

JOÃO PAULO CAMPOLINA LAMAS

**PREDIÇÃO DE CRIMES E OTIMIZAÇÃO DE AÇÕES DE SEGURANÇA
PÚBLICA PARA CIDADES DE PEQUENO PORTE UTILIZANDO
GEOTECNOLOGIAS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Civil, para obtenção do título *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2013

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

L217p
2013

Lamas, João Paulo Campolina, 1980-
Predição de crimes e otimização de ações de segurança
pública para cidades de pequeno porte utilizando geotecnologias
/ João Paulo Campolina Lamas. – Viçosa, MG, 2013.
xvi, 113 f. : il. (algumas color.) ; 29 cm.

Inclui anexo.

Orientador: Nilcilene das Graças Medeiros.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Sistemas de Informação Geográfica - Banco de Dados.
2. Sistemas de recuperação da informação - Geografia.
3. Segurança pública. I. Universidade Federal de Viçosa.
Departamento de Engenharia Civil. Programa de Pós-Graduação
em Engenharia Civil. II. Título.


CDD 22 ed. 005.74

JOÃO PAULO CAMPOLINA LAMAS

PREDIÇÃO DE CRIMES E OTIMIZAÇÃO DE AÇÕES DE
SEGURANÇA PÚBLICA PARA CIDADES DE PEQUENO PORTE
UTILIZANDO GEOTECNOLOGIAS

Tese apresentada à Universidade
Federal de Viçosa, como parte das
exigências do Programa de Pós-Graduação
em Engenharia Civil, para obtenção do título
Doctor Scientiae.


APROVADA: 13 de novembro de 2013.


Profª Maria Lúcia Calijuri


Prof. Crysttian Arantes Paixão


Prof. Alcione de Paiva Oliveira


Prof. Gérson Rodrigues dos Santos
Coorientador


Profª. Nilcilene das Graças Medeiros
Orientadora

Dedico à minha querida esposa Camila,
ao meu amado filho Davi,
ao meus Pais Joarês e Beatriz e
Irmãos Vitor e Lara.

AGRADECIMENTOS

A Deus, por sempre estar ao meu lado, me concedendo sabedoria e paz para enfrentar os desafios da vida.

À minha querida esposa Camila, pelo carinho, companheirismo, cumplicidade, que, sempre ao meu lado, soube compreender os momentos difíceis que enfrentei nesta empreitada, me ajudando e incentivando. A você Camila, o meu muito obrigado!

Ao meu querido filho Davi, que chegou durante o doutorado trazendo muita alegria.

Aos meus pais Joarês e Beatriz, pelos valores ensinados, exemplos de bondade, honestidade e fraternidade.

Aos meus irmãos Vitor e Lara que, embora distantes, sempre procuraram me apoiar e ajudar.

À minha orientadora Prof^a Nilcilene, pela amizade, apoio, atenção, estímulo, disponibilidade e credibilidade depositada em mim durante os anos em que estive na UFV.

Aos coorientadores, Prof. Jugurta e Prof. Géron, pelos valiosos ensinamentos, sugestões, disponibilidade e paciência que tiveram comigo nesta jornada.

A todos os professores do DEC/UFV, que são verdadeiros educadores, sempre buscando o melhor de cada aluno.

Aos componentes da banca examinadora de qualificação e tese.

Aos colegas e amigos do Programa de Pós-Graduação do DEC, pelos estudos em equipe, trocas de experiências e inesquecíveis momentos de convivência.

Ao Diretor do Instituto Federal do Sudeste de Minas Gerais Campus Rio Pomba, Prof. Arnaldo (Juninho), que me possibilitou cursar estes últimos dois anos com afastamento integral.

À Diretoria de Extensão do Campus Rio Pomba, por me conceder a bolsa de incentivo à qualificação.

Aos colegas do DACC do Campus Rio Pomba, que me ajudaram de alguma forma. Especialmente ao professor e amigo Alex, pela motivação e colaboração em meus estudos.

Aos alunos do LAMIF que de alguma forma contribuíram para a realização da pesquisa.

À Polícia Militar de Rio Pomba, pelos dados repassados para a pesquisa; em especial ao amigo Capitão Tagliate, que não mediu esforços para disponibilizar o que fora pedido.

Ao colega Gilberto dos Santos, funcionário da agência do IBGE de Rio Pomba que, com muita presteza e paciência, disponibilizou os dados cartográficos utilizados.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela auxílio financeiro parcial destinado a essa pesquisa com a concessão de bolsa de estudo.

A todos que, de alguma forma, contribuíram para a realização deste trabalho.

BIOGRAFIA

JOÃO PAULO CAMPOLINA LAMAS, filho de Joarês Lamas e Beatriz Campolina Lamas, nasceu no dia 06 de maio de 1980, em Rio Pomba, Minas Gerais.

Graduou-se em Tecnologia em Processamento de Dados pelo Centro de Ensino Superior de Juiz de Fora em 06 de dezembro de 2002.

Em janeiro de 2003 foi aprovado no concurso público para professor substituto da antiga Escola Agrotécnica Federal de Rio Pomba, trabalhando como professor no curso técnico em informática até 2005.

Em 2004 começou a lecionar disciplinas de informática na Universidade Presidente Antônio Carlos na cidade de Ubá – MG.

Em 2005, foi aprovado no mestrado em Modelagem Matemática e Computacional no Centro Federal de Educação Tecnológica de Minas Gerais e também obteve aprovação no concurso público para professor do ensino básico, técnico e tecnológico da UFV para o campus de Florestal, ficando em 2º lugar.

Em setembro de 2006, tomou posse como professor efetivo do CEFET-Rio Pomba, que atualmente é o Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerias – Campus Rio Pomba.

No 2º semestre de 2009, foi aprovado para cursar o doutorado na Universidade Federal de Viçosa em Engenharia Civil na área de concentração de Informações Espaciais com término previsto para 2º semestre de 2013.

SUMÁRIO

LISTA DE FIGURAS.....	ix
LISTA DE TABELAS.....	xii
LISTA DE GRÁFICOS.....	xiii
LISTA DE SIGLAS E ABREVIATURAS.....	xiv
RESUMO	xv
ABSTRACT	xvi
1 INTRODUÇÃO.....	1
1.1 Hipótese de Pesquisa.....	3
1.2 Objetivos.....	3
1.3 Organização do Trabalho	4
2 IMPLEMENTAÇÃO DE UM SISTEMA PARA CADASTRO GEORREFERENCIADO DE BOLETIM DE OCORRÊNCIA POLICIAL PARA CIDADES DE PEQUENO PORTE.	6
Resumo.....	6
Abstract.....	6
2.1 Introdução.....	7
2.2 Banco de Dados Geográficos	8
2.2.1 <i>O modelo conceitual UML-GeoFrame</i>	10
2.2.2 <i>Esquema Conceitual</i>	13
2.3 Materiais e métodos.....	15
2.3.1 <i>Materiais utilizados</i>	15
— <i>Área de estudo</i>	16
2.3.2 <i>Métodos</i>	16
2.3.3 <i>Aquisição e Tratamento dos Dados</i>	17
2.3.4 <i>A modelagem do Sistema de cadastro de ocorrências</i>	20
2.3.5 <i>Implementação do software</i>	21
2.3.6 <i>Modelagem do banco de dados geográficos</i>	22
2.4 Resultados.....	23
2.4.1 <i>Cadastro de ocorrências policiais</i>	23
2.4.2 <i>Resultados da implementação do banco de dados geográficos</i>	26
2.5 Conclusões.....	29
REFERÊNCIAS BIBLIOGRÁFICAS.....	30
Anexo – Relações da base de dados gerada.....	32

3 ANÁLISE DE DADOS DE ÁREAS E PREDIÇÃO DE ÍNDICES DE OCORRÊNCIAS POLICIAIS POR MEIO DE REDES NEURAIS ARTIFICIAIS.....	40
Resumo.....	40
Abstract.....	40
3.1 Introdução.....	41
3.2 Objetivo.....	42
3.3 Revisão de Literatura.....	43
3.3.1 <i>Algoritmo LinearRegression</i>	44
3.3.2 <i>Algoritmo SMOReg</i>	45
3.3.3 <i>Algoritmo Multilayer Perceptron</i>	47
3.4 Séries Temporais	49
3.5 Estatística Espacial.....	49
3.6 Materiais e Métodos.....	52
3.6.1 <i>Materiais Utilizados</i>	52
— <i>Área de estudo</i>	53
3.6.2 <i>Métodos</i>	53
— <i>A aquisição de dados</i>	54
— <i>Pré-processamento dos dados</i>	55
— <i>Determinação do “fator-ocorrência”</i>	56
3.6.3 <i>Mineração dos dados</i>	57
3.6.4 <i>A escolha do algoritmo</i>	58
3.6.5 <i>Estatística Espacial</i>	58
3.7 Resultados e Discussão.....	59
3.8 Conclusões.....	81
Referência Bibliográfica	82
4 SIMULAÇÃO DE OCORRÊNCIAS POLICIAIS VISANDO OTIMIZAR O TRABALHO DA POLÍCIA MILITAR E MINIMIZAR O ÍNDICE DE OCORRÊNCIAS POLICIAIS.....	85
Resumo.....	85
Abstract.....	85
4.1 Introdução.....	86
4.2 Objetivos.....	87
4.3 Revisão de literatura.....	87
4.3.1 <i>O algoritmo Multilayer Perceptron</i>	91
4.3.2 <i>O algoritmo Apriori</i>	92

4.4	Materiais e Métodos.....	96
4.4.1	<i>Materiais Utilizados</i>	96
—	<i>Área de estudo</i>	96
4.4.2	<i>Métodos</i>	97
—	<i>Coleta de dados</i>	98
—	<i>Determinação dos atributos “TrabPol e SomaCrime”</i>	100
4.5	Simulação dos dados	101
4.5.1	<i>Resultados Obtidos</i>	103
4.6	Extração de Padrões utilizando <i>Apriori</i>	106
—	<i>A Base de Instâncias</i>	106
4.6.1	<i>Resultados Obtidos</i>	107
4.7	Conclusões.....	109
	Referências Bibliográficas	109
5	CONSIDERAÇÕES FINAIS.....	111
	REFERÊNCIAS BIBLIOGRÁFICAS.....	113

LISTA DE FIGURAS

Figura 1 – Estrutura do <i>Framework</i> Geoframe dividida em 3 níveis: planejamento, metamodelo e representação espacial.....	11
Figura 2 - Estereótipos para generalização/especialização onde (a), é representado pelas classes <i>GeographicField</i> , <i>GeographicObject</i> e <i>ConventionalOBJ</i> . As classes <i>LandUse</i> , <i>River</i> e <i>NonGov Organization</i> são exemplos das classes acima respectivamente. (b) representa os estereótipos das classes de (a) e (c) são as classes de cada estereótipo.	14
Figura 3 - Estereótipos para representação espacial das classes <i>SpatialObject</i> e <i>FieldRepresentation</i>	14
Figura 4 – Área de estudos: cidade de Rio Pomba, situada na Zona da Mata de Minas Gerais.....	16
Figura 5 - Fluxograma das atividades desenvolvidas	17
Figura 6 – Planilhas eletrônica com as ocorrências policiais fornecida pela Polícia Militar.....	18
Figura 7 – Base digital de logradouros da cidade de Rio Pomba – MG	19
Figura 8 – Base digital dos setores do município de Rio Pomba – MG.....	19
Figura 9 – Ambiente da ferramenta case ArgoCaseGeo.....	22
Figura 10 - Modelagem conceitual do banco de dados geográfico.	23
Figura 11 – Imagem com o Módulo da tela Principal do sistema.....	24
Figura 12 – Imagem com o Módulo da tela de cadastro de ocorrências	25
Figura 13 - Imagem com o Módulo da tela de seleção de trechos	25
Figura 14 - Imagem com o Módulo da tela de cadastro de pessoas envolvidas na ocorrência.....	26
Figura 15 – Reta ajustada a nuvem de pontos correspondentes as variáveis.....	45
Figura 16 – Margem de perda do SVM.	46
Figura 17 - Conexões entre neurônios	48
Figura 18 – Relatório de estatística espacial no <i>Terraview</i>	52
Figura 19 - Área de estudos: cidade de Rio Pomba, situada na Zona da Mata de Minas Gerais.....	53
Figura 20 - Fluxograma de atividades desenvolvidas	54

Figura 21 – Plano de informação com apenas setores censitários urbanos	55
Figura 22 - Erros encontrados nas tabelas da Polícia Militar.....	55
Figura 23 - Figura (A) distribuição da variável IROP do ano de 2009 pelo método estatístico de Passos Iguais e Figura (B) distribuição da variável IROP do ano de 2009 pelo método estatístico Quartil	67
Figura 24 - Figura (A) distribuição da variável IROP do ano de 2010 pelo método estatístico de Passos Iguais e Figura (B) distribuição da variável IROP do ano de 2009 pelo método estatístico Quartil	68
Figura 25 - Figura (A) distribuição da variável IROP do ano de 2011 pelo método estatístico de Passos Iguais e Figura (B) distribuição da variável IROP do ano de 2009 pelo método estatístico Quartil	69
Figura 26 - Índice de <i>Moran</i> Local da variável IROP para o ano de 2009.....	72
Figura 27 – Índice de <i>Moran</i> da variável IROP para o ano de 2010.....	73
Figura 28 - Índice de <i>Moran</i> Local da variável IROP para o ano de 2011	73
Figura 29 - <i>BoxMap</i> para a variável IROP do ano de 2009.	74
Figura 30 - <i>BoxMap</i> para a variável IROP do ano de 2010.	75
Figura 31 - <i>BoxMap</i> para a variável IROP do ano de 2010.	76
Figura 32 - Média Móvel Local da variável IROP para o ano de 2009.....	77
Figura 33 - Média Móvel Local da variável IROP para o ano de 2010.....	78
Figura 34 - Média Móvel Local da variável IROP para o ano de 2011.....	78
Figura 35 - Figura (A) distribuição de ocorrências do ano de 2012 pelo método de Passos Iguais e Figura (B) distribuição de ocorrências do ano de 2012 pelo método de Quartil.....	80
Figura 36 – Geração dos itemsets candidatos (C) e dos itemsets frequentes (L)	94
Figura 37 - Área de estudos: cidade de Rio Pomba, situada na Zona da Mata de Minas Gerais.....	97
Figura 38 – Fluxograma das atividades desenvolvidas	98
Figura 39 - Interface <i>Weka</i> aba de classificação com algoritmo <i>Multilayer Perceptron</i>	101
Figura 40 - Arquivo de base para classificação no formato <i>.arff</i>	102
Figura 41 - Arquivo de simulação para classificação no formato <i>.arff</i>	103

Figura 42 - Instâncias representando parte do banco de dados de ocorrências de crimes.....	107
Figura 43 - Regras extraídas do banco de dados de ocorrências de crimes através do algoritmo <i>Apriori</i>	108

LISTA DE TABELAS

Tabela 1 - Relação com os atributos armazenados em uma ocorrência.....	26
Tabela 2 – Dados relacionados ao cadastro de natureza de crimes.....	27
Tabela 3 - - Relação dos dados de pessoas cadastradas no sistema.....	27
Tabela 4 - Relação de pessoa_ocorrencia é a ligação entre ocorrência e pessoa..	27
Tabela 5 - Relação dos trechos do mapa de Rio Pomba-MG.....	28
Tabela 6 - Representação de todos os atributos da Tabela logradouro.....	28
Tabela 7 - Relação de setores do município.....	28
Tabela 8 –Tabela de atributos relacionados aos pontos cadastrados.....	28
Tabela 9 – Elo de ligação entre ocorrência e localização de ocorrências.....	29
Tabela 10 - Relacionamento de dados XOR.....	48
Tabela 11 - Matriz de correlação entre tipos de ocorrências policiais do ano de 2009	61
Tabela 12 - Matriz de correlação entre tipos de ocorrências policiais do ano de 2010	63
Tabela 13 - Matriz de correlação entre tipos de ocorrências policiais ano de 2011.	64
Tabela 14 - Estatística global (Índice de <i>Moran</i>).....	70
Tabela 15 - Predição de ocorrências policiais mostradas por trimestres da cidade de Rio Pomba - MG.....	80
Tabela 16 - Dados fictícios do grupo de professores da UFV.....	93
Tabela 17 - Ocorrências dos conjuntos de atributos.....	95

LISTA DE GRÁFICOS

Gráfico 1 - Gráfico dos "Erros Médios" e " Erros Quadráticos Médios"	58
Gráfico 2 - Gráfico dos tipos de crimes de acordo com os trimestres de cada ano. 60	
Gráfico 3 – Gráfico de predições de tipologia de ocorrências utilizando o algoritmo <i>Multilayer Perceptron</i> para os 6 meses subsequentes a 2011 confrontando com os dados reais dos 6 meses após 2011.	65
Gráfico 4 – Os gráficos representam a relação entre TP (Trabalho da Polícia) e SC (Soma de Crimes) para o ano de 2009: (a) com os setores de 1 a 9 e (b) com os setores de 10 a 18.	104
Gráfico 5 - Os gráficos representam a relação entre TP (Trabalho da Polícia) e SC (Soma de Crimes) para o ano de 2010: (a) com os setores de 1 a 9 e (b) com os setores de 10 a 18.	105
Gráfico 6 - Os gráficos representam a relação entre TP (Trabalho da Polícia) e SC (Soma de Crimes) para o ano de 2011: (a) com os setores de 1 a 9 e (b) com os setores de 10 a 18.	105

LISTA DE SIGLAS E ABREVIATURAS

SIG	Sistemas de Informação Geográfica
TIN	<i>Triangulated Irregular Network</i>
BDG	Banco de Dados Geográficos
SGBD	Sistema de Gerenciamento de Banco de Dados
SGBDE	Sistema de Gerenciamento de Banco de Dados Espaciais
UML	<i>Unified Modeling Language</i>
CP	Chave Primária
CE	Chave Estrangeira
SVM	<i>Support Vector Machine</i>
SMO	<i>Sequential Minimal Optimization</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
FC	Fator-Ocorrência
<i>Uco</i>	<i>unidade de ocorrências criminais</i>
<i>E.M.</i>	<i>Erro Médio</i>
<i>E.M.Q</i>	<i>Erro Quadrático Médio</i>
IROP	Índice Relativo de Ocorrências Policiais
IM	Índice de <i>Moran</i>
IML	Índice de <i>Moran</i> Local
Q	Quadrantes
GP	Grupo de Professores
UFV	Universidade Federal de Viçosa
TP	Variável TrabPol
SC	Variável SomaCrime

RESUMO

LAMAS, João Paulo Campolina, D.Sc., Universidade Federal de Viçosa, novembro de 2013. **Predição de crimes e otimização de ações de segurança pública para cidades de pequeno porte utilizando geotecnologias.** Orientadora: Nilcilene das Graças Medeiros. Coorientadores: Jugurta Lisboa Filho e Gérson Rodrigues dos Santos

O presente trabalho teve como finalidade realizar estudos sobre a área de segurança pública voltada a cidades de pequeno porte. Para tanto foi implementado um Sistema de Informações Geográficas (SIG) e um banco de dados geográfico com o objetivo de organizar a base de dados correspondente às informações disponibilizadas pela polícia militar do município de Rio Pomba – MG, visto que esta cidade, assim como a maioria das cidades de pequeno porte, ainda realizam o cadastro das atividades da Polícia Militar e o mapeamento de crimes de forma rudimentar, com a utilização de mapas de pinos. Dessa forma, por meio de um SIG, é possível obter uma gama de informações espaciais, e, além disso, com a alimentação desses dados, também uma riqueza de informações temporais, o que possibilita avaliar o comportamento de índices de criminalidade num dado município. Este trabalho também visou o estudo da otimização de processos, com consequente diminuição de investimentos em segurança pública a partir de análises geoestatísticas e predições de ocorrências por meio de algoritmos de redes neurais. Ademais, utilizou-se algoritmos de redes neurais com a finalidade de classificar o melhor investimento da polícia para os diferentes setores urbanos do município de Rio Pomba – MG a fim de possibilitar a redução dos índices de criminalidade com as regras de associação geradas pelo algoritmo *Apriori*, que demonstram a associação espaço-temporal de ocorrências policiais, setores e turnos. A partir dos resultados, pode-se verificar que a implementação de uma base de dados geográfica pode auxiliar a Polícia Militar tanto na espacialização das informações criminais, por setores, bairros, etc., como verificar o comportamento temporal dos crimes do referido município, além de poder identificar padrões de ocorrências não antes perceptíveis numa grande base de dados.

ABSTRACT

LAMAS, Joao Paulo Campolina, D.Sc., Universidade Federal de Viçosa, November, 2013. **Prediction of crimes and optimizing public security actions for small cities using geotechnology.** Advisor: Nilcilene das Graças Medeiros. Co-Advisors: Jugurta Lisboa Filho and Gerson Rodrigues dos Santos

The present work had purpose studies on the public safety area related to the small cities. Therefore, it was implemented a Geographic Information System (GIS) and geographic database with the aim of organizing the database corresponding to the information provided by Rio Pomba's Military Police - MG, since this city, as well as most small towns still hold the record of Military Police's activities and mapping rudimentary crimes, using maps pins. Thus, through a SIG it is possible to obtain a range of spatial information, and, in addition, to supply these data, temporal also a wealth of information, which allows to evaluate the behavior of crime rates in a given city. This work also aimed, the study regarding the process optimization, with consequent investment's reduction in public safety by geostatistical analyzes and predictions of occurrences through neural networks algorithms. Furthermore, it was used neural network algorithms in order to classify the best police's investment to the different urban sectors of Rio Pomba - MG in order to able the crime rates reduction with the association rules generated by the Apriori algorithm which demonstrate the spatio-temporal association of police reports, sectors and shifts. From the results, it can see that the geographic database's implementation can help both the Military Police of criminal spatial information, by sectors, neighborhoods, etc, as to verify the temporal crime's behavior of the town, in addition identify the temporal behavior that could not be identified in a large database before.

1 INTRODUÇÃO

Atualmente no Brasil verifica-se a ocorrência de significativas transformações no espaço urbano que estão relacionadas, principalmente, ao desenvolvimento socioeconômico. As cidades passaram a receber um grande contingente de pessoas para trabalharem em diversas atividades urbanas, tais como os setores industriais, de transporte e de serviços, o que acarretou um crescimento desordenado e desprovido de controle efetivo das autoridades responsáveis.

Além disso, o aumento da taxa de criminalidade deixou de ser um problema apenas dos grandes centros, tendo passado a ocorrer, também, em cidades menores.

Segundo o anuário de informações criminais para o Estado de Minas Gerais (SEDS, 2012), no ano de 2010 houve uma redução de crimes violentos de um modo geral. Entretanto, as taxas desses crimes ainda apresentam-se muito altas quando comparadas às da década de 90, visto que, no início e meados dos anos 90 essa taxa tinha um crescimento pouco acentuado.

Em 1997 o índice de criminalidade apresentou aumento de 28,42%, em relação ao ano anterior, tendo crescido aproximadamente 3,5 vezes entre os anos de 1996 até 2003. Cumpre ressaltar que a taxa de criminalidade em Minas Gerias diminuiu consideravelmente no período compreendido entre 2003 e 2010, sendo, contudo, muito alta quando comparada às duas décadas anteriores (SEDS, 2012).

Neste cenário, ações para conter a taxa de criminalidade são implementadas pelos órgãos de segurança pública de estados e municípios. Todavia, releva frisar que tais ações não devam ficar restritas ao campo da repressão, devendo abranger, principalmente, a prevenção de crimes, com o intuito de garantir a segurança, melhorando a qualidade de vida da população.

Dessa forma, destaca-se a necessidade de que novas metodologias de combate ao crime sejam estudadas e implementadas, considerando-se

as diferentes formas de cometimento de delitos que dificultam a ação por parte dos agentes de segurança pública.

Na maioria das cidades de pequeno porte o mapeamento de crimes é realizado de forma rudimentar, furando-se pinos em mapas. Porém, a utilização de um banco de dados geográfico relacionado às ocorrências policiais de um dado município, proporciona, além de uma riqueza de informação espacial, também uma riqueza de informação temporal.

Com a organização dos dados geográficos é possível gerar informações relevantes no que diz respeito à criminalidade de uma cidade. Assim, o uso de tecnologias da informação, por meio de dados estatísticos gerados pelas ocorrências policiais e também por modelos de simulação computacional, apresenta-se como opção de grande utilidade na prevenção de ocorrência de ações criminosas.

De acordo com Ebdon (1985), existem simples técnicas de estatística espacial que podem ser aplicadas utilizando métodos de cálculos manuais que são de aprendizado intuitivo. Estes métodos são medidas de tendência central, dispersão, forma, padrão e relações espaciais.

Modelos de simulação computacional de criminalidade são ferramentas por meio das quais o setor de segurança pública pode simular a ocorrência de crimes em determinados locais, tentando solucioná-los com o mínimo de contingente policial. Além disso, permitem a prevenção de crimes por meio de um policiamento direcionado na cidade.

A informatização e análise dos dados por meio de softwares adequados proporcionam informações mais detalhadas do que aquelas obtidas pela simples visualização de um mapa com pinos. As soluções de softwares apropriadas são comumente referidas como Sistemas de Informação Geográfica (SIG), que representam dados espaciais em diversas formas, tais como, pontos, linhas, polígonos, *Triangulated Irregular Network* (TIN), grade de linhas, isolinhas, dentre outras representações.

Em um ambiente de SIG, os dados descritivos são armazenados em forma de tabelas, enquanto um índice mantém uma ligação entre as tabelas e as representações espaciais. Assim, um ponto em uma base digital pode indicar o local de um roubo, enquanto seus atributos listam o tipo de crime, a data do crime, o valor dos bens roubados, dentre outros. Uma das principais

características de um SIG é a capacidade de pesquisar e selecionar informações de um determinado crime como, por exemplo, mapear roubos de um dado modelo de carro.

A aplicação de pena ao criminoso não é suficiente no combate ao crime, visto ser uma ação repressiva, posterior à ocorrência do delito, sendo de grande importância ações preventivas a fim de se evitar a prática delituosa.

Não basta recuperar um bem roubado, ou até mesmo separar uma briga. É necessário ter a percepção do antes, trabalhando-se de modo a evitar que o delito aconteça. Dessa forma, sistemas simuladores de crimes, tais como, *JackCounter* é uma ferramenta que pode atuar de modo a retratar a dinâmica da criminalidade nas cidades.

Para o adequado funcionamento dos sistemas simuladores são necessários dados dos índices de criminalidade dos setores da cidade, que podem ser obtidos junto à Polícia Militar e organizados por meio de um Banco de Dados Geográficos (BDG), que, além de ser um banco de dados convencional, possui uma extensão para dados espaço-temporais.

Um BDG tem como base modelos de dados geográficos que podem ser usados para descrever a estrutura (esquema) e as operações em um banco de dados. A modelagem de BDGs é uma tarefa complexa, pois envolve a descrição do espaço como constituinte do processo de abstração, obtendo-se assim representações adequadas aos fenômenos geográficos.

1.1 Hipótese de Pesquisa

É possível representar o comportamento espacial da criminalidade de uma cidade de pequeno porte, através da determinação e análise de padrões estabelecidos por um inventário temporal dos tipos de crimes e das ações de atuação e prevenção de ocorrências praticadas pela Polícia Militar.

1.2 Objetivos

O objetivo geral desta pesquisa é realizar um estudo sobre o comportamento de ocorrências criminais por meio de algoritmos de

aprendizagem e estatística espacial, com o intuito de verificar padrões espaciais e prever índices de criminalidade para cidades de pequeno porte.

Para alcançar este objetivo, optou-se por desenvolver um estudo de caso para a cidade de Rio Pomba, uma cidade de aproximadamente 17000 habitantes localizada na Zona da Mata do estado de Minas Gerais. Desta forma, os seguintes objetivos específicos foram estabelecidos:

- Elaborar um esquema conceitual de banco de dados geográfico para armazenar informações de ocorrências policiais;
- Implementar um banco de dados geográfico utilizando informações obtidas pelos órgãos de segurança pública da área urbana do município de Rio Pomba, referentes aos dados de criminalidade num período de 3 anos;
- Projetar e implementar um software de cadastro de ocorrências policiais para alimentar e organizar os dados que já estão cadastrados e aqueles que ainda serão cadastrados no banco de dados;
- Realizar um estudo estatístico de correlação linear dos tipos de ocorrências e apresentar um estudo de estatística espacial com base nos setores censitários;
- Utilizar algoritmos de redes neurais artificiais para a predição de tipologia de ocorrências e predição de ocorrências por setor censitário;
- Realizar uma simulação por meio de redes neurais artificiais com a finalidade de encontrar o melhor investimento, ou seja, redução de gastos e contingente da polícia militar em cada setor da cidade;
- Extrair padrões de crimes espaço-temporais por meio de algoritmos de mineração de dados;

1.3 Organização do Trabalho

O segundo capítulo corresponde ao primeiro artigo produzido, o qual descreve a implementação de um software para cadastro de ocorrências policiais espaço-temporais. O terceiro capítulo (segundo artigo) apresenta a

predição por tipologia de ocorrências e também a predição de ocorrências em setores censitários da cidade, utilizando redes neurais artificiais, além disso, faz uma análise espacial dos dados reais de ocorrências criminais. O quarto capítulo (terceiro artigo) descreve as simulações realizadas sobre as ocorrências criminais ressaltando o trabalho da polícia militar e a quantidade de ocorrências, permitindo, por meio de redes neurais artificiais, um melhor investimento dos órgãos de segurança pública. O quinto capítulo apresenta as conclusões desta pesquisa.

2 IMPLEMENTAÇÃO DE UM SISTEMA PARA CADASTRO GEORREFERENCIADO DE BOLETIM DE OCORRÊNCIA POLICIAL PARA CIDADES DE PEQUENO PORTE.

Resumo

Este artigo apresenta estudos relacionados às áreas de informações espaciais e segurança pública, visando a implementação de um Sistema de Informação Geográfica (SIG) voltado para gerência de ocorrências policiais. A definição desse sistema teve como objetivo a organização de uma base de dados geográfica correspondente às informações da polícia militar da cidade de Rio Pomba – MG, visto que, a maioria das cidades de pequeno porte, como Rio Pomba, ainda realizam o cadastro das atividades da Polícia Militar e o mapeamento de crimes de forma rudimentar, utilizando mapas de pinos. Com a organização e tratamento dos dados numa base de dados geográfica é possível gerar informações relevantes a respeito da criminalidade de uma cidade. Além disso, com a utilização de um banco de dados geográfico pode-se obter uma riqueza de informações espaciais, e com a alimentação desses dados, também uma riqueza de informações temporais, o que possibilita avaliar o comportamento de índices de criminalidade num dado município. Com a base de dados inicial sobre ocorrências criminais da cidade de Rio Pomba foi possível visualizar e espacializar as ocorrências criminais associadas aos trechos de logradouros da cidade, além do boletim de pessoas envolvidas de modo automatizado, o que pode auxiliar as práticas da segurança pública deste município.

Palavras Chave: Sistemas de Informação Geográfica, Banco de dados geográfico, Segurança pública.

Abstract

This article presents the studies related to the area of spatial information and public safety, aimed at the implementation of a Geographical Information System (GIS) facing management of police incidents. The definition of it had as objective aims to organize a database corresponding geographic information military police of Rio Pomba city-MG, once the small towns, as Rio Pomba, in the most of the cases still perform the registration of activities of the Military Police and crime mapping in a rudimentary way, using maps pins. With the organization and processing of data in a geographic database is possible to have relevant information about the criminality of the city. Besides of it, with the use of a geographic database it can get a wealth of information space, with the feed data also a wealth of temporal information, which it can able to evaluate the behavior of crime rates at a given municipality. From the database on crime in the city of Rio Pomba you can view and store the records of criminal activities associated with streets of the city, beyond the bulletin people involved in an automated way, which can aid the practice of public safety of the city.

Key Words: Geographical Information System, Geographic Database and Public Safety.

2.1 Introdução

Nas últimas décadas, as discussões sobre o tema segurança pública giraram em torno de três grandes temas, quais sejam: criminalidade urbana; o espaço urbano e seus aspectos de marginalidade e desorganização; crise nas unidades de polícia. (BEATO, 2008).

Em suas concepções, as cidades foram criadas para proporcionar segurança aos habitantes, visando encontrarem nesse ambiente um espaço de proteção e de liberdade.

Com o desenvolvimento dos grandes centros urbanos, tais locais passaram a ser sinônimos de insegurança, restringindo de diversas formas a liberdade de seus habitantes e erodindo a sensação de segurança (DAVIS, 1998). Porém, atualmente, tal fato não se aplica a apenas aos grandes centros, mas também a cidades de pequeno e médio porte.

Diante deste contexto, entidades governamentais vêm buscando soluções para reduzir a criminalidade em uma tentativa de tornar as cidades mais seguras. Neste sentido, o uso das geotecnologias tem se tornado um aliado de grande interesse por parte destas entidades, por propiciar condições de análises específicas do comportamento da criminalidade em dado local, bem como possíveis ações de prevenções a partir de dados simulados.

Uma tecnologia que pode ser utilizada para auxiliar a segurança pública é a criação de um banco de dados geográficos, com intuito de mapear a criminalidade, visando propiciar a geração de informações espaciais de determinado tipo de crime a um dado lugar. Com isto, pode-se, por exemplo, realizar análises espaciais sobre os eventos de criminalidade, como a definição de regiões onde ocorrem maior números de certos tipos de crimes. Além disso, as informações temporais também são de vital importância para atuação da polícia militar, por exemplo, para averiguar que em determinadas épocas do ano ocorre um comportamento diferente da criminalidade em comparação a outras épocas. Assim, percebe-se que informações espaço-temporais são de grande relevância para a prevenção e combate à criminalidade.

Diversas são as propostas de modelos de dados geográficos como o MODUL-R (Bédard *et al.*, 1996) que é um modelo de banco de dados geográfico onde utiliza o modelo de entidade e relacionamento para mapear dados espaciais, GMOD (Oliveira *et al.*, 1997) que é capaz de integrar com a modelagem de banco de dados geográficos com diversos SIGs, OMT-G (Borges *et al.*, 2001) que incluem primitivas, tais como classes e relacionamentos espaciais para mapear a geometria e a tipologia dos dados geográficos, GeoFrame (Lisboa Filho e Iochpe, 2008) que faz a modelagem conceitual do banco de dados geográfico por meio de classes primitivas e classes geográficas, MADS (Parent *et al.*, 1999) que também trabalha com a ideia de modelagem conceitual de entidade e relacionamento para objetos espaciais.

