



**UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO  
UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA  
COMPUTAÇÃO**



**FRANCISCO PAULO DE FREITAS NETO**

**BUSINESS INTELLIGENCE APLICADA NO APOIO À  
TOMADA DE DECISÕES EM SEGURANÇA PÚBLICA**

**MOSSORÓ – RN**

**2014**

**FRANCISCO PAULO DE FREITAS NETO**

**BUSINESS INTELLIGENCE APLICADA NO APOIO À  
TOMADA DE DECISÕES EM SEGURANÇA PÚBLICA**

Dissertação apresentada ao Mestrado de Ciência da Computação – associação ampla entre a Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semiárido, para a obtenção do título de Mestre em Ciência da Computação.

Orientadora: Profa. Dra. Angélica Félix de Castro – UFERSA.

Co-orientador: Francisco das Chagas Lima Junior - UERN

**MOSSORÓ – RN**

**2014**

**O conteúdo desta obra é de inteira responsabilidade de seus autores**

**Dados Internacionais de Catalogação na Publicação (CIP)  
Biblioteca Central Orlando Teixeira (BCOT)  
Setor de Informação e Referência**

F866b Freitas Neto, Francisco Paulo.

Business intelligence aplicada na tomada de decisões em segurança pública / Francisco Paulo Freitas Neto. -- Mossoró, 2014.

91f.: il.

Orientadora: Prof<sup>a</sup>. Dra. Angélica Félix de Castro.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal Rural do Semi-Árido. Pró-Reitoria de Graduação.

1. Análise de dados. 2. Inteligência de negócios. 3. Segurança pública. I. Título.

RN/UFERSA/BCOT /185-14

CDD: 343.8

Bibliotecária: Vanessa Christiane Alves de Souza Borba  
CRB-15/452

**FRANCISCO PAULO DE FREITAS NETO**

**BUSINESS INTELLIGENCE APLICADA NO APOIO À  
TOMADA DE DECISÕES EM SEGURANÇA PÚBLICA**

Dissertação apresentada ao Mestrado em Ciência  
da Computação para a obtenção do título de Mestre  
em Ciência da Computação.

APROVADA EM: \_\_\_ / \_\_\_ / \_\_\_\_.

**BANCA EXAMINADORA**

---

Prof. Dra. Angélica Félix de Castro – UFERSA  
Presidente

---

Prof. Dr. Marcelino Pereira dos Santos Silva – UERN  
Primeiro Membro

---

Prof. Dr. Paulo Cesar Moura da Silva – UFERSA  
Segundo Membro

---

Prof. Dra. Marcia Jacyntha Nunes Rodrigues Lucena – UFRN  
Membro Externo

## **DEDICATÓRIA**

A todos que contribuíram direta ou indiretamente para a realização deste trabalho. Em especial à minha família pelo total apoio durante o período de realização do estudo.

## AGRADECIMENTOS

A Deus, pois sem ele não seria possível alcançar qualquer objetivo;

À professora orientadora, Profa. Dra. Angélica Félix de Castro, que me acompanha desde a graduação, pelo apoio, amizade e pela cobrança que me auxiliou a concluir o presente trabalho;

Ao coorientador Prof. Dr. Francisco das Chagas de Lima Júnior, por se dispor a auxiliar no desenvolvimento do estudo.

Aos professores da banca, que se dispuseram não somente a avaliar, mas também a contribuir para a melhora do presente trabalho;

À minha família, que desde cedo incentivaram os estudos, mostrando que este é o melhor caminho a ser trilhado. Em especial à minha mãe Edilma, meu pai Adjenilson e meus irmãos Gustavo e Alexandre;

À minha namorada Ana Paula Lima de Oliveira, pelo companheirismo nos momentos difíceis durante a realização do trabalho;

Aos amigos da cidade de Apodi, pelos momentos de descontração, que tornaram essa caminhada menos árdua, e por entenderem as várias “faltas” durante o período. Em especial a Raoni Freitas, Duann Aires, Paulo Pinto, Marcos Paulo, Marcos Gurgel, Lindemberg Rodrigues (o Grande), Jaquesson William e Junior Carvalho;

Aos colegas da graduação, dos quais muitos também foram colegas de mestrado, e que foram fundamentais em diversos momentos da minha vida acadêmica. Em especial a: Jarbas Filho, Roberto Alves, Danilo Câmara, Victor Hugo e Adriana Mara;

Aos colegas de república, com os quais convivi por vários anos, nos tornando praticamente irmãos, em especial a: Lyneker Moraes, Porlhan Hiphollyson, José Rodrigues e Rildo Denizier;

Ao CIOSP-RN, pela confiança ao ceder seus dados para a realização dessa pesquisa, e pelo apoio na realização desse trabalho. Em especial ao Major Macedo (Supervisor do CIOSP-RN), Tenente Jácome (Gerente do CIOSP-Mossoró), Soldado Andrey (Gerente do Banco de Dados do CIOSP-Mossoró) e Soldado Fernando;

Ao Programa de Pós-Graduação em Ciência da computação (PPgCC – UERN/UFERSA) e à CAPES, pela bolsa de mestrado concedida.

## EPÍGRAFE

Por que caminhos você vai e volta? Aonde você nunca vai? Em que esquinas você nunca para? A que horas você nunca sai? Há quanto tempo você sente medo? Quantos amigos você já perdeu? Entrincheirado, vivendo em segredo e ainda diz que não é problema seu.

Herbert Vianna

## RESUMO

O presente documento apresenta a ferramenta CIOSP-BI, desenvolvida para validar a aplicação da técnica de *Business Intelligence* (BI) em dados provenientes de ferramentas CAD (*Computer Aided Dispatch*), utilizadas para apoiar o atendimento às chamadas de emergência em Centros Integrados. A ferramenta foi desenvolvida em parceria entre o Programa de Pós-Graduação em Ciência da Computação (PPgCC – UERN/UFERSA) e o Centro Integrado de Operações em Segurança Pública (CIOSP). As informações utilizadas no desenvolvimento e testes são originárias do município de Mossoró, localizado na região Oeste do estado do Rio Grande do Norte (RN). No texto é discutida a problemática da violência a nível nacional, estadual e municipal, além de fundamentada a teoria que apoia o software, em especial *BI e Data Warehouse*. São também descritas cada uma das funcionalidades da aplicação, bem como seus mecanismos de proteção e otimização utilizados, além da sua real utilidade ao Centro. O objetivo principal do trabalho é validar a aplicabilidade da técnica de BI em órgãos públicos, em especial em órgãos de segurança pública, que apresentam grandes dificuldades nos dias atuais. Além disso, a ferramenta pode vir a ser utilizada no dia a dia da corporação como um software de apoio à tomada de decisões. Para isso, foram utilizadas, na sua maioria, bibliotecas e softwares de apoio gratuitos visando não gerar altos custos.

**Palavras-Chave:** Inteligência de Negócios, Análise de Dados, Segurança Pública

## ABSTRACT

This document presents the CIOSP-BI, tool developed to validate the technique of Business Intelligence (BI) on data from tools CAD (Computer Aided Dispatch), used to aims the service of assist emergency calls in Integrated Centers. The tool was developed in partnership between the Graduate Program in Computer Science (PPgCC) and the Integrated Centre of Operations in Public Safety (CIOSP). The information used is originating from the city of Mossoró, located in the western region of Rio Grande do Norte. In the text, the violence at national, state and local level is discussed, as well as grounded theory that supports the software, especially BI and Data Warehouse. Are also described each of the features of the application, as well their mechanisms of protection and optimization used beyond its actual usefulness to CIOSP. The main objective is validating the applicability of the technique of BI in public bodies, in particular the public safety organs, which presents difficulties today. Furthermore, the tool can subsequently be used in daily operations of the corporation as software to support decisions making. For that, were used free software library and support, aimed not generating high costs for the CIOSP.

**Keywords:** Business Intelligence, Data Analysis, Public Safety

## **LISTA DE TABELAS**

Tabela 1 – versões de business intelligence .....	34
---	----

## LISTA DE FIGURAS

Figura 1 – Variação da taxa de homicídios entre 1980 e 2010.....	19
Figura 2 – Taxas de homicídio no RN em 2000 e 2010 .....	21
Figura 3 – Esquematização de modelos para integração de dados.....	28
Figura 4 – Arquitetura típica de um sistema baseado em Business Intelligence.....	34
Figura 5 – Diagrama de Casos de Uso do CIOSP-BI.....	49
Figura 6 – Tela de login da aplicação CIOSP-BI .....	58
Figura 7 – Mensagem de saída da proteção contra SQL Injection .....	59
Figura 8 – Tela inicial da aplicação.....	62
Figura 9 – Gráficos da funcionalidade: Resumo das Ocorrências.....	64
Figura 10 – Gráficos da funcionalidade: Comparar Intervalos .....	68
Figura 11 – Exemplo de uso da funcionalidade Análise Espacial das Ocorrências .....	72

## LISTA DE SIGLAS

ALP – Preditor da Localização de Atratores

BI – *Business Intelligence*

CAD – *Computer Aided Dispatch*

CBM – Corpo de Bombeiros Militar

CIOSP – Centro Integrado de Operações de Segurança Pública

CriMM – Modelo do Movimento Criminal

CSU – Caso de Uso

DCDB – Descoberta de Conhecimento em Banco de Dados

DW – Data Warehouse

IBGE – Instituto Brasileiro de Geografia e Estatística

ITEP – Instituto Técnico Científico de Polícia

KDD – *Knowledge Discovery in Databases*

OLAP – *Online Analytical Processing*

OMS – Organização Mundial de Saúde

PC – Polícia Civil

PIB – Produto Interno Bruto

PM – Polícia Militar

RN – Rio Grande do Norte

SAMU – Serviço de Atendimento Médico de Urgência

SESED – Secretaria de Estado da Segurança Pública e da Defesa Social

SGBD – Sistema Gerenciador de Banco de Dados

SIG – Sistema de Informações Geográficas

SIOP – Sistema Integrado de Planejamento e Orçamento

SUS – Sistema Único de Saúde

SVM – *Support Vector Machine*

TI – Tecnologia da Informação

UERN – Universidade do Estado do Rio Grande do Norte

UFERSA – Universidade Federal Rural do Semi-Árido

UML – Linguagem de Modelagem Unificada

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>15</b>
1.1 CONTEXTUALIZAÇÃO.....	16
1.2 PROBLEMÁTICA.....	18
<b>1.2.1 Números da violência no Brasil.....</b>	<b>19</b>
<b>1.2.2 Os custos da violência no Brasil e a impunidade.....</b>	<b>22</b>
1.3 OBJETIVOS .....	23
1.4 ORGANIZAÇÃO DA DISSERTAÇÃO .....	24
<b>2 REVISÃO BIBLIOGRÁFICA .....</b>	<b>25</b>
2.1 MINERAÇÃO DE DADOS.....	25
2.2 DATA WAREHOUSE.....	27
2.3 INTELIGÊNCIA DE NEGÓCIOS ( <i>BUSINESS INTELLIGENCE</i> ).....	30
<b>2.3.1 Histórico da Inteligência de Negócios.....</b>	<b>32</b>
<b>2.3.2 Arquitetura básica de um sistema de Business Intelligence.....</b>	<b>34</b>
2.4 CRIMINOLOGIA AMBIENTAL.....	36
<b>3 TRABALHOS RELACIONADOS .....</b>	<b>39</b>
3.1 BI – INTELIGÊNCIA PARA O NEGÓCIO DE COMBATE AO CRIME.....	39
3.2 AK-MODES: UM ALGORITMO DE AGRUPAMENTO PONDERADO PARA BUSCA DE CASOS SEMELHANTES.....	40
3.3 ANALISANDO O TRAJETO DO CRIMINOSO AO CRIME: UM MODELO DO MOVIMENTO CRIMINAL .....	41
3.4 UM MODELO COMPUTACIONAL PARA PREDIÇÃO DA LOCALIZAÇÃO DE ATRADORES CRIMINAIS EM UMA REDE DE ESTRADAS .....	42
3.5 PREVISÃO DE CRIMES UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS .....	43
3.6 DETECÇÃO E MAPEAMENTO DE PONTOS QUENTES CRIMINAIS BASEADOS EM CLUSTERIZAÇÃO ORIENTADA AO MELHOR ATRIBUTO .....	44
<b>4 ASPECTOS METODOLÓGICOS .....</b>	<b>46</b>
4.1 DEFINIÇÃO DOS CASOS DE USO.....	48

4.2	MODELAGEM DOS CASOS DE USO .....	52
4.3	CARACTERÍSTICAS GERAIS DO CIOSP-BI .....	57
4.4	TECNOLOGIAS DA INTELIGÊNCIA DE NEGÓCIOS UTILIZADAS.....	60
<b>5</b>	<b>RESULTADOS E DISCUSSÕES .....</b>	<b>62</b>
5.1	RESUMO DAS OCORRÊNCIAS .....	63
5.2	COMPARAR INTERVALOS .....	66
5.3	ANÁLISE DO ATENDIMENTO .....	68
5.4	ANÁLISE ESPACIAL DAS OCORRÊNCIAS .....	70
<b>6</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>74</b>
	<b>REFERÊNCIAS .....</b>	<b>76</b>
	<b>ANEXOS .....</b>	<b>80</b>

## 1 INTRODUÇÃO

“A segurança pública está entre as maiores preocupações da sociedade brasileira nos dias atuais. Disputa, com a saúde e a educação, a prioridade na atenção de autoridades e imprensa.” Waiselfisz (2011, p. 5).

Qualquer cidadão que acompanha algum meio de comunicação, seja ele de qual natureza for, confirma a veracidade dessa informação. Diariamente é comum deparar-se com situações que denotam o crescimento de um problema tão sério que é a violência.

Em outra vertente tem-se o problema dos altos investimentos e custos gerados pela violência. Monteiro (2012) apresenta que, para o quadriênio de 2008-2011, foram alocados R\$ 23,92 bilhões e que incrementados por créditos adicionais resultaram em um montante de R\$ 43,70 bilhões aplicados em segurança pública.

Por outro lado, os custos da violência vão muito mais além, pois nesses valores não estão incluídos os gastos com saúde, previdência social, além dos prejuízos no setor privado, que também é diretamente afetado pela violência.

Desse modo, percebe-se que é necessária a criação de mecanismos que venham a amenizar a violência no Brasil, ou auxiliar no combate à mesma. Além disso, preferencialmente esses mecanismos ser de baixo custo, tendo em vista os inúmeros investimentos que o governo já realiza visando combater este mal.

