM, Luc et al. Inteligência obtida pela aplicação de data mining em bases de teses francesas sobre o Brasil. **Ciência da Informação**, Brasília, v. 30, n. 2, p.20-28, 01 ago. 2001. Quadrimestral. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000200004&lng=pt&nrm=iso&tlng=pt>. Acesso em: 17 jun. 2012.

ROMÃO, Cide Ferreira. **O que é Inteligência Policial - Discutindo um conceito.** Disponível em: <<http://www2.forumseguranca.org.br/node/24404>>. Acesso

em: 06 mar. 2012.

RUD, Olivia Parr. **Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management**. New York: John Wiley & Sons. Inc., 2001.

SECRETARIA NACIONAL DE SEGURANÇA PÚBLICA. Portaria nº 22, de 22 de julho de 2009. Aprova a Doutrina Nacional de Inteligência de Segurança Pública – DNISP.

SECRETARIA NACIONAL DE SEGURANÇA PÚBLICA. Resolução nº 1, de 15 de julho de 2009. Regulamenta o Subsistema de Inteligência de Segurança Pública - SISP, e dá outras providências.

SILVA, Glauco Carlos. **Mineração de Regras de Associação Aplicada a Dados da Secretaria Municipal de Saúde de Londrina - PR**. 2005. 94 f. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.

SOLANGE OLIVEIRA REZENDE (Org.). **Sistemas Inteligentes: fundamentos e aplicações**. Barueri: Manole, 2005. 525 p.

TAN, Pang-ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao Data Mining**. Rio de Janeiro: Editora Ciência Moderna Ltda, 2009. 896 p.

WENDT, Emerson. **Inteligência de Segurança Pública e DNISP - Aspectos iniciais**. Disponível em: <<http://www.inteligenciapolicia.com.br/2010/03/inteligencia-de-seguranca-publica-e.html>>. Acesso em: 07 mar. 2012.

APÊNDICE

TERMO DE COMPROMISSO

Eu, ALESSANDRO ROBERTO LUZ, RG nº 9.607.310-0/PR, CPF nº 267.724.218-41, Delegado de Polícia 4ª Classe da Polícia Civil do Estado do Paraná, declaro para os devidos fins de direito ser responsável pela custódia e utilização dos dados extraídos e replicados das bases de dados utilizadas nos sistemas informatizados da Polícia Civil, declarando que serão adotadas todas as cautelas para garantir o sigilo e a confidencialidade das informações, asseverando que os dados serão utilizados somente para fins de pesquisa acadêmica como trabalho de conclusão de curso em Sistemas de Informação.

Curitiba, 1º de dezembro de 2.011.

Alessandro Roberto Luz,
Delegado de Polícia.

1 Script de criação do banco de dados no Windows:

```
CREATE DATABASE bounificado
WITH OWNER = postgres
ENCODING = 'WIN1252'
TABLESPACE = pg_default
TEMPLATE = template0;
SET client_encoding = 'WIN1252';
SET check_function_bodies = false;
SET client_min_messages = warning;
COMMENT ON SCHEMA public IS 'Standard public schema';
CREATE SCHEMA sa_boletim;
ALTER SCHEMA sa_boletim OWNER TO postgres;
CREATE SCHEMA sa_historico;
ALTER SCHEMA sa_historico OWNER TO postgres;
```

2 Script de criação do banco no Linux:

```
CREATE DATABASE bounificado
WITH OWNER = postgres
ENCODING = 'LATIN1'
TABLESPACE = pg_default;
SET client_encoding = 'LATIN1';
SET check_function_bodies = false;
SET client_min_messages = warning;
COMMENT ON SCHEMA public IS 'Standard public schema';
CREATE SCHEMA sa_boletim;
ALTER SCHEMA sa_boletim OWNER TO postgres;
CREATE SCHEMA sa_historico;
ALTER SCHEMA sa_historico OWNER TO postgres;
```