Este artigo descreve o desenvolvimento de um software com banco de dados geográfico cuja finalidade é cadastrar ocorrências policiais de uma cidade de pequeno porte. Como estudo de caso utilizou-se dados sobre ocorrências criminais da cidade de Rio Pomba-MG, com a finalidade de identificar e apresentar a localização de crimes em determinados intervalos de tempo.

O restante do artigo está organizado como segue. A seção 2 faz uma revisão sobre métodos de projeto de banco de dados geográficos. A seção 3 apresenta os materiais e métodos utilizados. A seção 4 mostra os resultados obtidos do estudo. A seção 5 apresenta as conclusões do trabalho e a seção 6 apresenta o referencial bibliográfico utilizado.

2.2 Banco de Dados Geográficos

De acordo com Elmasri e Navathe (2011), um modelo de dados é um conjunto de conceitos que descrevem a estrutura e operações de um banco de dados. Em um Banco de Dados Geográfico (BDG), estruturas possuem objetos e fenômenos geográficos que são complexos demais para permitir uma representação convencional completa. Para isso, é necessário um Sistema de Gerenciamento de Banco de Dados (SGBD) para gerenciar os dados sobre objetos e fenômenos geográficos. Para uma melhor interação com estes objetos e fenômenos geográficos, é necessário construir uma

abstração para que esses dados sejam representados da melhor forma possível.

Modelos de dados voltados às aplicações geográficas têm necessidades específicas em relação à abstração de conceitos e entidades, e quanto a seu tipo de representação e seu inter-relacionamento.

De acordo com Lisboa Filho e Lochpe (2008), um sistema de gerenciamento de banco de dados espaciais (SGBDE) gerencia estruturas e operações para manipulação de dados espaciais. Integrado ao SGBDE, a ferramenta SIG provê um mecanismo de análise e visualização geográfica dos dados.

Segundo Davis JR. e Queiroz.(2005), as funções de um sistema de gerência de bancos de dados espaciais, em especial operações sobre representações vetoriais (pontos, linhas e polígonos), estão subjacentes a situações típicas, tais como: seleção de um determinado objeto através da interface gráfica, determinação do relacionamento espacial entre dois objetos para garantir o estabelecimento de restrições de integridade espaciais no banco de dados, criação de mapas de distância e solução de problemas de proximidade, sobreposição de polígonos para operações de análise espacial.

A definição da estrutura de um banco de dados geográficos segue abordagens tradicionais de projeto quanto à geração de banco de dados, composta por três etapas: o projeto conceitual, o projeto lógico e o projeto físico.

Na etapa do projeto conceitual, uma forte tendência na modelagem de um banco de dados espacial é o uso da UML (*Unified Modeling Language*). Tal linguagem adota paradigmas de orientação a objetos para o desenvolvimento do esquema por meio do diagrama de classes.

No entanto, na concepção de banco de dados geográficos, é necessário estender a UML com novos elementos que permitam a modelagem de dados com características espaço-temporais dos fenômenos geográficos.

Lisboa Filho e Lochpe (2008) ressaltam que a UML é uma linguagem que possui construções que possibilitam a sua extensão. Dessa forma, o conceito de esteriótipo adequada a UML, por meio de extensões que permitem

a definição de novos elementos de modelos específicos, como o modelo conceitual *UML-GeoFrame*, que estende a UML gerando um perfil de estereótipos para apoiar a modelagem conceitual de BDGs.

Dessa forma, as aplicações geográficas necessitam de uma extensão no modelo para que estes consigam capturar a semântica dos dados geográficos, oferecendo mecanismos de abstração mais elevados e independência de implementação. Essa modelagem estendida se deve às peculiaridades dos dados geográficos, como o espaço e o tempo.

Neste trabalho o esquema conceitual foi elaborado usando o modelo *UML-GeoFrame*, proposto por Lisboa Filho e Lochpe (2008).

2.2.1 O modelo conceitual *UML-GeoFrame*¹

De modo geral, um modelo é uma abstração, uma representação simplificada de uma parcela do mundo real, composta por objetos reais.

Assim, o esquema conceitual de dados é uma descrição dos dados que devem ser armazenados em um banco de dados, ou seja, é a descrição formal da estrutura do banco de dados que compreende a descrição e a relação entre os dados, a semântica e as restrições impostas aos dados.

De acordo com Lisboa Filho e Lochpe (2008), um modelo conceitual de dados fornece uma base formal (notacional e semântica) para ferramentas e técnicas utilizadas na modelagem de dados.

Entende-se por modelagem de dados, o processo de abstração no qual somente os elementos essenciais da realidade observada são enfatizados, descartando-se os elementos não essenciais.

O modelo conceitual *UML-GeoFrame* baseia-se no paradigma da orientação a objetos, formado por um conjunto de classes, podendo ser definido como uma estrutura a ser usada como modelo de uma aplicação em um determinado domínio.

Assim, enquanto o diagrama de classes UML fornece construções semânticas para um modelo conceitual de dados, o *GeoFrame* oferece os elementos fundamentais necessários para modelar qualquer BDG. O

¹ <http://www.dpi.ufv.br/projetos/argocasegeo/download.htm>

resultado dessa integração apresenta um diagrama de fácil compreensão pelos usuários.

A principal contribuição do modelo UML-GeoFrame consiste em fornecer um conjunto de construtores que permite ao projetista de BDG realizar a modelagem dos fenômenos geográficos.

O diagrama de classes GeoFrame possui três níveis de abstração: o nível de planejamento, que compreende a classe *GeographicRegion*, cujas instâncias correspondem às áreas de interesse de aplicação, e a classe “Tema”, descrevendo os diversos temas que retratam esta área; o nível de metamodelo, que abarca as classes mais genéricas da realidade geográfica, divididas em duas categorias, quais sejam, os objetos convencionais (sem representação espacial) e os fenômenos geográficos; e o último nível, inclui as classes de objetos que permitem ao projetista abstrair o tipo de representação espacial específico para cada tipo de fenômeno geográfico, sendo possível múltiplas representações.

A Figura 1 ilustra a estrutura do GeoFrame.

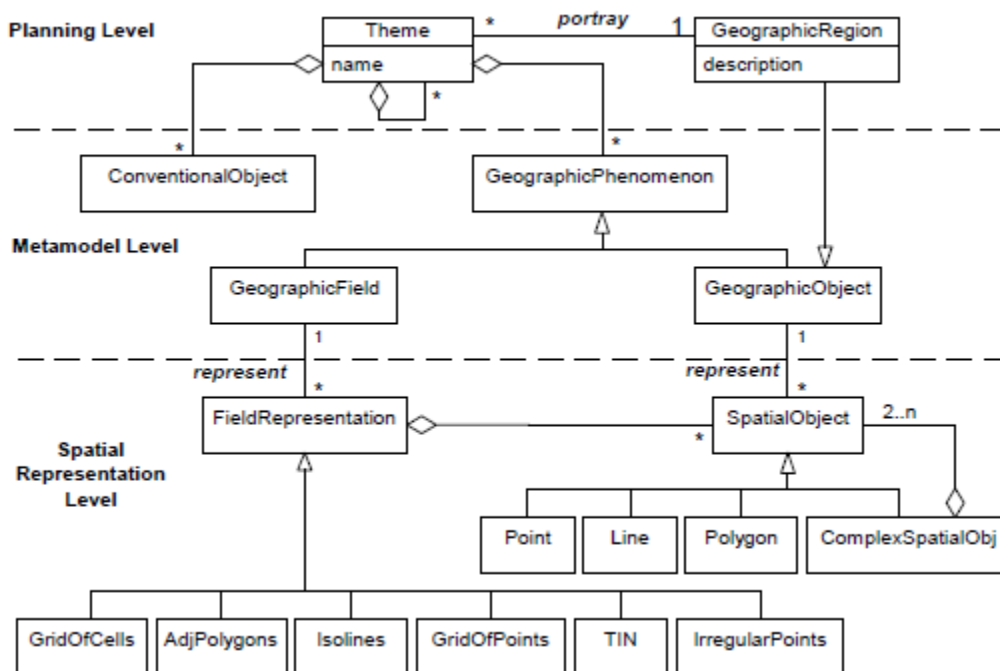


Figura 1 – Estrutura do *Framework* Geoframe dividida em 3 níveis: planejamento, metamodelo e representação espacial.

Fonte: Lisboa Filho e Lochpe, 2008.

A Figura 1 apresenta a classe *GeographicRegion* que está situada no nível de planejamento, define as áreas de interesse da aplicação.

Um exemplo da classe *GeographicRegion* é o município de Rio Pomba – MG. Uma classe *GeographicRegion* é descrita por uma coleção de temas. O agrupamento de classes que são usadas para descrever os fenômenos geográficos em temas, funciona como mecanismo para reduzir a complexidade em grandes esquemas de dados (Lisboa Filho, 2000).

Na classe *Theme* (Figura 1), cada *GeographicRegion* pode ser retratada por várias classes *Theme*. Um exemplo que representa esta classe pode ser definido pela setorização urbana. Nesta classe *Theme* é possível hierarquizar os temas malha viária e setores do município. Estas duas classes (*Theme* e *GeographicRegion*) são principais na modelagem conceitual de qualquer aplicação geográfica com modelo *UML-GeoFrame*.

Quando um modelo conceitual possui classes de temas afins, estas podem ser agrupadas em um tema mais genérico, formando uma hierarquia de temas.

No nível metamodelo a classe *GeographicPhenomenon* pode ser vista como um campo ou objeto geográfico. Estas visões apresentam diferentes maneiras de realizar a modelagem dos fenômenos geográficos.

Ademais, uma classe *GeographicPhenomenon* é composta de atributo, espaço e tempo. Além disso, a classe *GeographicPhenomenon* é especializada pelas classes *GeographicField* e *GeographicObject*.

A classe *GeographicObject* generaliza todas as classes da aplicação que são percebidas como objetos, estando incluídas as classes representativas de fenômenos geográficos individualizados, que possuem características passíveis de serem descritas como objeto, tais como parques, ruas, bares, etc.

A classe *GeographicField* generaliza os fenômenos geográficos que se enquadram na visão de campo. Referida classe é modelada por meio de funções como, por exemplo, altimetria, temperatura, população, etc.

O nível metamodelo possui outra classe denominada *ConventionalObject*, que generaliza as classes sem representação espacial. Um exemplo utilizado na pesquisa é a classe Pessoa, que armazena atributos relacionados a uma pessoa.

O nível de representação espacial prioriza “o que” e não “como”, e possui o objetivo de modelar como o componente espacial de um fenômeno

geográfico é percebido pelo usuário. Dessa forma, as classes do GeoFrame no nível de representação espacial são *SpatialObject* e *FieldRepresentation* (Figura 1).

A classe *SpatialObject* generaliza todas as classes que são usadas para a representação de um componente espacial de um objeto geográfico.

Segundo Lisboa Filho e Lochpe (2008), o objetivo da inclusão no esquema conceitual de informações espaciais relacionadas a fenômenos geográficos, é permitir a modelagem dos componentes espaciais de cada fenômeno.

Ainda de acordo com a Figura 1, as representações espaciais da classe *SpatialObject* são as classes *Point*, *Line*, *Polygon* e *ComplexSpatialObj*, que podem ser constituídos por dois ou mais objetos espaciais.

A classe *ComplexSpatialObj* é utilizada quando o fenômeno geográfico apresenta uma característica composta ou complexa, como, por exemplo, um arquipélago.

Para a abstração dos componentes espaciais são utilizados modelos geográficos adaptados à modelagem de fenômenos na visão de campo.

A classe *FieldRepresentation* possui as seguintes subclasses: *GridOfCells* que representa um grid de células, *AdjPolygons* representado por um conjunto de polígonos, *isolines* representado por um conjunto de linhas, *GridOfPoints* que é um conjunto de pontos, *TIN* que é a rede irregular de triângulos e *IrregularPoints* conjuntos de pontos irregulares.

2.2.2 Esquema Conceitual

O esquema conceitual de um banco de dados geográfico é elaborado a partir da especialização das classes abstratas do *framework* GeoFrame. Cada classe identificada na aplicação deve ser modelada como subclasse das classes Objeto Geográfico, Objeto Não Geográfico e Campo Geográfico.

Dessa forma, o GeoFrame possui três tipos de estereótipos que são utilizados para facilitar o entendimento dos esquemas de dados gerados.

A Figura 2 mostra os três estereótipos que fazem a substituição dos relacionamentos de generalização entre as classes da aplicação e as do GeoFrame.

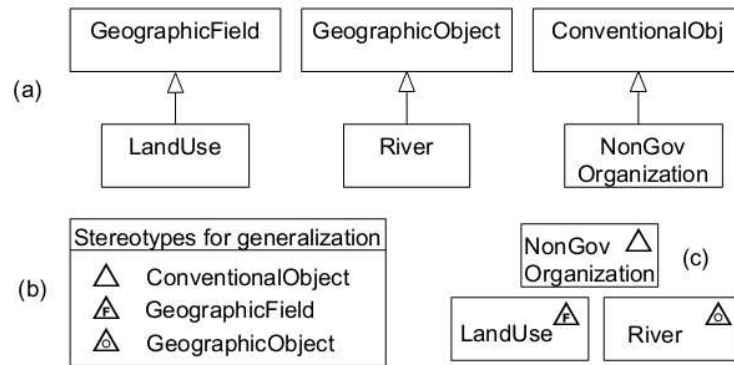


Figura 2 - Estereótipos para generalização/especialização onde (a), é representado pelas classes *GeographicField*, *GeographicObject* e *ConventionalOBJ*. As classes *LandUse*, *River* e *NonGov Organization* são exemplos das classes acima respectivamente. (b) representa os estereótipos das classes de (a) e (c) são as classes de cada estereótipo.

Fonte: LISBOA FILHO *et al.* (2008)

Segundo Lisboa Filho e Iochpe (2008), todo campo e objeto geográficos podem ser representados por diversas instâncias das classes representação campo e objeto espacial.

Um fenômeno geográfico pode possuir diversas representações por vários motivos. Por exemplo, a necessidade de múltiplas escalas, usuários com diferentes visões de um mesmo fenômeno e versões temporais.

Dessa forma, múltiplas representações de um mesmo fenômeno geográfico podem ser modeladas por meio de associações diferentes entre o fenômeno geográfico e as suas possíveis formas de abstração (Figura 3).

SpatialObject	FieldRepresentation
Point Polygon	GridOfCells Isolines TIN
Line Complex	AdjPolygons GridOfPoints IrregularPoints

Figura 3 - Estereótipos para representação espacial das classes *SpatialObject* e *FieldRepresentation*.

Fonte: LISBOA FILHO *et al.* (2008).

De acordo com Lisboa Filho (2000), para a implementação do modelo conceitual de um banco de dados geográficos no *framework* GeoFrame é utilizada uma abordagem *top-down* com três fases:

- Construção do diagrama de Temas - Uma região geográfica é representada por nenhum ou vários temas. Para aumentar a legibilidade dos esquemas a serem gerados, os temas são modelados

como subclasses da classe Tema. Em UML, o Tema é chamado *package*, que pode ser definido como um conjunto de elementos de qualquer tipo, como, por exemplo, classes e associações.

- Identificação dos fenômenos geográficos e objetos não geográficos - Nesta fase ocorre o refinamento dos temas que foram encontrados. As classes identificadas devem ser modeladas como subclasses das classes do *framework* GeoFrame, podendo ser um objeto não geográfico, campo geográfico e objeto geográfico.
- Representação das componentes espaciais dos fenômenos geográficos - Um fenômeno geográfico pode possuir diversas representações. Assim, a modelagem destas representações é formada por diferentes associações entre o fenômeno geográfico e as possíveis representações de sua componente espacial. Estas componentes são definidas por meio de uma livre combinação de diferentes estereótipos de uma mesma classe.

2.3 Materiais e métodos

2.3.1 Materiais utilizados

Para o desenvolvimento da pesquisa foram utilizados os seguintes materiais:

- Sistema de Gerência de Banco de Dados – PostgreSQL;
- Módulo PostGis;
- Ferramenta *Case ArgoCaseGeo*;
- *ArcGis* versão 10.1 (ESRI, 2011);
- Dados espaciais da base digital do IBGE da agência de Rio Pomba;
- Dados de Boletins de Ocorrência policial fornecidos pelo quartel de Polícia Militar de Rio Pomba;
- *Visual Studio* 2010 (Linguagem CSharp).

— **Área de estudo**

A área de estudo deste trabalho compreende toda a parte urbana do município de Rio Pomba – MG, que está situado na mesorregião da Zona da Mata mineira e na microrregião de Ubá.

Com população atual de aproximadamente 17000 habitantes, Rio Pomba - MG está passando por um processo de crescimento populacional que foi de, aproximadamente, 20% nos últimos 10 anos, de acordo com os censos de 2000 e 2010 (IBGE, 2013).

Sua data de fundação foi em 25 de dezembro de 1767, sendo, portanto, uma das cidades mais antigas da região.

Além disso, ela é uma cidade típica de pequeno porte, apresentado em sua arquitetura uma região central com comércio e bairros residenciais.

A Figura 4 ilustra a localização do município de Rio Pomba, caracterizando-a dentro da Zona da Mata e do estado de Minas Gerais.

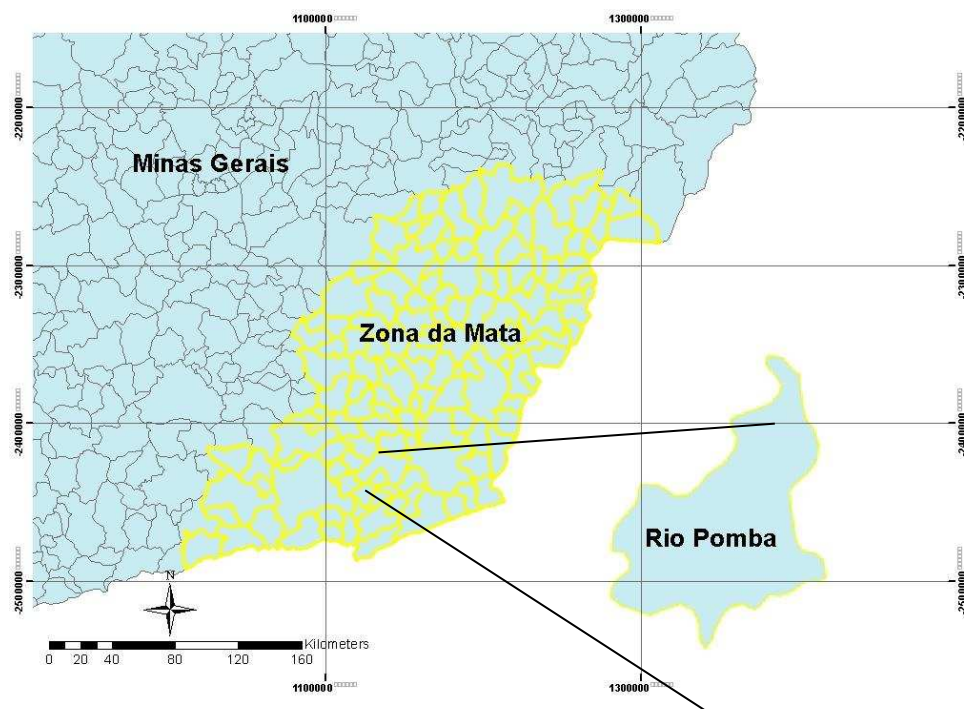


Figura 4 – Área de estudos: cidade de Rio Pomba, situada na Zona da Mata de Minas Gerais.

Fonte: Adaptado de IBGE (2011).

2.3.2 Métodos

O objetivo deste estudo foi a implementação de um sistema de informações geográficas voltado para dados de criminalidade em cidades de

pequeno porte, em especial a cidade de Rio Pomba. Desta forma, o trabalho foi dividido em algumas etapas, como a aquisição e tratamento dos dados, seguida da modelagem do banco de dados geográficos e, por fim, a implementação do sistema. Com o intuito de melhorar a compreensão das etapas descritas, um fluxograma contendo um esquema dos principais passos relacionados ao trabalho é apresentado na Figura 5.

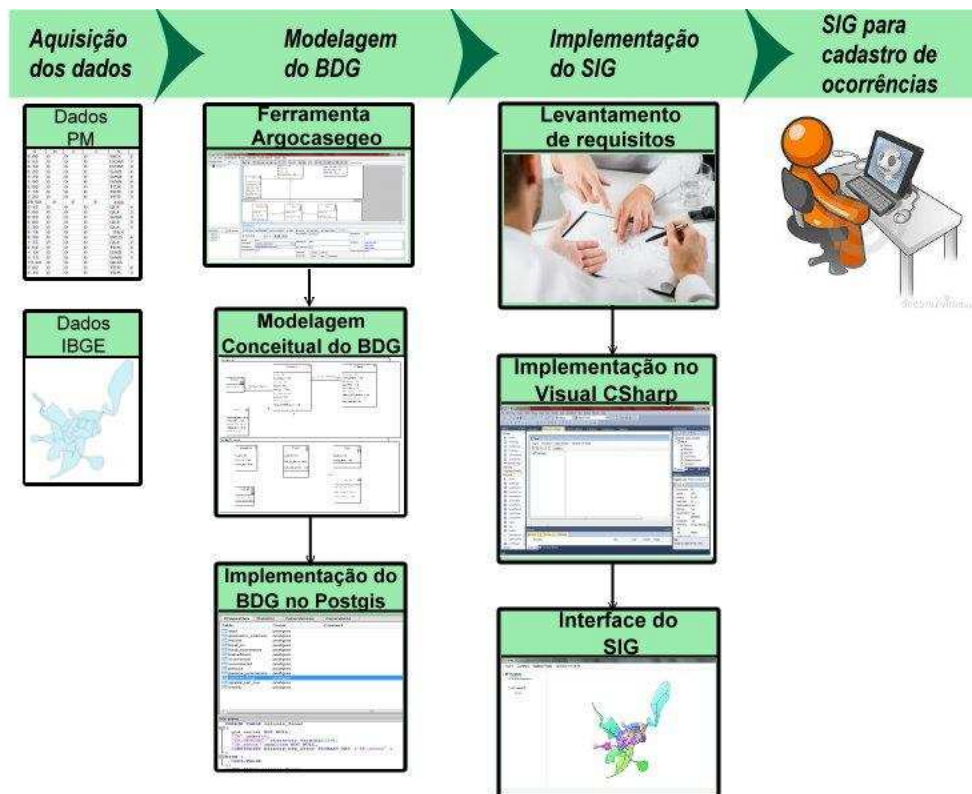


Figura 5 - Fluxograma das atividades desenvolvidas

2.3.3 Aquisição e Tratamento dos Dados

Os dados de ocorrências policiais referentes ao município de Rio Pomba foram fornecidos pelo quartel da Polícia Militar deste município, e a base cartográfica digital na escala de 1:50000 foi adquirida em uma das unidades de agência do IBGE, localizada no município de Rio Pomba.

Os dados de ocorrência provenientes do quartel da polícia militar encontravam-se no formato de planilhas, sem qualquer tipo de tratamento.

A Figura 6 mostra parte desses dados, ilustrando um exemplo de erro de inconsistência, contido nesta base.

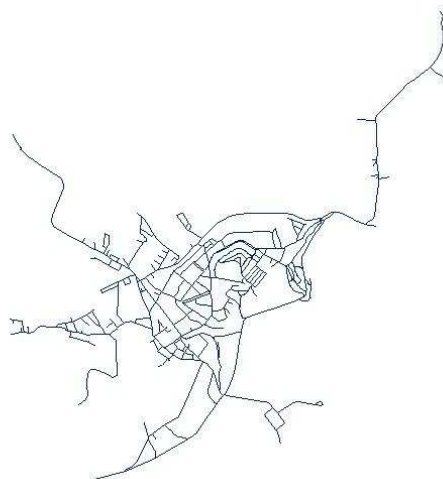


Figura 7 – Base digital de logradouros da cidade de Rio Pomba – MG
Fonte: Adaptado de IBGE (2011).

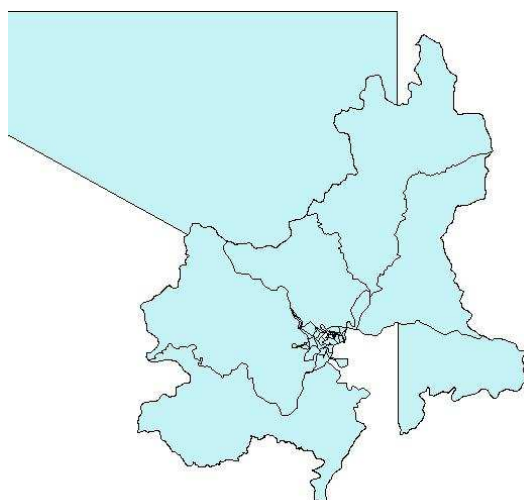


Figura 8 – Base digital dos setores do município de Rio Pomba – MG
Fonte: Adaptado de IBGE (2011).

A etapa de edição dos referidos planos de informação foi realizada no software *ArcGis* na versão 10.1.

A partir da seleção apenas dos setores censitários da área urbana, a cidade de Rio Pomba ficou subdividida em 18 setores, enquanto que o município totaliza 33 setores (incluindo a zona rural). Houve a necessidade de verificar se cada linha do plano de informação eixo_logradouro estava fragmentada ou representada em sua totalidade. Foi verificado que determinadas ruas estavam divididas em dois ou mais trechos de linhas representando o mesmo logradouro. Dessa forma, foi realizada uma junção,

de modo manual, utilizando a ferramenta “*merge*” do *ArcGis* unindo os trechos de ruas em seus respectivos logradouros.

Assim, após o pré-processamento do plano de informação eixo_logradouro, pôde-se identificar todos os trechos de cada rua da cidade.

Cada trecho de rua compreende o espaço entre dois cruzamentos de rua. Para isso, a identificação destes trechos foi realizada por meio de recortes nas linhas que representam os logradouros de forma automática pela ferramenta “*planarise*” da caixa de ferramentas *Topology* do *ArcGis*. A partir daí, pôde-se identificar todos os trechos de rua da cidade de Rio Pomba, que foram utilizados como agentes, para quantizar o número de ocorrências de cada trecho de rua e também de cada setor da cidade, uma vez que cada trecho está situado em um determinado setor.

2.3.4 A modelagem do Sistema de cadastro de ocorrências

Para implementação do módulo de cadastro das informações criminais do município de Rio Pomba, primeiramente foi necessário fazer o levantamento de requisitos, com a finalidade de verificar quais são os dados relevantes que precisam ser cadastrados e também a maneira que são geradas as informações sobre ocorrências criminais.

O esquema conceitual tem como características prioritárias o espaço, o tempo e o cadastro de ocorrências criminais. Para a etapa de modelagem é necessário realizar a análise dos requisitos. Para isso, foram realizadas várias entrevistas no quartel da Polícia Militar de Rio Pomba no intuito de projetar o aplicativo considerando que o mesmo apresentasse em sua configuração, as ferramentas adequadas para a realidade estudada.

As entrevistas foram realizadas no sentido de compreender o cotidiano do quartel da polícia, de acordo com as prioridades espaço-temporais. Os requisitos levantados nas entrevistas estão listados a seguir:

- Existência de um software integrado com o governo do estado que sirva para a integração das ocorrências policiais com a secretaria de segurança do Estado de Minas Gerais. Da forma em que se encontra, tem-se um grande número de dados que são requisitados ao sistema e assim, uma ocorrência para ser cadastrada leva em torno de 50 minutos;

- Outra deficiência encontrada no quartel é a forma de espacialização dos dados no mapa, que é feita manualmente por meio de mapas de pinos coloridos, de acordo com a prioridade da ocorrência. Dessa forma, todo mês o mapa, que é um quadro de vidro indicando os logradouros de Rio Pomba é apagado para que novas ocorrências possam ser representadas. Com isso não se consegue estabelecer informações temporais, o que dificulta a determinação de regiões com maiores índices de criminalidade (*hotspots*);
- Ficou evidente, por meio das entrevistas, a dificuldade na manutenção dos dados gerados nas ocorrências para efeito de estatísticas futuras. Tais dados são gravados em planilhas dificultando o acesso a determinada ocorrência. Ainda não existe uma normalização dos dados nessas planilhas, onde encontram-se dados relacionados a ocorrências, pessoas e natureza de crimes. Vale ainda ressaltar que é necessário realizar uma varredura minuciosa nos dados gravados nas planilhas, uma vez que existem muitas inconsistências como, por exemplo, o logradouro “Praça Doutor Último de Carvalho” está escrito de várias formas: “Pr. Último de Carvalho”, “Pr. Dr. Último de Carvalho”. Dessa forma, a criação de um banco de dados normalizado é uma opção viável para o armazenamento correto dos dados. Com isso, evita-se a redundância e preserva-se a integridade dos dados;
- O levantamento de requisitos mostrou ainda que não existe uma base digital dos dados, que é de fundamental importância no combate à criminalidade.

Após o levantamento dos requisitos definiu-se que o cadastro de ocorrências policiais.

2.3.5 Implementação do software

Para a implementação do software, definiu-se inicialmente os componentes a serem utilizados, como as telas do sistema, impressões do

sistema e consultas à base de dados, além dos códigos fontes no ambiente de programação. Assim a definição da linguagem de programação é um fator importante nesse processo, uma vez que esta pode agilizar ou retardar o processo de desenvolvimento.

A linguagem escolhida para a implementação do sistema foi o *CSharp* fornecida pela empresa *Microsoft* na ferramenta chamada *Visual Studio*. Tal escolha foi determinante pelo fato da linguagem *CSharp* possuir referências para tratar arquivos do tipo *shapefile*, que são os arquivos associados à base digital da cidade de Rio Pomba. Além disso, a linguagem de programação *CSharp* possui uma parte visual amigável, visto que o usuário com poucos conhecimentos em informática, consegue interagir com o sistema mais facilmente. Ela proporciona aos desenvolvedores uma variedade de ferramentas capazes de criar telas com um alto padrão de recursos para o usuário final.

2.3.6 Modelagem do banco de dados geográficos

Para realizar a modelagem do banco de dados geográfico relacionado à criminalidade do município de Rio Pomba, utilizou-se a ferramenta *ArgoCaseGeo* (Lisboa Filho *et al.*, 2004) para realizar a modelagem conceitual do banco de dados geográfico. A Figura 9 apresenta o ambiente de desenvolvimento da ferramenta *ArgoCaseGeo* e a Figura 10 a modelagem conceitual do BDG do sistema de cadastro de ocorrências criminais.

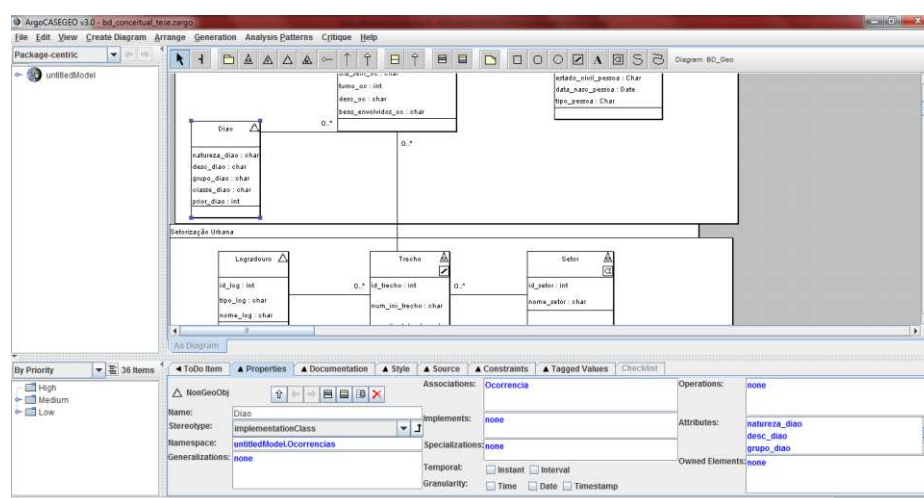


Figura 9 – Ambiente da ferramenta case ArgoCaseGeo

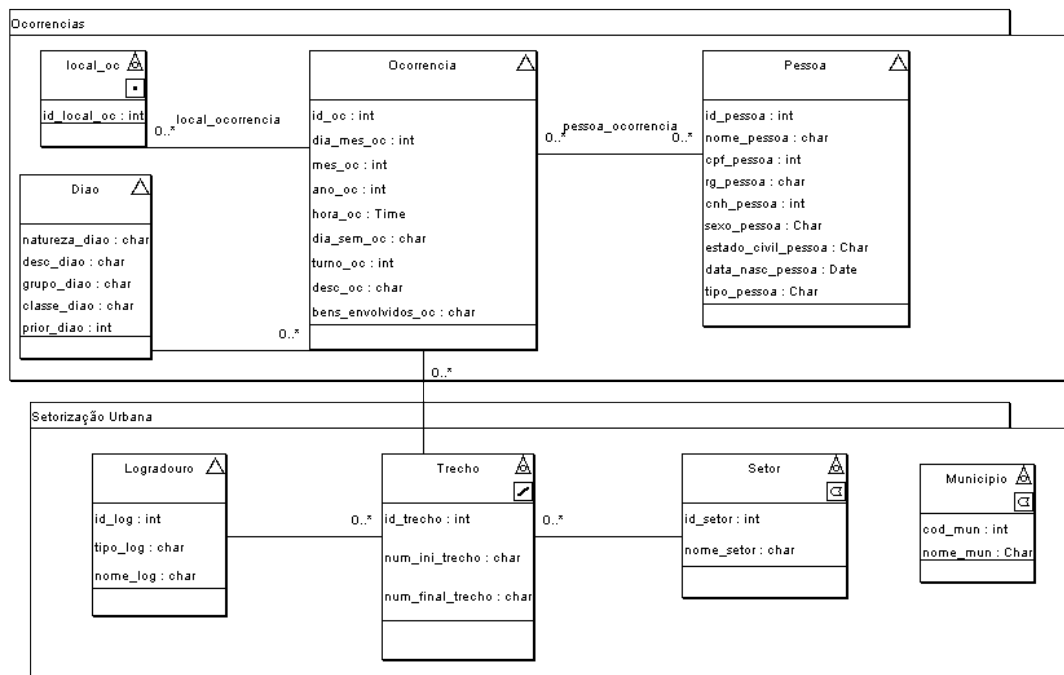


Figura 10 - Modelagem conceitual do banco de dados geográfico.