Alguns desses mecanismos, que podem ser desenvolvidos, são ferramentas e técnicas computacionais que auxiliam a polícia em uma melhor execução das suas tarefas melhorando, assim, o uso dos recursos disponíveis e de certa forma reduzindo os custos operacionais.

O presente trabalho parte desses princípios e visa aplicar a técnica de *Business Intelligence* (BI) em dados provenientes de ferramentas CAD (*Computer Aided Dispatch*), que são programas que auxiliam no atendimento às ocorrências por órgãos de segurança pública.

Como estudo de caso, foi desenvolvida uma ferramenta baseada na técnica, em uma parceria entre o Centro Integrado de Operações de Segurança Pública do Rio Grande do Norte (CIOSP-RN) e o Programa de Pós-Graduação em Ciência da Computação (PPgCC), associação ampla entre Universidade Federal Rural do Semi-Árido (UFERSA) e Universidade do Estado do Rio Grande do Norte (UERN).

A ferramenta desenvolvida visa validar a técnica de BI nesse cenário, além de contribuir com uma melhor utilização dos recursos disponíveis no CIOSP, ao mesmo tempo

em que proporciona aos pesquisadores do Programa de Pós-Graduação o contato com um problema real enfrentado por diversas cidades brasileiras.

## 1.1 CONTEXTUALIZAÇÃO

Segundo Souza Neto (2007), há duas concepções principais quanto à Segurança Pública. A primeira possui um foco no combate ao crime em si, no embate entre polícia e criminalidade. Nela, as políticas de segurança pública são tidas praticamente como estratégias de guerra. Na segunda, a segurança é tida como um serviço a ser oferecido à população pelo Estado. Não existe a figura do inimigo, apenas o cidadão a ser servido.

Essa dicotomia se dá por não haver um texto que defina claramente o termo. A Constituição Federal (Brasil, 1988), trata no artigo 144 especificamente do tema Segurança Pública. Por outro lado, segundo Souza Neto (2007), no texto não há uma definição precisa sobre o termo, gerando muitas vezes conflitos conceituais.

Desse modo, no decorrer deste texto, a Segurança Pública será abordada como a junção das duas definições apresentadas, tendo a mesma como um serviço prestado à população pelas polícias e outros órgãos de segurança, mas sabendo que esse serviço envolve, além da prevenção, o combate à criminalidade.

Outro fator é em relação ao atendimento às ocorrências. A legislação brasileira (Brasil, 1988) dá poderes às polícias para atuarem no território nacional. Porém, não há definições exatas de como deve ser o atendimento às ocorrências da população prestado pelas mesmas, cabendo a cada corporação definir a forma que realiza esse atendimento. Por isso, o modo de atender as ocorrências podem apresentar diferenças entre uma corporação e outra.

Quanto ao meio de solicitar o atendimento desses órgãos, também ainda não existe um padrão definido na legislação, assim o mesmo varia entre cada órgão. Comumente o atendimento é solicitado por meio de telefone.

Não existe no Brasil um único número de atendimento às chamadas de emergências, mas sim, números distintos para cada órgão: 190 para polícia, 192 para serviço de atendimento à emergência, 193 para bombeiros, dentre outros.

Em alguns países, dentre eles os Estados Unidos, existem números únicos como o 911 que pode ser acionado em qualquer tipo de ocorrência, seja um incêndio, uma ocorrência médica, ocorrência policial, dentre outras.

Como isso não existe no Brasil, e visando criar mecanismos para integrar o atendimento às chamadas de emergência, estão sendo instalados em vários estados brasileiros, os Centros Integrados (Silva, 2004). Esses centros integram os serviços de atendimento às chamadas de emergências de diversos órgãos como Polícia Militar, Polícia Civil, Corpo de Bombeiros dentre outros.

Sapori e Andrade (2008) ressaltam um grande nível de aceitação da integração de serviços das polícias, tomando como referência a integração policial ocorrida no Estado de Minas Gerais. Segundo os autores, Sapori e Andrade (2008 *apud* Mendes 2007), 80% dos comandantes de companhias e delegacias afirmaram que a integração tem contribuído para a melhoria na gestão.

O presente trabalho foi realizado no estado do Rio Grande do Norte (RN). O RN desde 2010 também passou a integrar o atendimento às chamadas de emergência, instituindo no âmbito da Secretaria de Estado da Segurança Pública e da Defesa Social (SESED), o Centro Integrado de Operações de Segurança Pública (CIOSP) pela lei complementar número 442 (Rio Grande do Norte, 2010).

A lei complementar 442 institui o CIOSP no âmbito da SESED para atuar na região metropolitana da capital do RN, Natal. Posteriormente, o Centro foi também instalado na cidade de Mossoró e há planos de expansão para outras cidades interioranas.

O CIOSP integra os serviços dos seguintes órgãos: Polícia Militar (PM), Polícia Civil (PC), Corpo de Bombeiros Militar (CBM), Instituto Técnico-Científico de Polícia (ITEP), Serviço de Atendimento Médico de Urgência (SAMU), Guarda Municipal Metropolitana e Serviço de Fiscalização de Trânsito Municipal.

O centro funciona como um *call center* e tem por função utilizar o seu aparato tecnológico para coordenar o atendimento às ocorrências dos órgãos que o integram, na região ao qual está inserido. Para isso, dispõe de equipamentos funcionando em regime ininterrupto além de diversas ferramentas computacionais para auxiliar na realização das funções para as quais foi criado.

Dentre essas ferramentas utilizadas pelo CIOSP estão softwares CAD (*Computer Aided Dispatch*). Chan *et al* (2001) resalta dentre outros mecanismos computacionais, a importância da utilização de ferramentas CAD no atendimento às ocorrências, mostrando a sua contribuição para um atendimento mais eficiente.

Quando um cidadão realiza uma chamada de emergência para algum dos órgãos que integram o CIOSP-RN, todo o processo é auxiliado pela ferramenta CAD que foi implantada no centro, o iCAD. Essa ferramenta armazena informações referentes à ocorrência, como:

horário, localização, tipologia, localização das viaturas, despacho (envio de viaturas para atender a ocorrência) e conclusão. Desse modo, existe uma grande quantidade de informações no centro que caracterizam as ocorrências nas cidades onde o mesmo está presente.

Uma problemática citada em Hardoim (2007) é que apesar dos centros integrados disporem de um grande aparato tecnológico, o que gera uma grande quantidade de informações, muitas vezes essa tecnologia é subutilizada e a maioria das decisões é tomada empiricamente, o que nem sempre resulta na melhor situação, tendo em vista que normalmente existem informações implícitas em uma grande quantidade de dados.

Considerando a grande quantidade de dados gerados pelo CIOSP-RN, torna-se possível a utilização de técnicas de análise de dados que permitem a tomada de decisões baseada em dados, como a Inteligência de Negócios, do inglês *Business Intelligence* (BI), (Chaudhuri *et al*, 2011).

O estudo de caso do presente trabalho foi realizado com base em dados oriundos de uma ferramenta CAD utilizada pelo CIOSP. Foram utilizadas informações referentes ao município de Mossoró-RN, onde o Centro Integrado atua desde 2011.

## 1.2 PROBLEMÁTICA

Apesar de o CIOSP integrar diversos órgãos, que atuam em várias áreas, a problemática em foco no presente trabalho é a violência no Brasil, mais especificamente no município de Mossoró-RN. Porém, os resultados do trabalho podem ser úteis às áreas atendidas pelo Centro, tendo em vista que o mesmo se baseia no atendimento às ocorrências, e este é realizado igualmente em cada órgão.

Diversos estudos comprovam que a criminalidade está cada vez mais presente na vida do brasileiro. Diariamente é comum deparar-se com manchetes ou situações que denotam essa triste e alarmante realidade. As seções seguintes analisam alguns dados referentes à violência no Brasil.

### 1.2.1 Números da violência no Brasil

Conforme citado, a problemática da violência está se tornando cada vez mais evidente no Brasil. Waiselfisz (2011), com base em trinta anos de informações do Sistema Único de Saúde (SUS), demonstrou que entre 1980 e 2010, o número total de homicídios foi de 1.091.125. O estudo demonstra ainda que o número de homicídios cresceu de 13.910 em 1980 para 49.932 em 2010, um aumento de 259% no período.

Esse aumento poderia ser justificado pelo aumento populacional no período, mas, neste mesmo espaço de tempo, a população brasileira cresceu aproximadamente 63%, passando de 118.562.549 habitantes em 1980 para 193.252.604 em 2010 (projeção), segundo o Instituto Brasileiro de Geografia e Estatística (IBGE, 2008). A discrepância entre as taxas evidencia o grande crescimento do número de homicídios, tendo em vista que o crescimento desta foi muito superior ao da população.

Ainda segundo Waiselfisz (2011), a taxa de homicídios para cada 100 mil habitantes, parâmetro utilizado pela Organização Mundial de Saúde (OMS) para estimar a violência, mais que dobrou em trinta anos, passando de 11,7 em 1980 para 26,2 em 2010. Segundo a OMS uma localidade onde a taxa de homicídios para cada 100 mil habitantes é superior a 10, encontra-se em um estado epidêmico (Serrano-Berthet e Lopez, 2011).

Face ao exposto, percebe-se que o Brasil encontrava-se em 2010 com uma taxa próxima ao triplo do valor considerado epidêmico, sendo que esta não foi a maior taxa do período, como pode ser verificado na figura 1 que apresenta a variação das taxas de homicídios no Brasil no período de 1980 a 2010.

Figura 1 – Variação da taxa de homicídios entre 1980 e 2010



Fonte: Waiselfisz (2011)

Entre a população jovem, os dados são ainda mais alarmantes. Segundo Waiselfiz (2013), em 1980 a taxa de homicídios para cada 100 mil habitantes era de 17,2 na população jovem, já em 2011 esse número atingiu o valor de 53,4, um crescimento de 209,5%. Ainda segundo o autor, na população jovem os homicídios são responsáveis por 39,3% das mortes por causas externas, matando mais até que os acidentes de transportes. Enquanto isso, na população não jovem esse valor é de somente 3,1%, o que mostra uma grande diferença nos números.

Para um melhor dimensionamento desses números, Waiselfiz (2011) compara os homicídios no Brasil com vários conflitos armados que ocorreram em diversos países nas últimas décadas. Considerando-se os 30 anos de estudo (1980 a 2010), tem-se um total de 1.091.125 mortes, ou 36.371 mortes por ano. Enquanto conflitos como Chechênia/Rússia, Guerra do Golfo e Israel/Palestina tem-se respectivamente 25.000, 10.000 e 2.358 mortes por ano.

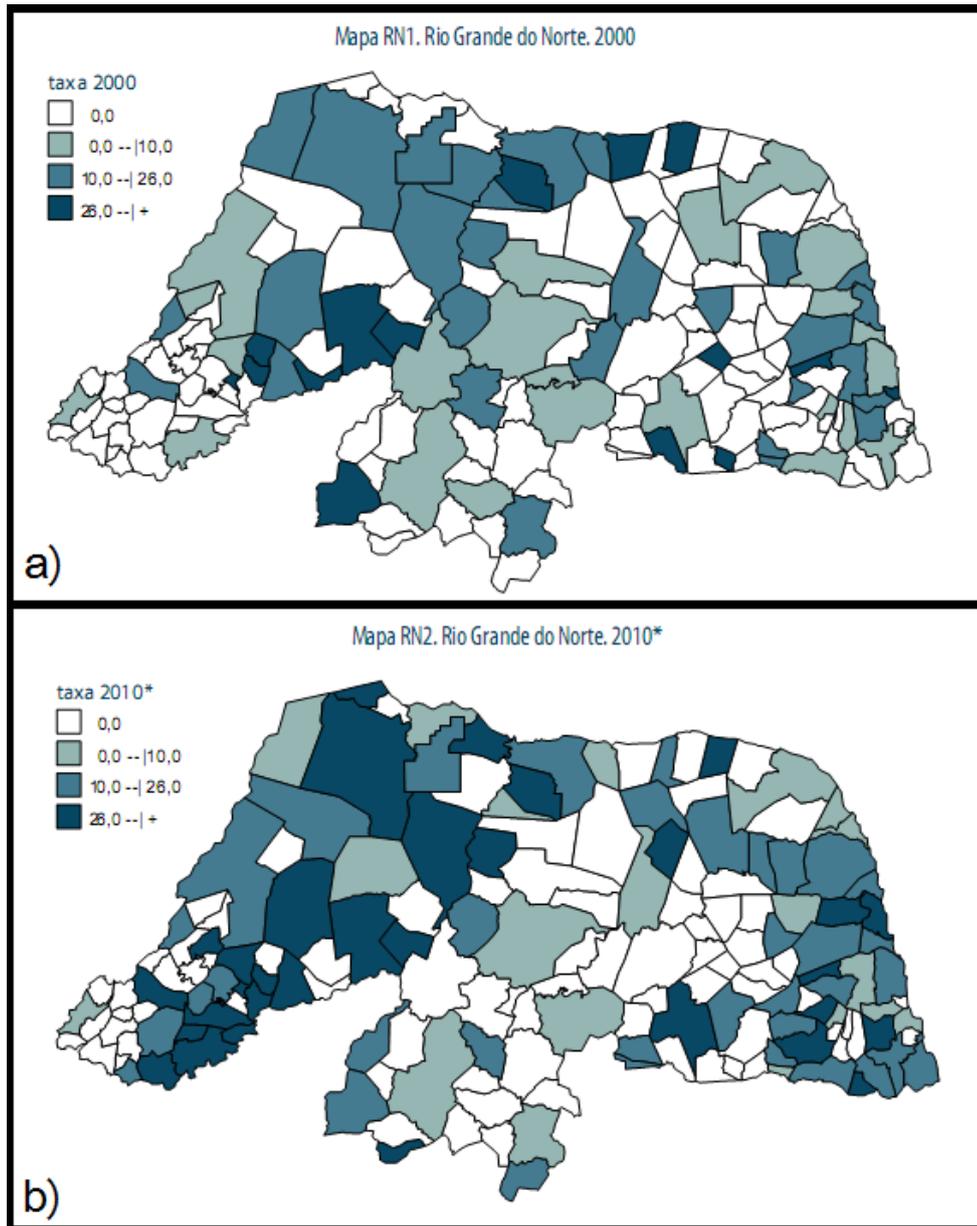
Waiselfiz (2011) ressalta ainda que não se pode atribuir os números de homicídios no Brasil às dimensões continentais do mesmo, pois até países como a Índia, com mais de 1,2 bilhão de habitantes e altos índices de pobreza, possui taxas menores.