3 Script de criação da tabela alvo de mineração e dados:

```
CREATE TABLE data_mining.TB_DATAMINING_2011_DEF (
tipoNatureza,
tituloPenal,
nomeJuridico,
cidadeOcorrencia,
sexoEnvolvido,
idadeEstimada,
grauInstrucao,
corCutis,
entorpecente,
especieArma,
calibreArma,
municipioEnvol,
situacaoEnvolvido,
tipoMeioEmpregado,
localOcorrencia,
estadoCivilEnvol,
provPolicial,
```

```

    grupoMeioEmpr,
    tipoLocal
) TABLESPACE "Tablespace_TCC"
AS SELECT
    tb_natureza.tipo,
    tb_titulopenal.titulo,
    tb_tipopenal.nomejuridico,
    tb_boletim.cidade,
    tb_envolvido.sexo,
    tb_envolvido.idadeestimada,
    tb_envolvido.grauinstrucao,
    tb_envolvido.corcutis,
    tb_entorpecente.descricao,
    tb_armaespecie.descricao,
    tb_armacalibre.descricao,
    tb_copelmunicipio.nome,
    tb_envolvido.situacaoenvolvido,
    tb_meioemptipo.descricao,
    tb_localespecifico.descricao,
    tb_envolvido.estadocivil,
    tb_provpoliciaaltipo.descricao,
    tb_meioempgrupo.descricao,
    tb_tipolocal.descricao
FROM
    sa_boletim.tb_boletim
LEFT JOIN sa_boletim.tb_natureza ON tb_natureza.codboletim =
tb_boletim.codboletim AND tb_natureza.anoboletim = tb_boletim.anoboletim
LEFT JOIN sa_boletim.tb_tipopenal ON tb_natureza.codtipopenal =
tb_tipopenal.codtipopenal
LEFT JOIN sa_boletim.tb_titulopenal ON tb_tipopenal.codtitulopenal =
tb_titulopenal.codtitulopenal
LEFT JOIN sa_boletim.tb_envolvido ON tb_boletim.codboletim =
tb_envolvido.codboletim AND tb_boletim.anoboletim = tb_envolvido.anoboletim
LEFT JOIN sa_boletim.tb_boentorpecente ON tb_boentorpecente.codboletim =
tb_boletim.codboletim AND
tb_boentorpecente.anoboletim = tb_boletim.anoboletim
LEFT JOIN sa_boletim.tb_entorpecente ON tb_boentorpecente.codentorpecente =
tb_entorpecente.codentorpecente
LEFT JOIN sa_boletim.tb_envolvidoentorpecente ON
tb_envolvidoentorpecente.codboletim = tb_boentorpecente.codboletim AND
tb_envolvidoentorpecente.anoboletim = tb_boentorpecente.anoboletim AND
tb_envolvidoentorpecente.codentorpecente = tb_entorpecente.codentorpecente
AND
tb_envolvidoentorpecente.codenvolvido = tb_envolvido.codenvolvido
LEFT JOIN sa_boletim.tb_boarma ON tb_boarma.codboletim =
tb_boletim.codboletim AND tb_boarma.anoboletim = tb_boletim.anoboletim
LEFT JOIN sa_boletim.tb_armaespecie ON tb_boarma.codarmaespecie =
tb_armaespecie.codarmaespecie
LEFT JOIN sa_boletim.tb_armacalibre ON tb_boarma.codarmacalibre =
tb_armacalibre.codarmacalibre
LEFT JOIN sa_boletim.tb_copelmunicipio ON tb_copelmunicipio.codmunicipio =
tb_envolvido.codmunicipio
LEFT JOIN sa_boletim.tb_envolvidoarma ON tb_envolvidoarma.codboletim =
tb_boletim.codboletim AND
tb_envolvidoarma.anoboletim = tb_boletim.anoboletim AND
tb_envolvidoarma.codarma = tb_boarma.codarma AND
tb_envolvidoarma.codenvolvido = tb_envolvido.codenvolvido
LEFT JOIN sa_boletim.tb_meiosempregado ON tb_meiosempregado.codboletim =
tb_boletim.codboletim AND
    tb_meiosempregado.anoboletim = tb_boletim.anoboletim
LEFT JOIN sa_boletim.tb_meioempgrupo ON tb_meiosempregado.codmeioempgrupo =