2.4 Resultados

A apresentação dos resultados está dividida em dois tópicos: no primeiro o resultado obtido foi a implementação do módulo de cadastro de ocorrências criminais; e o segundo tópico apresenta o resultado da implementação de um BDG com a finalidade de gerar informações a respeito dos dados da criminalidade.

2.4.1 Cadastro de ocorrências policiais

Foram implementados quatro módulos para o cadastro de ocorrências policiais que envolvem as seguintes telas: *Principal*, *Ocorrências*, *Pessoas* e *Trechos*. A tela *Principal* é o ambiente onde se carrega os planos de informação da base cartográfica digital a ser utilizada na associação com as ocorrências, que pode ser ampliada ou reduzida de tamanho na visualização, com o uso da ferramenta *zoom*, também tem-se a possibilidade de visualização em modo de tela cheia, ou ainda na escala original. Os menus constituintes da tela *Principal* são: Arquivo, Ocorrências, Operações de Mapas e Operações com Tabelas. A Figura 11 ilustra o módulo da tela *Principal* do sistema.

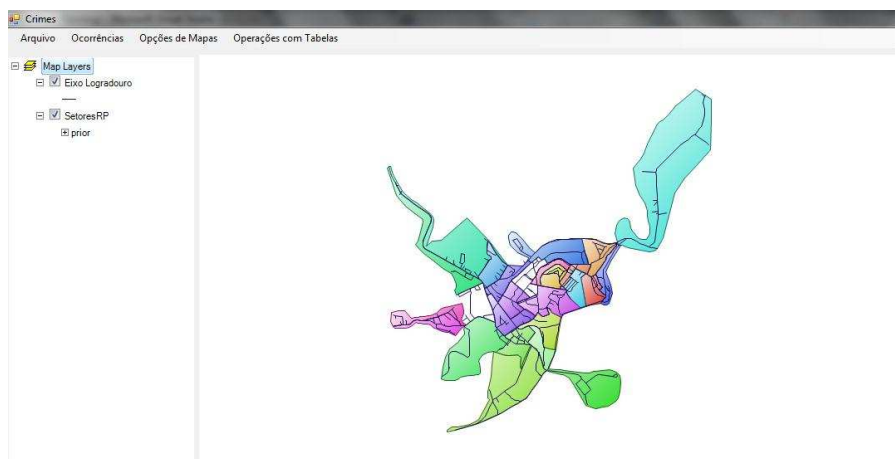


Figura 11 – Imagem com o Módulo da tela Principal do sistema.

Visando um cadastro com maior precisão comparado ao cadastro realizado anteriormente pela Polícia Militar de Rio Pomba, o menu de *Ocorrências* permite a geração de um novo arquivo *shapefile* com a geometria do tipo ponto. A ideia de gerar um arquivo pontual na malha viária da cidade de Rio Pomba se deve ao fato de existir lotes, residências ou pontos comerciais sem numeração, ficando inviável realizar uma geocodificação para associar a ocorrência a um local específico (ex. uma casa ou um bar). De acordo com o levantamento de requisitos, a ocorrência deve estar associada a um trecho de rua. Os arquivos de pontos gerados pelo sistema servem para uma melhor visualização do mapa de pinos, uma vez que eles armazenam dados temporais e espaciais. Ao selecionar o menu *Ocorrências* e o sub-menu *Cadastrar Ocorrências*, o cursor do mouse se transforma em formato de cruz que será utilizado para identificar visualmente o local onde aconteceu a referida ocorrência. Em sequência abre-se a tela *Ocorrência Policial* para o cadastro das ocorrências (Figura 12).

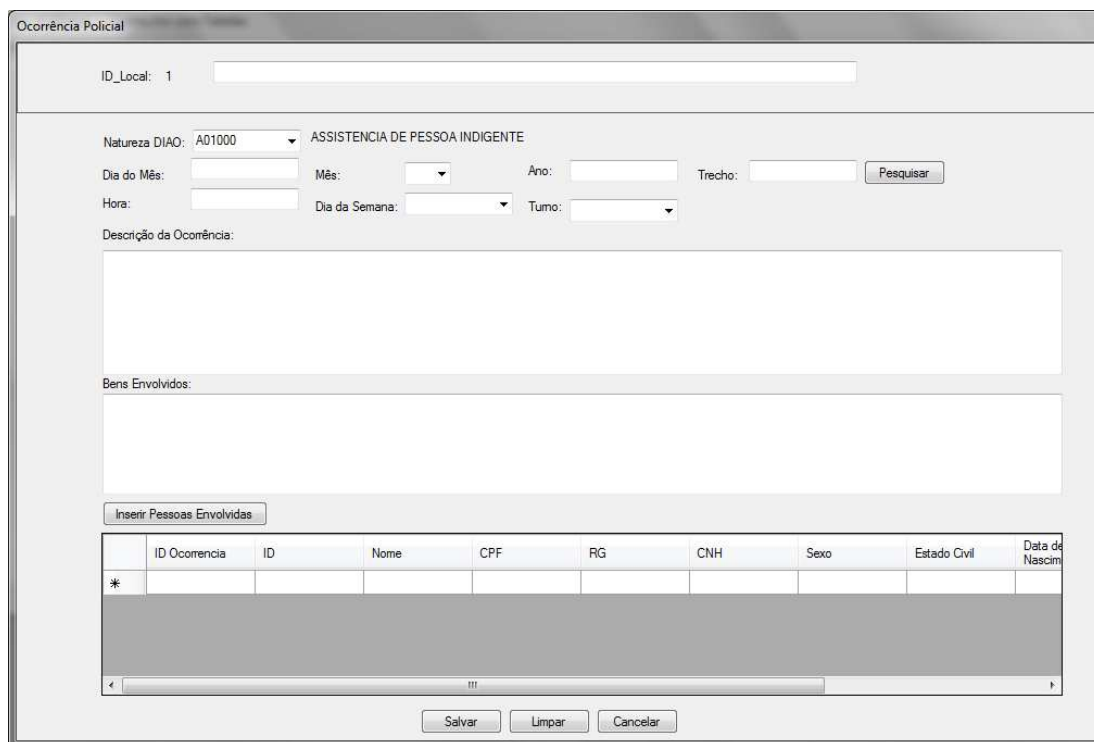


Figura 12 – Imagem com o Módulo da tela de cadastro de ocorrências

A partir da tela de ocorrências duas novas telas são apresentadas, a tela para seleção de trechos, que tem por finalidade a escolha do trecho de rua da referida ocorrência, e também a tela de pessoas envolvidas, com a finalidade de inserir os dados referentes as pessoas envolvidas na ocorrência. As Figuras 13 e 14 mostram, respectivamente, as telas de seleção de trechos e de pessoas envolvidas.

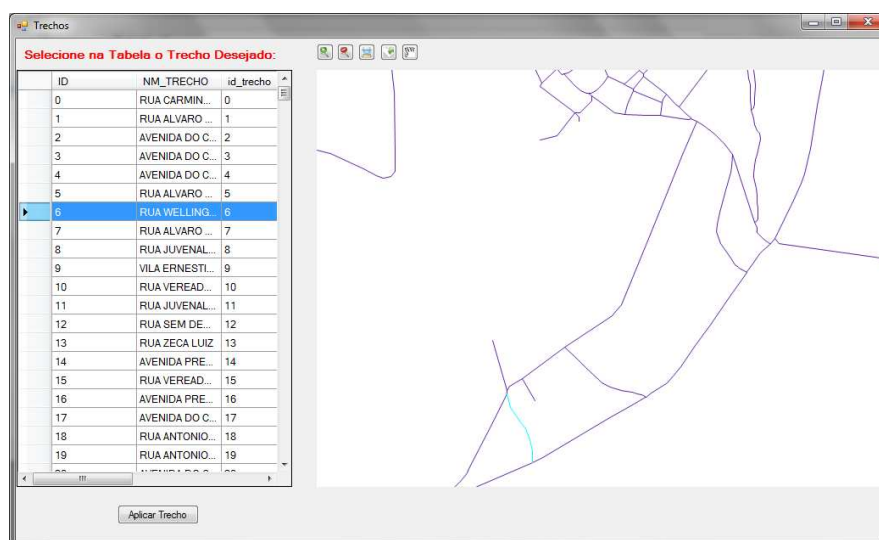


Figura 13 - Imagem com o Módulo da tela de seleção de trechos

Figura 14 - Imagem com o Módulo da tela de cadastro de pessoas envolvidas na ocorrência

2.4.2 Resultados da implementação do banco de dados geográficos

O esquema conceitual, foi transformado no modelo lógico no SGBD *PostGreSQL*. O SGBD *PostGreSQL* foi escolhido devido a sua funcionalidade de tratar dados geográficos com o módulo *PostGis*. A Figura 16 apresenta o modelo físico gerado no *PostGreSQL*.

As relações dos atributos criadas com os registros e campos apresentam-se ilustradas nas Tabelas 1, 2, 3, 4, 5, 6, 7, 8 e 9.

Tabela 1 - Relação com os atributos armazenados em uma ocorrência.

Atributo	Descrição	Tipo	Tam	Restrição
Id_oc	Identificação da Ocorrência	Integer	-	CP
Natureza_Diao	Natureza_Diao	Character	10	CE
Dia_mes_oc	Dia do mês	Integer	-	
Mes_oc	Mês da ocorrência	Integer	-	
Ano_oc	Ano da ocorrência	Integer	-	
Hora_oc	Hora da ocorrência	Time	-	
Dia_sem_oc	Dia da semana	Character	10	
Turno_oc	Turno da ocorrência	Integer	-	
Desc_oc	Descrição da ocorrência	Character	300	
Bens_envolvidos_oc	Bens envolvidos	Character	300	
	Identificação do	Integer	-	

De acordo com a Tabela 1, o atributo *Id_oc* é uma restrição de chave primária (CP) da Tabela 1. Além disso, a Tabela 1 possui outras restrições de chave estrangeira (CE), como, *Natureza_Diao* que faz referência à Tabela de DIAO, e *Id_trecho* que referencia a Tabela de Trecho.

A Tabela 2 relata os dados relacionados ao cadastro da natureza de cada crime.

Tabela 2 – Dados relacionados ao cadastro de natureza de crimes.

Atributo	Descrição	Tipo	Tam	Restrição
Natureza_Diao	Natureza_Diao	Character	10	CP
Desc_diao	Descrição da DIAO	Character	50	
Grupo_diao	Grupo da DIAO	Character	10	
Classe_diao	Classe da DIAO	Character	10	
	Prioridade da DIAO	Integer	-	

A Tabela DIAO faz ligação por meio da chave estrangeira com a Tabela ocorrência e possui uma restrição de chave primária da Tabela é Natureza_Diao.

Para fazer a ligação entre as pessoas envolvidas nas ocorrências, a Tabela pessoa_ocorrencia (Tabela 4) foi implementada com restrições de chave primária (id_oc, id_pessoa) que também são restrições de chave estrangeira, sendo que id_oc faz referência à Tabela ocorrência e cod_pessoa que referencia a Tabela pessoa.

Tabela 3 - - Relação dos dados de pessoas cadastradas no sistema

Atributo	Descrição	Tipo	Tam	Restrição
Id_pessoa	Identificador	Integer	-	CP
Nome_pessoa	Nome da pessoa	Character	50	
Cpf_pessoa	CPF da pessoa	Character	15	
Rg_pessoa	RG da pessoa	Character	15	
CNH_pessoa	Prioridade da DIAO	Character	15	
Sexo_pessoa	Sexo da pessoa	Character	1	
Est_civ_pessoa	Estado civil pessoa	Character	10	
Dt_nasc_pessoa	Data de nascimento	Date	-	
	Tipo de pessoa	Character	15	

Tabela 4 - Relação de pessoa_ocorrencia é a ligação entre ocorrência e pessoa.

Atributo	Descrição	Tipo	Tam	Restrição
Id_pessoa	Pessoa	Integer	-	CP e CE
Id_oc	Ocorrência	Integer	-	CP e CE

Os atributos de trecho estão representados na Tabela 5.

Tabela 5 - Relação dos trechos do mapa de Rio Pomba-MG.

Atributo	Descrição	Tipo	Tam	Restrição
Id_trecho	Identificador	Integer	-	CP
Nome_trecho	Nº inicial do	Character	10	
Linha_trecho	Linha no mapa do trecho	Geometry	-	
Id_setor	Identificador do setor	Integer	-	CE
Id_log	Identificador do logradouro	Integer	-	CE

O atributo Id_log é uma restrição de chave estrangeira referenciado na relação logradouro, e o campo Id_setor é uma restrição de chave estrangeira referenciada com a Tabela setor. A restrição de chave primária da Tabela trecho é o atributo Id_trecho. O atributo linha_trecho é um atributo geométrico representando uma feição linear.

Tabela 6 - Representação de todos os atributos da Tabela logradouro.

Atributo	Descrição	Tipo	Tam	Restrição
Id_log	Identificação do logradouro	Integer	-	CP
Tipo_log	Tipo de logradouro	Character	15	
Nome_log	Nome do logradouro	Character	50	

Na Tabela 6 são armazenados os atributos de todos os logradouros da cidade de Rio Pomba. Tal Tabela possui uma restrição de chave primária definida por Id_log.

Outra relação existente na modelagem é a Tabela 7 dos setores da cidade. O campo Tam_setor é um atributo geográfico que representa um polígono na Tabela e também possui uma restrição de chave primária Id_setor.

Tabela 7 - Relação de setores do município.

Atributo	Descrição	Tipo	Tam	Restrição
Id_setor	Identificação do setor	Integer	-	CP
Nome_setor	Nome do setor	Character	30	
Tam_setor	Polígono do setor	Geometry	-	

Para espacializar um ponto de ocorrência policial digitalmente no mapa, foi criado a Tabela 8 que se chama local_oc.

Tabela 8 –Tabela de atributos relacionados aos pontos cadastrados.

Atributo	Descrição	Tipo	Tam	Restrição
Id_local	Identificação do local	Integer	-	CP
Ponto_oc	Ponto no mapa	Geometry	-	

A relação local_oc possui uma restrição de chave primária e também um atributo do tipo geométrico para representação do ponto.

A interação das Tabelas ocorrência e local_oc é realizada por meio de outra Tabela chamada local_ocorrencia. A Tabela 9 representa a referida relação. A relação local_ocorrencia possui uma restrição de chave primária representada pelos atributos (id_local e Id_oc) que também são restrições de chave estrangeira, sendo que, id_local referencia a Tabela local_oc e id_oc a relação ocorrência.

Tabela 9 – Elo de ligação entre ocorrência e localização de ocorrências.

Atributo	Descrição	Tipo	Tam	Restrição
Id_local	Localização oc	Integer	-	CP e CE
Id_oc	Ocorrência	Integer	-	CP e CE

2.5 Conclusões

A implantação de um banco de dados geográficos para o cadastro de ocorrências policiais pode auxiliar de forma mais eficiente as atividades dos órgãos de Segurança Pública, quanto a organização desses dados, visto que, para algumas cidades, como Rio Pomba, tais informações ainda são armazenadas em planilhas de texto não normalizadas, sendo muito vulnerável a erros.

Dessa forma, um sistema de gerenciamento de banco de dados geográfico visa garantir a integridade dos dados que serão gravados. Além disso, também possui a função de evitar e/ou minimizar a redundância dos dados, que é um fato frequente verificado nas planilhas geradas e utilizadas pela Polícia Militar da cidade de Rio Pomba.

Deve-se considerar também, que uma visão espaço-temporal dos dados registrados em um BDG pode contribuir de modo significativo na avaliação do comportamento das ocorrências policiais em intervalos de tempo regulares, como dias, meses, trimestres ou até mesmo para anos.

Por fim, vale ressaltar que um BDG é sistema que de acordo com sua implementação, pode possuir uma interface amigável, contribuindo para uma maior facilidade de manuseio, permitindo usuários com poucos conhecimentos em informática, realizar a manipulação do sistema.

REFERÊNCIAS BIBLIOGRÁFICAS

BEATO, C.; SILVA, B.; TAVARES, R. Crime e estratégias de policiamento em espaços urbanos. **Dados**, 2008. v. 51, n. 3. Disponível em: <<http://www.scielo.br/pdf/dados/v51n3/05.pdf>>. Acesso em: 27 out. 2013.

BÉDARD, Y. *et al.*. Adapting data models for the design of spatio-temporal databases. **Computers, Environment and Urban Systems**, jan. 1996. v. 20, n. 1, p. 19–41. . Acesso em: 27 set. 2013.

BORGES, K. A. V.; DAVIS, C. A.; LAENDER, A. H. F. OMT-G: An Object-Oriented Data Model for Geographic Applications. **Geoinformatica**, 1 set. 2001. v. 5, n. 3, p. 221–260. . Acesso em: 27 set. 2013.

DAVIS, C. A.; QUEIROZ, G. R. Algoritmos geométricos e representações topológicas. **Banco de Dados Geográficos**. [S.l.]: MundoGeo, 2005.

DAVIS, M. **Ecology of Fear: Los Angeles and the Imagination of Disaster**. [S.l.]: [s.n.], 1998.

ELMASRI, R.; NAVATHE, S. **Fundamentals of database systems**. Boston: Addison-Wesley, 2011.

IBGE. **Censo Demográfico de 2010**. [S.l.]: Instituto Brasileiro de Geografia e Estatística, 2013. . Acesso em: 24 out. 2013.

LISBOA FILHO, J. **Projeto Conceitual de Banco de Dados Geográficos através da Reutilização de Esquemas, utilizando Padrões de Análise e um Framework Conceitual**. [S.l.]: Universidade Federal do Rio Grande do Sul, 2000. Disponível em: <<http://www.dpi.ufv.br/~jugurta/papers/tesejug.pdf>>. Acesso em: 27 set. 2013.

LISBOA FILHO, J.; IOCHPE, C. Modeling with a UML Profile. *In*: SHEKHAR, S.; XIONG, H. (Org.). **Encyclopedia of GIS**. [S.l.]: Springer US, 2008, p. 691–700.

LISBOA FILHO, J.; JÚNIOR, M. F. R.; DALTIO, J. ArgoCASEGEO-Uma Ferramenta CASE de Código-Aberto para o Modelo UML-GeoFrame. **Vilccosa: UFV/DPI**, 2004. Disponível em: <<http://www.dpi.ufv.br/~jugurta/papers/ideas04.pdf>>. Acesso em: 25 out. 2013.

OLIVEIRA, J. L. D.; PIRES, F.; MEDEIROS, C. B. An Environment for Modeling and Design of Geographic Applications. **Geoinformatica**, 1 abr. 1997. v. 1, n. 1, p. 29–58. . Acesso em: 27 set. 2013.

PARENT, C.; SPACCAPIETRA, S.; ZIMÁNYI, E. **Spatio-temporal conceptual models: data structures + space + time**. GIS '99. *In*: EIGHTH INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT. **Anais...** ACM. 1999. Disponível em: <<http://doi.acm.org/10.1145/320134.320142>>. Acesso em: 27 set. 2013.

Anexo – Relações da base de dados gerada.

```
---- PostgreSQL database dump
--
-- Dumped from database version 9.1.1
-- Dumped by pg_dump version 9.1.1
-- Started on 2013-10-21 13:23:27
SET statement_timeout = 0;
SET client_encoding = 'UTF8';
SET standard_conforming_strings = on;
SET check_function_bodies = false;
SET client_min_messages = warning;
SET search_path = public, pg_catalog;
SET default_tablespace = '';
SET default_with_oids = false;
--
-- TOC entry 173 (class 1259 OID 35984)
-- Dependencies: 5
-- Name: local_oc; Type: TABLE; Schema: public; Owner: postgres;
Tablespace:
--
CREATE TABLE local_oc (id_local_oc integer NOT NULL);
ALTER TABLE public.local_oc OWNER TO postgres;
--
-- TOC entry 172 (class 1259 OID 35982)
-- Dependencies: 173 5
-- Name: Local_oc_id_local_oc_seq; Type: SEQUENCE; Schema: public;
Owner: postgres
--
CREATE SEQUENCE "Local_oc_id_local_oc_seq"
START WITH 1
INCREMENT BY 1
NO MINVALUE
NO MAXVALUE
CACHE 1;
ALTER TABLE public."Local_oc_id_local_oc_seq" OWNER TO postgres;
--
-- TOC entry 2816 (class 0 OID 0)
-- Dependencies: 172
-- Name: Local_oc_id_local_oc_seq; Type: SEQUENCE OWNED BY; Schema:
public; Owner: postgres
--
ALTER SEQUENCE "Local_oc_id_local_oc_seq" OWNED BY
local_oc.id_local_oc;
--
-- TOC entry 167 (class 1259 OID 35617)
-- Dependencies: 5
-- Name: diao; Type: TABLE; Schema: public; Owner: postgres;
Tablespace:
--
CREATE TABLE diao (natureza_diao character(10) NOT NULL,
desc_diao character(150),
grupo_diao character(10),
classe_diao character(10),
prior_diao integer);
ALTER TABLE public.diao OWNER TO postgres;
--
-- TOC entry 177 (class 1259 OID 54326)
-- Dependencies: 5
```

```

-- Name: logradouro; Type: TABLE; Schema: public; Owner: postgres;
Tablespace:
--
CREATE TABLE logradouro (id integer NOT NULL,
nm_txtmemo character varying(255));
ALTER TABLE public.logradouro OWNER TO postgres;
--
-- TOC entry 180 (class 1259 OID 54418)
-- Dependencies: 5
-- Name: ocorrencial; Type: TABLE; Schema: public; Owner: postgres;
Tablespace:
--
CREATE TABLE ocorrencia (id_oc1 integer NOT NULL,
natureza_diao character(10) NOT NULL,
dia_mes_oc integer,
mes_oc integer,
ano_oc integer,
hora_oc time with time zone,
dia_sem_oc character(10),
turno_oc integer,
desc_oc character(1500),
bens_oc character(500),
id_log_oc integer);
ALTER TABLE public.ocorrencia OWNER TO postgres;
SET default_with_oids = true;
--
-- TOC entry 163 (class 1259 OID 16755)
-- Dependencies: 5
-- Name: geometry_columns; Type: TABLE; Schema: public; Owner:
postgres; Tablespace:
--
CREATE TABLE geometry_columns (f_table_catalog character
varying(256) NOT NULL,
f_table_schema character varying(256) NOT NULL,
f_table_name character varying(256) NOT NULL,
f_geometry_column character varying(256) NOT NULL,
coord_dimension integer NOT NULL,
srid integer NOT NULL,
type character varying(30) NOT NULL);
ALTER TABLE public.geometry_columns OWNER TO postgres;
SET default_with_oids = false;
--
-- TOC entry 171 (class 1259 OID 35671)
-- Dependencies: 5
-- Name: imovel; Type: TABLE; Schema: public; Owner: postgres;
Tablespace:
--
--
-- TOC entry 174 (class 1259 OID 35990)
-- Dependencies: 5
-- Name: local_ocorrencia; Type: TABLE; Schema: public; Owner:
postgres; Tablespace:
--
CREATE TABLE local_ocorrencia (
id_local_oc integer NOT NULL,
id_oc integer NOT NULL);
ALTER TABLE public.local_ocorrencia OWNER TO postgres;
--
-- TOC entry 179 (class 1259 OID 54416)
-- Dependencies: 180 5
-- Name: ocorrencial_id_oc1_seq; Type: SEQUENCE; Schema: public;

```

```

Owner: postgres
--
CREATE SEQUENCE ocorrencial_id_ocl_seq
    START WITH 1
    INCREMENT BY 1
    NO MINVALUE
    NO MAXVALUE
    CACHE 1;
ALTER TABLE public.ocorrencial_id_ocl_seq OWNER TO postgres;
--
-- TOC entry 2819 (class 0 OID 0)
-- Dependencies: 179
-- Name: ocorrencial_id_ocl_seq; Type: SEQUENCE OWNED BY; Schema:
public; Owner: postgres
--
ALTER SEQUENCE ocorrencial_id_ocl_seq OWNED BY ocorrencial.id_ocl;
--
-- TOC entry 169 (class 1259 OID 35622)
-- Dependencies: 5
-- Name: pessoa; Type: TABLE; Schema: public; Owner: postgres;
Tablespace:
--
CREATE TABLE pessoa (id_pessoa integer NOT NULL,
nome_pessoa character(30),
cpf_pessoa character(14),
rg_pessoa character(20),
cnh_pessoa character(20),
sexo_pessoa text,
estado_civil_pessoa text,
data_nasc_pessoa date,
tipo_pessoa character(15));
ALTER TABLE public.pessoa OWNER TO postgres;
--
-- TOC entry 175 (class 1259 OID 37345)
-- Dependencies: 5 169
-- Name: pessoa_estado_civil_pessoa_seq; Type: SEQUENCE; Schema:
public; Owner: postgres
--
CREATE SEQUENCE pessoa_estado_civil_pessoa_seq
START WITH 1
INCREMENT BY 1
NO MINVALUE
NO MAXVALUE
CACHE 1;
ALTER TABLE public.pessoa_estado_civil_pessoa_seq OWNER TO
postgres;
--
-- TOC entry 2820 (class 0 OID 0)
-- Dependencies: 175
-- Name: pessoa_estado_civil_pessoa_seq; Type: SEQUENCE OWNED BY;
Schema: public; Owner: postgres
--
ALTER SEQUENCE pessoa_estado_civil_pessoa_seq OWNED BY
pessoa.estado_civil_pessoa;
--
-- TOC entry 168 (class 1259 OID 35620)
-- Dependencies: 5 169
-- Name: pessoa_id_pessoa_seq; Type: SEQUENCE; Schema: public;
Owner: postgres
--
CREATE SEQUENCE pessoa_id_pessoa_seq

```

```

START WITH 1
INCREMENT BY 1
NO MINVALUE
NO MAXVALUE
CACHE 1;
ALTER TABLE public.pessoa_id_pessoa_seq OWNER TO postgres;
--
-- TOC entry 2821 (class 0 OID 0)
-- Dependencies: 168
-- Name: pessoa_id_pessoa_seq; Type: SEQUENCE OWNED BY; Schema:
public; Owner: postgres
--
ALTER SEQUENCE pessoa_id_pessoa_seq OWNED BY pessoa.id_pessoa;
--
-- TOC entry 176 (class 1259 OID 37601)
-- Dependencies: 5
-- Name: pessoa_ocorrencia; Type: TABLE; Schema: public; Owner:
postgres; Tablespace:
--

CREATE TABLE pessoa_ocorrencia (
id_oc integer NOT NULL,
id_pessoa integer NOT NULL);
ALTER TABLE public.pessoa_ocorrencia OWNER TO postgres;
--
-- TOC entry 183 (class 1259 OID 63944)
-- Dependencies: 5
-- Name: setores_final; Type: TABLE; Schema: public; Owner:
postgres; Tablespace:
--
CREATE TABLE setores_final (gid integer NOT NULL,
"ID" numeric,
"CD_GEOCODI" character varying(254),
"ID_setor" smallint NOT NULL);
ALTER TABLE public.setores_final OWNER TO postgres;
--
-- TOC entry 182 (class 1259 OID 63942)
-- Dependencies: 183 5
-- Name: setores_final_gid_seq; Type: SEQUENCE; Schema: public;
Owner: postgres
--
CREATE SEQUENCE setores_final_gid_seq
START WITH 1
INCREMENT BY 1
NO MINVALUE
NO MAXVALUE
CACHE 1;
ALTER TABLE public.setores_final_gid_seq OWNER TO postgres;
--
-- TOC entry 2822 (class 0 OID 0)
-- Dependencies: 182
-- Name: setores_final_gid_seq; Type: SEQUENCE OWNED BY; Schema:
public; Owner: postgres
--
ALTER SEQUENCE setores_final_gid_seq OWNED BY setores_final.gid;
--
-- TOC entry 162 (class 1259 OID 16747)
-- Dependencies: 5
-- Name: spatial_ref_sys; Type: TABLE; Schema: public; Owner:
postgres; Tablespace:
--

```

```

CREATE TABLE spatial_ref_sys (srid integer NOT NULL,
auth_name character varying(256),
auth_srid integer,
srtext character varying(2048),
proj4text character varying(2048));
ALTER TABLE public.spatial_ref_sys OWNER TO postgres;
--
-- TOC entry 178 (class 1259 OID 54362)
-- Dependencies: 5
-- Name: trecho; Type: TABLE; Schema: public; Owner: postgres;
Tablespace:
--
CREATE TABLE trecho (id double precision NOT NULL,
nm_trecho character varying(50),
id_setor integer,
id_log integer);
ALTER TABLE public.trecho OWNER TO postgres;
--
-- TOC entry 2780 (class 2604 OID 35674)
-- Dependencies: 170 171 171
-- Name: id_imovel; Type: DEFAULT; Schema: public; Owner: postgres
--
ALTER TABLE imovel ALTER COLUMN id_imovel SET DEFAULT
nextval('imovel_id_imovel_seq'::regclass);
--
-- TOC entry 2781 (class 2604 OID 35987)
-- Dependencies: 173 172 173
-- Name: id_local_oc; Type: DEFAULT; Schema: public; Owner:
postgres
--
ALTER TABLE local_oc ALTER COLUMN id_local_oc SET DEFAULT
nextval('"Local_oc_id_local_oc_seq"::regclass);
--
-- TOC entry 2782 (class 2604 OID 54421)
-- Dependencies: 179 180 180
-- Name: id_ocl; Type: DEFAULT; Schema: public; Owner: postgres
--
ALTER TABLE ocorrencial ALTER COLUMN id_ocl SET DEFAULT
nextval('ocorrencial_id_ocl_seq'::regclass);
--
-- TOC entry 2779 (class 2604 OID 35625)
-- Dependencies: 168 169 169
-- Name: id_pessoa; Type: DEFAULT; Schema: public; Owner: postgres
--
ALTER TABLE pessoa ALTER COLUMN id_pessoa SET DEFAULT
nextval('pessoa_id_pessoa_seq'::regclass);
--
-- TOC entry 2783 (class 2604 OID 63947)
-- Dependencies: 182 183 183
-- Name: gid; Type: DEFAULT; Schema: public; Owner: postgres
--
ALTER TABLE setores_final ALTER COLUMN gid SET DEFAULT
nextval('setores_final_gid_seq'::regclass);
--
-- TOC entry 2795 (class 2606 OID 35989)
-- Dependencies: 173 173
-- Name: Local_oc_pkey; Type: CONSTRAINT; Schema: public; Owner:
postgres; Tablespace:
--
ALTER TABLE ONLY local oc ADD CONSTRAINT "Local oc pkey" PRIMARY

```

```

KEY (id_local_oc);
--
-- TOC entry 2787 (class 2606 OID 16762)
-- Dependencies: 163 163 163 163 163
-- Name: geometry_columns_pk; Type: CONSTRAINT; Schema: public;
Owner: postgres; Tablespace:
--
ALTER TABLE ONLY geometry_columns ADD CONSTRAINT
geometry_columns_pk PRIMARY KEY (f_table_catalog, f_table_schema,
f_table_name, f_geometry_column);
--
-- TOC entry 2797 (class 2606 OID 35994)
-- Dependencies: 174 174 174
-- Name: local_ocorrencia_pkey; Type: CONSTRAINT; Schema: public;
Owner: postgres; Tablespace:
--
ALTER TABLE ONLY local_ocorrencia ADD CONSTRAINT
local_ocorrencia_pkey PRIMARY KEY (id_local_oc, id_oc);
--
-- TOC entry 2799 (class 2606 OID 37605)
-- Dependencies: 176 176 176
-- Name: pessoa_ocorrencia_pkey; Type: CONSTRAINT; Schema: public;
Owner: postgres; Tablespace:
--
ALTER TABLE ONLY pessoa_ocorrencia ADD CONSTRAINT
pessoa_ocorrencia_pkey PRIMARY KEY (id_oc, id_pessoa);
--
-- TOC entry 2789 (class 2606 OID 35735)
-- Dependencies: 167 167
-- Name: pk_diao; Type: CONSTRAINT; Schema: public; Owner:
postgres; Tablespace:
--
ALTER TABLE ONLY diao ADD CONSTRAINT pk_diao PRIMARY KEY
(natureza_diao);
--
-- TOC entry 2803 (class 2606 OID 54396)
-- Dependencies: 178 178
-- Name: pk_id_trecho; Type: CONSTRAINT; Schema: public; Owner:
postgres; Tablespace:
--
ALTER TABLE ONLY trecho ADD CONSTRAINT pk_id_trecho PRIMARY KEY
(id);
--
-- TOC entry 2801 (class 2606 OID 54347)
-- Dependencies: 177 177
-- Name: pk_log; Type: CONSTRAINT; Schema: public; Owner: postgres;
Tablespace:
--
ALTER TABLE ONLY logradouro ADD CONSTRAINT pk_log PRIMARY KEY (id);
--
-- TOC entry 2805 (class 2606 OID 54547)
-- Dependencies: 180 180
-- Name: pk_ocl; Type: CONSTRAINT; Schema: public; Owner: postgres;
Tablespace:
--
ALTER TABLE ONLY ocorrencia ADD CONSTRAINT pk_ocl PRIMARY KEY
(id_ocl);
--
-- TOC entry 2791 (class 2606 OID 35686)

```

```

-- Dependencies: 169 169
-- Name: pk_pessoa; Type: CONSTRAINT; Schema: public; Owner:
postgres; Tablespace:
--
ALTER TABLE ONLY pessoa ADD CONSTRAINT pk_pessoa PRIMARY KEY
(id_pessoa);
--
-- TOC entry 2807 (class 2606 OID 63959)
-- Dependencies: 183 183
-- Name: primary_key_setor; Type: CONSTRAINT; Schema: public;
Owner: postgres; Tablespace:
--
ALTER TABLE ONLY setores_final ADD CONSTRAINT primary_key_setor
PRIMARY KEY ("ID_setor");
--
-- TOC entry 2785 (class 2606 OID 16754)
-- Dependencies: 162 162
-- Name: spatial_ref_sys_pkey; Type: CONSTRAINT; Schema: public;
Owner: postgres; Tablespace:
--
ALTER TABLE ONLY spatial_ref_sys ADD CONSTRAINT
spatial_ref_sys_pkey PRIMARY KEY (srid);
--
-- TOC entry 2812 (class 2606 OID 54397)
-- Dependencies: 178 177 2800
-- Name: fk_log; Type: FK CONSTRAINT; Schema: public; Owner:
postgres
--
ALTER TABLE ONLY trecho ADD CONSTRAINT fk_log FOREIGN KEY (id_log)
REFERENCES logradouro(id);
--
-- TOC entry 2813 (class 2606 OID 63960)
-- Dependencies: 2806 178 183
-- Name: fk_setor; Type: FK CONSTRAINT; Schema: public; Owner:
postgres
--
ALTER TABLE ONLY trecho ADD CONSTRAINT fk_setor FOREIGN KEY
(id_setor) REFERENCES setores_final("ID_setor");
--
-- TOC entry 2808 (class 2606 OID 35995)
-- Dependencies: 2794 174 173
-- Name: local_ocorrencia_id_local_oc_fkey; Type: FK CONSTRAINT;
Schema: public; Owner: postgres
--
ALTER TABLE ONLY local_ocorrencia ADD CONSTRAINT
local_ocorrencia_id_local_oc_fkey FOREIGN KEY (id_local_oc)
REFERENCES local_oc(id_local_oc);
--
-- TOC entry 2809 (class 2606 OID 36000)
-- Dependencies: 174 166
-- Name: local_ocorrencia_id_oc_fkey; Type: FK CONSTRAINT; Schema:
public; Owner: postgres
--
ALTER TABLE ONLY local_ocorrencia ADD CONSTRAINT
local_ocorrencia_id_oc_fkey FOREIGN KEY (id_oc) REFERENCES
ocorrencia(id_oc);
--
-- TOC entry 2810 (class 2606 OID 37606)
-- Dependencies: 166 176
-- Name: pessoa_ocorrencia_id_oc_fkey; Type: FK CONSTRAINT; Schema:
public; Owner: postgres

```

```
--
ALTER TABLE ONLY pessoa_ocorrencia ADD CONSTRAINT
pessoa_ocorrencia_id_oc_fkey FOREIGN KEY (id_oc) REFERENCES
ocorrencia(id_oc);
--
-- TOC entry 2811 (class 2606 OID 37611)
-- Dependencies: 176 2790 169
-- Name: pessoa_ocorrencia_id_pessoa_fkey; Type: FK CONSTRAINT;
Schema: public; Owner: postgres
--
ALTER TABLE ONLY pessoa_ocorrencia ADD CONSTRAINT
pessoa_ocorrencia_id_pessoa_fkey FOREIGN KEY (id_pessoa) REFERENCES
pessoa(id_pessoa);
-- Completed on 2013-10-21 13:23:29
--
-- PostgreSQL database dump complete
--
```

3 ANÁLISE DE DADOS DE ÁREAS E PREDIÇÃO DE ÍNDICES DE OCORRÊNCIAS POLICIAIS POR MEIO DE REDES NEURAI ARTIFICIAIS.

Resumo

Através dos dados de ocorrências policiais coletados no município de Rio Pomba/MG, este artigo produz uma gama de artefatos de software para otimizar os processos e diminuir os investimentos na segurança pública. Para tanto, realiza uma análise estatística tipológica, uma análise de dados da setorização urbana e uma predição de ocorrência de crimes. A estatística convencional, pelo método de *spearman* mostra a correlação de tipologias de ocorrências, que é utilizada para verificar sua correlação linear. A estatística espacial de ocorrências policiais é a forma escolhida para a compreensão do inter-relacionamento dos eventos espaciais. A predição de ocorrências é realizada por meio do algoritmo Redes Neurais para o cálculo futuro das localização das ocorrências.