Em nível estadual, o presente estudo foi desenvolvido no estado do Rio Grande do Norte, localizado na região Nordeste do Brasil e composto por 167 municípios. Em 2013 o Estado possui a população estimada em 3.373.959 habitantes (IBGE, 2010b). A economia norte rio-grandense baseia-se principalmente na exploração de petróleo em terra, produção de sal, exploração vegetal, dentre outros.

Quando se trata de homicídios, a realidade do estado é semelhante ao restante do país. Segundo Waiselfiz (2011), até o ano de 2002, o RN manteve-se abaixo da taxa de 10 homicídios para cada 100 mil habitantes. Por outro lado, em apenas oito anos, atingiu o patamar de 19,9.

Para uma maior percepção do crescimento da violência no Estado, a figura 2 apresenta dois mapas com a taxa de homicídios por município, o mapa a é relativo ao ano de 2000 e o b à 2010. As áreas mais escuras do mapa denotam uma taxa de homicídios maior que nas áreas claras, e é evidente o crescimento das áreas escuras no mapa b em relação ao mapa a.

Figura 2 – Taxas de homicídio no RN em 2000 e 2010



Fonte: Waiselfisz (2011)

Em nível municipal, os dados cedidos pelo CIOSP para a realização desse trabalho são referentes ao município de Mossoró. Localizada na região oeste do Estado do Rio Grande do Norte, a cidade possuía em 2013 uma população estimada em 280.314 habitantes (IBGE, 2010a). Nos últimos anos, a cidade vem tendo um crescimento acentuado decorrente das vagas de emprego nas áreas petrolífera, salineira, industrial e construção civil, além de possuir uma grande população flutuante devido às universidades que recebem alunos de vários municípios próximos.

Em se tratando de violência, os números na cidade são alarmantes. Segundo Waiselfisz (2011), nos anos de 2008, 2009 e 2010, foram registrados 126, 118 e 137 homicídios respectivamente. Atingindo a taxa de 50,7 homicídios para cada 100 mil habitantes, quase o dobro da média nacional, que já representa um valor epidêmico, como ressaltado anteriormente.

### **1.2.2 Os custos da violência no Brasil e a impunidade**

Outra face da problemática são os custos da violência. Segundo Monteiro (2012) o valor investido na Segurança Pública foi da ordem de R\$ 43,70 bilhões no quadriênio 2008-2011. Ainda no relatório de Monteiro (2012), entre 2007 e 2010, a dotação orçamentária nessa função quase dobrou, saltando de R\$ 5,97 bilhões em 2007 para R\$ 10,81 bilhões em 2010. Já no ano de 2011 houve uma pequena queda, resultando em R\$ 10,22 bilhões.

Os valores citados anteriormente tratam de investimentos na área de Segurança Pública e sequer podem ser visualizados como custos. Os custos propriamente ditos, gerados a partir da violência são bem maiores, talvez até incalculáveis (Cerqueira *et al*, 2007).

A violência causa mudanças de hábitos de consumo por parte da população, que muitas vezes deixa de adquirir um material com um custo maior para não atrair a atenção dos criminosos. A população também passa a evitar frequentar certas localidades, o que gera vários prejuízos tanto no setor público quanto no privado.

Segundo Cerqueira *et al* (2007 *apud* Cohen *et al*, 1994), os custos com a violência se dividem em duas categorias: custo social e custo externo. O primeiro é aquele que afeta o bem estar da sociedade, no momento em que as pessoas passam a ser privadas da sua liberdade em função da violência; já o segundo, é imposto a outras pessoas de forma involuntária, mas acarretam prejuízos a ela, como por exemplo, aos comerciantes de determinadas localidades que têm uma diminuição das suas vendas devido a violência.

Segundo Cerqueira *et al* (2007) não existe uma grande quantidade de trabalhos que analisam os custos da violência no Brasil, e há também diferenças metodológicas nesse cálculo. No estudo, utilizando uma metodologia própria, os autores estimaram que em 2004 a violência custou R\$ 92,202 bilhões ou 5,09% do Produto Interno Bruto (PIB) daquele ano. Destes, R\$ 31,889 bilhões foram custos arcados pelo setor público e R\$ 60,313 bilhões pelo setor privado.

Outro fator relevante a ser discutido é a impunidade no Brasil devido à ineficiência nas investigações, que contribuem ainda mais para o aumento da violência devido à impunidade. No texto de ENASP (2012), é ressaltado que o homicídio é o crime mais grave, pois é cometido contra o bem maior a ser protegido: a vida.

Por outro lado, ENASP (2012) ressalta também que o índice de solução de homicídios no Brasil é muito baixo, variando entre 5% a 8%. Enquanto que em outros países esse percentual é bastante superior: 65% nos Estados Unidos, 90% no Reino Unido e 80% na França.

Uma forma de evitar essa impunidade seria o flagrante, mas para isso, a polícia teria que possuir um tempo de resposta muito alto, o que é difícil nos dias de hoje mediante os vários problemas estruturais que as mesmas possuem. Entretanto, mecanismos computacionais podem auxiliar na melhoria desse tempo de resposta.

Números tão contundentes deixam claro o grave problema que o país se encontra, e sendo a Segurança Pública, segundo o artigo 144 da constituição federal (Brasil, 1988): “*dever do Estado, direito e responsabilidade de todos*”, torna-se notória a necessidade de cooperação entre todos os setores da sociedade para uma resposta efetiva a esse problema. E isso inclui as universidades, centros de pesquisa, dentre outros.

### 1.3 OBJETIVOS

O viés científico deste trabalho consiste no fato de apesar da técnica de *Business Intelligence* ser amplamente utilizada no mundo dos negócios, existirem poucos trabalhos que aplicam essa técnica nos serviços públicos, principalmente em Segurança Pública. Assim, o presente trabalho apresenta uma aplicação dessa técnica e visa incentivar a utilização dessa tecnologia na segurança pública. Além disso, são objetivos do presente trabalho:

- Objetivos Gerais:

- Validar a aplicação da técnica de BI em ferramentas CAD;
- Contribuir na melhora do atendimento do CIOSP-RN;
- Diminuir o tempo de criação dos relatórios;
- Aplicar a técnica de BI em segurança Pública.

- Objetivos Específicos:

- Desenvolver uma aplicação com base no banco de dados do CIOSP;
- Automatizar a produção de relatórios estatísticos do CIOSP;
- Identificar padrões de ocorrência nas localidades atendidas pelo CIOSP;
- Utilizar ferramentas livres de análise de dados.

#### 1.4 ORGANIZAÇÃO DA DISSERTAÇÃO

A presente dissertação está organizada da seguinte maneira: o capítulo 2 apresenta uma revisão bibliográfica acerca dos temas abordados no documento; o capítulo 3 apresenta trabalhos relacionados ao que foi desenvolvido; no capítulo 4 são apresentadas algumas funcionalidades da ferramenta desenvolvida, demonstrando onde a técnica de BI foi aplicada; o capítulo 5 apresenta resultados do trabalho bem como discussões acerca dos mesmos; e por fim, a seção 6 apresenta as considerações finais e as propostas de trabalhos futuros.

## 2 REVISÃO BIBLIOGRÁFICA

O presente capítulo apresenta os principais conceitos para o desenvolvimento da ferramenta CIOSP-BI que foi desenvolvida para validar o presente estudo. A principal técnica utilizada é a de Business Intelligence, porém, visando uma melhor compreensão da técnica, serão abordadas duas outras técnicas de que integram e apoiam a mesma: Mineração de Dados e *Data Warehouse*, além de algumas teorias de criminologia ambiental que reforçam a capacidade de sucesso do presente trabalho.

### 2.1 MINERAÇÃO DE DADOS

“*A procura por padrões em dados não é um fato recente*” (Witten *et al.* 2011). O homem sempre buscou encontrar padrões no mundo que o cerca, de modo a vir a auxiliá-lo na realização de suas tarefas. O pescador busca entender o comportamento dos peixes, os agricultores procuram padrões no crescimento das lavouras, os políticos buscam entender o padrão de voto dos eleitores. Dentre muitos outros exemplos.

A técnica de Mineração de Dados, ou seu termo em inglês, mais comumente encontrado na literatura: *Data Mining*, se baseia no mesmo princípio citado, porém tenta automatizar ou semiautomatizar o processo de descoberta de padrões válidos em bancos de dados (Witten *et al.*, 2011). Atualmente, a utilização da técnica vem crescendo devido principalmente à redução dos custos da aquisição de memória para armazenamento de dados por parte das empresas.

Kantardzic (2011) descreve a mineração de dados, reforçando a participação humana no processo, como sendo: “*a busca por informações novas, valiosas e não triviais em grandes volumes de dados. Sendo um esforço colaborativo entre humanos e computadores*”.

Witten *et al.* (2011) apresenta uma descrição semelhante, para os autores, mineração de dados é: “*a extração de conhecimento implícito, previamente desconhecido e potencialmente útil de uma base de dados*”. Os autores reforçam o desenvolvimento de programas de computadores que analisem automaticamente os bancos de dados buscando padrões nos mesmos.

Kantardzic (2011) define como objetivos principais da mineração, a predição e a descrição dos dados. Desse modo, surgem duas classificações da *Data Mining*: a Mineração de Dados Preditiva e a Descritiva. A Mineração de Dados Preditiva utiliza valores conhecidos para criar modelos que permitem estimar dados futuros, como por exemplo, a regressão que encontra funções que permitem estimar dados futuros; já a Mineração de Dados Descritiva, busca padrões válidos e não triviais para descrever os dados a serem interpretados por humanos, como por exemplo, a sumarização que facilita a compreensão das informações utilizando unidades estatísticas.

Ainda segundo Kantardzic (2011), as tarefas primárias da mineração de dados são: classificação, regressão, clusterização, sumarização, modelagem de dependências e detecção de mudanças e desvios.

A classificação é, segundo Witten *et al.* (2011), a criação de modelos com base em dados existentes que podem ser úteis para a realização de predições de dados futuros, além de representar o comportamento dos dados.

A regressão, segundo Hall *et al.* (2009), é a descoberta de uma função de aprendizagem preditiva que mapeia os dados. O autor reforça que a técnica é tida como um preditor de classes contínuas, tendo em vista que ela pode ser utilizada para obter dados futuros em classes que apresentem comportamento semelhante a uma função contínua.

Clusterização ou agrupamento, é a busca por dados semelhantes em bancos de dados e estes são agregados em grupos criados. A principal vantagem na utilização da técnica é a possibilidade de se analisar dados em conjunto e não somente individualmente (Kantardzic, 2011).

A sumarização é uma tarefa descritiva que visa apresentar as informações de maneira compacta, de forma que facilite a sua compreensão por parte dos usuários. Exemplos dessas formas são gráficos e unidades estatísticas como médias, medianas, modas, desvios padrão, dentre outros (Witten *et al.*, 2011). Essa foi a principal técnica utilizada durante o desenvolvimento do CIOSP-BI, devido ser a que mais se assemelha às técnicas utilizadas no dia a dia do Centro Integrado.

A modelagem de dependências, segundo Hall *et al.* (2009), é a descrição de dependências significativas entre dados, permitindo uma melhor visualização de conceitos como hierarquias. Existem alguns algoritmos que podem ser utilizados para realizar essa modelagem e facilitar a identificação de dependências entre os dados.

Já a detecção de mudanças e desvios, diz respeito a analisar mudanças significativas nas informações, permitindo identificar padrões de modificações nos dados. Com isso, é

possível entender melhor como os dados se comportam ou mesmo, criar mecanismos que venham a prever valores futuros (Kantardzic, 2011).

Em um ponto de vista mais amplo, tem-se, segundo Miller e Han (2009), que a mineração de dados é uma das etapas integrantes do processo de Descoberta de Conhecimento em Bancos de Dados (DCDB) ou seu termo original em inglês pelo qual é mais popular: *Knowledge Discovery in Databases* (KDD).

Por outro lado, tem-se que a descoberta de conhecimento é uma tarefa mais ampla e vai além do domínio da técnica de mineração de dados e da estatística, englobando também conceitos de: processamento de dados, aprendizagem de máquina, reconhecimento de padrões e visualização científica dos dados.

O KDD tem por função a descoberta de conhecimento relevante e não trivial em grandes bancos de dados e representa todo o processo definido desde a extração dos dados até a sua validação por um especialista ou por alguma técnica computacional (Miller e Han, 2009).

A mineração de dados é uma técnica muito poderosa, e devido seus resultados em experimentos científicos e no mundo dos negócios, tem os seus conceitos expandidos para outros domínios, permitindo a análise de dados não convencionais, como exemplo, a mineração de dados espaço-temporais.

Segundo Miller e Han (2009) a mineração de dados espaço-temporais apresenta uma série de particularidades devido à natureza destes dados. Dentre essas mudanças, as principais são relativas aos tipos de regras geradas, que possuem forma diferente das convencionais principalmente quando se trata de associação e clusterização destes dados.

A mineração de dados espaço-temporais é também muito aplicada no combate ao crime, permitindo a tentativa de reconhecimento de padrões criminais que levam em consideração a localidade, além de artifícios temporais destes eventos. Em outras palavras: modelando onde e quando os crimes estão ocorrendo.

## 2.2 DATA WAREHOUSE

Nagabhushana (2006) define *Data Warehouse* (DW) ou Armazém de Dados (tradução literal) como: "*Uma coleção de bancos de dados projetada para fornecer informações necessárias e apoiar as tomadas de decisão*". Para uma melhor concepção da definição de

DW, os autores Kozielski e Wrembel (2009) utilizam os conceitos de fontes de dados (*Data Sources*) e repositórios de dados (*Data Marts*).