```

```

tb_meioempgrupo.codmeioempgrupo
LEFT JOIN sa_boletim.tb_meioemptipo ON tb_meioemptipo.codmeioempgrupo =
tb_meioempgrupo.codmeioempgrupo AND
tb_meioempregado.codmeioemptipo = tb_meioemptipo.codmeioemptipo
LEFT JOIN sa_boletim.tb_tiposambiente ON tb_tiposambiente.codboletim =
tb_boletim.codboletim AND
tb_tiposambiente.anoboletim = tb_boletim.anoboletim
LEFT JOIN sa_boletim.tb_localespecifico ON
tb_tiposambiente.codlocalespecifico = tb_localespecifico.codlocalespecifico
LEFT JOIN sa_boletim.tb_tipolocal ON tb_tipolocal.codtipolocal =
tb_localespecifico.codtipolocal AND
tb_tipolocal.codtipolocal = tb_tiposambiente.codtipolocal
AND tb_tiposambiente.codtipolocal = tb_localespecifico.codtipolocal
LEFT JOIN sa_boletim.tb_providencias ON tb_boletim.codboletim =
tb_providencias.codboletim AND
  tb_boletim.anoboletim = tb_providencias.anoboletim
LEFT JOIN sa_boletim.tb_provpoliciaaltipo ON
tb_providencias.codprovpoliciaaltipo =
tb_provpoliciaaltipo.codprovpoliciaaltipo
WHERE
  tb_boletim.anoboletim = 2011 ORDER BY RANDOM() LIMIT 300000;

```

4 Script de adição de chaves primárias:

```

ALTER TABLE data_mining.tb_datamining_2011_def ADD COLUMN id INTEGER;
CREATE SEQUENCE data_mining.dm2011_seq;
UPDATE data_mining.tb_datamining_2011_def SET id =
nextval('data_mining.dm2011_seq');
ALTER TABLE data_mining.tb_datamining_2011_def
  ALTER COLUMN id SET DEFAULT nextval('data_mining.dm2011_seq');
ALTER TABLE data_mining.tb_datamining_2011_def
  ALTER COLUMN id SET NOT NULL;
ALTER TABLE data_mining.tb_datamining_2011_def ADD UNIQUE (id);
ALTER TABLE data_mining.tb_datamining_2011_def ADD PRIMARY KEY (id);

```

5 Script de mapeamento de domínios:

```

--Acrescentando domínios no campo natureza
ALTER TABLE data_mining.tb_datamining_2011_def ADD COLUMN
natureza_descricao VARCHAR(15);
UPDATE data_mining.tb_datamining_2011_def SET natureza_descricao =
'CHAMADA' WHERE tiponatureza = '1';
UPDATE data_mining.tb_datamining_2011_def SET natureza_descricao =
'CONSUMADA' WHERE tiponatureza = '3';
UPDATE data_mining.tb_datamining_2011_def SET natureza_descricao =
'TENTADA' WHERE tiponatureza = '4';
ALTER TABLE data_mining.tb_datamining_2011_def DROP COLUMN tiponatureza;

--Acrescentando domínios no campo grau de instrução
ALTER TABLE data_mining.tb_datamining_2011_def ADD COLUMN grau_instrucao
VARCHAR(25);
UPDATE data_mining.tb_datamining_2011_def SET grau_instrucao = 'NÃO
INFORMADO' WHERE grauinstrucao = '1';