Palavras Chaves: Ocorrências Policiais, análise de dados de área, predição de ocorrências policiais, tipologia de ocorrências e setorização de ocorrências.

Abstract

Beyond of the policy reports collected at the Rio Pomba city, the article produce lots of kinds of sotware to otmize the process and reduce the investments in public safety. The study presents a statistical analysis typological, spatial analysis and a prediction of sectorization for crimes. The statistical conventional method of correlation spearman show types of occurrences, is used to check a linear correlation. Spatial analysis of police reports is the choice for understanding the interrelationship of spatial events. The prediction of events is performed using the Neural Networks algorithm for calculating the location of future occurrences.

Key Words: Occurrences Police, Statistical analysis predicting police incidents, occurrences and type of sectorization of occurrence.

3.1 Introdução

Atualmente, nosso país vem passando por grandes transformações no tocante à criminalidade, com mudanças significativas quanto a seu perfil social. Diferentemente do que ocorria no passado, cada vez mais, pessoas não pertencentes à classe social de baixa renda têm figurado como autores principais ou coadjuvantes dessa criminalidade.

O grande volume de ocorrências criminais está ocasionando sérios problemas em muitos estados no Brasil. Por exemplo, no anuário de informações criminais para o estado de Minas Gerais, (SEDS, 2012) em relação ao ano de 2010, consta, de modo geral, uma redução nos crimes considerados violentos, porém a taxa desses crimes ainda é bastante elevada, se comparada à taxa de ocorrência da década de 90. No início e meados de 90 essa taxa tinha um crescimento pouco acentuado. No entanto, no final da década de 90 houve um crescimento de aproximadamente 3,5 vezes, considerando os anos de 1996 até 2003.

Para o estado de Minas Gerais, cenário onde está sendo realizado esse trabalho, houve uma redução considerável de aproximadamente 40% da taxa de criminalidade no intervalo de 2004 a 2011. Contudo, ainda é muito elevada em relação às duas décadas anteriores.

Na maioria das vezes, em cidades de pequeno porte, os órgãos de segurança pública estão sujeitos a realizarem o cadastro das ocorrências criminais furando pinos de variadas cores (para diferenciar os tipos de crimes) em mapas analógicos, a fim de se ter um cenário das atividades criminais e um mapeamento, mesmo que estático, dos tipos de crimes, e os locais de sua ocorrência. O problema deste tipo de abordagem é que não há o armazenamento contínuo de informações, e desta forma, não se tem como definir e analisar o comportamento de uma dada variável, especificamente a criminalidade, nem como realizar uma análise espacial a respeito dessas informações, dificultando ações localizadas mais efetivas, bem como práticas de prevenção por parte da segurança pública.

À medida que os dados são informatizados e analisados por meio de software específicos, as informações geradas a partir de um banco de dados geográficos podem ser mais adequadas do que simplesmente uma

visualização por meio de mapa pinos. As soluções de software para dados geográficos são comumente referidas como Sistemas de Informação Geográfica (SIG), que representam dados espaciais em diversas formas geométricas: pontos, linhas, polígonos, TIN (*Triangular Irregular Network*), grade de linhas, isolinhas, dentre outras representações.

Porém, alguns usuários de SIG encontram dificuldades para fazer análises mais sofisticadas pelo fato do software não disponibilizar ferramentas para análise estatística de forma satisfatória. Por outro lado, isto pode ser resolvido por meio da estatística espacial, que possuem métodos de análises espaciais associados a áreas delimitadas por polígonos, como, por exemplo, o cálculo do índice de *Moran* de uma determinada variável para todas as cidades de um estado.

Outra ferramenta poderosa no apoio a tomada de decisão é a mineração de dados, que tem se mostrado bastante eficaz no auxílio quanto à tomada de decisão. Na pesquisa em questão, seu uso permitiu explorar com rapidez e eficiência as ocorrências policiais de cada setor censitário do município. Existe na literatura uma variedade de técnicas de mineração de dados voltados para análise da criminalidade com resultados promissores (Brown e Oxford, 2001, Chen *et al.*, 2004, Adderley, *et al.* 2007 e Vasconcelos e Furtado, 2005). No entanto, a metodologia da maioria das pesquisas enfatizou uma análise criminal numérico-quantitativa temporal, e poucos estudos consideram uma metodologia espaço-temporal.

Na mineração de dados foram utilizados três algoritmos para o processo de classificação dos dados: *SMOReg*, *MultilayerPerceptron* e *LinearRegression*. O primeiro algoritmo trabalha com a ideia de *Support Vector Machine* (SVM), o segundo com o conceito de redes neurais artificiais e o terceiro com regressão linear.

3.2 Objetivo

Nesta pesquisa, propõe-se uma metodologia de análise de ocorrências criminais por meio da mineração de dados, acoplado a um banco de dados geográfico das ocorrências policiais, com intuito de mapear o cenário de ocorrências policiais de uma cidade de pequeno porte, bem como realizar a predição destas ocorrências criminais.

3.3 Revisão de Literatura

Vários trabalhos relacionados ao estudo da criminalidade com auxílio de SIG e de mineração de dados têm sido realizados. Dentre eles, alguns estão descritos a seguir.

Chainey *et al.* (2008) utilizam a técnica de mapeamento de *hotspot* que é usada para auxiliar a determinação do local em que um crime pode acontecer, com base em dados antigos com intuito de informar ações futuras. Dessa forma, segundo os autores, é uma técnica para prever onde o crime pode ocorrer, usando a premissa de que os padrões de retrospectiva de crime são um indicador útil para os padrões futuros.

Na literatura existem técnicas de mapeamento que podem ser usadas para identificar e explorar padrões de crime, tais como, métodos estatísticos e algoritmos de redes neurais artificiais. Estas técnicas de mapeamento têm sido objeto de várias análises, mas essas avaliações têm demonstrado que elas apresentam resultados diferentes em termos de localização, tamanho e forma das áreas. Assim, deve-se considerar a diferença entre as técnicas de mapeamento *hotspot* para prever os padrões espaciais de crimes, permitindo que os setores de segurança pública determinem algum nível de precisão na utilidade da técnica para prever padrões de crimes futuros.

A visão do espaço (do grego *choros*) e do tempo (*chronos*) é uma experiência subjetiva do ser humano. De acordo com Dias *et al.* (2005), o espaço e o tempo se misturam ao se descrever uma realidade. Assim, pode-se modelar a superfície terrestre usando modelos de objetos espaciais, que podem corresponder a parcelas do solo, ou ainda utilizar os modelos de campos geográficos, que podem representar, por exemplo, a variação espacial da vegetação.

Para produzir uma representação do mundo real com o objetivo de elaborar um sistema de informação espaço-temporal muitas questões precisam ser investigadas e respondidas (FRANK, 2004). Tais questões envolvem a visão de mundo inerente ao sistema como, por exemplo, regras aplicáveis, comportamento dos objetos ao longo do tempo, interpretação da variação do tempo, natureza das mudanças e a influência dos processos de medida.

Em outra vertente, Palocsay *et al.* (2000), abordam o problema de prever a reincidência de crimes utilizando redes neurais artificiais. Diversos estudos (Corcoran *et al.* 2003, Li, *et al.* 2010, Keyvanpour, *et al.* 2011) sobre inteligência computacional usando mineração de dados apresentam-se como alternativa viável aos modelos estatísticos convencionais para problemas de classificação (WILSON e SHARDA, 1994). Neste sentido, o algoritmo *Multilayer Perceptron* relaciona a abordagem de reconhecimento estatístico e de padrões para os problemas de classificação e regressão por meio do aprendizado da base de dados (MURTAGH 1991).

Na mineração de dados, visando o processo de classificação de padrões, além do algoritmo *Multilayer Perceptron*, pode-se utilizar outras abordagens como *LinearRegression* e *SMOReg* que serão tratados nos subcapítulos a seguir.

3.3.1 Algoritmo LinearRegression

O algoritmo de regressão linear (*LinearRegression*) tem por princípio considerar a dependência de uma variável em relação à outra variável, geralmente denominadas X e Y , sendo Y dependente de X (PAULA, 2004). A equação 1 mostra a expressão da regressão linear que é dada por:

$$Y = a + bX + e, \tag{1}$$

Onde: a é a interseção da reta no eixo vertical e b representa a inclinação da reta.

Tais parâmetros (a e b) devem ser estimados, a fim de minimizar a diferença entre o valor calculado e o valor de referência (erro e), e também para que a reta descrita pela equação se ajuste a nuvem de dados do diagrama X-Y(Figura 15).

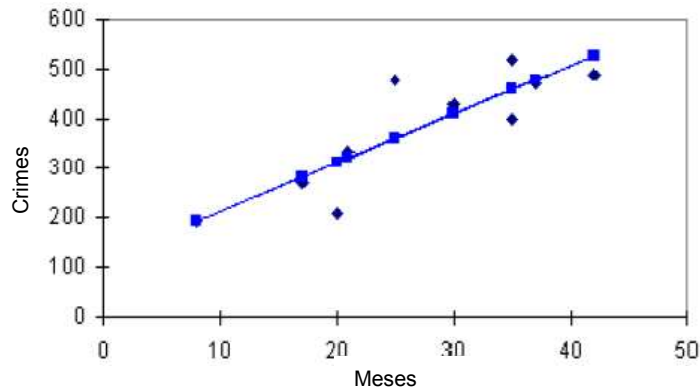


Figura 15 – Reta ajustada a nuvem de pontos correspondentes as variáveis

Como a tentativa de estimar os parâmetros a e b para minimizar cada erro seria algo inviável, uma função soma teria que ser atribuída, para adicionar a soma desses erros, a qual deve ser minimizada. Como os erros podem assumir valores negativos e positivos, sua soma tenderia a zero, desta forma, pode-se adotar a soma dos quadrados dos erros para não se trabalhar com somas que totalizem zero.

3.3.2 Algoritmo SMOReg

O algoritmo *SMOReg*, proposto por Smola e Schölkopf (1998), é uma extensão do algoritmo SMO (*Sequential Minimal Optimization*) proposto por Platt (1998) que utiliza o SVM (*Support Vector Machine*) para classificar design. O SVM é uma ferramenta para resolver problemas de reconhecimento de padrão e regressão linear.

A ideia básica do *SMOReg* é a seguinte: suponha que $\{(x_1, y_1), \dots, (x_l, y_l)\}$, $X \subset \mathbb{R}$, onde X denota o espaço de entrada dos padrões. A função linear $f(x)$ é descrita a seguir pela equação 2.

$$f(x) = \langle w, x \rangle + b \text{ com } w \in X \text{ e } b \in \mathbb{R}. \quad (2),$$

onde $\langle \cdot, \cdot \rangle$ denota o produto do ponto de suavização X . Se o ponto for igual a 1, significa que se procura um pequeno valor para w . Uma das formas de minimizar a norma seria $\|w^2\| = \langle w, w \rangle$ que pode ser escrito como um problema de otimização convexa:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 \\ \text{suposição} \quad & \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases} \end{aligned} \quad (3),$$

A suposição tácita descrita na equação (3) mostra que a função f se aproxima de todos os pares (x_i, y_i) existentes com uma precisão ε . Sendo assim, o problema de otimização convexa é viável. Porém, pode permitir alguns erros. Fazendo uma analogia à função de perda “*Soft Margin*”, proposta por Bennett e Mangasarian (1992); Cortes e Vapnik (1995) que fizeram uma adaptação para SVM, com intuito de introduzir variáveis de folga. Dessa forma pode-se chegar a equação 4 proposta por Vapnik (2000):

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{suposição} \quad & \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (4),$$

onde C é a constante que determina o limite entre o nivelamento e os maiores desvios tolerados ε . Isto corresponde a utilizar a função de perda ε -insensível $|\xi|_\varepsilon$ que é descrita pela equação 5:

$$|\xi|_\varepsilon := \begin{cases} 0 & \text{se } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{caso contrário} \end{cases} \quad (5).$$

Para melhor exemplificar as equações descritas, a Figura 18 descreve a situação graficamente.

De acordo com a Figura 16, apenas os pontos fora da região sombreada contribuem para o custo, uma vez que os desvios sejam penalizados de forma linear.

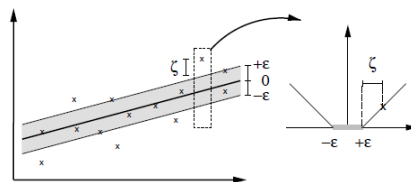


Figura 16 – Margem de perda do SVM.
Fonte: Smola e Scholköpf (1998)

3.3.3 Algoritmo Multilayer Perceptron

Apresentado por Rosenblatt (1958), o algoritmo *Perceptron* é um algoritmo de visão computacional, e tem apresentado resultados significativos de classificação supervisionada. O algoritmo *Multilayer Perceptron* é uma extensão do *Perceptron* que pode ser executado em diversas camadas, daí o nome *Multilayer Perceptron* apresentado por Murtagh (1991). Para entender o funcionamento do *Multilayer Perceptron*, primeiramente parte-se do estudo do aprendizado do *Perceptron*.

O aprendizado do *Perceptron*, de acordo com Murtagh (1991), é descrito da seguinte forma: seja x um vetor de valores binários, w um vetor de pesos e o uma saída escalar, o algoritmo *Perceptron* realiza os cálculos de acordo com a equação (6).

$$o = \sum_j w_j x_j \quad (6)$$

Alguns fatores devem ser levados em consideração:

- considere θ um limiar. Se $o \geq \theta$ e o esperado para a entrada de dados seria $o < \theta$, então a categorização ficaria incorreta, tendo que ajustar o limiar e os pesos. Dessa forma, atribui-se $\theta \leftarrow \theta + 1$ de forma a amenizar a categorização incorreta.
- se $x_j = 0$, w_j permanece o mesmo, porém se $x_j = 1$, então $w_j \leftarrow w_j - 1$ com intuito de diminuir a influência do peso.
- se a saída de dados foi configurada para ser menor que o limiar, quando deveria ter sido superior na entrada de dados, deve-se realizar uma atualização nos pesos e limiares de acordo com as equações 7, 8 e 9:

$$o = \sum_j w_j x_j \quad (7)$$

$$\Delta\theta = -(t_p - o_p) = -\delta_p \quad (8)$$

$$\Delta w_i = (t_p - o_p)x_{pi} = \delta_p x_{pi} \quad (9)$$

A equação 8 atualiza os limiares e a equação 9 atualiza os pesos. Após as atualizações (Equações 8 e 9) e, existindo pesos, o algoritmo *Perceptron* irá encontrá-los.

Para simular e encontrar o peso que cada entrada deve possuir, o algoritmo *Perceptron* utiliza como exemplo, operadores lógicos, tais como a porta lógica AND. Mas, um exemplo que contradiz isto é o problema porta exclusiva-OR (XOR), como mostra a Tabela 10 a seguir.

Tabela 10 - Relacionamento de dados XOR

Vetor de entrada de dados		Saída de Dados
0	0	0
0	1	1
1	0	1
1	1	0

Nota-se que não existe maneira de escolher os valores de w_1 e w_2 para poder discriminar o primeiro vetor e o quarto vetor (de um lado), e o segundo vetor e o terceiro vetor (do outro lado).

Dessa forma, Murtagh (1991), apresenta o Perceptron para várias camadas, denominado *Multilayer Perceptron*, que tem como característica principal todas as conexões entre os neurônios de uma camada com todos os neurônios na camada posterior. A Figura 17 mostra estas ligações.

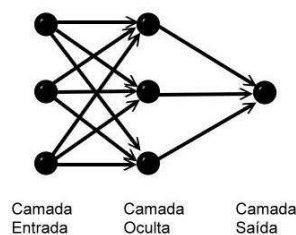


Figura 17 - Conexões entre neurônios
Fonte: Adaptado de Murtagh (1991).

De acordo com Murtagh (1991), o algoritmo *Multilayer Perceptron* é definido por:

- sua configuração ou arquitetura - número de camadas, o número de neurônios por camada, etc,
- métodos de ativação das funções de transferência usados em vários neurônios,
- especificação do método de aprendizagem-gradiente descendente, utilizando a regra delta generalizada, gradiente conjugado, ou outro método.

De acordo com Caulkins *et. al*, (1996), a previsão do comportamento individual humano não é uma tarefa fácil. Na medida em que medidas

existentes para prever reincidência criminal permanecem limitadas, modelos de redes neurais tem se mostrado um grande aliado na previsão de reincidências criminais, devido sua capacidade para detectar e estimar relações que envolvem não-linearidades, interações complexas, e descontinuidades.

3.4 Séries Temporais

A mineração de dados pode ser definida como o processo de explorar grandes quantidades de dados à procura de padrões consistentes, como regras de associação ou sequências temporais (KEOGH, KASETTY, 2003).

Uma sequência temporal (ou série temporal) é uma sequência de dados, tipicamente medidos por determinados dados, espaçados em intervalos de tempo uniformes (SHUMWAY, 1988). Séries temporais são frequentemente representadas por Gráficos de linha, e suas aplicações são inúmeras, estando presentes em diversas áreas, como na estatística, processamento de sinais, reconhecimento de padrões, economia, matemática financeira, previsão do tempo, previsão de terremotos, controle de engenharia e engenharia de comunicações.

O processo de predição a partir de séries temporais se dá com o uso de técnicas para prever valores futuros com base em valores anteriormente observados.

Para prever o comportamento de algum tipo de variável através de uma série temporal, pode-se utilizar técnicas como *Multilayer Perceptron*, *linearRegression* e *SMOReg*.

3.5 Estatística Espacial

A estatística espacial é um ramo da estatística que estuda os fenômenos considerando a dependência espacial das variáveis, possuindo aplicações em diversas áreas, tais como, epidemiologia, estudos de criminalidade, agronomia, demografia, geologia, dentre outras.

Dessa forma, de acordo com Câmara *et al.* (2004), métodos de análise de dados espaciais podem estar associados a áreas delimitadas por polígonos. Um exemplo disto são os dados coletados dos setores censitários

do IBGE, como, por exemplo, a renda média familiar dos setores censitários de um determinado município.

Segundo Câmara *et.al.* (2004) a autocorrelação espacial visa identificar a estrutura de correlação espacial que melhor descreva os dados.

A ideia principal seria estimar a magnitude da autocorrelação espacial e para isto, o índice global de *Moran*, o índice de Geary e o variograma são ferramentas utilizadas com a finalidade de gerar tais estatísticas. Para mostrar estas estruturas e evidenciar seus dados espaciais, devem-se utilizar os indicadores locais de autocorrelação espacial e o mapa de espalhamento de *Moran*. Todas estas estatísticas dependem da definição de vizinhança adotada, que é chamada de matriz de proximidade ou matriz de vizinhança.

A matriz de vizinhança estima a variabilidade espacial de dados de uma determinada área. Seja n um conjunto de áreas $\{A_1, \dots, A_n\}$ e $W^{(1)}(n \times n)$ a matriz a ser construída, onde cada um dos elementos w_{ij} representa uma medida de proximidade entre A_i e A_j . Tal medida, segundo Câmara *et al.*(2004) pode ser calculada de acordo com os seguintes critérios: $w_{ij} = 1$, se o centroide de A_i está a uma determinada distância de A_j , caso contrário $w_{ij} = 0$; $w_{ij} = 1$, se A_i compartilha um lado comum com A_j , caso contrário $w_{ij} = 0$; $w_{ij} = l_i/l_j$, onde l_{ij} é o comprimento da fronteira entre A_i e A_j e l_i é o perímetro de A_i .

A partir do cálculo da matriz de proximidade, a próxima etapa na estatística espacial é determinar o índice de *Moran* que tem por objetivo a caracterização da dependência espacial, mostrando como os valores estão correlacionados no espaço. O índice de *Moran* presta-se a um teste cuja independência espacial é a hipótese nula, ou seja, seu valor igual a zero. Valores positivos (entre 0 e +1) indicam uma correlação direta e os negativos (entre 0 e -1) correlação inversa. Dessa forma, pode-se dizer que o índice de *Moran* tem a função de mostrar a correlação espacial, quando existente, entre polígonos em um determinado mapa.

Após o cálculo do índice de *Moran*, é importante estabelecer sua validade estatística. Dessa forma, para estimar a significância do índice, a abordagem mais comum é um teste de pseudo-significância. A equação 10

mostra o cálculo do índice de *Moran* considerando apenas os primeiros vizinhos.

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2} \quad (10)$$

onde N é o número de áreas, z_i o valor do atributo considerado na área i , \bar{z} é o valor médio do atributo na região de estudo e w_{ij} os elementos da matriz normalizada de proximidade espacial.

Vale ressaltar que indicadores globais de autocorrelação espacial, tais como o Índice de *Moran*, fornecem apenas um valor como medida da associação espacial para todo o conjunto de dados de área. Porém, pode haver a necessidade de examinar padrões numa escala de maior detalhe, com a finalidade de verificar se a hipótese de estacionariedade do processo verifica-se localmente.

Dessa forma, é necessário utilizar indicadores que possam ser associados às diferentes localizações de uma variável distribuída espacialmente. Assim, o Índice de *Moran* Local é utilizado com a finalidade de encontrar a correlação espacial dessas áreas. Por se tratar de um indicador local, tem-se um valor específico de correlação para cada área, permitindo assim a identificação de *clusters* de áreas e *outliers*. (CÂMARA, *et al.*, 2004). O índice de *Moran* local pode ser expresso para cada área i a partir dos valores normalizados z_i do atributo de acordo com a equação 11.

$$I = \frac{(n/S_0) \sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \quad (11)$$

Onde, se acordo com Anselin (1995), $S_0 = \sum_i \sum_j w_{ij}$.

A Figura 18 ilustra um exemplo de relatório com resultados obtidos a partir do índice de *Moran* local determinado por ferramentas de estatística espacial do software *Terraview* (INPE, 2013).

	RENDAZ	RENDAWz	RENDAMoranIndex	RENDABoxMap	RENDA LISASig	RENDA LISAMap	RENDA MoranMap	dist.COD_UP	ID_DIST
1	-404.497856452	-147.876731452	0.050019611	2	0.47	0	0	1	5
2	-444.220056452	-304.163256452	0.112987176	2	0.4	0	0	10	3
3	-698.027356452	-198.493356452	0.115862437	2	0.34	0	0	11	7
4	-486.229656452	-37.107656452	0.015087914	2	0.54	0	0	12	4
5	-871.521256452	1284.406593548	-0.936062296	4	0.02	1	4	13	2
6	132.885143548	-68.982976452	-0.007665537	3	0.36	0	0	14	5

Figura 18 – Relatório de estatística espacial no *Terraview*.

Na Figura 20, *RENDAz* é definido pelo vetor dos desvios dos valores observados; a coluna que possui *RENDAWz* corresponde ao vetor da média ponderada local; a coluna *RENDAMoranIndex* é Índice de *Moran* Local; e por fim a coluna *RENDA LISASig* é o valor da estatística *p*.

A coluna *RENDABoxMap* apresenta valores que correspondem a relação entre os valores de *RENDAZ* e *RENDAWz* em um Gráfico de dispersão dividido em quadrantes (Q). Os valores variam de 1 a 4, onde 1 corresponde aos valores de Q_1 (alto-alto – valores altos de *RENDAZ* e valores altos de *RENDAWz*), 2 (Q_2) baixo-baixo, 3 (Q_3) alto-baixo e 4 (Q_4) baixo-alto.

Outra maneira de avaliar as variações das tendências espaciais da variável, em termos globais, é estimar a média móvel. De acordo com Câmara (2004), este método estatístico pode se descrito através da equação 12:

$$\mu_i = \sum_{j=1}^n w_{i,j} Y_j \quad (12)$$

onde $w_{i,j}$ são os elementos da matriz de proximidade e Y_j é a variável de cada área.

3.6 Materiais e Métodos

3.6.1 Materiais Utilizados

Para o desenvolvimento da pesquisa foram utilizados os seguintes materiais:

- *ArcGis* versão 10.1; (ESRI, 2011)
- *Terraview*. versão 4.2.2;
- Software estatístico *R* versão 3.0.1;
- Dados espaciais da base digital do IBGE;

- Dados de boletins de ocorrência fornecidos pelo quartel de Polícia Militar de Rio Pomba.

— *Área de estudo*

A área de estudo compreende o município de Rio Pomba – MG, considerando apenas a zona urbana. Rio Pomba está situada na zona da mata mineira e vem passando por um processo de crescimento populacional em torno de 20% nos últimos 10 anos, de acordo com os censos de 2000 e 2010 (IBGE, 2010). Tal fato pode ser atribuído ao crescimento de indústrias, comércio e o Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais – Campus Rio Pomba. A Figura 19 ilustra a localização do município de Rio Pomba – MG.



Figura 19 - Área de estudos: cidade de Rio Pomba, situada na Zona da Mata de Minas Gerais.

Fonte: Adaptado de IBGE (2011).

3.6.2 *Métodos*

A fim de facilitar o entendimento quanto as atividades desenvolvidas nesta pesquisa, um fluxograma destas atividades pode ser visualizado na Figura 20.

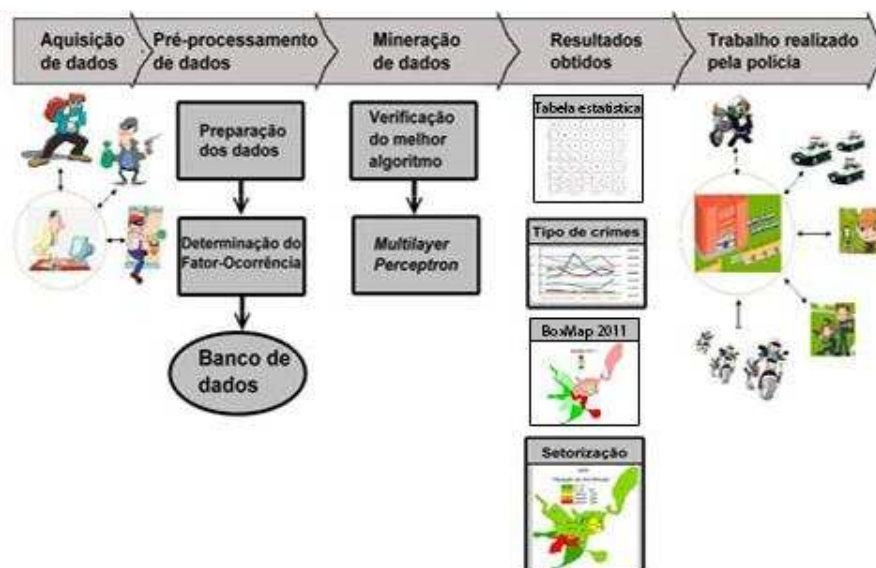


Figura 20 - Fluxograma de atividades desenvolvidas
 Fonte: Adaptado de Li, et. al (2010)

— **A aquisição de dados**

Os dados utilizados neste trabalho, fornecidos pela Polícia Militar da cidade de Rio Pomba, apresentam um cenário das ocorrências criminais ocorridas durante os anos de 2009, 2010 e 2011. Tais dados foram divididos em duas categorias: por tipo de crime e por setorização dos crimes na cidade.

A categoria por tipo de crime subdivide-se em: A – Ações de defesa social, B – Infrações contra pessoas, C – Infrações contra o patrimônio e a propriedade imaterial, D – Infrações contra a dignidade sexual e a família, E – Infrações contra a incolumidade pública e a paz pública, T - Infrações referentes ao trânsito, Y – Operações de defesa social. Portanto, um conjunto de 7X36 dados temporais foi estabelecido (7 grupos de ocorrências por 36 meses).

A setorização dos crimes na cidade subdivide-se em setores que estão numerados de 1 a 18, assim, um conjunto de 18X12 dados espaço-temporais foi formado (18 setores do município por 12 trimestres).

Também foi utilizada nesta pesquisa a base cartográfica digital na escala 1:50000 referente ao município de Rio Pomba disponibilizada pela agência do Instituto Brasileiro de Geografia e Estatística (IBGE) em formato *shapefile*. Como essa pesquisa se limita à região urbana do município, foi necessário realizar operações de edição para obter apenas dados

relacionados à cidade e não ao município. Assim, a cidade de Rio Pomba ficou subdividida em 18 setores, enquanto que o município totaliza 33 setores.

Verificou-se ainda que a base digital estivesse no sistema de projeção e referência UTM/SIRGAS2000. A Figura 21 mostra a carta digital com os 18 setores estudados.

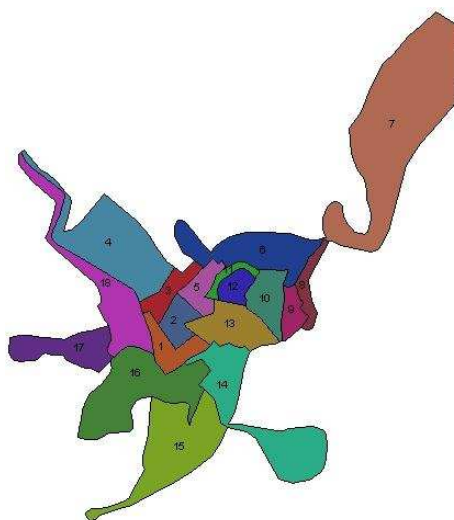


Figura 21 – Plano de informação com apenas setores censitários urbanos

— **Pré-processamento dos dados**

Os dados provenientes do quartel da polícia militar encontrava-se em formato de planilhas, sem qualquer tipo de tratamento, como mostra a Figura 22.

	G	H	I	J	K	L	M	N	O	P	R
118	09.00	0	0	0	SEX	2	Y	Y30	9	BOS	AV DR JOSE NEVES
119	05.55	0	0	0	DOM	1	C	C02	9	1034467	AV DR JOSE NEVES
120	20.30	0	0	0	DOM	4	H	H08	9	1072904	AV DR JOSE NEVES
121	20.20	0	0	0	SAB	4	H	H19	10	1175036	AV DR JOSE NEVES
122	23.30	0	0	0	SAB	4	H	H19	10	1175727	AV DR JOSE NEVES
123	23.50	0	0	0	SAB	4	T	T0	10	XXXXXXXXXX	AV DR JOSE NEVES
124	16.00	0	0	0	TER	3	Y	Y21	10	XXXXXXXXXX	AV DR JOSE NEVES
125	17.10	0	0	0	TER	3	H	H19	10	1188427	AV DR JOSE NEVES
126	17.20	0	0	0	TER	3	H	H19	10	XXXXXXXXXX	AV DR JOSE NEVES
127	20.00	0	0	0	SAB	4	T	T09	11		
128	20.42	0	0	0	QUI	4	T	T01	11		
129	15.00	0	0	0	QUI	3	Y	Y21	11		
130	08.00	0	0	0	SAB	3	T	T02	11		
131	15.00	0	0	0	QUI	3	Y	Y21	12		
132	15.30	0	0	0	QUI	3	C	C16	12		
133	19.18	0	0	0	SEX	4	B	B03			
134	18.30	0	0	0	SEG	4	A	A19	3		
135	11.35	0	0	0	QUI	2	Y	Y30	11		
136	22.50	0	0	0	TER	4	T	T02	11		
137	01.14	0	0	0	SAB	1	H	H19	11		
138	03.13	0	0	0	SAB	1	H	H19	11		
139	13.55	0	0	0	QUA	3	Y	Y14	2	XXXXXXXXXX	AV DR JOSE NEVES (BANCO BRADESCO)
140	07.00	0	0	0	TER	4	A	A19	5	XXXXXXXXXX	AV DR JOSE NEVES (ESCOLA SÃO JOSE)
141	12.30	0	0	0	TER	3	A	A19	5	XXXXXXXXXX	AV DR JOSE NEVES (ESCOLA SÃO JOSE)

Figura 22 - Erros encontrados nas tabelas da Polícia Militar.