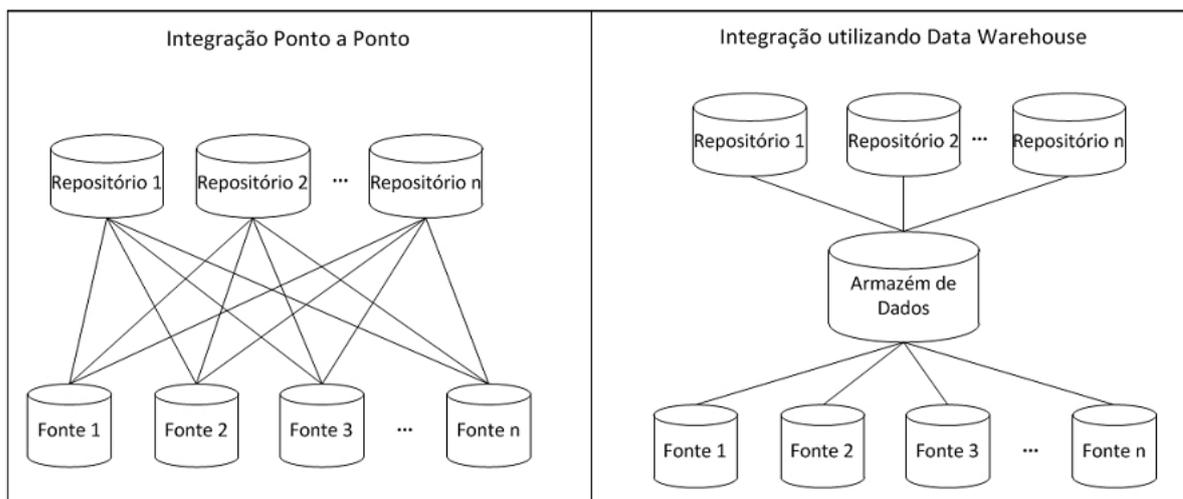
*Data Sources*, como o nome sugere, são entidades em que são gerados os dados. Pode ser: um banco de dados, uma planilha eletrônica, dentre outros. Já os *Data Marts* agregam informações sobre algum determinado tema, como por exemplo, uma empresa que possua um repositório para armazenar informações referentes à vendas, outro sobre estoque e outro para controle de pessoal.

Partindo desses conceitos, Kozielski e Wrembel (2009) apresentam dois modelos de integração entre fontes e repositórios que são a (I) Ponto a Ponto e a (II) integração usando *Data Warehouse*. Na integração ponto a ponto (I), que era utilizada antes do surgimento dos armazéns de dados, as fontes de dados eram integradas conforme a necessidade de maneira direta (*Ad-Hoc*), ou seja, partindo de um ponto de destino a um ponto alvo.

A principal dificuldade do modelo *Ad-Hoc* está relacionada à heterogeneidade das fontes de dados, ou seja, cada conexão tem que ser implementada de uma maneira diferente, pois as fontes normalmente não seguem um mesmo padrão para representar as informações.

Já a integração baseada em *Data Warehouse* (II) surgiu para tentar solucionar as deficiências da integração ponto a ponto. No modelo, os dados de determinada fonte são extraídos, transformados e inseridos no DW e após isso, são colocados nos repositórios de dados. A figura 3 ilustra os conceitos de integração de dados citados.

Figura 3 – Esquematização de modelos para integração de dados



Fonte: Kozielski e Wrembel (2009)

Uma das principais vantagens da utilização dos armazéns de dados, além da integração, encontra-se no fato que estes facilitam a análise dessas informações. Existem diversas técnicas de análise de dados que atuam em conjunto com os DW, dentre elas o Processo Analítico Online (*On-Line Analytical Processing – OLAP*).

A técnica datada do fim do ano de 1994 é muito utilizada até os dias atuais por gestores de conhecimento em empresas, principalmente em conjunto com a técnica de *Data Warehouse*, fazendo surgir a sigla DW+OLAP (*Data Warehouse+OLAP*).

Segundo Nagabhushana (2006) o termo OLAP não possui uma definição exata, porém, a maioria dos sistemas baseados na técnica tem características em comum. Assim, o autor utiliza um esquema de cinco palavras chave para descrever a técnica, que são: *Fast Analysis of Shared Multidimensional Information* (FASMI) que podem ser traduzidas literalmente como Análise Rápida de Informações Multidimensionais Compartilhadas.

O termo Rápido diz respeito ao sistema ter que responder rapidamente às consultas dos usuários. Não é usual o solicitante ter que esperar horas de processamento para receber os resultados das suas consultas. Existem estratégias que podem ser utilizadas para diminuir o tempo das consultas e o que demoraria horas pode vir a durar segundos.

A Análise está diretamente relacionada à finalidade de utilização da técnica de OLAP, o termo refere-se ao fato de que o sistema pode lidar com a lógica de negócios ao qual está inserido e também se refere à análise estatística que é relevante ao usuário.

O termo compartilhado é implícito à natureza dos armazéns de dados e implica que o sistema deve implementar requisitos de segurança que garantam a confidencialidade das informações, tendo em vista que estas estão compartilhadas com outros usuários e alguém mal intencionado poderia causar danos a esses dados.

Multidimensional é um dos termos chave do OLAP dada à natureza dos dados utilizada nos bancos de dados das empresas, o termo refere-se ao fato que o sistema deve fornecer uma visão multidimensional dos dados incluindo suporte a hierarquias e hierarquias múltiplas, que são muito comuns no mundo dos negócios.

Por fim, tem-se a Informação, que diz respeito a todos os dados e informações derivadas necessárias, as informações são extremamente úteis em OLAP, tendo em vista que são com base nelas que os sistemas realizam as suas atividades.

Além disso, não se pode falar de OLAP sem citar a Inteligência de Negócios. Um dos principais motivos para se desenvolver um DW e utilizar técnicas de análise de dados é, segundo Kozielski e Wrembel (2009), utilizar os resultados das análises para apoiar decisões futuras.

A obtenção de conhecimentos e inteligência que expliquem a lógica de funcionamento de determinado negócio pode prover vantagens incalculáveis para as empresas que utilizam essas técnicas. Provendo ganhos de tempo e economia de dinheiro. Essa é uma das premissas que permeiam o BI.

No presente trabalho não foi necessária a implantação de um Data Warehouse, nem a implementação de outros mecanismos de integração de dados, devido ao fato do Centro já contar com tecnologias dessa natureza, implementadas juntamente com a ferramenta CAD utilizada. Caso contrário seria necessário haver essa integração antes de iniciar o desenvolvimento da ferramenta de BI.

### 2.3 INTELIGÊNCIA DE NEGÓCIOS (*BUSINESS INTELLIGENCE*)

Chaudhuri *et al.* (2011) reforça o exposto em Witten *et al.* (2011) no que diz respeito aos fatores que levaram ao crescimento da utilização de técnicas de análise de dados por empresas, ao citar que um dos principais fatores que contribuíram para esse fato foi a redução no custo da aquisição de memória para armazenamento de dados.

A maioria das empresas hoje utiliza sistemas que geram diariamente uma grande quantidade de informações com granularidade fina, como transações de clientes no setor bancário, histórico de vendas, *e-mails*, *logs* de consultas em *sites*, análise de produtos, e até cliques em páginas da internet. Esse aumento dramático no volume, variedade e principalmente na velocidade com a qual as informações são geradas, fez surgir o conceito de *Big Data*.

McAfee e Brynjolfsson (2012) mostram que diariamente são gerados aproximadamente 2,5 *exabytes* de dados e esse número duplica a cada 40 meses. Estima-se, por exemplo, que só o site de vendas eletrônicas *Walmart* colete a cada hora, 2,5 *petabytes* de dados referentes a transações dos seus clientes.

Logo, é necessário o desenvolvimento de tecnologias capazes de realizar uma análise eficiente desses dados, tendo em vista que os mesmos possuem grande valor, já que representam as regras do negócio no qual estão inseridos.

No texto (McAfee e Brynjolfsson, 2012), a *Big Data* é tratada como sendo uma revolução no mundo corporativo, segundo os autores em pesquisa realizada em 330 empresas públicas americanas, apontou que em média as empresas que utilizam a tomada de decisões

orientada a dados são 5% mais produtivas e 6% mais rentáveis do que as que não utilizam. Existem técnicas que permitem a tomada de decisões baseadas em dados, dentre elas a Inteligência de Negócios.

Gopalkrishnan *et al.* (2012) informam que a análise de negócios encontra-se na interseção dos mundos da administração, ciência da computação e estatística. Segundo os autores, é grande o sucesso de muitas empresas ao realizar uma análise eficiente dos seus dados, fazendo com que as mesmas tenham de certa forma uma vantagem no mercado sobre as outras, e impulsionando cada vez mais as corporações a darem mais valor aos seus dados.

Segundo Chaudhuri *et al.* (2011) *Business Intelligence* é uma coleção de tecnologias de apoio à tomada de decisões que visa permitir aos gestores de determinada empresa, que tomem decisões melhores e mais rápidas com base no histórico de dados armazenados nesta.

Nas últimas duas décadas, a técnica vem tendo um largo crescimento principalmente no tocante ao número de produtos e serviços oferecidos e devido à adoção da mesma pela indústria. A expansão na área de estudo de negócios é tamanha, que em um estudo realizado com 1.200 profissionais de Tecnologia da Informação (TI), pesquisadores e estudantes de 13 países, o *IBM Trends Tech Report 2012* (IBM, 2012), ranqueou a análise de negócios como uma das quatro principais tecnologias do ano de 2011, juntamente com tecnologia móvel, computação em nuvem e *social business*.

Segundo Chen *et al.* (2012), 97% das empresas com receitas superiores a 100 milhões de dólares usam algum tipo de ferramenta para análise de negócios.

Por outro lado, a técnica de BI não é utilizada apenas no mundo dos negócios em si, existem algumas aplicações bem sucedidas da técnica em outras áreas como no combate ao crime, que serão abordados no capítulo 3, que trata sobre alguns trabalhos relacionados ao proposto.

Além disso, o próprio governo brasileiro utiliza algumas aplicações baseadas na técnica, só que voltadas à gestão. Dentre elas, existe o Sistema Integrado de Planejamento e Orçamento (SIOP) e há planos de implantação de alguns outros, como pode ser verificado no plano diretor de Tecnologia da Informação 2012-2014 (Brasil, 2011).

Diferente do que se possa imaginar, as técnicas citadas neste trabalho não excluem o componente humano na tomada de decisões, muito pelo contrário. Visa-se com a utilização dessas técnicas, apoiar os gestores no processo de gestão e não substituí-los (McAfee e Brynjolfsson, 2012). Os autores complementam dizendo que sempre será necessária a presença da figura do gestor que apoia, cobra, conversa, pensa criativamente e integra as equipes da empresa.

### 2.3.1 Histórico da Inteligência de Negócios

A técnica de BI vem se desenvolvendo ao longo dos anos, pois à medida que os novos desafios surgem, são criadas novas tecnologias para tentar solucioná-los. Czernicki (2009) e alguns outros autores defendem que a inteligência de negócios pode ser dividida em duas versões, mas outros autores como Chen *et al.* (2012) defendem a existência de uma terceira versão. As versões são classificadas em três perspectivas principais: o público alvo, aplicativo e arquitetura dos sistemas.

A primeira versão da tecnologia de BI, conhecida também como BI 1.0 teve início no final da década de 90, o principal fator que diferencia essa versão das outras é a interface. Czernicki (2009) mostra que as aplicações de BI nessa versão tinham interfaces mais complexas de se utilizar, dessa forma exigia que os usuários além de entender o sistema tinham também que entender os mecanismos computacionais por trás do mesmo.

No BI 1.0, os softwares não eram intuitivos, também não eram projetados de forma a abstrair as tarefas comuns dos usuários. Desse modo eram basicamente utilizados por profissionais de tecnologia da informação e os outros usuários da empresa praticamente só recebiam os resultados sem ter contato com os sistemas em si.

Quanto às aplicações propriamente ditas, os softwares da primeira versão eram desenvolvidos para *desktop*, instalados nas estações de trabalho e desenvolvidos em linguagens de alto nível como *Java*, *C++* e *Visual Basic*.

Segundo Chen *et al.* (2012) dentre os treze recursos essenciais para uma plataforma de *Business Intelligence* no relatório da *Gartner* (Sallan *et al.*, 2011), as aplicações de BI 1.0 apresentam nove desses recursos, que são: relatórios, painéis, consultas *ad hoc*, pesquisa baseada em BI, OLAP, visualização iterativa, tabela de indicadores, modelagem preditiva e mineração de dados.

Segundo Czernicki (2009) três fatores principais levaram à evolução de BI 1.0 para BI 2.0, que são: o advento da *WEB 2.0* que se refere principalmente a utilizar a internet como uma plataforma de *software*, que teve início por volta de 2004; metodologias ágeis de desenvolvimento, principalmente ao possibilitar um maior suporte a mudanças no sistema; *software* orientado a serviços, que diz respeito aos sistemas serem disponibilizados na forma de serviços, o que permite também que uma série de serviços independentes possam comunicar sua lógica de negócio com outros.

Uma das principais mudanças no BI 2.0 é a filosofia, Czernicki (2009) diz que a tecnologia na segunda geração tenta levar a técnica além dos profissionais com experiência em gestão de negócios para um público alvo maior. Na essência, o BI 2.0 é projetado para um usuário não técnico. As aplicações são simples com interfaces claras e de fácil compreensão.

Outro fator citado no texto Czernicki (2009), para o BI 2.0, é em relação à computação em nuvem, já que a técnica começa a também ser direcionada para esse tipo de dados. O artigo Ouf e Nasr (2011) trata o assunto de *Business Intelligence* aplicado na computação em nuvem de forma incremental, e propõe um modelo de desenvolvimento nessa arquitetura.

Chen *et al.* (2012) defende uma terceira geração de aplicações de inteligência de negócios. O BI 3.0, que está voltado para o mercado de aplicativos móveis, que se iniciou principalmente devido ao crescimento expressivo no número de aparelhos desse tipo.

Segundo artigo da revista *The Economist* (Economist, 2011), desde 2011 o número de *smartphones* e *tablets* já superam o número de *desktops* e *laptops*. Espera-se que em 2020 o número de aparelhos dotados de plataformas móveis atinja a marca de 10 bilhões de unidades.

Desse modo, para acompanhar as tendências de mercado, as aplicações de BI estão agora migrando também para os dispositivos móveis, o que permite uma maior praticidade no uso, tendo em vista que podem ser acessadas a qualquer hora e em qualquer lugar, seja durante uma reunião, no escritório, ou mesmo em casa.

Por outro lado, a terceira geração das aplicações de BI traz grandes desafios, principalmente no que diz respeito a poder de processamento, de armazenamento e segurança de dados nos dispositivos móveis.

A tabela 1 apresenta uma comparação entre as três versões de BI. A tabela é uma adaptação mostrada em Chen *et al.* (2012), e destaca os pontos principais de cada uma das versões do BI.