```

```

UPDATE data_mining.tb_datamining_2011_def SET grau_instrucao = 'NÃO
ALFABETIZADO' WHERE grauinstrucao = '2';
UPDATE data_mining.tb_datamining_2011_def SET grau_instrucao = '1o GRAU
INCOMPLETO' WHERE grauinstrucao = '3';
UPDATE data_mining.tb_datamining_2011_def SET grau_instrucao = '1o GRAU
COMPLETO' WHERE grauinstrucao = '4';
UPDATE data_mining.tb_datamining_2011_def SET grau_instrucao = '2o GRAU
INCOMPLETO' WHERE grauinstrucao = '5';
UPDATE data_mining.tb_datamining_2011_def SET grau_instrucao = '2o GRAU
COMPLETO' WHERE grauinstrucao = '6';
UPDATE data_mining.tb_datamining_2011_def SET grau_instrucao = '3o GRAU
INCOMPLETO' WHERE grauinstrucao = '7';
UPDATE data_mining.tb_datamining_2011_def SET grau_instrucao = '3o GRAU
COMPLETO' WHERE grauinstrucao = '8';
ALTER TABLE data_mining.tb_datamining_2011_def DROP COLUMN grauinstrucao;

--Acrescentando dominios no campo cor da cútis
ALTER TABLE data_mining.tb_datamining_2011_def ADD COLUMN cor_cutis
VARCHAR(15);
UPDATE data_mining.tb_datamining_2011_def SET cor_cutis = 'NÃO INFORMADO'
WHERE corcutis = '1';
UPDATE data_mining.tb_datamining_2011_def SET cor_cutis = 'AMARELA' WHERE
corcutis = '2';
UPDATE data_mining.tb_datamining_2011_def SET cor_cutis = 'BRANCA' WHERE
corcutis = '3';
UPDATE data_mining.tb_datamining_2011_def SET cor_cutis = 'PARDA' WHERE
corcutis = '4';
UPDATE data_mining.tb_datamining_2011_def SET cor_cutis = 'NEGRA' WHERE
corcutis = '5';
ALTER TABLE data_mining.tb_datamining_2011_def DROP COLUMN corcutis;

--Acrescentando dominios no campo situação do envolvido
ALTER TABLE data_mining.tb_datamining_2011_def ADD COLUMN
situacao_envolvido VARCHAR(30);
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido = 'VÍTIMA'
WHERE situacaoenvolvido = '1';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido = 'AUTOR'
WHERE situacaoenvolvido = '2';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =
'NOTICIADO' WHERE situacaoenvolvido = '3';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =
'DESAPARECIDO' WHERE situacaoenvolvido = '4';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =
'TESTEMUNHA' WHERE situacaoenvolvido = '5';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =
'ENCONTRADO' WHERE situacaoenvolvido = '6';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =
'NOTICIANTE' WHERE situacaoenvolvido = '7';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =
'ASSISTIDO' WHERE situacaoenvolvido = '8';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =
'DESAPARECIDO ENCONTRADO' WHERE situacaoenvolvido = '9';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =
'NOTICIANTE É VÍTIMA' WHERE situacaoenvolvido = '10';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =
'ADVERTIDO' WHERE situacaoenvolvido = '11';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =
'NOTICIADO NÃO IDENTIFICADO' WHERE situacaoenvolvido = '12';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido = 'ESTADO
É VÍTIMA' WHERE situacaoenvolvido = '13';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =

```

```
'ADOLESCENTE INFRATOR' WHERE situacaoenvolvido = '14';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =
'POLICIAL CIVIL' WHERE situacaoenvolvido = '15';
UPDATE data_mining.tb_datamining_2011_def SET situacao_envolvido =
'NOTICIANTE/NOTICIADO' WHERE situacaoenvolvido = '16';
ALTER TABLE data_mining.tb_datamining_2011_def DROP COLUMN
situacaoenvolvido;

--Acrescentando domínios no campo estado civil do envolvido
ALTER TABLE data_mining.tb_datamining_2011_def ADD COLUMN
estado_civil_envol VARCHAR(20);
UPDATE data_mining.tb_datamining_2011_def SET estado_civil_envol = 'NÃO
INFORMADO' WHERE estadocivilenvol = '1';
UPDATE data_mining.tb_datamining_2011_def SET estado_civil_envol = 'CASADO'
WHERE estadocivilenvol = '2';
UPDATE data_mining.tb_datamining_2011_def SET estado_civil_envol =
'DIVORCIADO' WHERE estadocivilenvol = '3';
UPDATE data_mining.tb_datamining_2011_def SET estado_civil_envol =
'SEPARADO' WHERE estadocivilenvol = '4';
UPDATE data_mining.tb_datamining_2011_def SET estado_civil_envol =
'SOLTEIRO' WHERE estadocivilenvol = '5';
UPDATE data_mining.tb_datamining_2011_def SET estado_civil_envol = 'VIÚVO'
WHERE estadocivilenvol = '6';
UPDATE data_mining.tb_datamining_2011_def SET estado_civil_envol =
'CONVIVENTE' WHERE estadocivilenvol = '7';
UPDATE data_mining.tb_datamining_2011_def SET estado_civil_envol = 'OUTROS'
WHERE estadocivilenvol = '8';
ALTER TABLE data_mining.tb_datamining_2011_def DROP COLUMN
estadocivilenvol;
```