Geralmente, a base de dados é suscetível a algumas inconsistências, que podem ser atribuídas a diferentes operadores para o cadastro da ocorrência e também pela falta de conhecimento técnico com relação a banco de dados, por parte dos operadores. A baixa qualidade desses dados

pode produzir padrões pouco convincentes. Assim, o tratamento dos dados é um fator crucial para a geração de informações confiáveis.

Dessa forma, os registros foram editados a fim de evitar inconsistências. Os problemas encontrados foram: um mesmo local apresentava denominações diferentes, tais como, “AV DR JOSÉ NEVES” e “AV DR JOSE NEVES”. A etapa de correção se procedeu de forma manual onde, foi analisado registro por registro totalizando 9009 registros. Todos os demais campos também foram verificados e corrigidos de forma a realizar uma predição de crimes utilizando algoritmos de mineração de dados. Esta é uma fase de pré-processamento, que de acordo com Agrawal *et al.* (1993) possui as seguintes etapas:

1. Seleção de dados: determina quais os dados que realmente devem ser considerados durante o processo;
2. Limpeza dos dados: etapa que a garantir e certificar a veracidade e completude dos dados;
3. Codificação dos dados: adequar os dados em formatos que possam ser compreendidos por softwares e algoritmos que serão utilizados no processo;
4. Enriquecimento dos dados: agrega-se informações aos registros existentes, com a finalidade de reconhecer novos padrões.

A análise dos dados é uma etapa fundamental para o sucesso no processo da mineração dos dados. Durante tal etapa, pode-se observar que muitos dados não possuem relevância para a informação que se deseja obter.

Outro fator importante para mineração de dados é o formato em que eles se encontram, para que um software qualquer, ao utilizar um algoritmo específico, possa realizar corretamente a leitura dos dados (Bigus,1996), extraindo estatísticas relevantes relacionadas à pesquisa.

— **Determinação do “fator-ocorrência”**

Para realizar a predição de ocorrências futuras considerando os dados dos anos de 2009, 2010 e 2011, foi definida a variável “fator-ocorrência”, com auxílio da polícia militar do município de Rio Pomba, a partir de várias entrevistas, a fim de armazenar o somatório de crimes, de

acordo com as categorias mencionadas. O fator-ocorrência é definido de acordo com a prioridade de cada natureza de crime. Por exemplo, a ocorrência de homicídio tem prioridade 1 para polícia, enquanto que a ocorrência de um carro estacionado em frente a uma garagem tem prioridade 4. As prioridades variam de 1 a 4 sendo a prioridade 1 a mais máxima e a prioridade 4 a prioridade mínima.

A variável “Fator-Ocorrência (FC)” foi normalizada de acordo com as prioridades, variando de 0,25 a 1 (Equação 13).

$$FC = \sum_{i=0}^{np1} p1_i + \sum_{i=0}^{np2} (0,75 * p2_i) + \sum_{i=0}^{np3} (0,5 * p3_i) + \sum_{i=0}^{np4} (0,25 * p4_i) \quad (13)$$

onde “ $p1_i$ ” é o número de ocorrências com prioridade igual a 1 para cada numero de ocorrência (i), o “ $p2_i$ ” é o número de ocorrências com prioridade igual a 2 para cada numero de ocorrência (i), “ $p3_i$ ” o número de ocorrências com prioridade igual a 3 para cada numero de ocorrência (i), “ $p4_i$ ” número de ocorrências com prioridade igual a 4 para cada numero de ocorrência (i) e “ $np1$ ”, “ $np2$ ”, “ $np3$ ” e “ $np4$ ” são os números máximos de ocorrências de acordo com as prioridades 1,2,3,4, respectivamente.

Dessa forma, foram criadas duas tabelas dinâmicas (com campos calculados de valores) de atributos com o tipo de crime e de setores, que com a aplicação da equação 13, faz o cálculo do somatório de ocorrência por setor e pela tipologia.

3.6.3 Mineração dos dados

Para realizar a mineração de dados foi utilizado o *framework Weka* versão 3.7.8 (Hall, *et al.*, 2009) que possui algoritmos que podem ser aplicados com intuito de gerar informações a partir da manipulação de dados. Como a pesquisa visa a predição de ocorrências, foi utilizada a ferramenta do *framework Weka* chamada “*forecast*”, que por meio de algoritmos classificadores traçam um histórico baseado em anos, meses ou dias e realizam a predição de ocorrências de acordo com os parâmetros determinados.

3.6.4 A escolha do algoritmo

Dentre os três algoritmos classificadores citados anteriormente os quais foram testados e analisados, verificou-se que o algoritmo classificador *Multilayer Perceptron* apresentou o menor erro médio e o menor erro médio quadrático para a categoria de setorização. Enquanto para a categoria de tipologia de crimes, o algoritmo *SMOReg* apresentou melhores resultados.

O resultado do erro médio e erro médio quadrático das duas categorias pode ser visualizado no Gráfico 1.

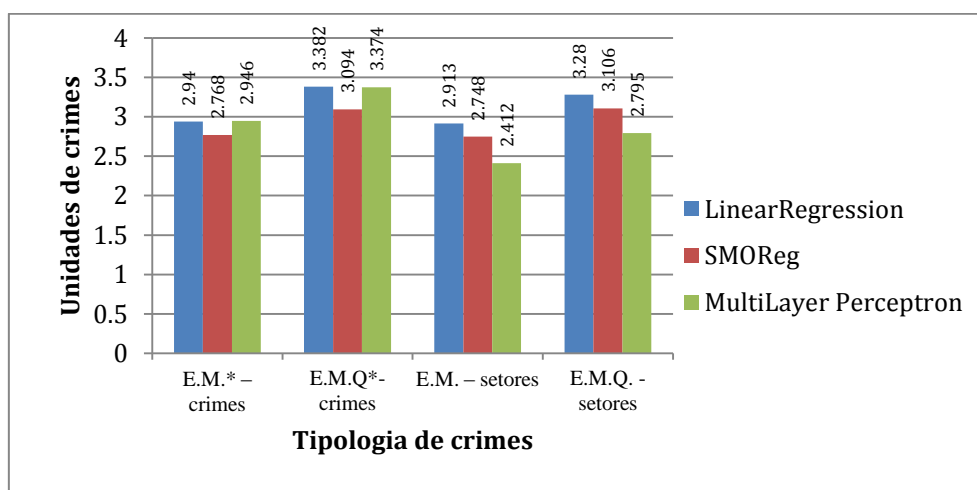


Gráfico 1 - Gráfico dos "Erros Médios" e " Erros Quadráticos Médios"
E.M.* – erro médio; E.M.Q* - erro quadrático médio

Pode-se perceber que os dois algoritmos (*SMOReg* e *Multilayer Perceptron*) apresentaram erros muito parecidos na predição de ocorrências policiais. A escolha do *Multilayer Perceptron* se deu devido ao fator de não apresentar "outliers" na predição. Enquanto o *SMOReg* apresentou alguns resultados com este tipo de erro, que para o estudo não deve ser desconsiderado.

3.6.5 Estatística Espacial

Para aplicar a estatística espacial sobre os dados de criminalidade, foi necessário criar uma variável de Índice Relativo de Ocorrências Policiais (IROP) para os anos de 2009, 2010 e 2011, com a finalidade de diminuir o viés dos dados de ocorrências policiais. O IROP é dado pela equação 14.

$$IROP = \frac{FC}{População Residente} \times 100 \quad (14)$$

onde, FC é fator ocorrência, a variável população residente é a população residente em cada setor censitário da cidade de Rio Pomba.

Após identificar o $IROP$ de cada setor em cada ano, foi acrescentado ao arquivo shapefile que contém a tabela de setores do município de Rio Pomba, os atributos $IROP2009$, $IROP2010$ e $IROP2011$.

Os procedimentos descritos acima fazem parte da preparação dos dados para realizar a estatística espacial. Dessa forma, o próximo passo foi determinar a matriz de proximidade e posteriormente, o Índice de *Moran*, que é um índice que retrata uma estatística global, ou seja, considerando todos os setores da cidade.

Os resultados obtidos visando a predição dos dados foram determinados por meio do *framework Weka-3-7-8* com a aba "*forecast*" e o algoritmo *Multilayer Perceptron*. Utilizaram-se dois tipos de dados de acordo com as categorias citadas. A predição dos dados para a categoria tipo de crime foi de seis meses e para a categoria setor foi de um ano variando a cada trimestre. Os dados da categoria tipo de crime eram um somatório das ocorrências de cada mês dos anos de 2009, 2010 e 2011. Dessa forma, tem-se um total de 36 meses para realizar a predição do semestre de 2012. Já os dados para categoria setores eram um somatório de ocorrências a cada trimestre dos anos de 2009, 2010 e 2011. Dessa forma, tem-se um total de 12 trimestres para predição dos setores. A predição da categoria setores foi realizada pelos trimestres dos anos e não pelos meses, devido à alta capacidade de processamento de memória que o algoritmo ocupa, não sendo possível realizar tal teste em um computador desktop.

3.7 Resultados e Discussão

Para a categoria tipo de crime, obteve-se os seguintes resultados mostrados no Gráfico 2, que para fins de melhor visualização apresenta o eixo x por trimestres.

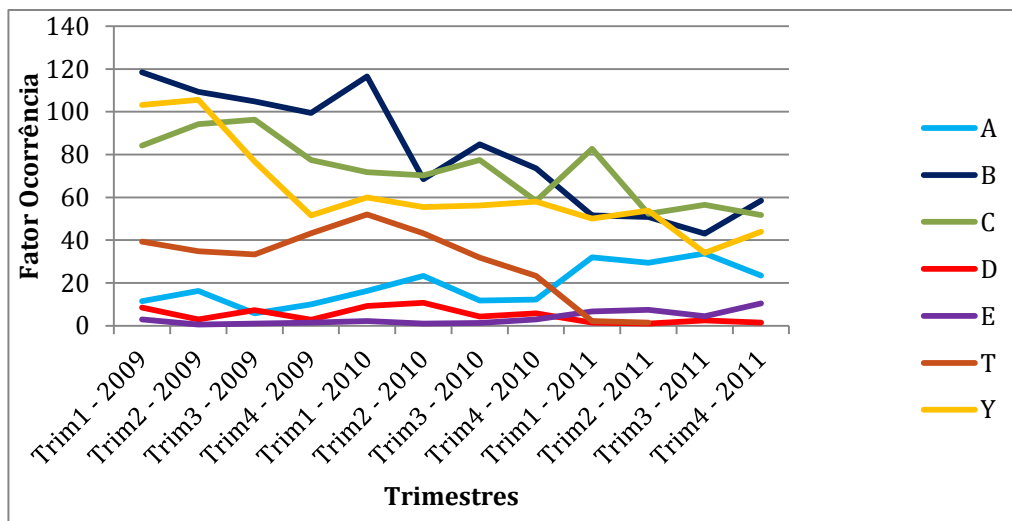


Gráfico 2 - Gráfico dos tipos de crimes de acordo com os trimestres de cada ano.

A partir de uma análise dos dados ilustrados no Gráfico 2, pode-se perceber, por exemplo, um aumento significativo no somatório dos fatores-ocorrências do tipo A, que seriam ações sociais da polícia militar. Além disso, perceber-se ainda que alguns tipos de ocorrências seguem o mesmo padrão no decorrer dos trimestres, como, por exemplo, o tipo E.

Com o aumento da ação da polícia militar, diminuiu-se consideravelmente as ocorrências da tipologia B (crimes contra pessoa), da tipologia C (crimes contra o patrimônio), da tipologia Y (operações de defesa social) e da tipologia T (ocorrências de trânsito).

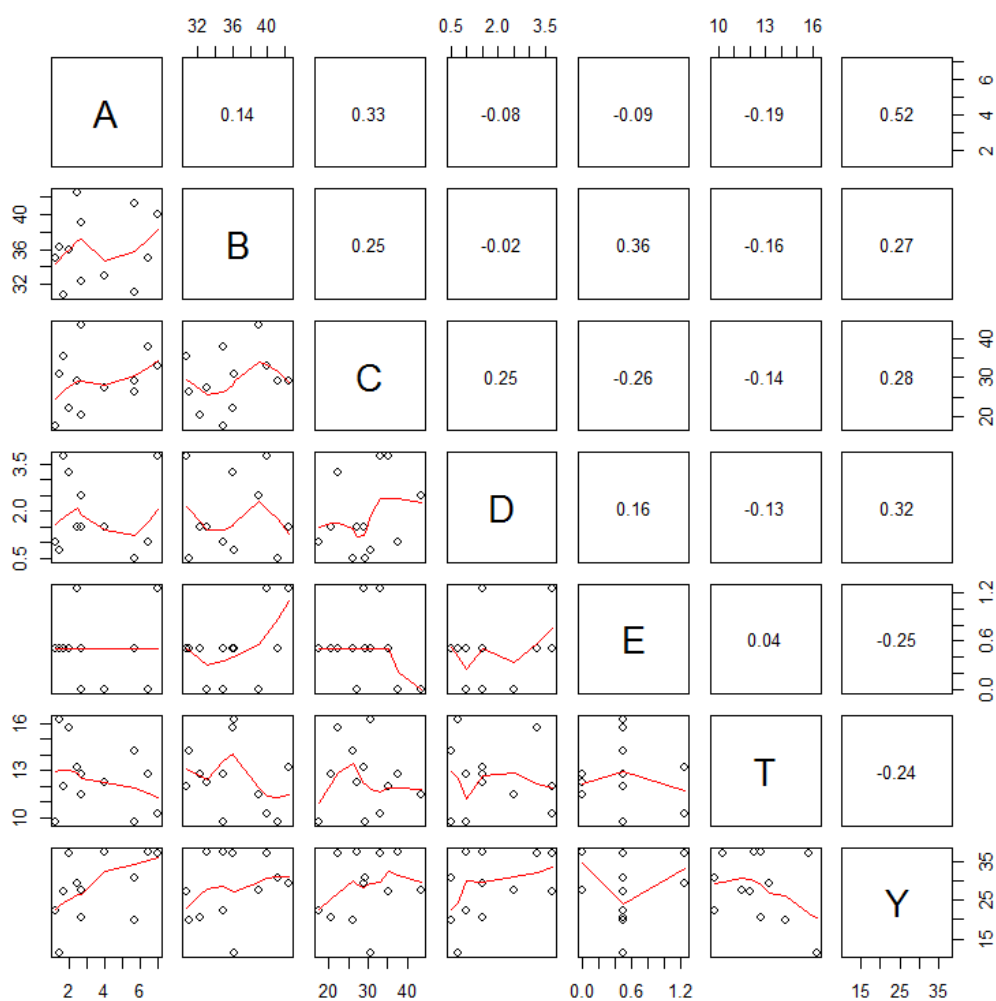
Pode-se perceber que no final do ano de 2009, ou seja, último trimestre de 2009, o valor das ocorrências do tipo A foi de aproximadamente 10. No ano de 2010, o FC para este mesmo tipo de ocorrência aumentou aproximadamente para 15 no mesmo intervalo de tempo e em 2011 aumentou ainda mais, para aproximadamente 30. Dessa forma, percebe-se claramente um crescimento das ações da polícia chegando a quase triplicar de ocorrências com ações sociais.

Ainda analisando o Gráfico 2, com o crescimento de ocorrências do tipo A, pode-se notar que houve uma redução nas ocorrências do tipo B, C e Y. Em 2009 o FC do último trimestre para ocorrências do tipo B era de aproximadamente 99, enquanto que, para 2010, o FC reduziu para aproximadamente 85 e 2011 diminuiu ainda mais para aproximadamente 50

ocorrências. Dessa forma, considerando apenas as ocorrências contra pessoa (tipologia B) a polícia militar conseguiu reduzir mais da metade dos casos. E o valor do FC para o tipo C do 4 trimestre de 2009 foi de aproximadamente 90, enquanto para o quarto trimestre de 2010 foi de aproximadamente 70 e o último trimestre de 2011 de aproximadamente 60, reduzindo assim, cerca de 30% o valor do FC.

Afim de avaliar estatisticamente se um determinado tipo de crime possui uma correlação direta ou inversa com outro tipo de crime, a matriz de correlação entre as ocorrências foi gerada, utilizando o software R, a qual é mostrada na Tabela 11. A correlação mostrada nas Tabelas 11, 12 e 13 é a correlação dos anos de 2009, 2010 e 2011 utilizando o método *spearman*.

Tabela 11 - Matriz de correlação entre tipos de ocorrências policiais do ano de 2009

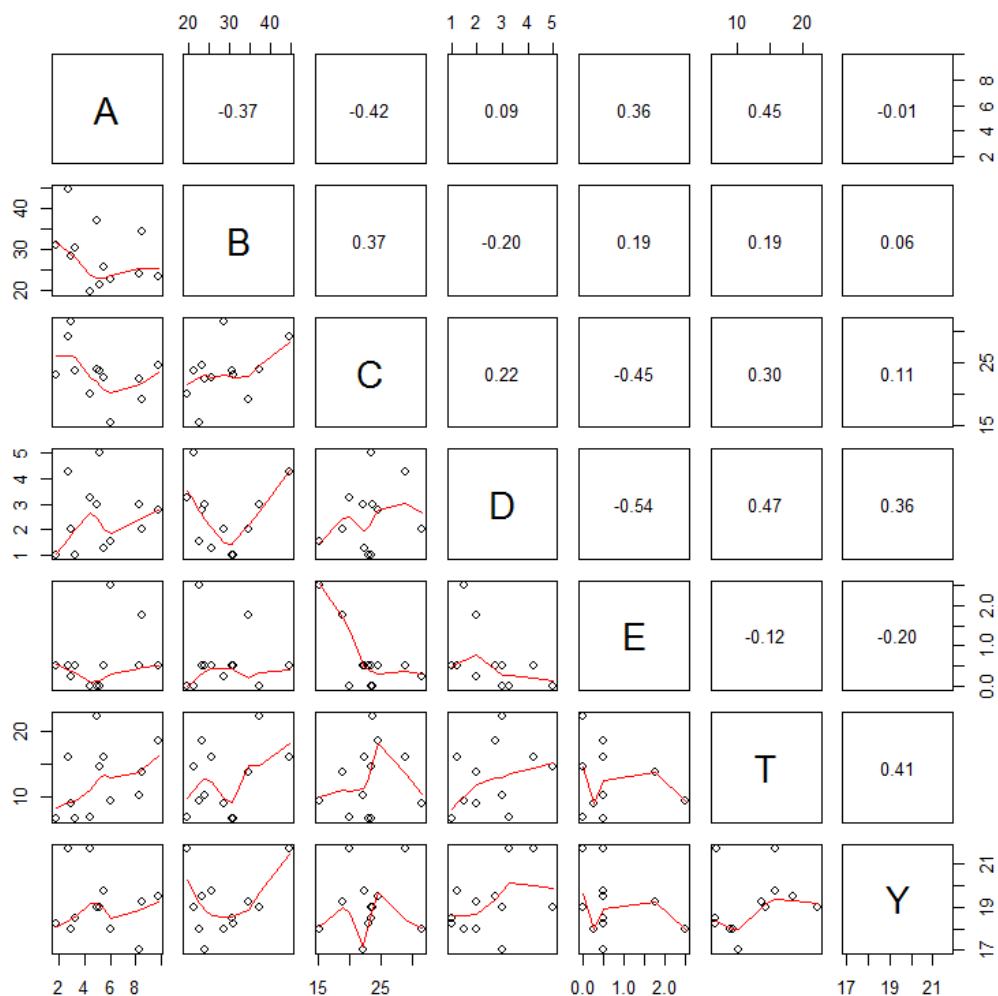


Pode-se perceber a partir da matriz de correlação (Tabela 11) que quanto maior for o número de ocorrências do tipo B (ocorrências contra

pessoas), maior será o número de ocorrências do tipo Y (operações da polícia) e também as ocorrências do tipo A (Ações de prevenção da polícia). Para as ocorrências do tipo C, as ocorrências dos tipos Y e A comportam-se de maneira semelhantes, ou seja, quanto maior for o número de ocorrências do tipo C, maior será as ocorrências de A e Y. Isto indica a influência do trabalho da polícia para combater as ocorrências criminais da cidade. Outro fato interessante é que quanto maior o numero de ocorrências do tipo Y, menor serão as ocorrências do tipo T (ocorrências relacionadas ao trânsito). Além disso, pode-se perceber outra relação muito importante para estudos da criminalidade, que seria a relação entre as ocorrências do tipo B e E (Infrações contra a incolumidade pública e a paz pública). Quanto maior as ocorrências do tipo B, maiores serão as ocorrências do tipo E.

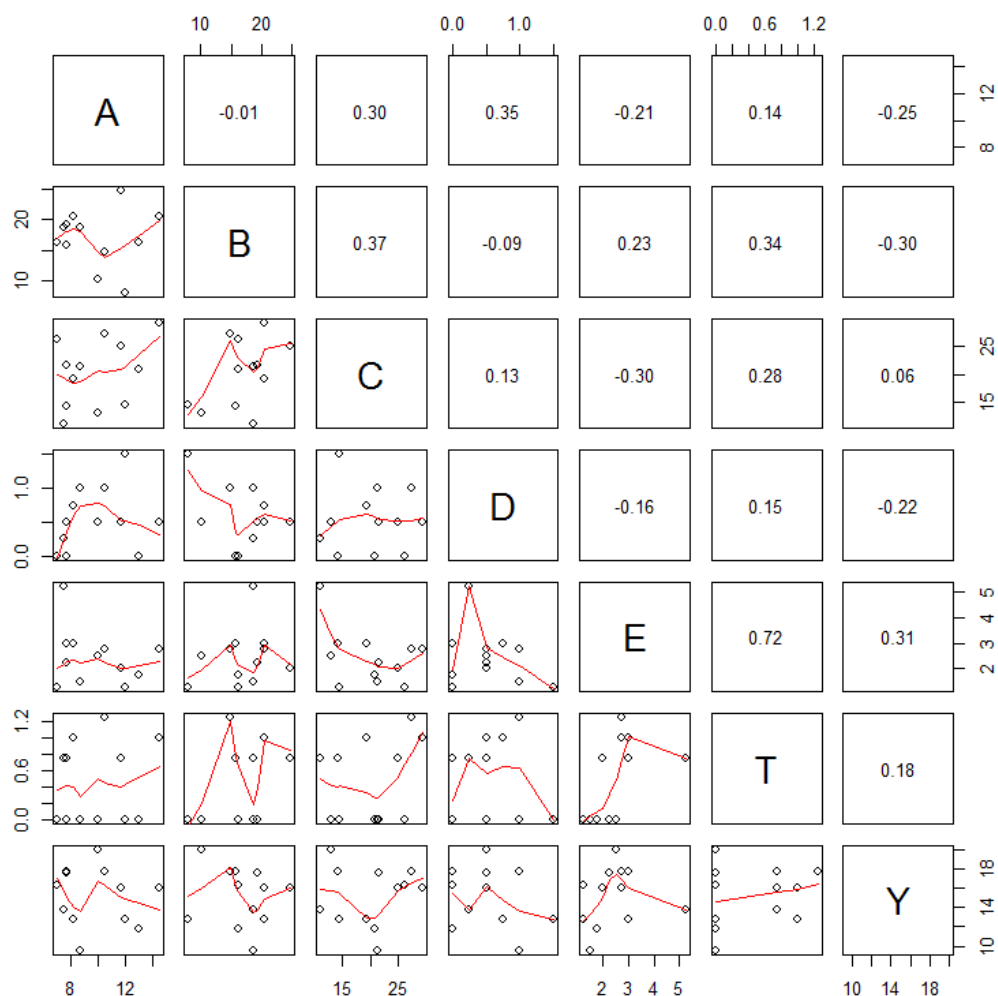
Uma relação importante é entre o tipo A e o tipo Y que possui uma correlação de 0,52, onde os dois casos são ações de combate à criminalidade. Neste caso, percebe-se um trabalho da polícia militar de modo ostensivo, ou seja, aumentando as ações sociais e operações da polícia para o ano de 2009. Isto pode ser notado devido a queda do índice de criminalidade dos anos subsequentes, como pode ser visto no gráfico 2.

Tabela 12 - Matriz de correlação entre tipos de ocorrências policiais do ano de 2010



A Tabela 12 mostra a correlação de *Spearman* dos tipos de ocorrências policiais para o ano de 2010. Pode-se notar que quando ocorre o aumento do número de ocorrências do tipo A, diminui o índice de ocorrências do tipo B e ocorrências do tipo C. Observa-se ainda que neste ano as ações da polícia militar (ocorrências do Tipo A) obtiveram mais efeito do que as operações (ocorrências do tipo Y) para ocorrências do tipo B e C. Nota-se ainda que quanto maior for as ocorrências do tipo T (relacionadas ao trânsito) maiores serão as ocorrências A e Y.

Tabela 13 - Matriz de correlação entre tipos de ocorrências policiais ano de 2011

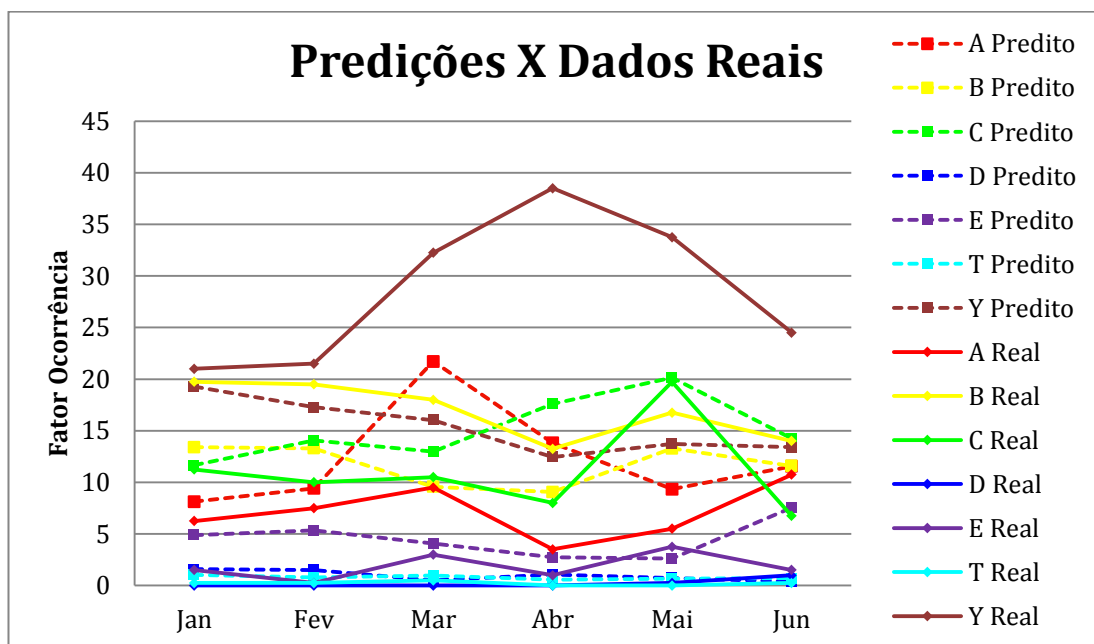


Pode-se perceber a partir da matriz de correlação (Tabela 13), quando aumentam as ocorrências do tipo Y diminuem as ocorrências do tipo B; outra análise que pode ser percebida é uma correlação negativa entre as ocorrências do tipo A e Y que foi de -0,25 indicando que com o aumento das ocorrências do tipo A, diminui-se as ocorrências do tipo Y. Para as ocorrências E (Infrações contra a incolumidade pública e a paz pública) e T percebe-se uma correlação forte positiva atingindo um valor de 0,72.

A partir das análises, considerando as matrizes de correlação, percebe-se que estas informações são de relevante importância, permitindo a polícia militar uma melhor leitura de suas ocorrências.

O próximo passo foi realizar etapa de predição utilizando o algoritmo *Multilayer Perceptron* das ocorrências apresentadas no Gráfico 2 como entrada de dados, para os 6 meses subsequentes apresentados no Gráfico 3.

Gráfico 3 – Gráfico de previsões de tipologia de ocorrências utilizando o algoritmo *Multilayer Perceptron* para os 6 meses subsequentes a 2011 confrontando com os dados reais dos 6 meses após 2011.



O Gráfico 3 relata a previsão da categoria tipo de ocorrência onde as linhas tracejadas representam as tipologias de ocorrências preditas e as linhas contínuas representam os dados reais coletados dos seis primeiros meses do ano de 2012.

Pode-se verificar a partir dos dados do Gráfico 3 que, de modo geral, ocorre o mesmo comportamento observado nos dados preditos e reais. Um exemplo seria o comportamento do tipo A, que obedeceu a mesma curva de FC com o passar dos meses.

Para a tipologia B pode-se perceber que os dados preditos também obedeceram o mesmo padrão comportamental dos dados reais, ou seja, o comportamento da série de dados preditas de tipologia B possui comportamento semelhante à série de dados reais da mesma tipologia. Vale ressaltar que, apesar de apresentarem o mesmo comportamento, essas séries apresentam proporções pouco diferentes, uma vez que os dados reais foram um pouco mais altos que os dados preditos.

Para os dados de tipologia C pode-se dizer que ambos possuem o comportamento e proporções semelhantes, com uma variação muito pequena.

Os dados de tipologia D também apresentaram comportamento e proporções semelhantes, sendo que no mês de Junho, os dados preditos foram maior que os dados reais.

As curvas dos dados de tipologia E mostraram-se pouco semelhantes com relação a comportamento e proporção, visto que alguns meses, como, por exemplo, no mês de Junho a curva de dados preditos está numa crescente, enquanto a curva de dados reais está descendo.

Para os dados de tipologia T tanto a curva de dados reais quanto a curva de dados preditos possuem comportamento e proporção semelhantes.

Para os dados de tipologia Y as curvas de dados reais e dados preditos apresentam comportamento e proporção pouco semelhantes. Pode-se perceber que no mês de abril ocorreu um salto nos valores de ocorrências do tipo Y. Acredita-se que tal fato pode ter ocorrido devido a uma operação especial da polícia militar no mês de abril do ano de 2012, visando conscientizar a população a respeito da segurança pública, além de uma fiscalização mais intensa na cidade, a fim de transparecer uma de segurança à população. De modo geral, percebe-se que as operações policiais tiveram valores de ocorrências maiores que os valores preditos, porém houve uma queda a partir do mês de abril no comportamento temporal dessas ações.

A outra análise realizada na pesquisa é a de ocorrências por setor. Dessa forma, foi utilizado o mesmo cálculo usado nas ocorrências por tipos de crimes. O fato interessante é que na primeira análise, visou-se perceber se uma ocorrência influenciava em outra e poder prever, por meio do algoritmo *Multilayer Perceptron*, os comportamento dos próximos 6 meses. Neste caso, muda-se o foco, pois além de trabalhar com os dados temporais, os dados espaciais serão também analisados.

Devido à necessidade de alto processamento de memória para executar o algoritmo, foi necessário realizar a predição para cada trimestre e não para cada mês conforme realizado anteriormente na análise por tipo de crime.

As Figuras 23, 24 e 25 apresentam a variável IROP com o comportamento de cada setor da cidade quanto à criminalidade, para os anos de 2009, 2010 e 2011, respectivamente. A variação das ocorrências de

cada setor foi classificada por passo iguais em 4 fatias e também editada por quartil.

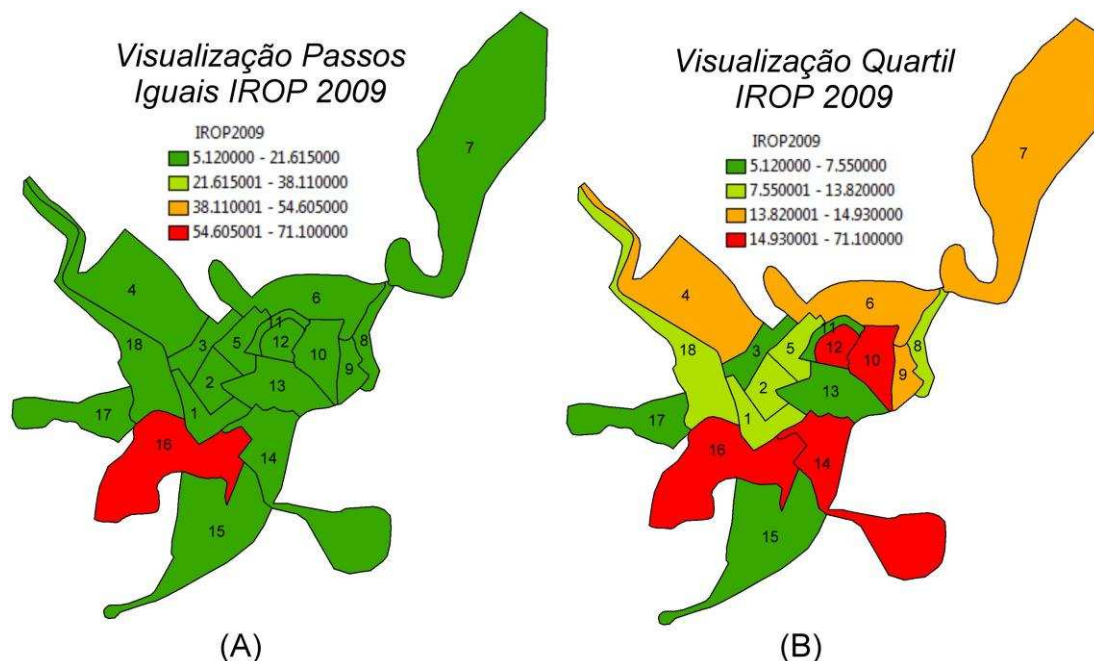


Figura 23 - Figura (A) distribuição da variável IROP do ano de 2009 pelo método estatístico de Passos Iguais e Figura (B) distribuição da variável IROP do ano de 2009 pelo método estatístico Quartil

Apresentam-se na Figura 23 duas formas de visualização dos dados, em relação a ocorrências policiais do ano de 2009. A Figura 23 (A) mostra a distribuição de ocorrências pelo método de passos iguais, ou seja, a diferença entre o valor máximo e o valor mínimo da variável é dividida pelo número de classes desejado, obtendo-se o mesmo intervalo em cada classe. Se a variável tem uma distribuição muito concentrada de um lado (assimetria), esta visualização deixa apenas um número pequeno de regiões com IROP mais elevado ressaltando a detecção de valores discrepantes. Considerando as classes de cores apresentadas na Figura 23 como índice da variável IROP baixo, moderado, moderado-alto e alto para as cores verde, verde claro, laranja e vermelho, respectivamente, pode-se notar que nesta visualização, a maioria dos setores (94,44%, 17 setores) está com baixo índices da variável IROP, e o setor de número 16 com índice da variável IROP alto. Já a Figura 23 (B) representa a distribuição da variável IROP pelo método quartil, onde o intervalo da quantidade de ocorrências de cada classe é calculado através do ordenamento dos valores da variável (crescente ou decrescente). Assim, cada classe conterá mesmo número de

valores da variável. É uma forma de separar as classes em percentual de valores da variável. Dessa forma, 27,77% (setores 3, 11, 13, 16 e 17) apresentam índices da variável IROP baixo. Os setores que apresentaram índices da variável IROP moderado-alto foram os de número 4, 6, 7 e 9. E 22,22% dos setores, ou seja, setores 1, 2, 5, 8 e 18 apresentaram índices da variável IROP moderado. Da mesma forma, pode-se perceber que para o índice da variável IROP alto foi de 22,22%, contemplando os setores 10, 12, 14 e 16.