Tabela 1 – Versões de *Business Intelligence*

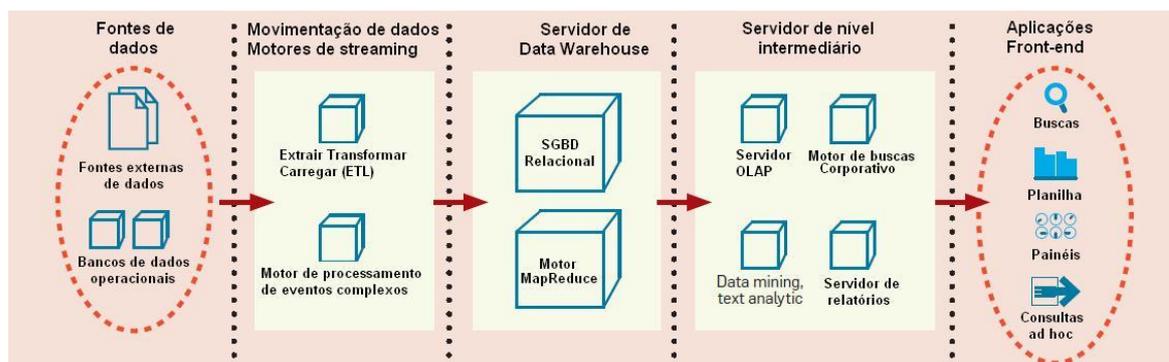
Versão	Características Chave
BI 1.0	Baseada em SGBD, Conteúdo Estruturado, Interfaces Complexas, Sumarização de Dados, Relatórios e Paineis
BI 2.0	Software Orientado a Serviços, WEB 2.0, Interfaces simples, Computação em nuvem
BI 3.0	Plataformas móveis, Recursos limitados

Fonte: Chen *et al.*(2012)

### 2.3.2 Arquitetura básica de um sistema de Business Intelligence

Segundo Chaudhuri *et al.* (2011), a arquitetura de um sistema de *Business Intelligence* é tipicamente semelhante à mostrada na figura 4. Nela é descrito de forma resumida todo o processo realizado desde a coleta dos dados até as aplicações de interface (*Front-end*).

Figura 4 – Arquitetura típica de um sistema baseado em *Business Intelligence*



Fonte: Chaudhuri *et al.* (2011)

Quanto às *Fontes de Dados*, tem-se que normalmente os dados de uma empresa têm origem em diferentes departamentos, são gerados por diversas pessoas, estão em vários formatos e com qualidades variáveis. Então, problemas como integração, limpeza, e padronização dos dados são alguns dos principais desafios em se utilizar técnicas de análise de dados.

Kozielski e Wrembel (2009) no seu primeiro capítulo abordam dentre outros assuntos, o problema da integração relacionado à *Data Warehousing*. No texto os autores citam que em

muitas empresas há diversos tipos de dados como: bancos de dados relacionais, dados textuais em documentos, *e-mails*, dados de GPS, dentre outros.

Em relação à *Movimentação dos Dados*, trata-se de uma das tarefas mais importantes em análise de dados, já que as bases de dados normalmente são incrementais. Ou seja, a quantidade de dados é sempre maior o que torna imprescindível a existência de mecanismos eficientes de manipulação de dados como consultas, dentre outros.

Esse problema torna-se maior ainda quando se trata de aplicações em tempo real, pois precisam responder às consultas rapidamente. Logo, ferramentas para realizar o extração-transformação-carregamento (*Extract Transform Load- ETL*) e manipulação de eventos complexos se tornam mais que um incremento, para virem a ser uma exigência para que esses sistemas venham a funcionar de maneira adequada. Bons mecanismos de carregamento de dados podem fazer o que era realizado em horas vir a ser feito em minutos ou segundos.

Segundo Kozielski e Wrembel (2009), uma das formas mais modernas de gestão atualmente utilizada em muitas empresas é basear as ações em conhecimento, que por sua vez pode ser obtido através da análise dos dados dessas empresas. A tecnologia que materializa esse conceito é *Data Warehouse* conceituada na subseção 2.2.

Esses armazéns integram dados de diferentes sistemas em um repositório central. Essa integração facilita, por exemplo, a construção de relatórios, históricos, dentre outros. Existem diversas soluções em *Data Warehousing* no mercado.

Segundo Chaudhuri *et al.* (2011), os servidores de *Data Warehouse* são complementados pelos servidores de nível intermediário. Para realizar suas funções, diversas tecnologias são utilizadas. O OLAP (*Online Analytic Processing*) expõe as visões multidimensionais dos dados para aplicativos ou usuários e permitem a realização de serviços comuns em BI, como filtragem, agregação, dentre outros.

Dentre os servidores de níveis intermediários têm-se diversas opções de utilização, dependendo da aplicação na qual a solução está inserida. São nesses servidores que o processamento de dados em si é realizado, aplicando-se algoritmos de mineração de dados ou outras técnicas. Dentre esses servidores têm-se os servidores de relatórios, motores de busca corporativo, mecanismos de mineração de dados e mecanismos de análise textual.

Os *Servidores de relatórios* permitem a criação eficiente de relatórios, como por exemplo, relatório de vendas de uma empresa em um mês, ou no caso da segurança pública, relatório de assassinatos em determinada região em um dado período, por exemplo.

*Motores de busca corporativos* permitem a realização de buscas por palavras chaves em documentos corporativos, como: *e-mails*, mensagens, documentos, dentre outros. No

presente trabalho não foram utilizadas tecnologias dessa natureza, tendo em vista que esse tipo de informação é sigiloso dentro de órgãos militares.

*Mecanismos de mineração de dados* permitem a realização de análises mais aprofundadas nos dados que vão além do conjunto *Data Warehouse+OLAP*, permitindo também a criação de modelos que podem ajudar a prever dados futuros.

*Mecanismos de análise textual* são capazes de com base em textos extraírem informações relevantes para a companhia, o que é de grande ajuda, pois se o trabalho fosse realizado manualmente seria além de lento, bastante desgastante.

Por fim, têm-se as aplicações *front-end* que são aquelas com as quais os usuários finais interagem diretamente. Pode-se citar como exemplo: mecanismo corporativo de buscas, aplicações de análise de desempenho, planilhas, painéis de indicadores, dentre outros.

Vale ressaltar que essa arquitetura não é uma regra, desse modo, nem todos os sistemas de inteligência de negócio seguem o modelo apresentado, pois a estrutura varia de acordo com a aplicação. E é possível encontrar sistemas que possuam algumas, todas ou até mais técnicas que as apresentadas.

Outro ponto fundamental é que essas aplicações sejam desenvolvidas com interfaces claras, simples e objetivas para auxiliar o usuário final na sua utilização. Principalmente quando se tem em vista que quem as utilizará normalmente serão os responsáveis pela gestão das empresas e baseará suas decisões nas saídas desses programas.

## 2.4 CRIMINOLOGIA AMBIENTAL

De acordo com Wortley e Mazerolle (2008) a criminologia ambiental é "*um conjunto de teorias que compartilham um interesse comum em eventos criminais e as circunstâncias imediatas em que ocorrem*".

Essas teorias tentam explicar os crimes como um conjunto de fatores composto por: infrator, vítima e as leis. Ou seja, elas analisam os padrões criminais e tentam explicar o que leva ao acontecimento dos crimes. São analisados fatores como: fatores biológicos, experiências de desenvolvimento ou forças sociais.

Segundo os autores Wortley e Mazerolle (2008): "*A ocorrência de crime é largamente entendida como uma expressão do desvio adquirido do agente, que pode ser uma função dos acontecimentos que ocorreram muitos anos antes*".

Ainda segundo Wortley e Mazerolle (2008), a perspectiva ambiental do crime é baseada em três premissas. A primeira diz que o comportamento criminal é influenciado pela natureza do meio ambiente em que ocorre. Assim, o ambiente interfere diretamente no crime, moldando o seu curso.

A segunda premissa informa que a distribuição espaço-temporal dos crimes não é aleatória, devido à ocorrência dos crimes dependerem de fatores situacionais, isso é visível quando se percebe que existem regiões e horários em que a ocorrência de crimes é maior que em outros.

Já a terceira premissa diz que havendo a compreensão do comportamento criminal obtém-se uma arma poderosa no combate a esses crimes, e é justamente nesta terceira premissa que o presente trabalho se baseia, pois o mesmo visa entender como os crimes ocorrem para um melhor combate aos mesmos.

Existem diversas teorias com base nessa perspectiva, dentre elas, Wortley e Mazerolle (2008) abordam quatro que são: a perspectiva da escolha racional, precipitadores situacionais de crime, abordagem de atividades rotineiras e a teoria padrão criminal.

A *Perspectiva da Escolha Racional*, ao invés de ver os crimes como o resultado de estáveis motivações criminais, vê o desejo, preferências e motivações. Em suma, o comportamento do criminoso é semelhante com o dos outros cidadãos, e os mesmos ao planejarem as suas decisões, ponderam os custos e os benefícios das mesmas.

Assim, ao optar por seguir o caminho do crime, a constante obtenção de sucesso na realização dos eventos criminais leva à intensificação deste comportamento por parte dos indivíduos, da mesma maneira que o fracasso leva a uma diminuição ou mesmo desistência das atividades criminosas, tendo em vista que benefícios já não são mais vantajosos em relação aos malefícios.

Na teoria da *Perspectiva da Escolha Racional*, o ambiente imediato é a fonte de informação que um indivíduo utiliza para decidir se quer ou não cometer um determinado crime.

Os *Precipitadores Situacionais de Crime* argumentam que a perspectiva da escolha racional apresenta apenas metade dos conceitos sobre o ambiente imediato. A teoria apresenta alguns incrementos relativos a fatores situacionais como uso de álcool ou entorpecentes além de fatores emocionais.

Fatores como pressão, estresse ou provocação são também levados em consideração nesta teoria, tendo em vista que nem todos os crimes são totalmente premeditados. Logo, para

a ocorrência de um crime primeiramente se tem os precipitadores seguidos da oportunidade, em caso da presença dos dois fatores é que o crime poderá acontecer.

A *Abordagem de Atividades Rotineiras* foi criada na década de 1970 e tem seu foco voltado para a situação dos crimes. A teoria afasta-se de explicar os crimes por razões sociais como pobreza ao citar o exemplo dos Estados Unidos depois da segunda guerra mundial, que iniciou uma fase de grande enriquecimento e ao mesmo tempo uma onda de violência tomou conta do país.

A teoria afirma que para a ocorrência de determinado crime é necessária a junção de três fatores: um criminoso motivado a cometer determinado crime, um alvo apropriado e a ausência de um guardião. O guardião não necessariamente é a figura do policial, pode ser um pai de uma criança, um sistema de segurança ou mesmo os moradores de determinada residência.

Já a *Teoria do Padrão Criminal*, aborda o crime como um evento complexo, mas assume que existem padrões de ocorrência dos mesmos. Ela explica que não existe fundamento em acreditar na aleatoriedade completa das ocorrências, tendo em vista que visivelmente percebe-se que determinadas regiões sofrem com uma maior quantidade de crimes que outras, determinado grupo de pessoas cometem mais crimes e que determinados horários são mais propícios à ocorrência de crimes. Assim, a teoria apresenta diversas regras para explicar o comportamento criminal individual e em grupos de indivíduos, além de regras espaço temporais.

Desse modo, a utilização de ferramentas computacionais para análise de dados de ocorrências, pode levar a polícia a entender como funciona o comportamento dos criminosos de determinada localidade, além do modo com o qual os mesmos agem.

Um fato a ser considerado é que apesar do CIOSP compreender vários órgãos, a maioria dos atendimentos realizados pelo Centro é acerca de ocorrências criminais, como será abordado no capítulo 4 que trata dos resultados. Sendo assim, o combate ao crime tem um enfoque maior neste trabalho, mas as técnicas podem ser úteis para todos os órgãos do Centro.

É interessante citar também que a análise de dados pode ser em alguns casos discriminatória com determinados grupos (Witten *et al.*, 2011). Logo, a interpretação dos resultados das análises de dados deve ser feita de maneira cautelosa e principalmente responsável. Essa preocupação também foi levada em consideração no desenvolvimento deste trabalho, blindando informações de vítimas, ofensores e localidades.

### 3 TRABALHOS RELACIONADOS

Conforme ressaltado anteriormente, a maioria dos artigos relacionados às técnicas de análise de dados, aborda principalmente o mundo dos negócios. Apenas uma pequena parcela relaciona-se à utilização dessas técnicas por órgãos públicos e dentre esses, uma parte está relacionada na aplicação dessas técnicas no combate ao crime.

A seguir são listados alguns trabalhos recentes utilizando técnicas de análise de dados no combate ao crime, visando mostrar o estado da arte dessa área. Embora somente alguns conceitos apresentados tenham sido diretamente utilizados no presente trabalho, os mesmos serviram para demonstrar exemplos de sucesso da aplicação de técnicas de análise de dados criminais.

Existem poucos trabalhos nacionais abordando o tema. A maior parte do trabalhos concentram-se em eventos internacionais, então, as buscas foram realizadas em bases de pesquisa internacionais.

#### 3.1 BI – INTELIGÊNCIA PARA O NEGÓCIO DE COMBATE AO CRIME

O trabalho apresentado por Warden (2010) descreve resumidamente algumas das utilidades do software utilizado pelo serviço de polícia da cidade de *Edmonton* no Canadá. A aplicação utiliza a técnica de *Business Intelligence* e permite aos utilizadores terem uma visão geral do negócio, que nesse caso é o combate à criminalidade.

Segundo o autor do texto, que é funcionário do departamento de polícia, o sistema auxilia os policiais ao gerar relatórios para responder a questões como: onde os crimes estão ocorrendo, quando estão ocorrendo, quem estão cometendo-os e quem são as vítimas. Permitindo assim, um maior planejamento das ações, além de proporcionar aos gestores indicadores para analisar os resultados das ações realizadas pela polícia.

No artigo é descrito uma ocorrência que demonstra como o software pode ser útil, a ferramenta indicou uma região onde estava havendo um crescimento no número de assaltos e arrombamentos a residências, então os comandantes resolveram enviar uma equipe para realizar uma averiguação na área em busca de obter pistas sobre os suspeitos. Ao chegar ao

local, um suspeito, ao avistar a viatura e os policiais, evadiu-se, vindo a ser perseguido e capturado, sendo depois reconhecido como quem efetuou os arrombamentos.

Obviamente essa ou qualquer outra ferramenta não informará o local e horário exato em que os crimes irão acontecer, mas só pelo fato de visualizarem-se as localidades em que estão sendo mais assoladas com a criminalidade já é suficiente para a polícia se organizar para combater o crime nessa localidade.

Apesar do texto não apresentar detalhes técnicos acerca do desenvolvimento da ferramenta, o mesmo serviu como uma das bases para o desenvolvimento deste trabalho, tendo em vista que demonstra que a técnica está em plena utilização por um renomado departamento de polícia.