6 Mineração perfil do envolvido

Association Rules

```
[situacao_envolvido = ASSISTIDO] --> [sexoenvolvido = M] (confidence:
0.764)
[idadeestimada = 18 a 30] --> [sexoenvolvido = M] (confidence: 0.833)
[grau_instrucao = 1o GRAU INCOMPLETO] --> [sexoenvolvido = M] (confidence:
0.843)
[estado_civil_envol = SOLTEIRO] --> [sexoenvolvido = M] (confidence: 0.846)
[situacao_envolvido = AUTOR] --> [sexoenvolvido = M] (confidence: 0.920)
```

7 Mineração arma x idade do envolvido

Association Rules

```
[calibrearma = 038,00] --> [especiearma = REVOLVER] (confidence: 0.893)
[idadeestimada = 18 a 30, calibrearma = 038,00] --> [especiearma =
REVOLVER] (confidence: 0.895)
```

8 Mineração droga x arma

Association Rules

```
[entorpecente = CRACK, calibrearma = 032,00] --> [especiearma = REVOLVER]
(confidence: 0.728)
[entorpecente = MACONHA, especiearma = REVOLVER] --> [calibrearma = 038,00]
(confidence: 0.739)
[especiearma = REVOLVER, entorpecente = COCAINA] --> [calibrearma = 038,00]
(confidence: 0.822)
[calibrearma = 028,00] --> [especiearma = ESPINGARDA] (confidence: 0.847)
[calibrearma = 036,00] --> [especiearma = ESPINGARDA] (confidence: 0.868)
[calibrearma = 038,00] --> [especiearma = REVOLVER] (confidence: 0.893)
[entorpecente = MACONHA, calibrearma = 038,00] --> [especiearma = REVOLVER]
(confidence: 0.896)
[entorpecente = CRACK, calibrearma = 038,00] --> [especiearma = REVOLVER]
(confidence: 0.915)
[calibrearma = 038,00, entorpecente = COCAINA] --> [especiearma = REVOLVER]
(confidence: 0.921)
```