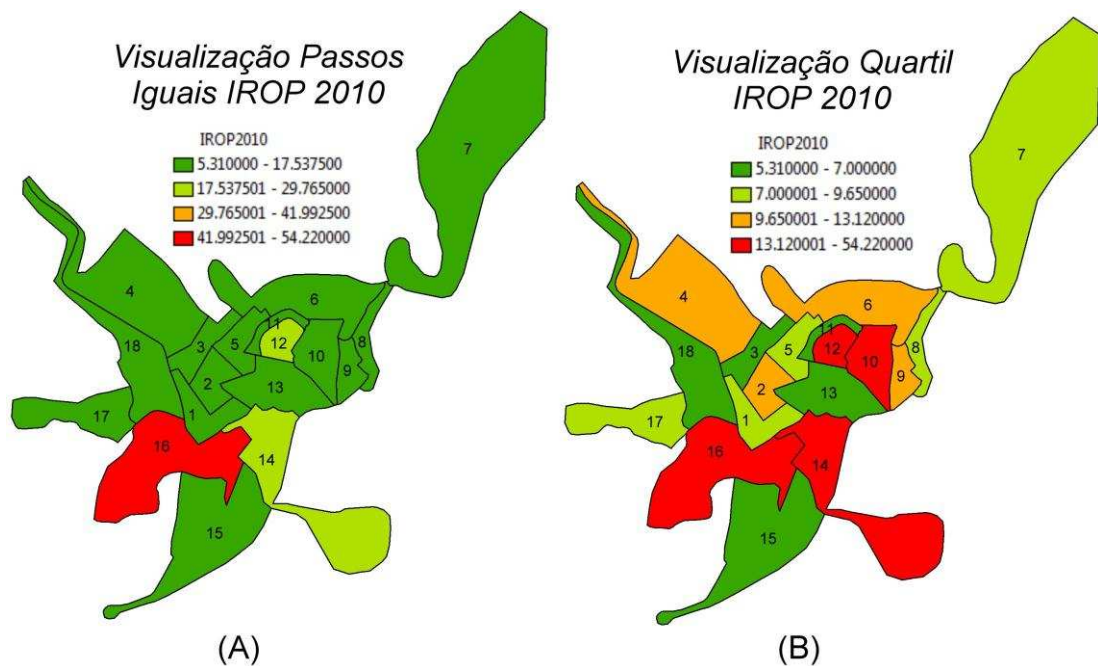


Figura 24 - Figura (A) distribuição da variável IROP do ano de 2010 pelo método estatístico de Passos Iguais e Figura (B) distribuição da variável IROP do ano de 2009 pelo método estatístico Quartil

A Figura 24 representa a variável IROP do ano de 2010. Pode-se notar que nesta visualização, a maioria dos setores (83,33%, 15 setores) está com baixo índices da variável IROP, 2 setores (14 e 12) possuem índices da variável IROP moderado, e o setor 16 com índice variável IROP alto. A Figura 24 (B) representa a distribuição dos índices da variável IROP pelo método quartil. Dessa forma, 27,77% (setores 3, 11, 15, 13 e 18) apresentam índices da variável IROP baixo. O mesmo número de setores apresentam índices da variável IROP moderado (setores 1, 5, 7, 8 e 17). E 22,22% (setores 2, 4, 6 e 9) apresentam índices da variável IROP

moderado-alto. Pode-se perceber ainda que para o índice da variável IROP alto foi de 22,22% (setores 10, 12, 14 e 16).

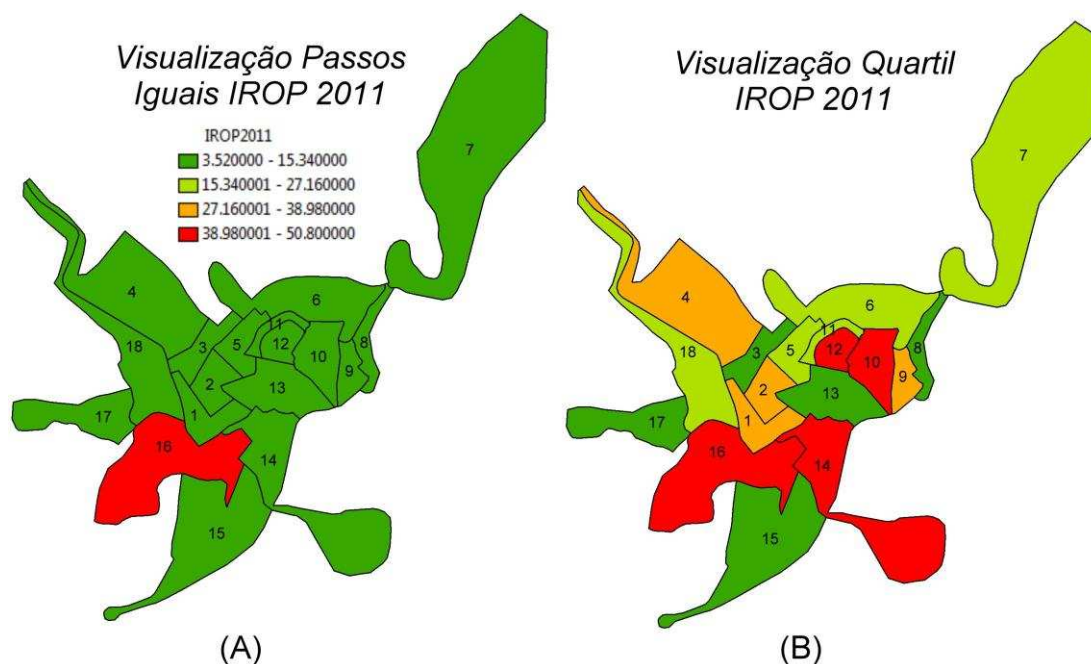


Figura 25 - Figura (A) distribuição da variável IROP do ano de 2011 pelo método estatístico de Passos Iguais e Figura (B) distribuição da variável IROP do ano de 2009 pelo método estatístico Quartil

Pode-se notar que nesta visualização (Figura 25 (A)), a maioria dos setores (94,44%, 17 setores) está com índice da variável IROP baixo e o setor 16 com índice de IROP alto. A Figura 25 (B) apresenta a distribuição de índices de IROP pelo método quartil, Assim, 27,77% (3, 8, 13, 15 e 17) apresentam IROP baixo. O mesmo número de setores (5, 6, 7, 11 e 18) apresentam índices de IROP moderado. E 22,22% (setores 1, 2, 4 e 9) apresentam índices de IROP moderado-alto. Pode-se perceber que para o índice de IROP alto foi de 22,22% (setores 10, 12, 14 e 16).

A análise da Figura 25 é satisfatória quanto a redução do índice IROP, uma vez que em relação aos anos de 2009 e 2010, o ano de 2011, apresentou um resultado muito significativo no combate à criminalidade da cidade de Rio Pomba.

Pelo método de passos iguais, nota-se que eles sempre realçam a quantidade de setores que fizeram parte de uma determinada classe. Já pelo método de visualização por quartis, pode-se perceber visualmente que estes sempre realçam a mudanças nas classes.

Por meio das Figuras 23, 24 e 25, pode-se perceber um declínio de ocorrências nos setores da cidade a cada ano. Nota-se que a região com maior número de ocorrências policiais é a região central da cidade, pois se concentram todos os bancos da cidade, é local de muitos bares e também de vida noturna mais intensa. Além disso, vale a pena ressaltar que estes dados corroboram com Beato (2008), onde a maior parte das ocorrências criminais de Belo Horizonte aconteceu na região central da cidade.

A redução das ocorrências policiais pode envolver diversos fatores que aconteceram na cidade durante esses três anos. Um fator que vem trazendo resultados significativos é o programa de combate às drogas que a polícia militar vem desenvolvendo. Outro fator interessante é o remodelamento das ruas da cidade, diminuindo para quase zero o número de ocorrências de trânsito, como pode ser visto na análise por tipo de ocorrência.

Realizando uma análise estatística, pode-se perceber uma significância nos dados estudados. A Tabela 14 mostra o nível de significância para o Índice de *Moran*.

Tabela 14 - Estatística global (Índice de *Moran*)

IROP	IM (Índice de <i>Moran</i>)	p-valor
2009	-0,2114	0,05
2010	-0,1659	0,07
2011	-0,1603	0,06

Os valores do índice IM de *Moran* global para a variável IROP foram, de -0,2114, -0,1659 e -0,1603 para os anos de 2009, 2010 e 2011, respectivamente. Uma vez que os valores do índice de *Moran* variam de -1 a 0 para autocorrelação inversa, sendo que -1 representa uma forte autocorrelação inversa e 0 representa ausência de autocorrelação, e de 0 até 1, onde o 1 representa uma forte correlação direta. Dessa forma, os valores obtidos para os anos de 2009, 2010 e 2011 são baixos para concluir se houve, ou não, uma autocorrelação espacial inversa, contudo suficientes, caso o nível de significância seja de 10%, por exemplo.

Para isso, foi realizado o teste de permutação aleatória do nível de significância de IM, para as variáveis IROP de cada ano, chegando a 99

permutações. Dessa forma, obtiveram-se os p -valores = 0,05, 0,07 e 0,06 para os anos de 2009, 2010 e 2011, respectivamente. Assim, foram testadas duas hipóteses: H_0 que seria a ausência de autocorrelação espacial e H_a que seria a presença de autocorrelação espacial. Dessa forma, rejeitou-se a hipótese nula de ausência de autocorrelação espacial entre os setores da cidade de Rio Pomba em relação a variável IROP ao nível de 10% de significância.

Como o valor de IM foi negativo e o teste foi significativo a 10% de probabilidade, então existe autocorrelação espacial negativa entre os setores da cidade de Rio Pomba, ou seja, o valor da variável IROP de um setor tende a ser inversamente proporcional aos valores dos seus vizinhos. Vale ressaltar que o IM é uma estatística de toda a região e que pode-se, por meio do Índice de *Moran* Local (IML), calcular a estatística de *Moran* em vários setores da mesma região. A Figura 28 retrata o IML da variável IROP para o ano de 2009, utilizando-se o teste da permutação aleatória (99 permutações).

Em estudo similar, Martarole *et al.*, (2009) apresentaram o índice global de *Moran* para criminalidade da cidade de Porto Velho – RO. Analisando as 99 permutações dos dados sobre roubos para o ano de 2006, e um p -valor igual a 0,15 e um IM igual a 0,0904827. Dessa forma, com o valor de índice tão próximo a zero e a alta significância os autores concluíram que não houve correlação entre os dados, diferentemente dos resultados apresentados neste trabalho.

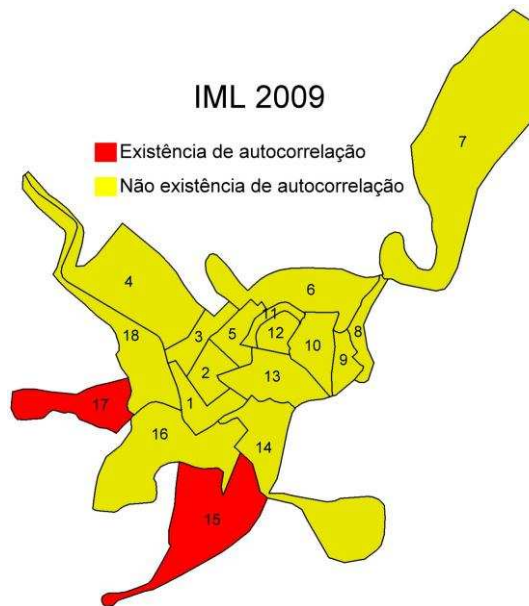


Figura 26 - Índice de *Moran* Local da variável IROP para o ano de 2009

Na Figura 26, utilizando os teste de permutações aleatórias (99), observou-se que a menor parte dos setores da cidade (os setores de número 15 e 17) apresentaram o IML significativo, indicando a significativa existência de autocorrelação desses setores com seus vizinhos para a variável IROP, 2009. Enquanto, os setores de número 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, apresentaram valores não significativos para o IML.

Para o ano de 2010, utilizando o mesmo teste de permutações aleatórias, a figura 27 apresenta o resultado para o referido ano.

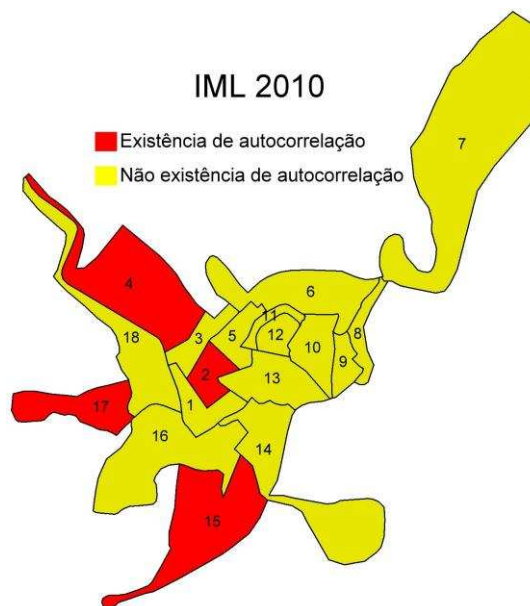


Figura 27 – Índice de *Moran* da variável IROP para o ano de 2010.

No ano de 2010, além dos setores que apresentaram correlação espacial no ano de 2009, mais 2 setores ficaram com o IML abaixo de 0,10, ou seja 10% de significância. Na Figura 27 observou-se que a menor parte dos setores da cidade (setores 2, 4, 15, 17) apresentaram o IML significativo, indicando a existência de autocorrelação desses setores com seus vizinhos para a variável IROP, 2010. Enquanto, os setores 1, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16 e 18, apresentaram valores não significativos para o IML.



Figura 28 - Índice de *Moran* Local da variável IROP para o ano de 2011

A Figura 28 retrata uma pequena mudança com relação ao ano de 2010. Observa-se que uma pequena parte dos setores da cidade (setores 5, 7 e 15) apresentou o IML significativo, indicando a existência de autocorrelação desses setores com seus vizinhos para a variável IROP, 2011. Enquanto, a outra parte dos setores composta por 1, 2, 3, 4, 6, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18 apresentaram valores não significativos para o IML.

O próximo passo foi calcular o *BoxMap* que é um mapa que representa uma extensão do gráfico de espalhamento de *Moran*, onde os elementos de cada quadrante do gráfico de espalhamento de *Moran* são representados por uma cor específica com seus respectivos polígonos. A Figura 29 apresenta o *BoxMap* da variável IROP para o ano de 2009.

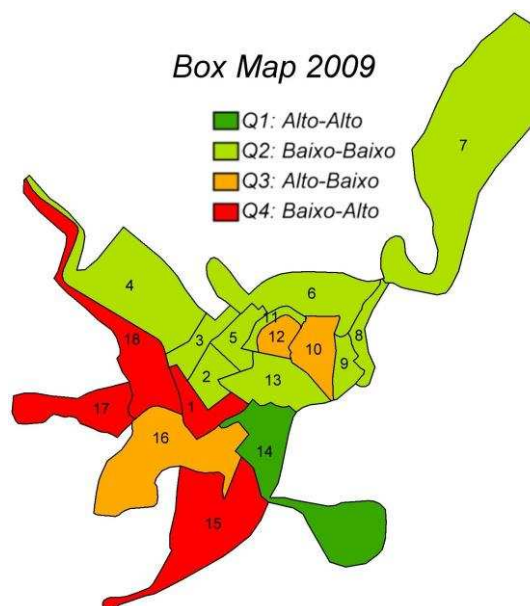


Figura 29 - *BoxMap* para a variável IROP do ano de 2009.

Valores mostrados na Figura 29 correspondem a relação entre os valores de Z que é definido pelo vetor dos desvios dos valores observados e W_z que corresponde ao vetor da média ponderada local em um gráfico de dispersão dividido em Quadrantes (Q). Dessa forma, esses valores variam nas cores verde, verde-claro, laranja e vermelho, onde verde corresponde aos valores de Q1 (alto-alto – valores altos de Z e valores altos de W_z), verde-claro (Q2) baixo-baixo, laranja (Q3) alto-baixo e vermelho (Q4) baixo-alto.

O *BoxMap* para o ano de 2009, apresenta 1 setor (14) com valores de Z e W_z alto-alto, 10 setores (2, 3, 4, 5, 6, 7, 8, 9, 11 e 13) com valores de Z e W_z baixo-baixo, 3 setores (10, 12 e 16) com valores de Z e W_z com alto-baixo e 4 setores (1, 15, 17 e 18) com valores Z e W_z baixo-alto.

A Figura 30 apresentam o *BoxMap* para o ano de 2010.

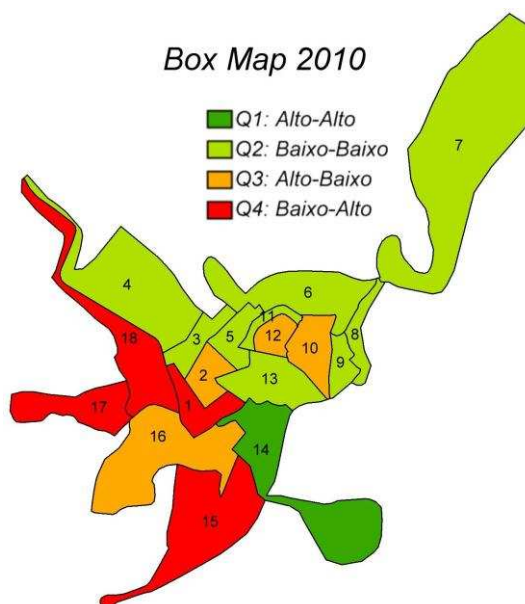


Figura 30 - *BoxMap* para a variável IROP do ano de 2010.

O *BoxMap* representado no ano de 2010, apresenta 1 setor (14) com valores de Z e W_z alto-alto, 9 setores (3, 4, 5, 6, 7, 8, 9, 11 e 13) com valores de Z e W_z baixo-baixo, 4 setores (2, 10, 12 e 16) com valores de Z e W_z com alto-baixo e 4 setores (1, 15, 17 e 18) com valores Z e W_z baixo-alto.

Para o ano de 2011, pode-se perceber que a Figura 31 apresenta 2 setores (1 e 14) no quadrante Q1, 9 setores (2, 3, 4, 5, 6, 7, 8, 9, 11 e 13) com legenda na cor verde-claro, 4 setores (4, 10, 12 e 16) com legenda na cor laranja e 3 setores (15, 17 e 18) na cor vermelha.

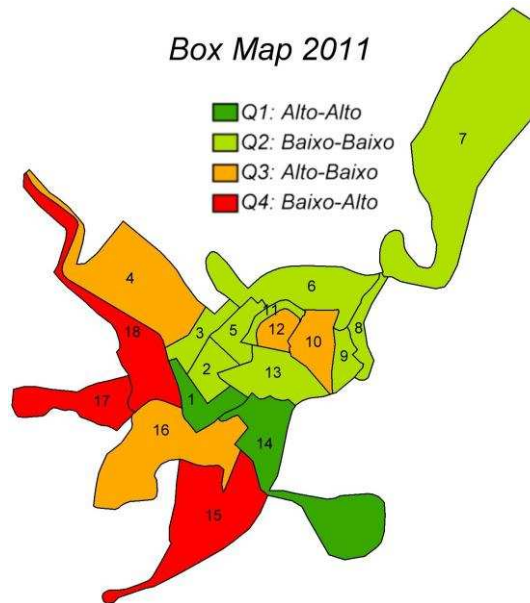


Figura 31 - *BoxMap* para a variável IROP do ano de 2010.

Com a análise acima, tornou-se pertinente o uso de uma técnica apta a identificar padrões e tendências espaciais. Para isto, o método da média móvel local seria necessário, pois resulta de um processo estocástico. Esse método baseia-se nos denominados efeitos de primeira ordem que expressam as variações do valor médio do processo no espaço (BAILEY e GATRELL, 1995; CÂMARA *et al.*, 2004). Ao explorar o valor médio da variável IROP para os anos de 2009, 2010 e 2011 essa técnica produzirá uma superfície mais suavizadas do que as estimativas originais. Para o cálculo da média móvel, foi desconsiderado o valor do IROP do setor 16, visto que este setor estava distorcendo os resultados dos seus vizinhos, uma vez que ele apresentava valores de IROP muito maiores do que seus vizinhos, e quando era realizado o teste da média móvel local, os setores vizinhos ao setor 16 ficavam com valores muito altos, não condizendo com a realidade dos mesmos. Dessa forma, o valor de IROP calculado no setor 16 foi a média aritmética dos seus vizinhos. Assim este setor não influenciará o cálculo da média móvel para as variáveis IROP 2009, 2010 e 2011. A Figura 32 apresenta a média móvel da variável IROP para o ano de 2009.

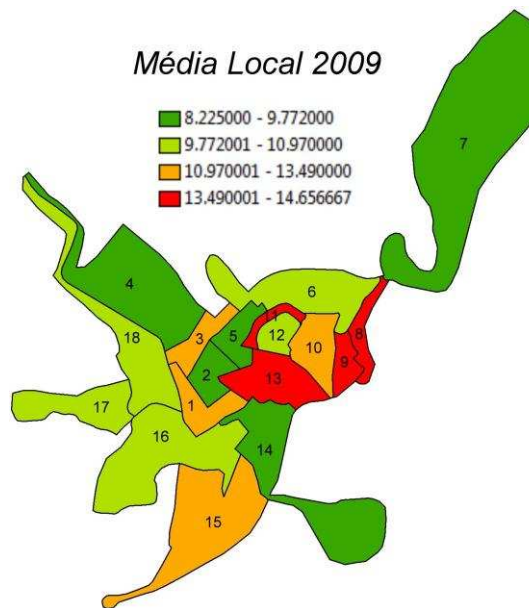


Figura 32 - Média Móvel Local da variável IROP para o ano de 2009.

Nota-se que o cálculo da média móvel local da variável IROP do ano de 2009 apresenta uma distribuição homogênea dos setores e também uma pequena variação entre o setor com menor média e o setor com maior média. Pode-se perceber que os setores 2, 4, 5, 7 e 14 apresentaram médias de IROP entre 8,22 e 9,77, os setores 6, 12, 16, 17 e 18 com média entre 9,77 e 10,97, os setores 1, 3, 10 e 15 com média entre 10,97 e 13,49 e os setores 8, 9, 11 e 13 variando sua média entre 13,49 e 14,65. Isto significa que o mapeamento da média de ocorrências na cidade, possibilitando a polícia militar um maior investimento em determinados locais, tais como, a região nordeste da cidade, onde concentram-se os setores 8, 9, 11 e 13.

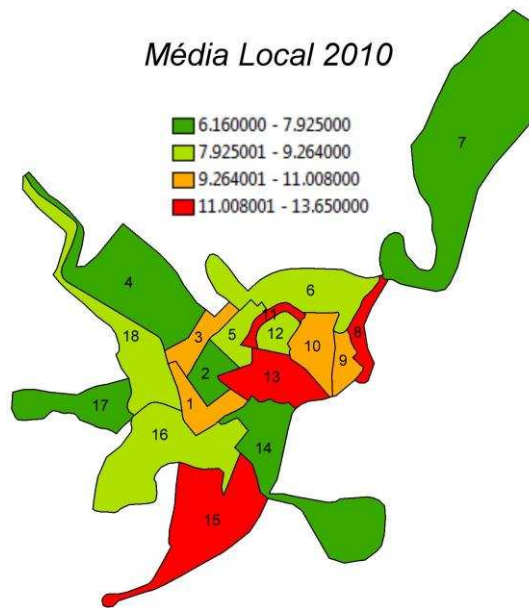


Figura 33 - Média Móvel Local da variável IROP para o ano de 2010.

A Figura 33 apresenta a média local da variável IROP para o ano de 2010. Vale ressaltar que foi realizado o mesmo cálculo com relação ao setor 16, que apresentava um IROP discrepante em relação aos outros setores. Dessa forma, os setores 2, 4, 7, 14 e 17 apresentam média local de IROP para o ano de 2010 variando entre 6,16 e 7,92, os setores 5, 6, 12, 16 e 18 variando entre 7,92 e 9,26, os setores 1, 3, 9 e 10 variando entre 9,26 e 11,00 e os setores 8, 11, 13 e 15 com variação de 11,00 a 13,65.

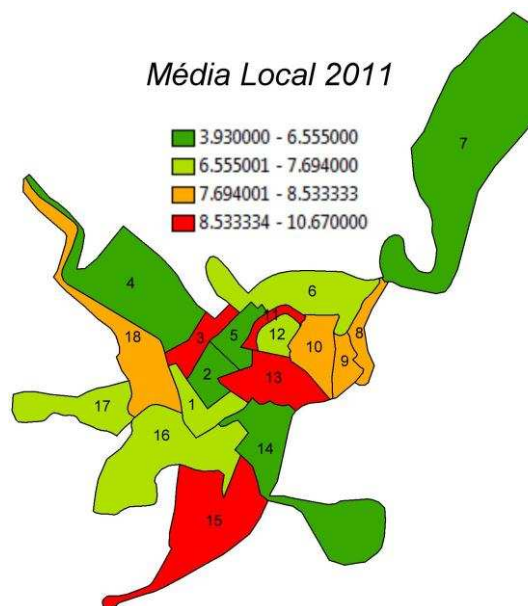


Figura 34 - Média Móvel Local da variável IROP para o ano de 2011.

A média local da variável IROP para o ano de 2011 é retratada na Figura 34. O método estatístico apresentado nas Figuras 33, 34 e 35 é o método de quartil, onde os setores 2, 4, 5 e 7 apresentam média local de IROP para o ano de 2011 variando entre 3,93 e 6,55, os setores de números 1, 6, 12, 16 e 17 variando entre 6,55 e 7,69, os setores 8, 9, 10 e 18 variando entre 7,69 e 8,53, e os setores 3, 11, 13 e 15 com variação de 8,53 e 10,67.

Após a análise espacial dos dados reais, foi realizada a predição de ocorrências policiais dos setores para os próximos quatro trimestres (Figura 35). Além disso, a Tabela 15 relata números empíricos preditos nos trimestres.

Tabela 15 - Predição de ocorrências policiais mostradas por trimestres da cidade de Rio Pomba - MG

	1º Trim	2º Trim	3º Trim	4º Trim	TOTAL
Setor 1	17.51	19.73	18.22	19.27	74.73
Setor 2	11.89	11.24	10.81	8.56	42.50
Setor 3	6.54	8.52	7.83	8.41	31.30
Setor 4	11.62	12.67	13.62	14.10	52.01
Setor 5	7.81	11.86	8.34	8.92	36.93
Setor 6	10.42	10.60	12.17	7.75	40.94
Setor 7	5.82	5.54	3.92	5.72	21.00
Setor 8	5.31	5.95	7.23	4.44	22.93
Setor 9	15.45	13.58	15.08	14.04	58.15
Setor 10	18.93	19.39	20.30	19.60	78.22
Setor 11	5.01	6.72	4.41	4.89	21.03
Setor 12	9.86	12.11	9.39	10.30	41.66
Setor 13	4.07	5.55	4.43	3.61	17.66
Setor 14	13.70	8.24	22.55	8.49	52.98
Setor 15	5.73	6.29	6.34	4.25	22.61
Setor 16	30.00	44.56	37.21	34.38	146.15
Setor 17	4.16	4.18	3.57	3.79	15.70
Setor 18	7.12	5.81	4.50	4.44	21.87

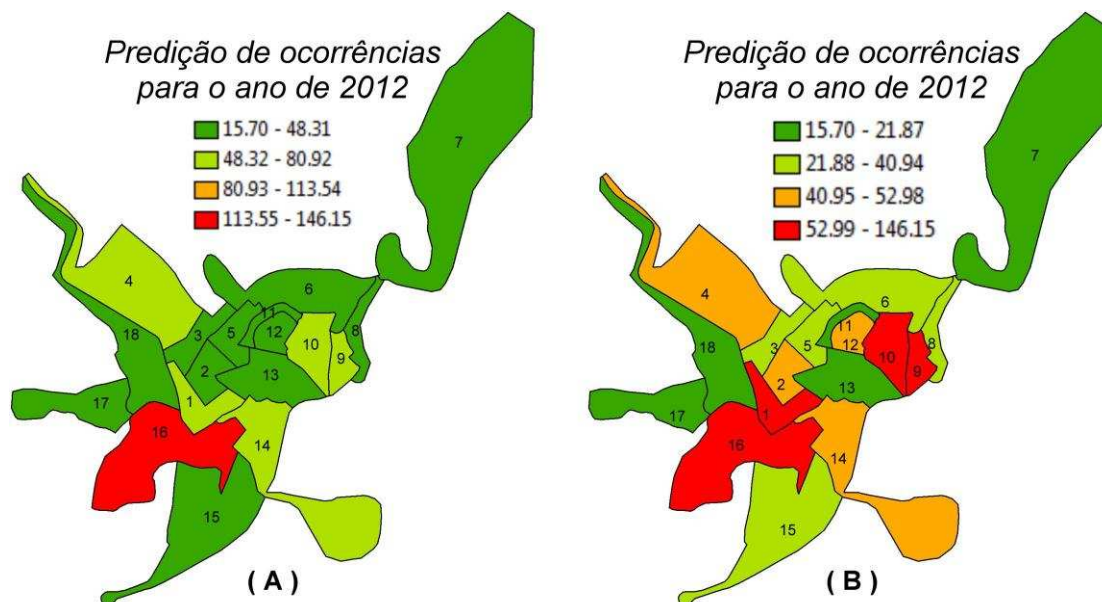


Figura 35 - Figura (A) distribuição de ocorrências do ano de 2012 pelo método de Passos Iguais e Figura (B) distribuição de ocorrências do ano de 2012 pelo método de Quartil

Pode-se notar que ao passar dos anos, a criminalidade do município vem diminuindo gradativamente e a predição caracterizou esta redução apresentando valores empíricos com taxa de aprendizado do algoritmo de

0,02, porém, consistentes se comparados à redução que vem acontecendo no decorrer dos anos.

O “Setor16” ainda continua sendo o alvo de crimes. Fazendo uma comparação dos anos com dados reais e os dados preditos, pode-se perceber que:

- De 2009 para 2010 – A redução de criminalidade no “Setor16” foi de aproximadamente 23,74%, o que significa uma alta redução em comparação aos anos de 2010 e 2011. Isto se deve ao fato de trabalho ostensivo na região, com intuito de diminuir a quantidade de ocorrências policiais do local.
- De 2010 para 2011 – Também houve uma redução de criminalidade, porém, nota-se que a redução foi menos acentuada em comparação ao ano anterior, pois foi apenas de 6,31% em relação ao ano de 2010.
- De 2011 para 2012 predito – Verifica-se que a predição para o ano de 2012 acompanhou a mesma variação do ano de 2010 para 2011, diminuindo aproximadamente de 8,37% em relação a 2011.

Realizando uma análise geral, a criminalidade da cidade de Rio Pomba vem diminuindo gradativamente devido ao trabalho da polícia com ações sociais, operações e conscientização da população. Sabe-se que nunca vai acabar, porém a tendência é chegar a um ponto mínimo de criminalidade com o passar dos anos.

3.8 Conclusões

Pode-se concluir que o índice de ocorrências policiais da cidade de Rio Pomba-MG vem reduzindo a cada ano. Isto se deve aos programas de ações sociais promovidas pela referida instituição.

Outro fato interessante é que as ocorrências do tipo A (ação social) reduziram também as operações da polícia (ocorrências do tipo Y). Com isso pode-se concluir que um dos fatores determinantes para a redução da criminalidade são as ocorrências de ações sociais.

Através de uma análise considerando o contexto espacial, pode-se concluir que na maioria dos setores da cidade o número de ocorrências

policiais diminuíram. A percepção espacial da polícia militar, combatendo as áreas de maiores ocorrências, através de rondas, palestras socioeducativas, programas sociais, pode ser um dos fatores para redução da criminalidade na cidade de Rio Pomba.

Vale ressaltar que os resultados estatísticos são de vital importância para um novo campo de análise para a Polícia Militar, mostrando informações muito relevantes que não poderiam ser percebidas, não considerando um contexto espacial.

Assim, o uso de ferramentas de mineração de dados em conjunto com a estatística espacial torna-se de grande importância para análises considerando a informação espaço-temporal, a fim contribuir no combate dos índices de criminalidade.

Referência Bibliográfica

ADDERLEY, R.; TOWNSLEY, M.; BOND, J. Use of data mining techniques to model crime scene investigator performance. **Knowledge-Based Systems**, 2007. v. 20, n. 2, p. 170–176. . Acesso em: 24 out. 2013.

AGRAWAL, R.; IMIELIŃSKI, T.; SWAMI, A. Mining association rules between sets of items in large databases. **SIGMOD Rec.**, jun. 1993. v. 22, n. 2, p. 207–216.

ANSELIN, L. Local Indicators of Spatial Association—LISA. **Geographical Analysis**, 1995. v. 27, n. 2, p. 93–115. . Acesso em: 25 nov. 2013.

BEATO, C. **Compreendendo e avaliando: projetos de segurança pública**. [S.I.]: Editora UFMG, 2008.

BENNETT, K. P.; MANGASARIAN, O. L. Robust linear programming discrimination of two linearly inseparable sets. **Optimization methods and software**, 1992. v. 1, n. 1, p. 23–34. . Acesso em: 25 out. 2013.