### 3.2 AK-MODES: UM ALGORITMO DE AGRUPAMENTO PONDERADO PARA BUSCA DE CASOS SEMELHANTES

Ma *et al.* (2010) apresentam um algoritmo de clusterização de duas fases chamado *AK-Modes*, para buscar automaticamente subconjuntos de casos criminais semelhantes em grandes bases de dados. A finalidade de agrupar casos similares é a possibilidade de solucionar ambos de maneira conjunta, visto que um determinado crime pode gerar evidências para a solução de outros. Além disso, eles podem alertar da possibilidade de algum conjunto de crimes terem sido cometidos pela mesma pessoa ou por uma mesma facção criminosa.

O algoritmo é dividido em duas fases: na primeira, são atribuídos pesos a cada atributo proveniente da base de dados, através do conceito de ganho de informação. Esse conceito é muito importante na teoria da informação, e muito utilizado em árvores de decisão, visando nesse caso, refletir a relevância de um atributo.

A segunda fase é responsável pela clusterização com base nos resultados da fase anterior. O diferencial do *AK-Modes* para algoritmos mais tradicionais de agrupamento é a utilização de modos ao invés de médias, ao integrar a pesagem de atributos no processo.

O texto também apresenta os resultados de alguns experimentos realizados a fim de validar o algoritmo proposto. Nos primeiros testes foram comparados os resultados do algoritmo de clusterização sem a pesagem com o proposto no texto e os resultados foram satisfatórios, pois na maioria dos casos apresentou precisão maior ou semelhante à obtida com o algoritmo sem a utilização da atribuição de pesos aos atributos.

No segundo conjunto de testes, o algoritmo foi testado na base de dados do departamento de segurança pública de *Zhejiang*, província da China, com base em casos de onze cidades desde 2004. Os resultados foram validados com o auxílio de profissionais da área e foram satisfatórios tendo saídas como, por exemplo, em um dos grupos gerados onde 100 % dos casos tratavam de roubo de motocicletas e todos os roubos foram realizados pela ligação dos fios da ignição.

Com base nos resultados, os gestores da polícia podem intensificar as investigações nesse determinado tipo de crime, ficar em atenção ou alertar a população sobre a modalidade criminal.

### 3.3 ANALISANDO O TRAJETO DO CRIMINOSO AO CRIME: UM MODELO DO MOVIMENTO CRIMINAL

O artigo Iwanski *et al* (2011) propõe, com base na teoria do padrão criminal, um modelo chamado Modelo do Movimento Criminal (CriMM) baseado nas localizações dos crimes e em informações pessoais dos criminosos.

Segundo os autores, a teoria do padrão criminal informa que normalmente os crimes são cometidos dentro de uma área onde o criminoso costuma frequentar de forma rotineira, como por exemplo: residência, local de trabalho e local de recreação. Desse modo, o modelo visa construir as rotas que os delinquentes rotineiramente utilizam das suas residências até os locais onde praticam seus delitos, com base em algoritmos computacionais.

O trabalho foi feito baseado em dados reais obtidos durante cinco anos pelo departamento de polícia do Distrito Regional da Grande *Vancouver* na província de *British Columbia*, Canadá. Outra particularidade do trabalho é que o estudo baseia-se apenas em crimes contra residências, pois estas têm localizações fixas, diferentemente de crimes contra pessoas que muitas vezes não têm relação direta com o local onde ocorreram os delitos.

O modelo CriMM analisa o histórico de localizações de incidentes e cruza as informações com dados pessoais de criminosos. É levada em conta a localização da residência do criminoso, bem como locais comumente frequentados por estes, como centros comerciais, por exemplo.

No texto, esses locais recebem o nome de atratores. Dado um determinado crime, o algoritmo então calcula as distâncias euclidianas e de blocos entre a localidade desta

ocorrência com relação à casa e aos outros atratores, na tentativa de identificar qual caminho este indivíduo utilizou para cometer o crime. Então, o algoritmo traça uma rota que provavelmente foi tomada pelo delinquente, com base em algoritmos como *Dijkstra*.

A validação do estudo foi realizada com base em dados reais e reforçaram a teoria do padrão criminal ao demonstrar que os crimes normalmente estão nas proximidades de rotas que os bandidos utilizam diariamente, e que normalmente eles não saem muito dessas rotas para praticar seus delitos.

O principal benefício do trabalho é auxiliar as autoridades a entenderem como funciona a mente do criminoso no tocante à malha urbana, e assim possibilitar a adoção de algumas políticas públicas visando o combate à criminalidade.

As principais limitações do trabalho estão relacionadas a focar em um único tipo de crime (crimes contra residências), além do fato do caminho utilizado pelos criminosos nem sempre se iniciam nas suas residências, mas em outros locais. Outra limitação foi relacionada aos atratores utilizados no estudo, que poderiam ser adicionados uma maior variedade destes, visando obter melhores resultados.

#### 3.4 UM MODELO COMPUTACIONAL PARA PREDIÇÃO DA LOCALIZAÇÃO DE ATRADORES CRIMINAIS EM UMA REDE DE ESTRADAS

O texto de Iwanski *et al.* (2012) é basicamente um aperfeiçoamento do artigo (Iwanski *et al.*, 2011), apresentado na seção 3.3 Ele descreve um modelo chamado Preditor da Localização de Atratores (ALP) que localiza quais os atratores chave para os quais os criminosos se dirigiam quando cometeram seus crimes.

O modelo parte da teoria que os bandidos quando cometem seus crimes estão em rotas com as quais têm mais familiaridade, que normalmente estão entre sua casa e os atratores, que podem ser shoppings, seu local de trabalho, locais de recreação, escolas, casa de amigos, dentre outros.

A modificação diz respeito ao fato de em Iwanski *et al.* (2011) considerar cada estrada como tendo um peso igual na escolha dos criminosos, enquanto em Iwanski *et al.* (2012) é levada em consideração também a frequência com a qual cada trecho de estrada é utilizado. Isso é feito analisando os padrões de movimentação da cidade que variam de acordo com o estado do trânsito naquele local.

Os resultados apresentados no texto para a região da Grande *Vancouver*, Canadá, mostram que as rotas mais utilizadas pelos bandidos são as que levam aos centros comerciais, principalmente aos *shoppings centers*. As que têm uma taxa de utilização também muito grande são as que ligam um centro a outros.

Os resultados serviram para validar o modelo computacional proposto, haja vista que já eram esperados que os caminhos que levam aos centros comerciais viessem a ser muito utilizadas por criminosos, principalmente pela grande quantidade de pessoas que as utilizam e seriam assim, vítimas em potencial para os criminosos.

A importância deste trabalho diz respeito a auxiliar os gestores a identificar rotas em que a fiscalização deve ser intensificada, além de informar que deve haver um aumento da fiscalização nos entornos desses grandes centros comerciais. Além disso, pode auxiliar os gestores públicos em tarefas comuns de planejamento urbano, como definir locais onde serão inseridos novos shoppings, prédios comerciais, dentre outros.

### 3.5 PREVISÃO DE CRIMES UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

O artigo Yu *et al.* (2011) parte da teoria que os crimes não acontecem de maneira aleatória e propõe a aplicação de técnicas de mineração de dados no intuito de prever os locais e horários em que ocorrerão crimes. Na metodologia utilizada pelos autores, a cidade é dividida como uma grade para tratar as questões geográficas dos crimes.

Os locais onde as células da grade são vazias são desconsiderados já que indicam que não há ocorrências nessa área, pois na sua grande maioria trata-se de regiões afastadas ou sem presença de pessoas, o que torna a ocorrência de crimes muito baixa ou nula.

Na metodologia existe também o conceito de "pontos quentes" (*Hot Spots*) que são regiões com grande incidência de crimes e "pontos em aquecimento", que são os locais em que o número de ocorrências vem aumentando. Com relação ao tempo, os crimes são divididos em meses e são estudadas também as relações entre um mês e outro, por exemplo: os crimes ocorridos no mês de janeiro podem influenciar os de fevereiro além de também poderem ter sido influenciados por dezembro.

Para realizar os experimentos, foram comparados alguns algoritmos de classificação de dados: Um Vizinho Mais Próximo (*One Nearest Neighbor*) (Cover e Hart, 1967), Árvores de Decisão J48 (Witten et al., 2011), *Support Vector Machine* (SVM) (HEARST *et al.*, 1998)

com função de base radial como tipo de *Kernel*, Redes Neurais com múltiplas camadas e Redes Bayesianas (Witten *et al.*, 2011). Sendo que na maioria dos casos, as redes neurais e as redes bayesianas apresentaram resultados mais precisos em relação às outras técnicas, mas salientando que isso depende diretamente da entrada.

Então, para prever localidades que apresentam tendências a serem pontos quentes, é utilizada uma espécie de votação entre os classificadores, se a maioria informar que naquele local será um ponto quente, então a saída informa isto. No presente trabalho os três classificadores que realizam a "votação" são os que apresentaram melhores resultados individualmente: SVM, rede neural e rede de *Bayes*.

O texto informa algumas dificuldades enfrentadas para a sua realização que dizem respeito principalmente à falta de integração dos dados policiais e à falta de organização dos mesmos. Além de sugerir como trabalho futuro a construção de um mecanismo capaz de prever aumentos na ocorrência de determinado tipos de crimes, já que permitiria às autoridades realizar um melhor preparo para o combate a este tipo de crime.

### 3.6 DETECÇÃO E MAPEAMENTO DE PONTOS QUENTES CRIMINAIS BASEADOS EM CLUSTERIZAÇÃO ORIENTADA AO MELHOR ATRIBUTO

O texto Zhang *et al.* (2010) apresenta um método para detecção e mapeamento de pontos quentes (*Hot Spots*) criminais. Esses pontos são locais onde há grande ocorrência de incidentes ou onde há uma grande tendência à vitimização.

No método proposto há inicialmente um pré-processamento das informações espaciais dos dados originais, depois é aplicado o agrupamento orientado ao melhor atributo e por fim o mapeamento dos pontos extraídos para mostrar os pontos quentes. Ou seja, diferentemente dos métodos tradicionais que basicamente utilizam como base para o mapeamento apenas a localização dos incidentes, o método realiza uma análise prévia dos crimes a fim de entender a natureza dos mesmos, para então gerar o mapeamento dos pontos quentes.

No método proposto, há uma análise taxonômica dos eventos em relação a locais, categorias de crimes e tempo. Após a análise e agrupamento dos crimes utilizando o algoritmo proposto, os dados gerados serão representados graficamente no mapa a fim de informar os pontos quentes.

Para a realização dos testes, foram utilizados dados da capital de uma província chinesa que não teve seu nome divulgado. Nos resultados do agrupamento, alguns resultados interessantes, como por exemplo, a relação entre algumas palavras, como: {local de recreação, assaltos, noite, verão} ou {locais de transporte, roubo, dias da semana, feriados}.

Os três principais tipos quentes de crime são: roubos, assaltos e fraude. Com relação às vítimas, os resultados mostraram um maior agrupamento para: {Locais de recreação, todos os crimes, noite, verão, mulheres, jovens} e {Locais de trabalho, fraudes, dias da semana, todos os dias do ano, todos os gêneros, pessoas idosas}. O que demonstra respectivamente que nas noites de verão, mulheres jovens tendem a sofrer um maior número de roubos e que pessoas idosas são mais vítimas de fraudes em seus locais de trabalho.

Como resultado do trabalho, foi gerado um modelo para geração de mapas criminais que realiza o mapeamento dos pontos quentes levando em consideração não só dados geográficos dos crimes, como também informações referentes à natureza dos crimes, como vítimas, dia da semana, período do ano, dentre muitos outros. O que pode auxiliar a polícia no planejamento estratégico das suas ações para combater esses crimes.

## 4 ASPECTOS METODOLÓGICOS

Conforme discutido anteriormente, na maioria dos Centros Integrados, apesar da grande quantidade de dados gerados pelas ferramentas utilizadas, muitas decisões ainda são tomadas com base no empirismo.

Por outro lado, aparentemente essa realidade vem sendo modificada. Nas visitas realizadas ao CIOSP-RN, percebeu-se que, em diversas situações, o comando faz uso de relatórios estatísticos para basear as suas ações, o que demonstra uma preocupação dos gestores em fazer bom uso dos seus dados.

Em ambas as unidades do CIOSP visitadas durante a realização do presente trabalho, Mossoró e Natal, existem funcionários que têm por função produzir os relatórios estatísticos quando solicitados pelo comando.

O problema na forma de produção dos relatórios estatísticos como estão, consiste no fato do processo de produção dos mesmos ser praticamente de forma manual, dispendendo um grande tempo para a produção destes. Por exemplo, um relatório estatístico referente a um ano, pode levar quase uma semana para ser construído, custando tempo e pessoal para o Centro que já é carente destes.

Visando não se distanciar dessa realidade, e com o intuito de desenvolver uma ferramenta funcional que possa vir a ser realmente utilizada pelo CIOSP, optou-se que a maioria das saídas da ferramenta fosse dada em forma de relatórios.

Os relatórios produzidos pela ferramenta são baseados nos que já são utilizados pela corporação, com algumas modificações: o acréscimo de algumas características, principalmente relativas ao atendimento às chamadas, composição temporal, tipificação criminal e georreferenciamento das ocorrências.

A maioria dos relatórios da ferramenta CIOSP-BI é baseada na técnica de sumarização de dados, que é uma das atividades primárias da mineração de dados. Segundo Witten *et al.* (2011), a técnica visa resumir os dados organizando as informações mais relevantes e apresentando ao usuário uma informação mais clara e objetiva. Métodos estatísticos podem ser utilizados no desenvolvimento.

O processo de desenvolvimento da ferramenta começou com o estudo prévio do funcionamento do centro como um todo, visando entender como se dão os processos no mesmo. Assim, as visitas iniciais foram realizadas para colher as informações acerca do protocolo de atendimento às ocorrências do Centro.

Durante as visitas, foi também realizado o levantamento de requisitos do sistema, que é tido como o ponto inicial para se desenvolver qualquer aplicação, tendo em vista que nesse processo são identificadas as principais necessidades do cliente, neste caso o CIOSP, e como a ferramenta pode vir a ser útil para o mesmo.

Os requisitos são divididos comumente em duas classes principais: Requisitos funcionais, que são os que estão diretamente ligados à funcionalidade da aplicação; e Requisitos não funcionais, que são aqueles que não estão diretamente associados às funcionalidades, mas que podem auxiliar ou melhorar as mesmas (Bezerra, 2006).