9 Mineração meio empregado x idade do envolvido

Association Rules

```
[sexoenvolvido = M, tipomeioempregado = HOMEM/MENOR] --> [idadeestimada =
adolescente] (confidence: 0.700)
[idadeestimada = 51 a 60] --> [sexoenvolvido = M] (confidence: 0.735)
[tipomeioempregado = PESSOAS, idadeestimada = 31 a 40] --> [sexoenvolvido =
M] (confidence: 0.736)
[tipomeioempregado = HOMEM/MENOR] --> [sexoenvolvido = M] (confidence:
0.743)
[idadeestimada = 41 a 50] --> [sexoenvolvido = M] (confidence: 0.743)
[idadeestimada = acima de 61] --> [sexoenvolvido = M] (confidence: 0.766)
[idadeestimada = 31 a 40] --> [sexoenvolvido = M] (confidence: 0.772)
[tipomeioempregado = DROGA ILICITA] --> [sexoenvolvido = M] (confidence:
0.783)
[tipomeioempregado = AUTOMOVEL, idadeestimada = 41 a 50] --> [sexoenvolvido
= M] (confidence: 0.802)
[idadeestimada = 18 a 30, tipomeioempregado = PESSOAS] --> [sexoenvolvido =
M] (confidence: 0.809)
[tipomeioempregado = PESSOAS, idadeestimada = adolescente] -->
[sexoenvolvido = M] (confidence: 0.810)
[idadeestimada = adolescente] --> [sexoenvolvido = M] (confidence: 0.823)
[tipomeioempregado = HOMEM/ADULTO, idadeestimada = 41 a 50] -->
[sexoenvolvido = M] (confidence: 0.824)
[tipomeioempregado = REVOLVER] --> [sexoenvolvido = M] (confidence: 0.828)
[idadeestimada = 18 a 30] --> [sexoenvolvido = M] (confidence: 0.833)
[tipomeioempregado = AUTOMOVEL, idadeestimada = 31 a 40] --> [sexoenvolvido
= M] (confidence: 0.838)
[tipomeioempregado = HOMEM/ADULTO, idadeestimada = 31 a 40] -->
[sexoenvolvido = M] (confidence: 0.843)
[idadeestimada = 18 a 30, tipomeioempregado = AUTOMOVEL] --> [sexoenvolvido
= M] (confidence: 0.871)
[idadeestimada = 18 a 30, tipomeioempregado = DROGA ILICITA] -->
[sexoenvolvido = M] (confidence: 0.875)
[idadeestimada = 18 a 30, tipomeioempregado = HOMEM/ADULTO] -->
[sexoenvolvido = M] (confidence: 0.898)
[idadeestimada = 18 a 30, tipomeioempregado = MOTOCICLETA] -->
[sexoenvolvido = M] (confidence: 0.913)
[tipomeioempregado = MOTOCICLETA, idadeestimada = adolescente] -->
[sexoenvolvido = M] (confidence: 0.917)
[idadeestimada = adolescente, tipomeioempregado = HOMEM/MENOR] -->
[sexoenvolvido = M] (confidence: 0.950)
```

10 Mineração situação do envolvido x droga

Association Rules

```
[idadeestimada = 18 a 30, entorpecente = MACONHA] --> [situacao_envolvido =
AUTOR] (confidence: 0.727)
[idadeestimada = 31 a 40, entorpecente = COCAINA] --> [situacao_envolvido =
AUTOR] (confidence: 0.761)
[idadeestimada = 18 a 30, entorpecente = COCAINA] --> [situacao_envolvido =
AUTOR] (confidence: 0.805)
[idadeestimada = 18 a 30, entorpecente = CRACK] --> [situacao_envolvido =
AUTOR] (confidence: 0.840)
[entorpecente = MACONHA, situacao_envolvido = NOTICIADO] --> [idadeestimada
= 18 a 30] (confidence: 0.865)
[situacao_envolvido = ADOLESCENTE INFRATOR] --> [idadeestimada =
adolescente] (confidence: 0.985)
[situacao_envolvido = ADOLESCENTE INFRATOR, entorpecente = CRACK] -->
[idadeestimada = adolescente] (confidence: 0.995)
[situacao_envolvido = ADOLESCENTE INFRATOR, entorpecente = MACONHA] -->
[idadeestimada = adolescente] (confidence: 0.999)
[situacao_envolvido = ADOLESCENTE INFRATOR, entorpecente = COCAINA] -->
[idadeestimada = adolescente] (confidence: 1.000)
```