BIGUS, J. P. **Data mining with neural networks: solving business problems from application development to decision support**. [S.I.]: McGraw-Hill, Inc., 1996.

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. Essex: Longman Scientific & Technical, 1995.

BROWN, D. E.; OXFORD, R. B. **Data mining time series with applications to crime analysis**. 2001 Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=973487>. Acesso em: 24 out. 2013.

CÂMARA, G. *et al.*. Análise espacial e geoprocessamento. **Análise espacial de dados geográficos**. [S.l.]: EMBRAPA, 2004, p. 1–25.

CAULKINS, J. *et al.*. Predicting criminal recidivism: A comparison of neural network models with statistical methods. **Journal of Criminal Justice**, 1996. v. 24, n. 3, p. 227–240. . Acesso em: 25 out. 2013.

CHAINEY, S.; TOMPSON, L.; UHLIG, S. The utility of hotspot mapping for predicting spatial patterns of crime. **Security Journal**, 2008. v. 21, n. 1, p. 4–28. . Acesso em: 24 out. 2013.

CHEN, H. *et al.* Crime data mining: a general framework and some examples. **Computer**, 2004. v. 37, n. 4, p. 50–56. . Acesso em: 24 out. 2013.

CORCORAN, J. J.; WILSON, I. D.; WARE, J. A. Predicting the geo-temporal variations of crime and disorder. **International Journal of Forecasting**, 2003. v. 19, n. 4, p. 623–634. . Acesso em: 24 out. 2013.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, 1995. v. 20, n. 3, p. 273–297. . Acesso em: 25 out. 2013.

DIAS, TACIANA L.; CÂMARA, G.; DAVIS JR, C. A. MODELOS ESPAÇO-TEMPORAIS. **Banco de Dados Geográficos**. [S.l.]: MundoGeo, 2005, p. 506.

ESRI. **ArcGIS desktop. Release 10**. [S.l.]: Environmental Systems Research Institute Redlands, California, 2011.

FRANK, A.; RAPER, J.; CHEYLAN, J.-P. **Life and Motion of Socio-Economic Units: GISDATA**. [S.l.]: CRC Press, 2004.

HALL, M. *et al.* The WEKA data mining software: an update. **ACM SIGKDD Explorations Newsletter**, 2009. v. 11, n. 1, p. 10–18. . Acesso em: 28 out. 2013.

INPE. **Terraview**. São José dos Campos, SP: [s.n.], 2013.

KEOGH, E.; KASETTY, S. On the need for time series data mining benchmarks: a survey and empirical demonstration. **Data Mining and Knowledge Discovery**, 2003. v. 7, n. 4, p. 349–371. . Acesso em: 25 out. 2013.

KEYVANPOUR, M. R.; JAVIDEH, M.; EBRAHIMI, M. R. Detecting and investigating crime by means of data mining: A general crime matching framework. **Procedia Computer Science**, 2011. v. 3, p. 872–880. . Acesso em: 25 out. 2013.

LI, S.-T.; KUO, S.-C.; TSAI, F.-C. An intelligent decision-support model using FSOM and rule extraction for crime prevention. **Expert Systems with Applications**, 2010. v. 37, n. 10, p. 7108–7119. . Acesso em: 24 out. 2013.

MARTAROLE, T. L.; BUENO, L. F.; DALL'IGNA, L. G. Análise espacial dos dados da criminalidade infanto-juvenil no município de Porto Velho-RO. **Anais XIV Simpósio Brasileiro de Sensoriamento Remoto**, abr. 2009. n. INPE, p. 4035–4042. . Acesso em: 29 out. 2013.

MURTAGH, F. Multilayer perceptrons for classification and regression. **Neurocomputing**, 1991. v. 2, n. 5, p. 183–197. . Acesso em: 25 out. 2013.

PALOCSAY, S. W.; WANG, P.; BROOKSHIRE, R. G. Predicting criminal recidivism using neural networks. **Socio-Economic Planning Sciences**, 2000. v. 34, n. 4, p. 271–284. . Acesso em: 24 out. 2013.

PAULA, G. A. **Modelos de Regressão: com apoio computacional**. [S.l.]: IME-USP Sao Paulo:, 2004.

PLATT, J. Sequential minimal optimization: A fast algorithm for training support vector machines. 1998. Disponível em: <http://www.bradblock.com/Sequential_Minimal_Optimization_A_Fast_Algorithm_for_Training_Support_Vector_Machine.pdf>. Acesso em: 25 out. 2013.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, 1958. v. 65, n. 6, p. 386. . Acesso em: 25 out. 2013.

SEDS. **Estatísticas de Criminalidade Violenta – Janeiro a Novembro de 2012**. Belo Horizonte: Secretaria de Estado de Defesa Social, 2012. . Acesso em: 24 out. 2013.

SHUMWAY, R. H. Applied statistical time series analysis. **Prentice Hall Series in Statistics, Englewood Cliffs, NJ: Prentice-Hall,| c1988**, 1988. v. 1. Disponível em: <<http://adsabs.harvard.edu/abs/1988asts.book.....S>>. Acesso em: 25 out. 2013.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and computing**, 2004. v. 14, n. 3, p. 199–222. . Acesso em: 25 out. 2013.

VAPNIK, V. **The nature of statistical learning theory**. [S.l.]: springer, 2000.

VASCONCELOS, E.; FURTADO, V. Um Simulador Tutorial Multi-Agente para Treinamento da Alocação de Equipes Policiais. **Anais do XVIII Encontro Nacional de Inteligência Artificial**, 2005. p. 892–901. . Acesso em: 25 nov. 2013.

WILSON, R. L.; SHARDA, R. Bankruptcy prediction using neural networks. **Decision support systems**, 1994. v. 11, n. 5, p. 545–557. Acesso em: 24 out. 2013.

4 SIMULAÇÃO DE OCORRÊNCIAS POLICIAIS VISANDO OTIMIZAR O TRABALHO DA POLÍCIA MILITAR E MINIMIZAR O ÍNDICE DE OCORRÊNCIAS POLICIAIS

Resumo

O crescimento da criminalidade na última década para o estado de Minas Gerais fortalece a necessidade de medidas proativas da segurança pública. Agravando este problema, a maioria das cidades de pequeno porte não conta com um sistema de gerenciamento voltado ao cadastro de ocorrências criminais. Considerando que a solução ao combate ao crime não seja somente a aplicação de penas ao criminoso, mas também de tentar evitar que a ocorrência de um crime aconteça, este trabalho propõe a utilização de um modelo de dados espaço-temporal de ocorrências policiais para prever, por meio de redes neurais, qual o melhor investimento da polícia em cada setor urbano do município de Rio Pomba – MG. Além disso, para tentar reduzir os índices de criminalidade, regras de associação foram geradas, por meio do algoritmo *Apriori*, com intuito de mostrar a associação espaço-temporal de ocorrências policiais, setores e turnos.

Palavras Chaves: *Multilayer Perceptron*, *Apriori*, espaço-temporal, simulação de ocorrências policiais.

Abstract

The growth of crime in the last decade for the state of Minas Gerais strengthens the need for proactive measures of public safety. Getting this problem, most small cities does not have a management system aimed at the registration of criminal occurrences. Considering that the solution to combat the crime is not the only application of the criminal penalties, but also to try to prevent the occurrence of a crime happening, this paper proposes the use of a data model spatial-temporal police incidents to predict, by neural networks, which is the best investment of the police in each urban sector of Rio Pomba. Also, to try to reduce crime rates, association rules were generated through the *Apriori* algorithm, in order to show the spatial and temporal association of police incidents, sectors and shifts;

Key Words: *Multilayer Perceptron*, *Apriori*, spatial-temporal, simulation of police incidents.

4.1 Introdução

Com o passar dos anos, as cidades brasileiras vêm sofrendo grandes transformações relacionadas com o desenvolvimento socioeconômico. Um grande contingente de pessoas passou a habitá-las com intuito de trabalharem em atividades relacionadas aos setores industriais, de transporte e de serviços. Dessa forma, grande parte das cidades, até mesmo as que foram planejadas, passaram a crescer de forma desordenada, sem um controle efetivo das autoridades responsáveis. Um exemplo do reflexo da falta de planejamento é o trânsito atual, que apresenta filas de engarrafamentos cada dia maiores.

Segundo Gomes (2007), em tempos remotos as cidades eram criadas com a principal finalidade de proteção de seus habitantes. Entretanto, nos dias atuais, as cidades tornaram-se um lugar inseguro, perigoso e fragmentado. Possivelmente, tal fato ocorreu devido a seu crescimento desordenado e, em consequência, a segurança pública passou a ser uma das grandes preocupações dos habitantes em um ambiente urbano.

O crescimento da criminalidade para o Estado de Minas Gerais foi de aproximadamente 3,5 vezes a mais desde 1996 até 2003, porém, deve-se ressaltar que houve uma redução da taxa de criminalidade no intervalo de 2003 a 2010, mas ainda é muito alta em relação as duas décadas anteriores (SEDS, 2012).

Na maioria das cidades de pequeno porte, ainda não se tem o um sistema de gerenciamento voltado ao cadastro de ocorrências criminais, que também permite a inserção de informações espaciais relacionadas a essas ocorrências. O uso desses sistemas pode fornecer varias vantagens, como a análise espacial, com a indicação, por exemplo, de locais com maior índice de ocorrência para um determinado tipo de crime, bem como a predição do cenário de crimes frente a ação da policia militar. No entanto, para que essas funções possam ser aplicadas, é necessário que se tenha um histórico com os dados de ocorrência, o que não ocorre quando o cadastro é realizado de forma rudimentar, realizando o mapeamento das ocorrências através de furos de pinos em mapas.

Uma das grandes evoluções no combate ao crime não é considerar somente a aplicação de penas ao criminoso, mas também de tentar evitar que a ocorrência de um crime aconteça.

Neste sentido, o uso de sistemas simuladores tais como, NetLogo, de ocorrências policiais pode ser de grande utilidade no auxílio à segurança pública, no sentido de poder analisar a dinâmica de uma cidade quanto a sua criminalidade.

4.2 Objetivos

Os objetivos gerais desta pesquisa são voltados ao estudo da criminalidade de uma cidade de pequeno porte, especificamente Rio Pomba-MG, utilizando dados de ocorrências fornecidos pela Polícia Militar e uma base cartográfica digital disponibilizada no *site* do IBGE. Análises computacionais foram determinadas por meio do *framework Weka 3.7.8* (Hall, *et al.*, 2009), que tem a possibilidade de simulação do investimento da polícia e o número de ocorrências de um determinado setor da cidade, na tentativa de reduzir a criminalidade de todos os setores e também reduzir o investimento da Polícia Militar.

4.3 Revisão de literatura

Atualmente, o país vem passando por grandes transformações no tocante à criminalidade, com mudanças relacionadas a seu perfil social.

Um dos fatores que acarretam essas transformações é o aumento populacional, que necessita de um maior contingente policial para manter a ordem no município.

Modelos de simulação computacional de criminalidade são ferramentas que auxiliam o setor de segurança pública no sentido de simular ocorrências de crimes em determinados locais e tentar resolvê-los com o mínimo de investimento possível, por parte da polícia e ainda prever o acontecimento de outros crimes através de um policiamento bem distribuído na cidade.

Vários trabalhos relacionados ao estudo da criminalidade com auxílio de sistemas simuladores têm sido realizados. Dentre eles, alguns estão descritos a seguir.

De acordo com Sherman *et al.* (1989)² *Apud* Weisburd *et al.* (2009) em um dos estudos pioneiros nessa área, constata que apenas 3% dos endereços em Minneapolis produziram 50% de todas as chamadas policiais. Quinze anos depois, em um estudo em Seattle, Washington, Weisburd *et al.* (2004) relatam que a porcentagem entre 4 e 5 segmentos de rua na cidade foi responsável por 50% dos incidentes de crimes registrados oficialmente para cada ano, por mais de 14 anos.

Groff (2007) desenvolveu um método para a implementação de movimentação de agentes em uma rede de ruas, onde os espaços de rotina são a única atividade para os indivíduos de uma população, ou seja, esses espaços são: empregos e comércio, serviços e oportunidades de recreação. Vale a pena ressaltar que em contraste com o modelo de ruas, esta pesquisa não se concentra em reproduzir o movimento de indivíduos dentro da cidade, mas sim na demonstração de como esse movimento está relacionado a um resultado específico social, o roubo em uma rua da cidade.

Groff (2008) apresenta uma abordagem de atividades de rotinas propostas por Cohen e Felson (1979) para estudar a criminalidade. O estudo apresenta atividades em dois níveis: micro e macro. As atividades que apresentam mudanças no nível macro correspondem as mudanças sociais. Tais mudanças começaram nas décadas de 60 e 70 como, por exemplo, o trabalho fora das residências, que seria uma das fontes de aumento da criminalidade. Já as mudanças no nível micro é o papel das atividades cotidianas de pessoas, ou seja, como facilitar ou dificultar a aproximação dos infratores e a capacidade de proteção em um determinado tempo e lugar. A pesquisa revelou que no nível macro, aspectos da teoria da atividade de rotina levaram à sua ampla aplicação em pesquisas empíricas, mas não no nível micro. Isto se deu pela dificuldade de obtenção de dados a

² SHERMAN, L. W.; GARTIN, P. R.; BUERGER, M. E. HOT SPOTS OF PREDATORY CRIME: ROUTINE ACTIVITIES AND THE CRIMINOLOGY OF PLACE*. *Criminology*, v. 27, n. 1, p. 27-56, 1989.

nível individual, bem como deficiências técnicas estatísticas têm impedido testes de nível micro.

Chainey *et al.* (2008) utiliza a técnica de mapeamento de *hotspot* que é usada para ajudar a determinar o local em que um crime pode acontecer, com base em dados antigos com intuito de informar ações futuras. Dessa forma, segundo o referido autor, ele age como uma técnica básica para prever onde o crime pode ocorrer, usando a premissa de que os padrões de retrospectiva de crime são um indicador útil para os padrões futuros.

Groff e McCord (2011) relatam a criminalidade em parques abertos ao público, ou seja, identificam características dos parques e as combinaram com dados oficiais de criminalidade com a finalidade de descrever empiricamente a relação entre o parque e os vizinhos adjacentes.

O crescimento acelerado da criminalidade está forçando os órgãos de segurança pública a tomarem providências mais eficazes de combate e até mesmo de prevenção a essas atividades. Neste sentido, os modelos de simulação computacionais, cuja origem advêm das áreas da ciência da computação e inteligência artificial, podem auxiliar no combate ao crime, cujo objetivo é conhecer a realidade de uma cidade, em um modelo empírico. Possibilitando ao setor de segurança pública ter informações que possam ser utilizadas visando conter tal criminalidade.

Lui e Eck (2008) descrevem uma metodologia capaz de observar e estudar o fenômeno do crime por meio da implementação de experimentos através de uma simulação que, ao contrário de análises tradicionais, permitem uma observação e medição mais apurada de crimes, fazendo com que forneçam novos modelos de comportamento e também um refinamento das teorias criminológicas existentes.

Em uma simulação, pesquisadores podem alterar fatores que não estão em seu controle, implementar intervenções, capacidade de gravar todas interações entre os componentes do modelo, bem como saber o resultado destas com alta precisão, sem se preocupar com restrições logísticas e financeiras.

As simulações são uma forma de experimento, e a replicação das simulações revela a sua validade. Existem dois tipos de replicação: a replicação controlada que ocorre quando os métodos são utilizados em

condições idênticas (mesmo laboratório, amostras e equipe de pesquisa, etc) e a replicação prática quando um experimento tenta imitar um estudo original o mais próximo possível (LUI e ECK, 2008).

Estas pequenas diferenças (laboratórios e amostras diferentes, por exemplo) permitem aos pesquisadores investigar se essas disparidades influenciam o tamanho e a magnitude do efeito experimental. Dessa forma, com a replicação prática, pode-se alterar propositalmente uma variável independente como, por exemplo, tráfico de drogas, e encapsular representações mais próximas do alvo estudado. A vantagem dos modelos de simulação é capacitar o pesquisador a ter o controle absoluto das ações, fazendo com que ele compreenda melhor a situação estudada.

Townsley e Birks (2008) descrevem um modelo de simulação simples como um meio para ilustrar as diferentes formas de replicar experimentos através de simulação de uma maneira sistemática. A simulação foi desenvolvida utilizando NetLogo, uma plataforma de modelagem de simulação multi-agente. O objetivo do modelo é fornecer uma visão sobre a interação alvo-criminoso descrita na teoria da atividade de rotina.

Em uma visão abstrata de aplicação, o mundo pode ser simulado por três grupos distintos de indivíduos: alvos, criminosos e policiais, considerando que todos os indivíduos se movem em torno do ambiente de uma forma aleatória; que um crime ocorre quando um criminoso está no mesmo local e ao mesmo tempo que um agente alvo, e que uma detecção ou prevenção ocorre quando um criminoso, um alvo, e um policial reúnem-se no espaço e no tempo.

Como pode ser percebido, os sistemas simuladores podem ser muito importantes para o conhecimento e desenvolvimento de cidades e regiões. Estes sistemas trabalham com a figura do agente, que tem com uma de funções, coletar informações do meio em que está inserido. A partir dessas informações, pode-se simular o ambiente virtual em que ele se encontra.

Groff e Birks (2008) ressaltam o processo de desenvolvimento de estratégias de prevenção de crimes. Salientam ainda que, são softwares extremamente caros para serem testados empiricamente. A ideia seria testar essas estratégias em um mundo virtual, com a finalidade de fazer ajustes para cada lugar, de acordo com a cultura, costumes e religião.

A prevenção considera que as secretarias de segurança pública sejam proativas, com a capacidade de prever *hotspots* relacionados à criminalidade. De acordo com Ratcliffe (2010), a previsão de eventos individuais é raramente possível, uma vez que existe uma ligação direta entre a prevenção e os padrões de crimes. Dessa forma, a prevenção requer previsibilidade de padrões. A importância de identificar padrões de crimes como um precursor eficaz para a prevenção da criminalidade, parte do princípio da primeira regra de Tobler (1970), que diz, "tudo está relacionado com tudo, porém as coisas próximas estão mais relacionadas do que coisas distantes".

Na literatura existem diferentes técnicas de mapeamento que podem ser usadas para identificar e explorar padrões de crimes. Estas técnicas de mapeamento têm sido objeto de várias análises, mas essas avaliações têm demonstrado que elas apresentam resultados diferentes em termos de localização, tamanho e forma das áreas.

4.3.1 O algoritmo *Multilayer Perceptron*

O algoritmo classificador *Multilayer Perceptron* é uma extensão do algoritmo *Perceptron* (desenvolvido por Rosenblatt (1958)), podendo-se trabalhar em diversas camadas (camadas intermediárias). a fim de estabelecer padrões a partir dos dados de entrada e do processamento interno, através das conexões ponderadas, que podem ser consideradas como extratoras de características.

Dentre os aspectos relevantes desse classificador, tem-se que tal algoritmo se configura como um modelo adaptativo treinável, capaz de aprender a partir de exemplos, e com isso melhorar seu desempenho, atingindo uma solução generalizada para uma determinada classe de problemas. Pode ser aplicado a problemas não linearmente separáveis, Além de processar informações espaço/temporais.

Devido a sua característica de não-linearidade ele possui diversas aplicações, tais como classificação, detecção de padrões e previsão de séries temporais.

4.3.2 O algoritmo Apriori

Um dos processos mais comuns para extração de informações em um banco de dados é a mineração de dados. O termo mineração de dados foi definido por Witten e Frank (2005) como o “processo de descobrir padrões em dados”. A partir deste pressuposto e com um grande número de ocorrências geradas durante meses e anos referentes à cidade de Rio Pomba, a mineração de dados será uma ferramenta valiosa para extração de padrões de crimes espaço-temporais.

Bases de dados estão suscetíveis a conter algumas inconsistências, devido a diferentes fontes de origem, diferentes operadores no cadastro da ocorrência e também pela falta de conhecimento técnico de banco de dados por parte dos operadores. Com isso, a baixa qualidade desses dados pode produzir padrões de dados pouco convincentes. Assim, o tratamento dos dados é um fator crucial para gerar informações mais confiáveis.

Uma análise dos dados pode ser muito importante para o sucesso da mineração dos dados. Durante esta etapa, pode-se observar que muitos dados não possuem relevância para a informação que se deseja obter em uma determinada aplicação.

De acordo com Bigus (1996), na fase de pré-processamento é importante saber quais os objetivos que se pretende alcançar. Assim, a ajuda de um especialista para administrar os dados no banco de dados é fundamental para consistência dos mesmos.

Outro fator importante para mineração de dados é que os dados precisam estar apresentados sob determinados tipos de formatos, para que um respectivo software, ao utilizar um algoritmo específico, possa realizar corretamente a leitura dos dados (Bigus,1996), extraindo estatísticas relevantes relacionada à pesquisa.

Um dos algoritmos utilizados para descobrir associações relevantes em um banco de dados é o algoritmo *Apriori*, que através de buscas sucessivas à base de dados gera informações com ótimo desempenho computacional (AGRAWAL e SRIKANT, 1994).

O algoritmo de mineração de dados *Apriori* é um dos mais conhecidos quando se trata de regras de associação. De acordo com Pitoni (2002), o

algoritmo *Apriori* minera conjuntos de itens frequentes para regras de associação booleanas. Ele realiza diversas leituras na base de transações com a finalidade de encontrar todos os conjuntos de itens frequentes. Para isto, primeiramente gera-se um conjunto de itens candidatos e depois percorre o banco de dados para determinar se os candidatos satisfazem o suporte mínimo estabelecido. Na primeira passagem, o suporte para cada item individual (conjunto-de-1-item) é contado e todos aqueles que satisfazem o suporte mínimo são selecionados. Na segunda iteração, conjuntos-de-2-itens candidatos são gerados pela união do grupo inicial, ou seja, dos conjuntos-de-1-item frequentes com os seus suportes no qual são determinados pela pesquisa no banco de dados, formando o conjunto-de-2-itens frequentes. O algoritmo prossegue iterativamente, até que o conjunto de $k - itens$ encontrados com frequência seja um conjunto vazio. Segue abaixo um exemplo do algoritmo *Apriori* adaptado de Agrawal e Srikant (1994).

Suponha um banco de dados formado somente por um Grupo de Professores (GP) da Universidade Federal de Viçosa (UFV). Suponha, também, que este grupo seja composto por cinco professores, associando-se a cada um deles cinco itens categóricos, conforme mostrado na Tabela 16.

Tabela 16 - Dados fictícios do grupo de professores da UFV

Registro (Professor)	Itens Categóricos				
	A	B	C	D	E
501	1	0	0	1	0
502	0	1	1	0	1
503	1	0	1	0	1
504	0	1	1	0	1
505	0	1	1	0	0
Frequência	2	3	4	1	3

Sendo: A = Professor estrangeiro; B = Professor brasileiro; C = Sexo feminino; D = Sexo masculino; E = Pesquisador com título de doutorado; 0 = Ausência de atributo; 1 = Presença de atributo.

Seja C_k o conjunto de $k - itemsets$ candidatos, onde $k = 5$. Cada membro C_k deste conjunto tem dois campos: itemset e contador de suporte, representados, respectivamente, por *its* e *cs* na Figura 36.

C1		L1		C2		L2		C3		L3	
<i>its</i>	<i>cs</i>	<i>its</i>	<i>cs</i>	<i>its</i>	<i>cs</i>	<i>its</i>	<i>cs</i>	<i>is</i>	<i>cs</i>	<i>its</i>	<i>cs</i>
{A}	2	{A}	2	{AB}	0	{BC}	3	{BCE}	2	{BCE}	2
{B}	3	{B}	3	{AC}	1	{BE}	2				
{C}	4	{C}	4	{AE}	1	{CE}	3				
{D}	1	{E}	3	{BC}	3						
{E}	3			{BE}	2						
				{CE}	3						

Figura 36 – Geração dos **itemsets** candidatos (C) e dos **itemsets** frequentes (L)

Seja L_k o conjunto dos k – *itemsets* frequentes. De modo análogo, cada membro deste conjunto também possui *its* e *cs*.

O primeiro passo do algoritmo conta a frequência com que os itens ocorrem para determinar os 1 – *itemsets* frequentes (última linha da Tabela 18). Posteriormente, obtém-se o conjunto de candidatos 1 – *itemsets*, C_1 , mostrado na Figura 36. Assumindo um suporte mínimo igual a dois, ou seja, $\text{minsup} = 40\%$, L_1 é composto pelos elementos de C_1 com suporte igual ou superior a 40%. No exemplo, somente o *itemset* D não atendeu a esta condição, ficando L_1 composto por $\{A\}, \{B\}, \{C\}$ e $\{E\}$.

Para descobrir o conjunto dos 2 – *itemsets* frequentes, de modo a continuar satisfazendo ao suporte mínimo, o *Apriori* usa a concatenação $L_1 \times L_1$ para gerar o conjunto candidato C_2 , que consiste de 2 – *itemsets*. Por exemplo, $\{C\}$ e $\{E\}$ geram $\{CE\}$. Mais uma vez, cada ocorrência é computada. No caso $\{CE\}$ ocorre três vezes em GP (registros 502, 503 e 504). L_2 é determinado com base no suporte de cada candidato de C_2 . Agora são excluídos $\{AB\}, \{AC\}$ e $\{AE\}$, pois têm suporte inferior ao mínimo estabelecido.

A geração de C_3 é obtida a partir de L_2 de uma maneira distinta. Os futuros *itemsets* candidatos devem manter uma ordem lexicográfica, tal que quando a concatenação $L_2 \times L_2$ for realizada, o primeiro item de um *itemset* seja idêntico ao primeiro item do outro *itemset* e assim sucessivamente.

Porém, o último item do *itemset* deve ser menor, lexicograficamente, que o último item do outro *itemset*. Esta regra é expressa pela equação 15:

$$\text{itemset}_p = \{p_1, p_2, \dots, p_n\}, \text{itemset}_q = \{q_1, q_2, \dots, q_n\} \quad (15)$$

sendo necessário que

$$p_1 = q_1, p_2 = q_2, \dots, p_n < q_n \quad (16)$$

Na Figura 36 o *itemset* candidato $\{BCE\}$, em C_3 , foi formado concatenando $\{BC\}$ com $\{BE\}$, pois $B = B e C < E$. Este foi o único conjunto que pôde ser formado, pois não há outra concatenação que satisfaça (X2). A concatenação $\{BC\}X\{CE\}$, por exemplo, não satisfaz (X2), pois, lexicograficamente, $p_1 = B$ é menor que $q_1 = C$.

O passo seguinte é descobrir as regras de associação. No caso do fictício grupo de professores, supondo uma confiança mínima de 60% e mantendo o suporte mínimo em 40%, uma regra provável seria $BC \Rightarrow E$. Para ela, a confiança é igual suporte $(BCE) / \text{suporte}(BC)$, cujo resultado é $2/3$, ou 66%, satisfazendo a condição imposta (ver Tabela 17).

Para a regra $BC \Rightarrow E$ ou $(\text{professor brasileiro; sexo feminino}) \Rightarrow (\text{professor doutor})$, seu suporte seria o percentual de ocorrências de BCE com relação ao total de professores do grupo, que resulta em 40% ($2/5$). Então, esta é uma regra válida. Isto equivale a dizer que, das professoras brasileiras, 66% têm doutorado, muito embora estas brasileiras portadoras do título de doutor correspondam a apenas 40% dos indivíduos do grupo.

Outra provável regra seria $B \Rightarrow CE$. Para esta situação, o valor da confiança seria idêntico, pois a razão $\text{suporte}(BCE) / \text{suporte}(B)$ também é igual a $2/3$.

Tabela 17 - Ocorrências dos conjuntos de atributos

Registro (Professores)	Itens Categóricos		
	B	C	E
501	0	0	0
502	1	1	1
503	0	1	1
504	1	1	1
505	1	1	0
Frequência	3	4	3

4.4 Materiais e Métodos

4.4.1 Materiais Utilizados

Para o desenvolvimento da pesquisa foram utilizados os seguintes materiais:

- *ArcGis* versão 10.1(ESRI, 2011);
- *Weka* versão 3.7.8 (Hall, *et al.*, 2009);
- Dados espaciais da base digital do IBGE da agência de Rio Pomba;
- Dados de boletins de ocorrência fornecidos pelo quartel de Polícia Militar de Rio Pomba.

— Área de estudo

A área de estudo da pesquisa foi o município de Rio Pomba, que teve sua origem no século XVIII, época em que ocorreu o processo de colonização de parte significativa da Zona da Mata e do leste do território da Capitania de Minas Gerais. Atualmente o município de Rio Pomba, de acordo com o censo de 2010, possui aproximadamente 17000 habitantes. O município de Rio Pomba teve um crescimento populacional em torno de 20% nos últimos 10 anos de acordo com os censos de 2000 e 2010.

Além disso é uma cidade típica do interior do estado de Minas Gerais, uma vez que a maioria do comércio está no centro da cidade e as residências em bairros.

A Figura 37 mostra a localização do município.

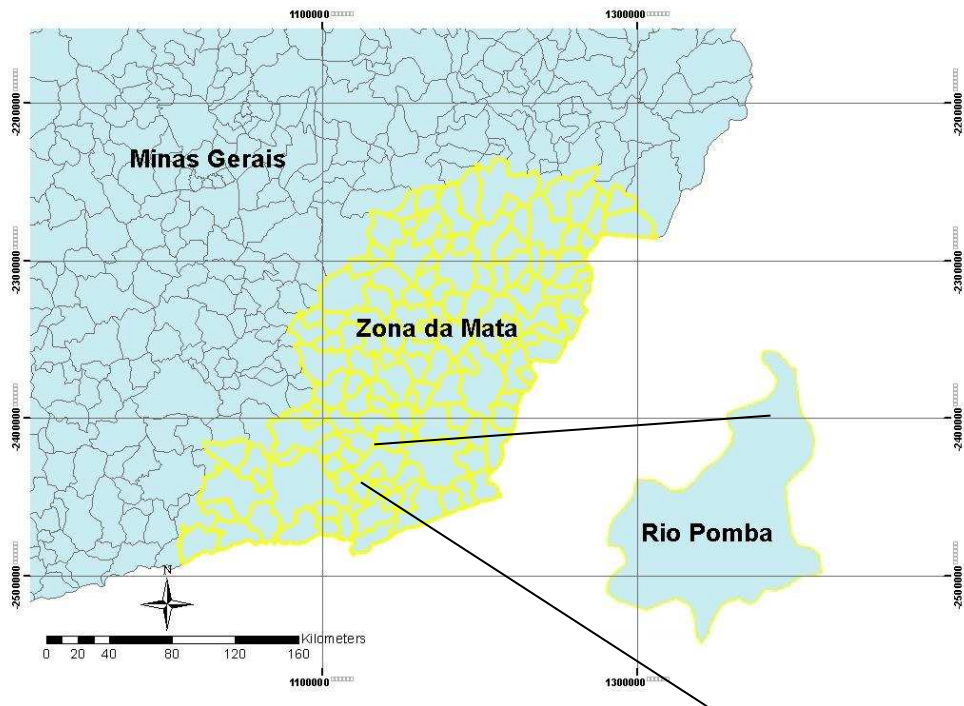


Figura 37 - Área de estudos: cidade de Rio Pomba, situada na Zona da Mata de Minas Gerais.

Fonte: Adaptado de IBGE (2011).

4.4.2 Métodos

A Figura 38 apresenta um resumo das atividades desenvolvidas na pesquisa considerando a simulação de ocorrências policiais e as associações de ocorrências.

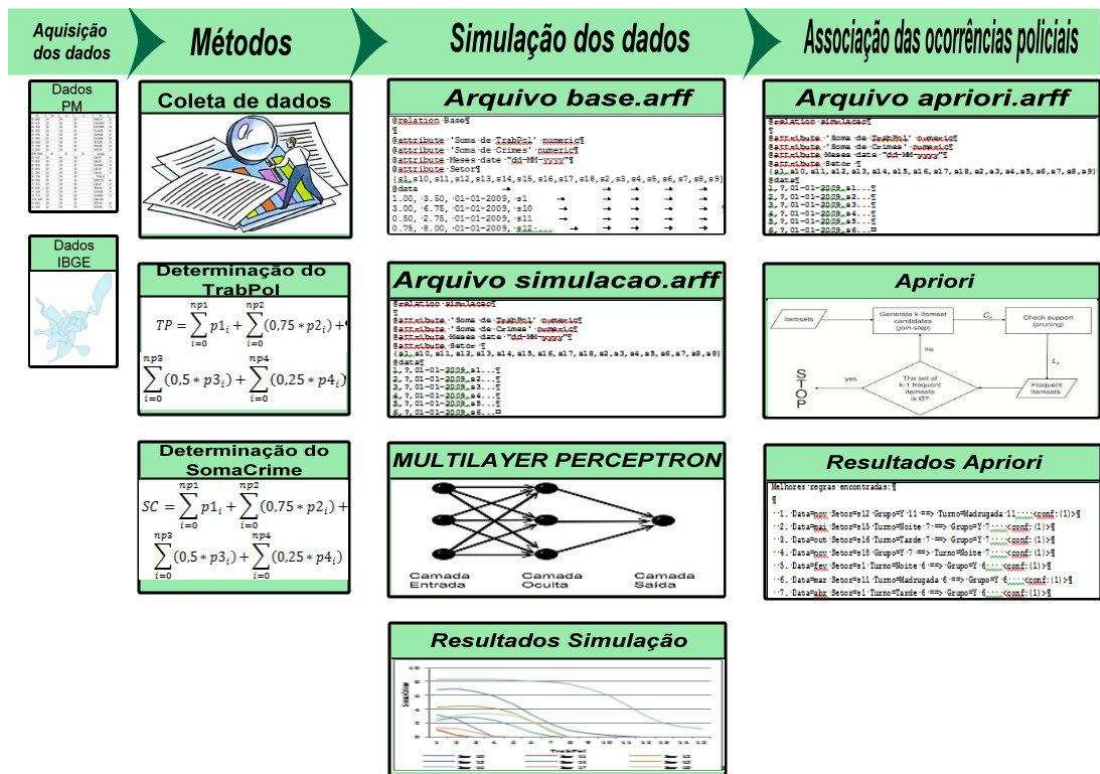


Figura 38 – Fluxograma das atividades desenvolvidas

— **Coleta de dados**

Os dados de ocorrências policiais foram disponibilizados pelo quartel da polícia militar da cidade de Rio Pomba, em formato de planilhas sem qualquer tipo de tratamento.