Os principais requisitos funcionais identificados foram:

- O sistema deve gerar relatórios acerca de um dado período de tempo, que sumarie as informações desse intervalo e que permita também as informações sejam limitadas a determinados tipos de ocorrência ou órgãos do CIOSP;
- O sistema deve ser capaz de comparar dois períodos distintos, e que também permita a aplicação de filtros visando especificar melhor o objeto de interesse;
- O sistema deve criar relatórios que resumam as informações acerca do atendimento em um dado período de tempo e que possibilite a identificação de gargalos no atendimento, sendo capaz também construir esses relatórios para determinado(s) órgão(s) do CIOSP;
- O sistema deve apoiar a criação de mapas de calor e que assim como nas outras funcionalidades, permita também que sejam aplicados filtros nas informações.

Quanto aos requisitos não funcionais, os identificados foram:

- Segurança - O sistema deve autenticar os usuários, tendo em vista que é necessário proibir o acesso de pessoas não autorizadas à aplicação, já que as informações presentes na mesma são sigilosas;
- Confiabilidade - O sistema deve analisar as entradas de dados informadas pelos usuários visando proporcionar uma maior segurança à aplicação além de eliminar entradas inconsistentes, como informações incorretas.
- Desempenho - O sistema deve apresentar baixo tempo de resposta, visto que com os relatórios em mãos mais rapidamente, diminui também o tempo da tomada de decisões como um todo.

Após as visitas e o levantamento dos requisitos, a tarefa seguinte foi compreender o funcionamento do *software* iCAD, desenvolvido pela SISGRAPH, que é a ferramenta CAD

utilizada pelo CIOSP para atender as chamadas de emergência. Para isso, foi necessário um estudo minucioso do dicionário de dados da aplicação que estava disponível no Centro.

Feito isso, a ferramenta passou a ser desenvolvida. Os casos de uso da aplicação são apresentados na seção 4.1, e têm um maior detalhamento na seção 4.2. A seção 4.3 apresenta alguns detalhes técnicos da aplicação e a seção 4.4 mostra onde a técnica de BI foi aplicada no desenvolvimento.

Vale ressaltar que no presente trabalho não são citados detalhes acerca da ferramenta CAD do Centro, bem como do seu banco de dados, tendo em vista a natureza do software utilizado. Dessa maneira normalmente são utilizados nomes genéricos para tabelas no banco de dados.

#### 4.1 DEFINIÇÃO DOS CASOS DE USO

Com base no levantamento de requisitos, foi definido que o sistema contaria com quatro funcionalidades principais: *Resumo das Ocorrências*, *Comparar Intervalos*, *Análise do Atendimento* e *Análise Espacial das Ocorrências*. Para cada uma delas foi definido um caso de uso, além de outro caso de uso chamado: *Exibir tela principal*.

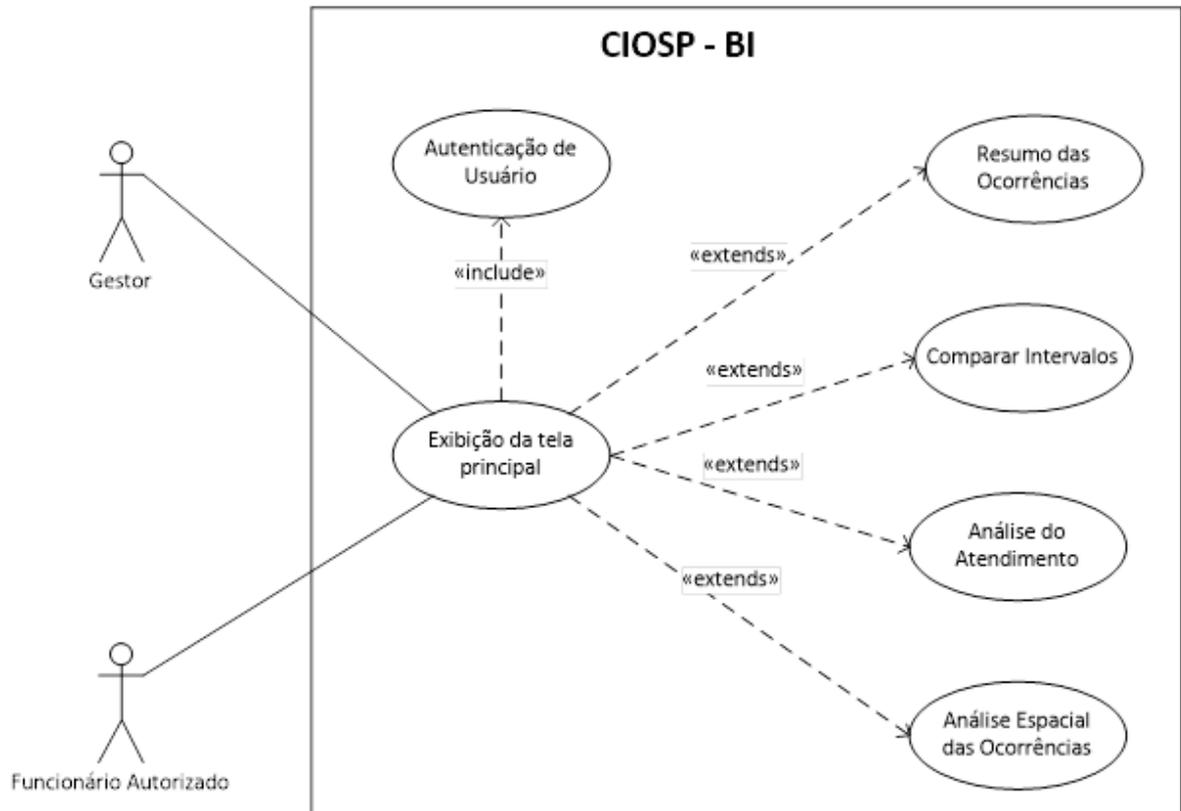
Foram identificados os seguintes atores do sistema, que são as entidades que farão uso da aplicação (Bezerra, 2006): *Gestor*, que é alguém do comando do CIOSP, que normalmente é o interessado nos resultados da aplicação, ou algum *Funcionário Autorizado* pelo mesmo para desempenhar essa tarefa.

Os casos de uso são regidos por regras de negócio, que representam as condições que devem ser satisfeitas durante a utilização do sistema. Desse modo, foram definidas as seguintes regras de negócio:

- **RN1** (Autenticação de usuário) – Só poderá ter acesso a qualquer funcionalidade do CIOSP-BI determinados usuários autorizados pelo comando do centro;
- **RN2** (Validação das entradas textuais) – As entradas textuais devem passar por um controle de verificação visando evitar erros nas consultas ao banco, bem como dificultar a ação de ataques de injeção de SQL;
- **RN3** (Validação das entradas não textuais) – As entradas não textuais devem também ser verificadas visando reduzir consultas inconsistentes ao banco de dados;

O diagrama de casos de uso da UML (Linguagem de Modelagem Unificada) é apresentado na figura 5, em seguida cada um dos casos de uso é brevemente apresentado. A modelagem de cada um destes é apresentada na seção 4.2.

Figura 5 – Diagrama de Casos de Uso do CIOSP-BI



Fonte: Autoria própria

O caso de uso *Exibição da tela principal* autentica o usuário através do caso de uso incluído, *Autenticar Usuário* e apresenta a tela de boas vindas à aplicação, além de permitir o acesso ao menu que possibilita o acesso aos outros casos de uso que são extensões deste caso de uso.

*Autenticação usuário* é um caso de uso que serve de base para as outras funcionalidades. O mesmo não é acessado diretamente pelo usuário, mas pelo caso de uso Exibir tela principal. Nenhum usuário deve ter acesso às informações sem que antes seja autenticado. Essa autenticação é realizada informando-se a matrícula na corporação e a senha utilizada na ferramenta CAD do CIOSP.

*Resumo de dados* cria um relatório acerca de um determinado período de tempo determinado pelo usuário, destacando os dias da semana e as horas do dia em que os crimes mais ocorrem. Além disso, como o CIOSP integra vários órgãos, são apresentadas as ocorrências por órgão e os tipos de crime mais comuns no intervalo.

*Comparar intervalos* é um caso de uso que abstrai uma necessidade rotineira do centro, que é a de comparar dois períodos de tempo. Assim, a funcionalidade basicamente constrói um relatório semelhante ao do *Resumo de dados*, porém colocando lado a lado informações referentes a dois intervalos de tempo para serem comparados.

O caso de uso *Análise do atendimento* cria um relatório voltado especificamente ao atendimento às ocorrências. No mesmo, é analisada a utilização de viaturas, a composição temporal das ocorrências e os códigos de finalização das mesmas, que caracterizam algumas ocorrências.

Por fim, *Análise espacial das ocorrências* gera um relatório em formato CSV (*Comma-Separated Values*) que pode ser importado pelo software SIG, Quantum GIS (QGIS, 2013), como uma camada a partir de texto, e assim permitir a visualização das ocorrências no mapa da cidade, além de possibilitar a criação do mapa de calor, utilizado um *plugin* do mesmo.

O funcionamento da aplicação é simplificado. Inicialmente é realizada a autenticação do usuário para possibilitar o acesso à aplicação. Após esse ponto é apresentada a tela de boas vindas ao usuário que pode escolher qualquer uma das funcionalidades através de um menu.

Quando o usuário opta por qualquer uma das quatro funcionalidades do CIOSP-BI, são apresentadas telas nas quais o mesmo define o que pretende obter com o uso dessa função. Em todas as funcionalidades, no primeiro campo, o usuário deve escolher um intervalo de tempo composto pelas datas e horas iniciais e finais.

Desse modo, o mesmo pode optar por intervalos de tempo diversos, sejam eles um mês, um ano, um dia, ou mesmo uma hora de determinado dia. Para evitar erros, são verificadas as entradas para evitar que as datas finais sejam anteriores às iniciais.

O CIOSP no município de Mossoró é composto por diversos órgãos, são eles: PMRN (Polícia Militar do RN), PCRN (Polícia Civil do RN), CBRN (Corpo de Bombeiros do RN), ITEP (Instituto Técnico-Científico de Polícia do RN) e GMM (Guarda Municipal Metropolitana).

Desse modo, o usuário pode estar interessado somente em informações relativas à PMRN, ou ao ITEP, ou a ambos. Assim, são exibidos *jCheckboxes* nos quais ele pode selecionar quais são os órgãos de interesse naquele relatório que deseja criar.

Por *default*, em todas as funcionalidades a opção TODOS está selecionada, o que cria um relatório de todos os órgãos, caso o usuário deseje somente algum(s) o mesmo deve especificar marcando as opções.

Outra característica é quanto aos tipos. No *software* CAD utilizado pelo CIOSP, existe uma tipologia fixa que se baseia nos tipos e subtipos de eventos. Por exemplo: para o tipo furto, existem diversos subtipos: de carga, em veículos, a estabelecimentos comerciais, à residência, à pessoa, dentre outros. No total tem-se mais de 1.000 combinações possíveis de tipos e subtipos.

Como cada tipo de ocorrência possui um código bem definido, o usuário também pode optar por criar o relatório sobre um tipo específico, ou um conjunto de tipos. Mas devido a enorme quantidade de combinações, optou-se por realizar essa entrada por meio textual dos códigos das ocorrências, pois os usuários já têm conhecimento prévio dos códigos mais utilizados por eles, além de haver a possibilidade de criação de novos tipos com o passar do tempo.

O banco de dados utilizado pelo centro é estruturado, composto por mais de 200 tabelas armazenando informações referentes a todo o atendimento às ocorrências e dados da corporação. Dados referentes às chamadas, comentários, despacho, uso de viaturas, pessoal, e vários outros são armazenados no mesmo.

Desse modo, o banco é completo e ao mesmo tempo complexo, tendo em vista a enorme variedade de informações armazenadas no mesmo. Após um exaustivo estudo prévio desse banco de dados, foram definidas quais informações são mais relevantes para cada um dos relatórios gerados pelo CIOSP-BI, no intuito de extrair e analisar essas informações com maior rapidez, tendo em vista que não seria vantajoso extrair e analisar informações irrelevantes. Por informações irrelevantes, nesse caso, compreendem-se as informações duplicadas, ocorrências sem finalização, trotes, dentre outros.

Segundo Silberschatz *et al.* (2006), existem diversas técnicas para otimizar as consultas em linguagem SQL. Isso se dá devido ao fato de uma mesma pesquisa poder ser escrita de diferentes formas, com diferentes custos de avaliação. Dentre as técnicas, algumas foram utilizadas no desenvolvimento no caso de teste deste estudo.

As operações de junção unem informações de diferentes tabelas e podem ser construídas de maneiras diversas. Assim, uma boa ordenação nessas operações podem reduzir os resultados temporários e virem a deixar a expressão mais rápida como um todo. Nas consultas do CIOSP-BI foram realizados testes visando analisar qual a forma de realizar a junção mais eficiente.

Outra técnica utilizada foi a de consultas aninhadas. Essa técnica realiza a consulta nos resultados de outras consultas SQL, otimizando os resultados, tendo em vista que caso ela fosse realizada pela aplicação, as respostas seriam mais lentas.

Além destas, outras técnicas comuns, que também melhoram as consultas, foram utilizadas, como: a utilização dos nomes reais das relações nas consultas, operações aritméticas inclusas nas consultas, seleções distintas, além de filtros nos dados a serem consultados.

Visando acelerar também o acesso às funcionalidades, foram definidos atalhos para realizar o acesso rápido às mesmas, evitando que o usuário necessite clicar nos menus para ter acesso à tela de entrada da informação. Os atalhos são: Ctrl+R para *Resumo das Ocorrências*, Ctrl+C para *Comparar Intervalos*, Ctrl+P para *Análise do Atendimento* e Ctrl+E para *Análise Espacial das Ocorrências*.

Existem também teclas de atalho. A tecla *Escape* (ESC) pode ser utilizada para encerrar a aplicação e a tecla de função F1 direciona para uma janela de ajuda que apresenta uma breve descrição de cada funcionalidade.

## 4.2 MODELAGEM DOS CASOS DE USO

A presente seção apresenta a modelagem para cada caso de uso apresentados na seção 4.1. São modelados os casos de uso acessados diretamente pelos usuários da aplicação. O modelo exposto é baseado no apresentado em Bezerra (2006) para a modelagem de casos de uso, baseado na UML e apresentando uma especificação formal para cada um deles.

Cada subseção seguinte refere-se a um caso de uso apresentados na figura 5. É apresentada a sequência de passos realizados para a execução de cada um dos casos de uso, bem como os fluxos de exceção que ocorrem quando há problemas na execução. As regras de negócio foram definidas no início deste capítulo e são representadas aqui pelos códigos das mesmas.