11 Mineração título penal x sexo do envolvido

Association Rules

```
[situacao_envolvido = ASSISTIDO, sexoenvolvido = F] --> [titulopenal =
OCORRENCIAS BOMBEIROS] (confidence: 0.713)
[sexoenvolvido = M, situacao_envolvido = ADVERTIDO] --> [titulopenal =
OCORRENCIAS BOMBEIROS] (confidence: 0.719)
[situacao_envolvido = ADVERTIDO] --> [titulopenal = OCORRENCIAS BOMBEIROS]
(confidence: 0.720)
[situacao_envolvido = ASSISTIDO] --> [titulopenal = OCORRENCIAS BOMBEIROS]
(confidence: 0.723)
[sexoenvolvido = F, situacao_envolvido = ADVERTIDO] --> [titulopenal =
OCORRENCIAS BOMBEIROS] (confidence: 0.724)
[sexoenvolvido = M, situacao_envolvido = ASSISTIDO] --> [titulopenal =
OCORRENCIAS BOMBEIROS] (confidence: 0.727)
[situacao_envolvido = ASSISTIDO] --> [sexoenvolvido = M] (confidence:
0.764)
[titulopenal = OCORRENCIAS BOMBEIROS, situacao_envolvido = ASSISTIDO] -->
[sexoenvolvido = M] (confidence: 0.768)
[titulopenal = OCORRENCIAS NAO DELITUOSAS, situacao_envolvido = ASSISTIDO]
--> [sexoenvolvido = M] (confidence: 0.769)
[titulopenal = LEI 11343 - NOVA LEI DE TÓXICOS] --> [sexoenvolvido = M]
(confidence: 0.784)
[titulopenal = LEI 10.826/03 - ARMAS DE FOGO] --> [sexoenvolvido = M]
(confidence: 0.819)
[situacao_envolvido = NOTICIADO] --> [sexoenvolvido = M] (confidence:
0.837)
[situacao_envolvido = AUTOR, titulopenal = LEI 11343 - NOVA LEI DE TÓXICOS]
--> [sexoenvolvido = M] (confidence: 0.857)
[titulopenal = OCORRENCIAS BOMBEIROS, situacao_envolvido = NOTICIADO] -->
[sexoenvolvido = M] (confidence: 0.859)
```

[titulopenal = OCORRENCIAS NAO DELITUOSAS, situacao_envolvido = NOTICIADO] --> [sexoenvolvido = M] (confidence: 0.863)
[situacao_envolvido = AUTOR, titulopenal = LEI 9.605/98 - CRIMES CONTRA O MEIO AMBIENTE] --> [sexoenvolvido = M] (confidence: 0.868)
[situacao_envolvido = ENCONTRADO] --> [sexoenvolvido = M] (confidence: 0.878)
[situacao_envolvido = AUTOR, titulopenal = LEI 4737/65 - CODIGO ELEITORAL] --> [sexoenvolvido = M] (confidence: 0.883)
[titulopenal = OCORRENCIAS BOMBEIROS, situacao_envolvido = ADVERTIDO] --> [sexoenvolvido = M] (confidence: 0.887)
[situacao_envolvido = ADVERTIDO] --> [sexoenvolvido = M] (confidence: 0.887)
[titulopenal = OCORRENCIAS BOMBEIROS, situacao_envolvido = ENCONTRADO] --> [sexoenvolvido = M] (confidence: 0.893)
[titulopenal = OCORRENCIAS NAO DELITUOSAS, situacao_envolvido = ADVERTIDO] --> [sexoenvolvido = M] (confidence: 0.895)
[situacao_envolvido = AUTOR, titulopenal = CRIMES CONTRA A PESSOA] --> [sexoenvolvido = M] (confidence: 0.902)
[situacao_envolvido = AUTOR, titulopenal = CRIMES CONTRA O PATRIMONIO] --> [sexoenvolvido = M] (confidence: 0.907)
[situacao_envolvido = AUTOR] --> [sexoenvolvido = M] (confidence: 0.920)
[titulopenal = OCORRENCIAS NAO DELITUOSAS, situacao_envolvido = AUTOR] --> [sexoenvolvido = M] (confidence: 0.921)
[situacao_envolvido = ADOLESCENTE INFRATOR] --> [sexoenvolvido = M] (confidence: 0.921)
[titulopenal = OCORRENCIAS BOMBEIROS, situacao_envolvido = AUTOR] --> [sexoenvolvido = M] (confidence: 0.923)
[titulopenal = OCORRENCIAS NAO DELITUOSAS, situacao_envolvido = ADOLESCENTE INFRATOR] --> [sexoenvolvido = M] (confidence: 0.925)
[titulopenal = OCORRENCIAS BOMBEIROS, situacao_envolvido = ADOLESCENTE INFRATOR] --> [sexoenvolvido = M] (confidence: 0.925)
[situacao_envolvido = AUTOR, titulopenal = LEI 10.826/03 - ARMAS DE FOGO] --> [sexoenvolvido = M] (confidence: 0.929)
[situacao_envolvido = AUTOR, titulopenal = CRIMES CONTRA A FAMILIA] --> [sexoenvolvido = M] (confidence: 0.931)