Assim, foi realizada a etapa de pré-processamento nos dados de forma a adequá-los, em termos de sua integridade, para o uso das ferramentas de predição e associação de crimes.

A escolha dos atributos da planilha se deu de acordo com os dados relevantes as implementações dos algoritmos. Assim, ficaram definidos os seguintes dados:

- Natureza da ocorrência – que pode ser definida como a ocorrência propriamente dita, como, por exemplo, um homicídio é definido por um determinado código. Este código é a natureza daquela ocorrência. Além disso, várias naturezas pertencem a um grupo de naturezas. Um exemplo disto seria o grupo de ocorrências do tipo B possui todos os crimes relacionados com a pessoa (homicídio, lesão corporal, etc).

- Data da ocorrência – refere-se à data que aconteceu a ocorrência.
- Prioridade da ocorrência – A prioridade da ocorrência é o fator que determina a ordem de atendimento da polícia. Ela se divide em 4 níveis. Ocorrências de prioridade 1 é a mais alta e ocorrências de prioridade 4 é a mais baixa.
- Setor da cidade – refere-se ao setor da cidade onde foi cadastrada a ocorrência.
- Grupo de ocorrência – é uma generalização das naturezas de ocorrências, como, por exemplo, todas os crimes contra a pessoa estão grupo B, todos os crimes contra o patrimônio estão no grupo C.
- Turno da ocorrência – é o turno que aconteceu a ocorrência. É dividido em 4 turno de 6 horas cada.
- “TrabPol” - que seria as ocorrências de trabalho da polícia, como, por exemplo, uma operação taxi, que fiscaliza todos os taxis da cidade por meio de blitz (ocorrências de operações do grupo Y). Este tipo de dado é o investimento que a polícia faz como tentativa de combater os índices de criminalidade.
- “SomaCrime” - que seriam as ocorrências criminais do grupo B e C (homicídio, roubo, etc). Esse tipo de dado seria um valor de criminalidade de cada setor da cidade.

Na etapa de pré-processamento, que de acordo com Agrawal *et.al.* (1993) inclui as seguintes funções: seleção de dados, limpeza dos dados, codificação dos dados e enriquecimento dos dados.

A função de seleção dos dados visa à seleção dos dados que devem ser efetivamente considerados durante o processo. A função de limpeza dos dados busca garantir a veracidade e integridade, e a certificação de completude dos dados.

Já na função de codificação dos dados, deve-se transformar os dados em formatos que possam ser compreendidos por softwares e algoritmos que serão utilizados no processo.

E, por fim, a função de enriquecimento dos dados que visa obter informações a serem agregadas aos registros existentes, a fim de enriquecer os dados e facilitar o reconhecimento de novos padrões.

A etapa de análise dos dados é fundamental importância para o sucesso no processo da mineração dos dados, visto que, é nela que se minimiza os dados sem relevância para a determinação da informação que se deseja obter.

— **Determinação dos atributos “TrabPol e SomaCrime”**

Para realizar a predição de ocorrências futuras considerando os dados dos anos de 2009, 2010 e 2011, foram definidas as variáveis “TrabPol e SomaCrime”, que representam as ocorrências geradas com o trabalho policial e ocorrências geradas por criminalidades, respectivamente.

Vale ressaltar que as duas variáveis foram definidas de acordo com a prioridade de cada natureza de crime, dando maior peso aos crimes mais hediondos. As prioridades variam de 1 a 4 sendo a prioridade 1 a mais importante e a prioridade 4 a menos importante, em termos de combate.

As variáveis “TrabPol (TP) e SomaCrime (SC)” foram normalizadas de acordo com as prioridades, variando de 0,25 a 1 (Equação 17 e 18).

$$TP = \sum_{i=0}^{np1} p1_i + \sum_{i=0}^{np2} (0,75 * p2_i) + \sum_{i=0}^{np3} (0,5 * p3_i) + \sum_{i=0}^{np4} (0,25 * p4_i) \quad (17)$$

$$SC = \sum_{i=0}^{np1} p1_i + \sum_{i=0}^{np2} (0,75 * p2_i) + \sum_{i=0}^{np3} (0,5 * p3_i) + \sum_{i=0}^{np4} (0,25 * p4_i) \quad (18)$$

Onde: “ $p1_i$ ” é o número de ocorrências com prioridade igual a 1 para cada numero de ocorrência (i), o “ $p2_i$ ” é o número de ocorrências com prioridade igual a 2 para cada numero de ocorrência (i), “ $p3_i$ ” o número de ocorrências com prioridade igual a 3 para cada numero de ocorrência (i), “ $p4_i$ ” número de ocorrências com prioridade igual a 4 para cada numero de ocorrência (i) e “ $np1$ ”, “ $np2$ ”, “ $np3$ ” e “ $np4$ ” são os números máximos de ocorrências de acordo com as prioridades 1,2,3,4 respectivamente.

Dessa forma, foi criada uma tabela dinâmica (com campos calculados de valores) de atributos com o somatório do trabalho realizado pela polícia em um mês de um determinado ano em um determinado setor censitário da

cidade e também o somatório de criminalidade gerada em um determinado mês de um ano e de um setor censitário.

4.5 Simulação dos dados

Para realização da simulação dos dados, foi utilizado o *framework Weka* versão 3.7.8 (Hall, *et al.*, 2009) que possui algoritmos que podem ser aplicados com intuito de gerar informações a partir da manipulação de dados.

Como a pesquisa visa a simulação de ocorrências, tentando retratar o quanto que a polícia militar deve atuar nas ações de combate e até mesmo nas ações de prevenção, para que a soma de índices de criminalidade seja mínima. Para isto utilizou-se uma classificação, que foi gerada a partir do *framework Weka* por meio do algoritmo classificador *Multilayer Perceptron*, que traça um histórico baseado em anos, meses ou dias e realiza a simulação de ocorrências de acordo com os parâmetros determinados. A Figura 39 apresenta a interface do *Weka*.

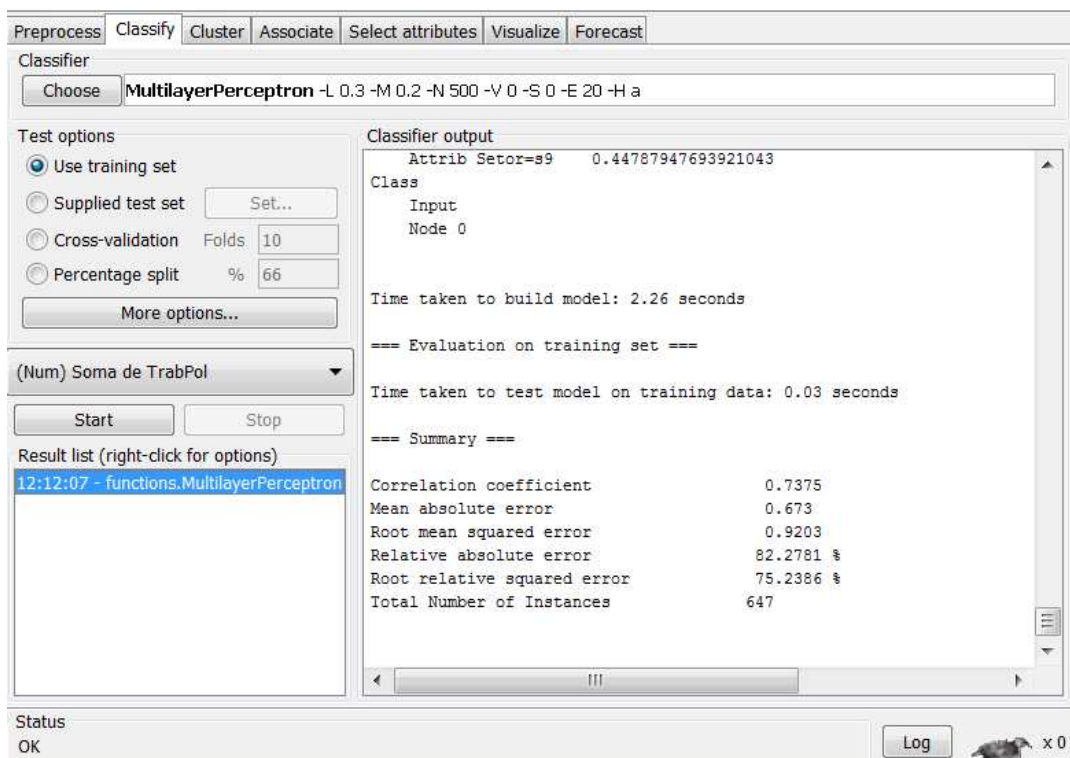


Figura 39 - Interface *Weka* aba de classificação com algoritmo *Multilayer Perceptron*.

Para utilizar a aba *classify* no *Weka*, primeiramente é necessário ativar a aba *preprocess* para carregar o arquivo objeto de estudo. Após, realizado o carregamento, já na aba *classify*, tem-se a escolha do algoritmo a ser utilizado como classificador.

A classificação dos atributos TP e SC foi realizada por meio dos arquivos *base2009.arff*, *base2010.arff* e *base2011.arff*, *simulacao2009.arff*, *simulacao2010.arff* e *simulacao2011.arff*. Os arquivos *base2009.arff*, *base2010.arff* e *base2011.arff* representam os dados reais obtidos junto a polícia militar do município de Rio Pomba. Já os arquivos *simulacao2009.arff*, *simulacao2010.arff* e *simulacao2011.arff* representam os registros a serem simulados pelo algoritmo.

Dessa forma, para que o algoritmo obtenha o conhecimento para a classificação dos dados, primeiramente ocorre uma varredura na base real e após aprender o comportamento dos dados, associa-se tal aprendizado com as instancias a serem classificadas no arquivo de simulação.

Os arquivos *base2009.arff*, *base2010.arff* e *base2011.arff* possuem essa configuração, como ilustrado na Figura 40.

```
@relation Base

@attribute 'Soma de TrabPol' numeric
@attribute 'Soma de Crimes' numeric
@attribute Meses date "dd-MM-yyyy"
@attribute Setor
{s1,s10,s11,s12,s13,s14,s15,s16,s17,s18,s2,s3,s4,s5,s6,s7,s8,s9}
@data
1.00, 3.50, 01-01-2009, s1
3.00, 6.75, 01-01-2009, s10
0.50, 2.75, 01-01-2009, s11
0.75, 8.00, 01-01-2009, s12 ...
```

Figura 40 - Arquivo de base para classificação no formato *.arff*.

Tais arquivos de base possuem 4 atributos como podem ser vistos na Figura 40. São eles a Soma de TrabPol do tipo numérico, Soma de Crimes do tipo numérico, Meses do tipo data, e Setor do tipo texto. A Figura 41 apresenta um dos arquivos de simulação utilizado na pesquisa.

```

@relation simulacao

@attribute 'Soma de TrabPol' numeric
@attribute 'Soma de Crimes' numeric
@attribute Meses date "dd-MM-yyyy"
@attribute Setor
{s1,s10,s11,s12,s13,s14,s15,s16,s17,s18,s2,s3,s4,s5,s6,s7,s8,s9}
@data
1,?,01-01-2009,s1...
2,?,01-01-2009,s2...
3,?,01-01-2009,s3...
4,?,01-01-2009,s4...
5,?,01-01-2009,s5...
6,?,01-01-2009,s6...

```

Figura 41 - Arquivo de simulação para classificação no formato .arff.

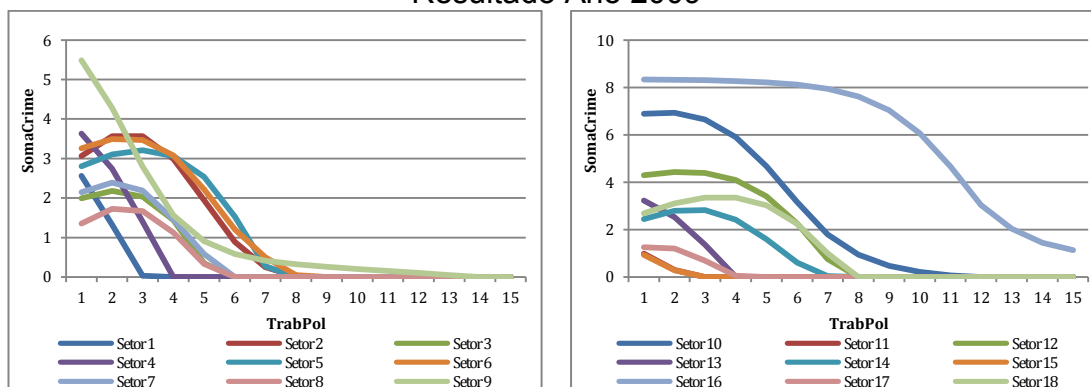
Os arquivos de simulação mantêm a mesma estrutura dos arquivos base com a diferença na variação do atributo Soma de TrabPol e um símbolo de interrogação no atributo Soma de Crimes. O atributo Soma de TrabPol varia de 1 a 15 em cada setor e em cada mês do ano. Dessa forma, os resultados da variação de Soma de TrabPol, que são os atributos que serão preditos, ou seja, os atributos que receberam o símbolo de interrogação (Soma de Crimes), serão mostrados por setor e por mês.

Para uma melhor visualização dos resultados, foi calculada a média de cada setor em determinado ano.

4.5.1 Resultados Obtidos

Os resultados da simulação dos dados obtidos através do algoritmo *Multilayer Perceptron*, foram apresentados por meio dos gráficos 4, 5 e 6. Tais gráficos representam os resultados dos anos de 2009, 2010 e 2011. Nestes experimentos visou-se prever o comportamento dos índices de criminalidade em cada setor em função da variação das ações da polícia. O Gráfico 4 apresenta os resultados para o ano de 2009.

Resultado Ano 2009



(a)

(b)

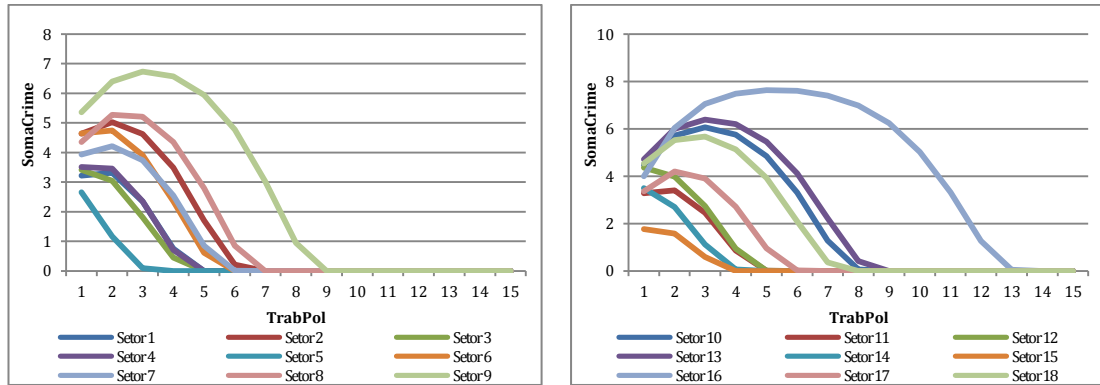
Gráfico 4 – Os gráficos representam a relação entre TP (Trabalho da Polícia) e SC (Soma de Crimes) para o ano de 2009: (a) com os setores de 1 a 9 e (b) com os setores de 10 a 18.

O Gráfico 4 foi dividido em (a) e (b) com a finalidade de visualizar os resultados com maior clareza. Verifica-se que todos os setores apresentaram resultados semelhantes, ou seja, a um determinado investimento da polícia o índice de criminalidade vai para próximo de 0. Nota-se que a partir de uma determinada ação da polícia, os índices de criminalidade começam a diminuir, corroborando com os dados reais, ou seja, quanto maior o trabalho da polícia, menor serão os índices de criminalidade.

No Gráfico 4 percebe-se dois casos distintos, ou seja, o setor 15 que possui um mínimo de investimento de polícia e o setor 16, que possui o maior investimento da polícia. No setor 16, mesmo simulando os dados até o trabalho máximo, este se apresenta para o ano de 2009 maior do que zero. É um setor que a polícia deve acompanhar com maior efetividade.

Além disso, o Gráfico 4 apresenta um trabalho da polícia tido como ideal de acordo com os seus dados históricos, como, por exemplo, o setor 1 que possui o trabalho da polícia ideal de aproximadamente 2,5.

Resultado Ano 2010



(a)

(b)

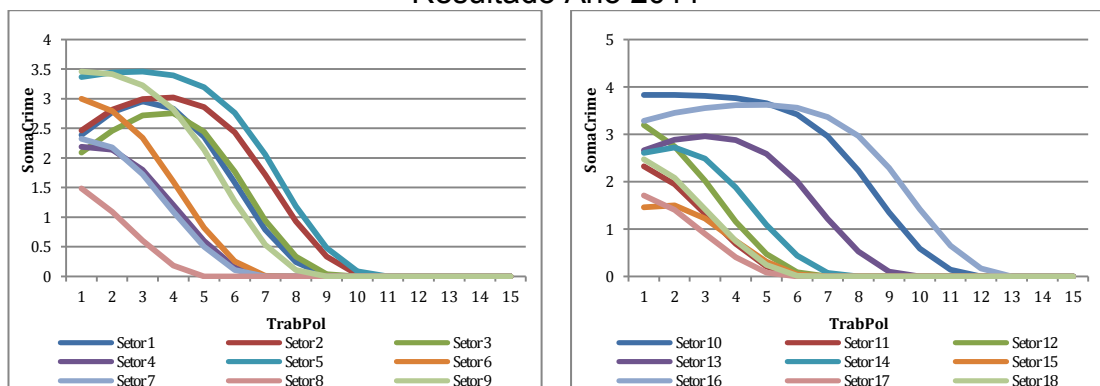
Gráfico 5 - Os gráficos representam a relação entre TP (Trabalho da Polícia) e SC (Soma de Crimes) para o ano de 2010: (a) com os setores de 1 a 9 e (b) com os setores de 10 a 18.

O Gráfico 5 mostra uma redução no atributo Soma de Crimes em relação ao ano de 2009. Além disso, os dados apresentados em 2010, corroboram com os dados de 2009, ou seja, a partir de um determinado trabalho da polícia, os índices de criminalidade começam a cair.

No setor 16, houve uma redução no trabalho máximo da polícia, que anteriormente não chegava a zero com investimento máximo de um número de 15 ações, agora com um investimento de aproximadamente 13 ações, torna este setor praticamente sem criminalidade.

Outro setor em que houve uma mudança significativa foi o setor 9 que em 2009 teve uma redução brusca de Soma de Crimes com trabalhos de ação da polícia entre 1 e 4, e no ano de 2010 houve um aumento no valor de Soma de Crimes para o mesmo intervalo.

Resultado Ano 2011



(a)

(b)

Gráfico 6 - Os gráficos representam a relação entre TP (Trabalho da Polícia) e SC (Soma de Crimes) para o ano de 2011: (a) com os setores de 1 a 9 e (b) com os setores de 10 a 18.

Pode-se verificar que o comportamento dos dados apresentados no Gráfico 6 possuem o mesmo padrão apresentado nos gráficos 4 e 5, porém com uma redução de índices de criminalidade considerável, com relação aos outros anos citados.

Com relação aos setores, destacam-se os de número 16 e 10 com maior exigência de investimento da polícia e o setor 15 com menor investimento da polícia.

4.6 Extração de Padrões utilizando *Apriori*

Para extrair padrões de crimes, foi utilizado o algoritmo *Apriori*, que faz a associação dos registros de uma tabela, encontrando a melhor regra para esses dados. Esse algoritmo é de grande importância para gerar informações relevantes que, às vezes, ficam escondidas na base de dados com muitos registros.

— A Base de Instâncias

Como pode ser visto no subcapítulo 4.5, foram utilizados os atributos da tabela de ocorrências policiais para os anos de 2009, 2010 e 2011. Assim, a escolha dos atributos para compor as regras é de vital importância para o trabalho de ação da polícia.

Os atributos foram relacionados de forma espaço-temporal. São eles: Data – representando o mês da ocorrência, sem se preocupar com o ano e dia em que a mesma ocorreu; Setor – que indica uma relação de espaço, compondo todos os setores censitários urbanos do município de Rio Pomba; Turno – que também traz a relação de tempo, sendo definido o dia em 4 turnos; e por fim, o atributo Grupo, que representa o tipo de ocorrência registrada. A Figura 42 mostra parte da base de instâncias no formato arff.

```

@relation Apriori

@attribute Data {jan,fev,mar,abr,mai,jun,jul,ago,set,out,nov,dez}
@attribute Setor
{s1,s10,s11,s12,s13,s14,s15,s16,s17,s18,s2,s3,s4,s5,s6,s7,s8,s9}
@attribute Turno {Manhã,Tarde,Noite,Madrugada}
@attribute Grupo {B,C,Y}

@data

jan,s16,Manhã,B
jan,s16,Manhã,B
jan,s10,Noite,B
jan,s10,Manhã,B
jan,s12,Tarde,B
jan,s9,Tarde,Y
fev,s15,Noite,B
fev,s2,Tarde,C
fev,s16,Tarde,Y
fev,s6,Tarde,Y
mar,s9,Noite,C
mar,s14,Tarde,Y
mar,s16,Tarde,Y
mar,s2,Tarde,Y
mar,s16,Noite,Y
mar,s16,Noite,Y
mar,s15,Madrugada,Y
abr,s12,Noite,B
abr,s4,Noite,C
abr,s16,Tarde,Y
dez,s16,Madrugada,Y
dez,s4,Tarde,C
dez,s14,Noite,C
dez,s14,Madrugada,C

```

Figura 42 - Instâncias representando parte do banco de dados de ocorrências de crimes.

4.6.1 Resultados Obtidos

Para a análise dos resultados procurou-se selecionar as regras de associação de maior interesse para o estudo de caso. Cabe ressaltar que as regras geradas são resultantes de um conjunto de atributos relevantes, cuja confiança mínima foi definida como 0,9 e o limiar de suporte é de 0,001. O número que precede o símbolo "=>" indica o apoio da regra. O próximo número que aparece na regra (após o símbolo), é o número dos itens válidos do próximo grupo de atributos relacionados com o primeiro grupo (antes de "=>"). Entre parênteses é a confiança da regra. A Figura 43 apresenta as regras geradas por meio do algoritmo de associação *Apriori*.

Melhores regras encontradas:

1. Data=nov Setor=s12 Grupo=Y 11 ==> Turno=Madrugada 11 <conf:(1)>
2. Data=mai Setor=s15 Turno=Noite 7 ==> Grupo=Y 7 <conf:(1)>
3. Data=out Setor=s16 Turno=Tarde 7 ==> Grupo=Y 7 <conf:(1)>
4. Data=nov Setor=s18 Grupo=Y 7 ==> Turno=Noite 7 <conf:(1)>
5. Data=fev Setor=s1 Turno=Noite 6 ==> Grupo=Y 6 <conf:(1)>
6. Data=mar Setor=s11 Turno=Madrugada 6 ==> Grupo=Y 6 <conf:(1)>
7. Data=abr Setor=s1 Turno=Tarde 6 ==> Grupo=Y 6 <conf:(1)>
8. Data=jun Setor=s15 Turno=Noite 6 ==> Grupo=Y 6 <conf:(1)>
9. Data=jul Setor=s8 Turno=Noite 6 ==> Grupo=B 6 <conf:(1)>
10. Data=jan Setor=s13 Turno=Madrugada 5 ==> Grupo=Y 5 <conf:(1)>
11. Data=fev Setor=s8 Grupo=Y 5 ==> Turno=Madrugada 5 <conf:(1)>
12. Data=mai Setor=s5 Turno=Noite 5 ==> Grupo=Y 5 <conf:(1)>
13. Data=jun Setor=s17 Turno=Madrugada 5 ==> Grupo=Y 5 <conf:(1)>
14. Data=out Setor=s10 Turno=Manhã 5 ==> Grupo=C 5 <conf:(1)>
15. Data=nov Setor=s3 Turno=Noite 5 ==> Grupo=C 5 <conf:(1)>
16. Data=dez Setor=s16 Turno=Noite 19 ==> Grupo=Y 18 <conf:(0.95)>
17. Data=out Setor=s9 Grupo=Y 14 ==> Turno=Madrugada 13 <conf:(0.93)>
18. Data=abr Setor=s10 Grupo=Y 12 ==> Turno=Madrugada 11 <conf:(0.92)>
19. Data=ago Setor=s8 Turno=Madrugada 11 ==> Grupo=Y 10 <conf:(0.91)>
20. Data=nov Setor=s9 Grupo=Y 11 ==> Turno=Madrugada 10 <conf:(0.91)>
21. Data=mar Setor=s7 Turno=Madrugada 10 ==> Grupo=Y 9 <conf:(0.9)>
22. Data=mai Setor=s8 Grupo=Y 10 ==> Turno=Madrugada 9 <conf:(0.9)>
23. Data=jun Setor=s1 Turno=Noite 10 ==> Grupo=Y 9 <conf:(0.9)>

Figura 43 - Regras extraídas do banco de dados de ocorrências de crimes através do algoritmo *Apriori*.

Pelos resultados apresentados pelo algoritmo *Apriori*, pode-se perceber que existem muitas operações da polícia em diversos setores, como, por exemplo, a regra:

- “Data = nov Setor = s12 Grupo = Y 11 ==> Turno = Madrugada 11 conf:(1)” que indica que no mês de novembro, no setor de número 12 ocorreram 11 operações da polícia e todas elas no turno da madrugada. Outra regra interessante é a regra de número 9:
- “Data = jul Setor = s8 Turno = Noite 6 ==> Grupo = B 6 conf:(1)” que indica que no mês de julho, no setor 8, no turno da noite aconteceram 6 ocorrências contra pessoas.

Pelos resultados apresentados, pode-se perceber padrões em determinados tipos de ocorrências, que não seriam facilmente perceptíveis em uma grande massa de dados de ocorrências. Dessa forma, estas regras geradas podem auxiliar a polícia militar a praticar operações em determinados turnos, meses e setores, afim de combater de forma mais eficiente a criminalidade da cidade.

4.7 Conclusões

Nesta pesquisa visou-se um estudo computacional com o apoio de informações espaciais para auxiliar no combate à criminalidade.

A predição de atividades de ação da policia militar, e a consequente minimização dos índices de ocorrência, frente as essas ações são de grande importância para uma melhor distribuição de investimentos da policia em cada setor da cidade.

As regras de associação de ocorrências policiais podem gerar grandes vantagens aos órgãos de segurança pública, uma vez que mostram regras de associações entre tipos de ocorrências que muitas vezes não tem como percebê-las no dia a dia, ou seja, certas informações, às vezes, passam despercebidas pelos órgãos de segurança pública.

Referências Bibliográficas

AGRAWAL, R.; IMIELIŃSKI, T.; SWAMI, A. Mining association rules between sets of items in large databases. **SIGMOD Rec.**, jun. 1993. v. 22, n. 2, p. 207–216.

AGRAWAL, R. *et al.* Fast algorithms for mining association rules. In:**Proc. 20th Int. Conf. Very Large Data Bases, VLDB.** 1994. p. 487-499.

BIGUS, J. P. **Data mining with neural networks: solving business problems from application development to decision support.** [S.l.]: McGraw-Hill, Inc., 1996.

CHANEY, S.; TOMPSON, L.; UHLIG, S. The utility of hotspot mapping for predicting spatial patterns of crime. **Security Journal**, 2008. v. 21, n. 1, p. 4–28. . Acesso em: 24 out. 2013.

COHEN, L. E.; FELSON, M. Social change and crime rate trends: A routine activity approach. **American sociological review**, 1979. p. 588–608. . Acesso em: 25 out. 2013.

ESRI. **ArcGIS desktop. Release 10.** [S.l.]: Environmental Systems Research Institute Redlands, California, 2011.

GOMES, M. A. A. DE F. Cultura urbanística e contribuição modernista: Brasil, anos 1930 - 1960. **Cadernos PPG-AU/UFBA**, 20 set. 2007. v. 4, n. 0. Disponível em: <<http://www.portalseer.ufba.br/index.php/ppgau/article/view/1691>>. Acesso em: 24 out. 2013.

GROFF, E.; BIRKS, D. Simulating crime prevention strategies: A look at the possibilities. **Policing**, 2008. v. 2, n. 2, p. 175–184. . Acesso em: 25 out. 2013.

GROFF, E.; MCCORD, E. S. The role of neighborhood parks as crime generators. **Security Journal**, 2011. v. 25, n. 1, p. 1–24. . Acesso em: 25 out. 2013.

GROFF, E. R. “Situating”Simulation to Model Human Spatio-Temporal Interactions: An Example Using Crime Events. **Transactions in GIS**, 2007. v. 11, n. 4, p. 507–530. . Acesso em: 25 out. 2013.

GROFF, E. R. Adding the temporal and spatial aspects of routine activities: A further test of routine activity theory. **Security Journal**, 2008. v. 21, n. 1, p. 95–116. . Acesso em: 25 out. 2013.

HALL, M. *et al.* The WEKA data mining software: an update. **ACM SIGKDD Explorations Newsletter**, 2009. v. 11, n. 1, p. 10–18. . Acesso em: 28 out. 2013.

LIU, L.; ECK, J. **Artificial Crime Analysis Systems: using computer simulations and geographic information systems**. [S.l.]: Information Science Reference, 2008.

PITONI, Rafael Moreira. Mineração de regras de associação nos canais de informação do direto. **Trabalho de Diplomação. Universidade Federal do Rio Grande do Sul. Porto Alegre**, 2002.

RATCLIFFE, J. Crime Mapping: Spatial and Temporal Challenges. *In*: PIQUERO, A. R.; WEISBURD, D. (Org.). **Handbook of Quantitative Criminology**. [S.l.]: Springer New York, 2010, p. 5–24.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, 1958. v. 65, n. 6, p. 386. . Acesso em: 25 out. 2013.

SEDS. **Estatísticas de Criminalidade Violenta – Janeiro a Novembro de 2012**. Belo Horizonte: Secretaria de Estado de Defesa Social, 2012. . Acesso em: 24 out. 2013.

TOBLER, W. R. A computer movie simulating urban growth in the Detroit region. **Economic geography**, 1970. v. 46, p. 234–240. . Acesso em: 28 out. 2013.

TOWNSLEY, M.; BIRKS, D. J. Building better crime simulations: systematic replication and the introduction of incremental complexity. **Journal of Experimental Criminology**, 2008. v. 4, n. 3, p. 309–333. . Acesso em: 25 out. 2013.

WEISBURD, D. *et al.* **TRAJECTORIES OF CRIME AT PLACES: A LONGITUDINAL STUDY OF STREET SEGMENTS IN THE CITY OF**

SEATTLE*. **Criminology**, 2004. v. 42, n. 2, p. 283–322. . Acesso em: 25 out. 2013.

WEISBURD, D.; MORRIS, N. A.; GROFF, E. R. Hot spots of juvenile crime: a longitudinal study of arrest incidents at street segments in Seattle, Washington. **Journal of Quantitative Criminology**, 2009. v. 25, n. 4, p. 443–467. . Acesso em: 25 out. 2013.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical machine learning tools and techniques**. [S.l.]: Morgan Kaufmann, 2005.

5 CONSIDERAÇÕES FINAIS

Com a realização da pesquisa, pode-se verificar que:

Um sistema de gerenciamento de banco de dados geográfico pode garantir a integridade dos dados que serão gravados e evita e/ou minimiza problemas relacionados à redundância, que é um fato frequente verificado nas planilhas advindas do quartel policial da cidade de Rio Pomba.

Através de algoritmos de previsões de séries temporais e análises estatísticas espaciais, conclui-se que na maioria dos setores da cidade, o número de ocorrências policiais diminuiriam. A percepção espacial da polícia militar, combatendo as áreas de maiores ocorrências, através de rondas, palestras socioeducativas, programas sociais, pode ser um dos fatores para redução da criminalidade na cidade de Rio Pomba.

Vale ressaltar que os resultados estatísticos são de vital importância para um novo campo de análise para a Polícia Militar, mostrando informações muito relevantes que não poderiam ser percebidas, sem considerar o contexto espacial.

Utilizando redes neurais, pôde-se perceber que a simulação de ocorrências policiais é de grande importância para uma melhor distribuição de investimentos da polícia em cada setor da cidade.

Utilizando o algoritmo *Apriori*, pôde-se gerar regras de associação de ocorrências policiais apresentando grandes vantagens aos órgãos de segurança pública, uma vez que mostram regras de associações entre tipos de ocorrências que muitas vezes não têm como ser percebidas no dia a dia.

Ou seja, certas informações, às vezes, passam despercebidas pelos órgãos de segurança pública por estarem implícitas dentro de um banco de dados.

Dessa forma, pode-se concluir que a pesquisa alcançou os objetivos e resultados significativos para a segurança pública.

Sugere-se, para trabalhos futuros, a integração dos três módulos apresentados na pesquisa, em um único software, tendo desta forma uma interface mais amigável, uma vez que o usuário não precisaria de utilizar vários módulos aplicativos separadamente, contribuindo com as análises por possibilitar a visualização dos resultados em conjunto.

Com relação à determinação da predição das ocorrências policiais, recomenda-se em estudos futuros a utilização, para análise temporal, de um menor intervalo para as medições de ocorrências, como, quinzenal ou mensal.

Recomenda-se também, que o processo de simulação para o investimento da Polícia Militar seja automatizado, a fim de minimizar o trabalho operacional.

Por fim, recomenda-se uma simulação multiagente de atividades rotineiras visando compreender melhor a dinâmica das cidades.

REFERÊNCIAS BIBLIOGRÁFICAS

EBDON, D. **Statistics in geography**. [S.l.]: B. Blackwell (Oxford Oxfordshire and New York, NY, USA), 1985.

GOMES, M. A. A. DE F. Cultura urbanística e contribuição modernista: Brasil, anos 1930 - 1960. **Cadernos PPG-AU/UFBA**, 20 set. 2007. v. 4, n. 0. Disponível em: <<http://www.portalseer.ufba.br/index.php/ppgau/article/view/1691>>. Acesso em: 24 out. 2013.

SEDS. **Estatísticas de Criminalidade Violenta – Janeiro a Novembro de 2012**. Belo Horizonte: Secretaria de Estado de Defesa Social, 2012. Disponível em: <https://www.seds.mg.gov.br/images/seds_docs/estatistnovo/12-12-18%20divulgacao%20mensal%20dados%20-%20novembro.pdf>. Acesso em: 24 out. 2013.