Após a modelagem de cada um dos casos de uso, que define todo o escopo da aplicação, é necessária a implementação do sistema, que significa escrever uma aplicação que faça exatamente o que foi delimitado na modelagem.

#### 4.2.1 Exibição da Tela Principal (CSU01)

**Sumário:** O usuário informa suas credenciais que são validadas no banco de dados

**Ator Primário:** Gestor.

**Atores Secundários:** Funcionário Autorizado.

**Precondições:** Banco de Dados acessível

**Fluxo Principal:**

- 1 – O usuário abre a aplicação;
- 2 – O usuário informa a matrícula e a senha;
- 3 – O banco de dados verifica se existe no banco de dados a combinação informada;
- 4 – O sistema direciona o usuário para a tela inicial da aplicação e o caso de uso é encerrado.

**Fluxo de Exceção 1:** Violação da RN02

O sistema informa ao usuário que não são permitidas a inserção de caracteres especiais nas entradas textuais, bem como trechos de instruções SQL e solicita um novo preenchimento desses campos.

**Pós-condições:** O usuário está autenticado e pode ter acesso a qualquer uma das funcionalidades da aplicação.

#### 4.2.2 Resumo das Ocorrências (CSU02)

**Sumário:** O usuário solicita do sistema um relatório sobre um determinado período de tempo.

**Ator Primário:** Gestor.

**Atores Secundários:** Funcionário Autorizado.

**Precondições:** O usuário foi previamente autenticado pelo sistema.

**Fluxo Principal:**

- 1 – O usuário acessa o menu Resumo das Ocorrências;
- 2 – O sistema apresenta um formulário a ser preenchido;
- 3 – O usuário seleciona o período de tempo, os órgãos e os tipos de ocorrência de interesse;

- 4 – O sistema analisa os campos do formulário;
- 5 – O sistema constrói a consulta SQL a ser realizada no banco de dados;
- 6 – O sistema realiza a consulta ao banco;
- 7 – O sistema cria o relatório;
- 8 – O sistema salva o relatório em PDF, o abre para visualização e o caso de uso termina.

**Fluxo de Exceção 1:** Violação da RN01

O sistema informa ao usuário que ele necessita realizar *login* para ter acesso à funcionalidade e encerra a aplicação.

**Fluxo de Exceção 2:** Violação da RN02

O sistema informa ao usuário que não são permitidas a inserção de caracteres especiais nas entradas textuais, bem como trechos de instruções SQL e solicita um novo preenchimento desses campos.

**Fluxo de Exceção 3:** Violação da RN03

O sistema informa ao usuário que o intervalo de tempo selecionado é inválido, já que o mesmo inseriu uma data final anterior à inicial. É solicitado um novo preenchimento desses campos.

**Pós-condições:** O relatório foi criado e encontra-se salvo no diretório da aplicação e pode ser utilizado na tomada de decisões do usuário.

### 4.2.3 Comparar Intervalos (CSU03)

**Sumário:** O usuário solicita do sistema um relatório comparativo entre dois intervalos distintos de tempo.

**Ator Primário:** Gestor.

**Atores Secundários:** Funcionário Autorizado.

**Precondições:** O usuário foi previamente autenticado pelo sistema.

**Fluxo Principal:**

- 1 – O usuário acessa o menu Comparar Intervalos;
- 2 – O sistema apresenta um formulário a ser preenchido;
- 3 – O usuário informa os dois períodos de tempo, os órgãos e os tipos de ocorrência de interesse;

- 4 – O sistema analisa os campos do formulário;
- 5 – O sistema constrói as duas consultas que serão realizadas no banco de dados;
- 6 – O sistema realiza as duas consultas;
- 7 – O sistema cria o relatório comparando os resultados das consultas com base nos seus valores relativos;
- 8 – O sistema salva o relatório em PDF, o abre para visualização e o caso de uso termina.

**Fluxo de Exceção 1:** Violação da RN01

O sistema informa ao usuário que ele necessita realizar *login* para ter acesso à funcionalidade e encerra a aplicação.

**Fluxo de Exceção 2:** Violação da RN02

O sistema informa ao usuário que não são permitidas a inserção de caracteres especiais nas entradas textuais, bem como trechos de instruções SQL e solicita um novo preenchimento desses campos.

**Fluxo de Exceção 3:** Violação da RN03

O sistema informa ao usuário que algum dos intervalos de tempo selecionados é inválido, e indica onde possivelmente ocorreu o erro, para facilitar a correção do mesmo.

**Pós-condições:** O relatório foi criado e encontra-se salvo no diretório da aplicação e pode ser utilizado na tomada de decisões do usuário.

#### 4.2.4 Análise do Atendimento (CSU04)

**Sumário:** O usuário solicita do sistema um relatório acerca de características do atendimento prestado pelo CIOSP em um dado intervalo de tempo.

**Ator Primário:** Gestor.

**Atores Secundários:** Funcionário Autorizado.

**Precondições:** O usuário foi previamente autenticado pelo sistema.

**Fluxo Principal:**

- 1 – O usuário acessa o menu do Atendimento;
- 2 – O sistema apresenta um formulário a ser preenchido;

- 3 – O usuário informa o intervalo de tempo do relatório e o(s) órgão(s) que comporão os mesmo;
- 4 – O sistema analisa as entradas não textuais, visando detectar inconsistências no intervalo;
- 5 – O sistema constrói a consulta a ser realizada no banco;
- 6 – O sistema cria um relatório contendo informações referentes ao atendimento;
- 7 – O sistema salva o relatório em PDF, o abre para visualização e o caso de uso termina.

**Fluxo de Exceção 1:** Violação da RN01

O sistema informa ao usuário que ele necessita realizar login para ter acesso à funcionalidade e encerra a aplicação.

**Fluxo de Exceção 3:** Violação da RN03

O sistema informa ao usuário que o intervalo de tempo selecionado é inválido e solicita a reinserção das informações.

**Pós-condições:** O relatório foi criado e encontra-se salvo no diretório da aplicação e pode ser utilizado na tomada de decisões do usuário.

#### 4.2.5 Análise Espacial das Ocorrências (CSU05)

**Sumário:** O usuário solicita do sistema um arquivo que será importado no software SIG QGIS para gerar o mapa de calor.

**Ator Primário:** Gestor.

**Atores Secundários:** Funcionário Autorizado.

**Precondições:** O usuário foi previamente autenticado pelo sistema.

**Fluxo Principal:**

- 1 – O usuário acessa o menu Análise Espacial das Ocorrências;
- 2 – O sistema apresenta um formulário a ser preenchido;
- 3 – O usuário seleciona o período de tempo, os órgãos e os tipos de ocorrência de interesse;
- 4 – O sistema analisa os campos do formulário;
- 5 – O sistema constrói a consulta SQL a ser realizada no banco de dados;
- 6 – O sistema realiza a consulta ao banco;

7 – O sistema cria o arquivo;

8 – O sistema salva o arquivo CSV, exibe uma mensagem de confirmação ao usuário e o caso de uso termina.

**Fluxo de Exceção 1:** Violação da RN01

O sistema informa ao usuário que ele necessita realizar login para ter acesso à funcionalidade e encerra a aplicação.

**Fluxo de Exceção 2:** Violação da RN02

O sistema informa ao usuário que não são permitidas a inserção de caracteres especiais nas entradas textuais, bem como trechos de instruções SQL e solicita um novo preenchimento desses campos.

**Fluxo de Exceção 3:** Violação da RN03

O sistema informa ao usuário que o intervalo de tempo selecionado é inválido, já que o mesmo inseriu uma data final anterior à inicial. É solicitado um novo preenchimento desses campos.

**Pós-condições:** O arquivo CSV foi criado e encontra-se no diretório da aplicação. O mesmo pode ser diretamente importado no QGIS para permitir a manipulação das informações espaciais.

#### 4.3 CARACTERÍSTICAS GERAIS DO CIOSP-BI

A aplicação desenvolvida para validar a utilização da técnica de *Business Intelligence* em dados referentes a ocorrências provenientes de ferramentas CAD foi intitulada CIOSP-BI. O *software* foi desenvolvido em linguagem de programação Java voltada para *desktop*.

O Sistema Gerenciador de Banco de Dados (SGBD) da aplicação é o Oracle 10G (Oracle, 2013), que embora não seja a versão mais recente do banco, é a utilizada pela ferramenta CAD do CIOSP. Desse modo, não houve necessidade de atualização do *software*. Foi somente necessário utilizar o JDBC apropriado para esta versão.

As atividades relacionadas ao georreferenciamento, são realizadas pelo Sistema de Informações Geográficas (SIG) QuantumGIS ou QGIS. O QGIS é um dos melhores SIGs da comunidade de *software* livre, apresentando uma interface gráfica intuitiva, o que facilita a sua utilização até mesmo por usuários não técnicos (QGIS, 2013). A versão utilizada é a 2.0.1. No CIOSP Mossoró não era utilizado outro SIG para realizar o processo até então.

Uma das principais preocupações durante o desenvolvimento foi relacionada à segurança da aplicação. O banco de dados do CIOSP apresenta informações de diversos tipos, que em mãos erradas podem gerar prejuízos incalculáveis, pois algumas envolvem até a vida de pessoas.

Então, para evitar que pessoas não autorizadas acessem a aplicação, o *login* é realizado com as mesmas credenciais utilizadas pelos usuários na ferramenta CAD do Centro, pois cada funcionário já possui matrícula e senha cadastradas.

Porém, como as saídas da aplicação devem ser acessíveis somente a pessoas autorizadas, foi criado um filtro que permite o acesso de somente algumas matrículas às informações, pois nem todo funcionário do centro deve ter acesso às mesmas. A tela de *login* é apresentada na figura 6.

Figura 6 – Tela de login da aplicação CIOSP-BI



Fonte: Autoria própria

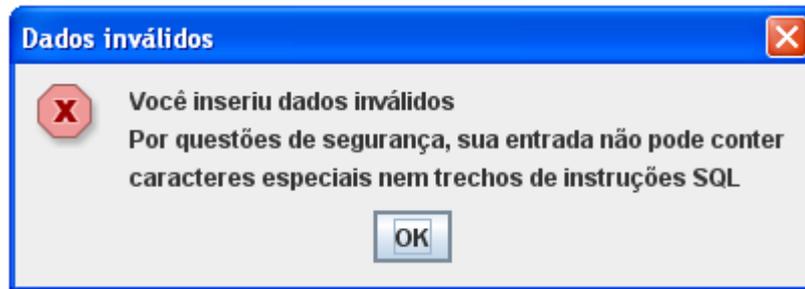
Um tipo de ataque que vem sendo cada vez mais utilizado contra aplicações, em especial em aplicações WEB, é o *SQL Injection*. Segundo Win e Htun (2013) esse tipo de ataque é um dos principais contra aplicações e funciona inserindo nas entradas da aplicação trechos de instruções SQL visando colher informações ou mesmo apaga-las dos bancos de dados. Mesmo o CIOSP-BI sendo desktop deve haver precaução quando a esse ataque.

Desse modo, em todas as entradas textuais da aplicação foi aplicado tratamento contra *SQL Injection*, visando evitar danos, caso alguém mal intencionado viesse a utilizar a aplicação.

Essa proteção é realizada por uma classe que analisa todas as entradas textuais, visando detectar a utilização de alguns caracteres especiais, bem como o uso de palavras reservadas pela linguagem SQL, como, por exemplo: SELECT, DELETE, dentre outras.

Essa proteção não interfere no funcionamento da aplicação, pois a maioria das entradas é numérica. A mensagem de erro apresentada quando algum caractere suspeito ou trechos de instruções SQL são inseridos é mostrada na figura 7.

Figura 7 – Mensagem de saída da proteção contra *SQL Injection*



Fonte: Autoria própria

Outra preocupação já relatada foi a de produzir as saídas em forma de relatórios. Na aplicação os mesmos são gerados no formato PDF. Existem diversas vantagens de se utilizar o formato PDF. Primeiramente tem-se o fato de ser um formato padrão e existirem diversos leitores gratuitos para o mesmo disponíveis, dispensando assim o uso de ferramentas pagas para isso. Esse formato também permite a visualização em diferentes plataformas, além de gerar arquivos compactos que podem ser facilmente enviados por e-mail, além de estarem prontos para serem impressos.

Para criar os relatórios nesse formato foi utilizada a biblioteca Java: *iText* (iText, 2013). A biblioteca é de código fonte aberta e pode ser usada gratuitamente, exceto em aplicações comerciais, onde se torna necessário um licenciamento específico.

O *iText* permite a criação dinâmica de arquivos PDF, RTF e XML. A mesma também se encontra disponível para *Java*, *.NET* e *Android*. Na aplicação foi utilizada a versão 2.1.5 para *Java*.

Visando uma melhor visualização dos dados, grande parte das saídas encontra-se na forma de gráficos. Para a construção destes foi utilizada a biblioteca de código fonte aberto *jFreeChart* (Jfree, 2013). A biblioteca Java começou a ser desenvolvida em 2000 e atualmente é bastante utilizada por desenvolvedores, principalmente devido à qualidade dos gráficos produzidos, bem como a facilidade em gerar os mesmos, pois já existem métodos que facilitam o processo. A versão utilizada foi a 1.0.15.

#### 4.4 TECNOLOGIAS DA INTELIGÊNCIA DE NEGÓCIOS UTILIZADAS

A técnica de BI foi apresentada na subseção 2.3.2, e os principais componentes de uma aplicação dessa natureza apresentados na figura 4 e explicados na mesma subseção. Esses componentes são: fontes de dados; movimentadores de dados e motores de streaming; servidores de *data warehouse*; servidores de nível intermediário; e aplicações *front-end*.

A presente seção relaciona a ferramenta produzida às tecnologias relacionadas à Inteligência de Negócios, destacando como as mesmas foram úteis no desenvolvimento da aplicação CIOSP-BI.

A **fonte de dados** da aplicação CIOSP-BI é o próprio banco de dados operacional do Centro. O banco de dados recebe em tempo real informações dos módulos da ferramenta CAD utilizada no CIOSP, sendo assim, foi necessário compreender esses dados, além de entender como os mesmos são gerados e como podem ser úteis.

Como já foi citado, a heterogeneidade é característica implícita de praticamente todos os bancos de dados corporativos, tendo em vista que as informações são geradas em diversos formatos, por diversos usuários, com diferentes precisões, dentre outros.

No CIOSP não é diferente, tem-se dados das chamadas telefônicas, dados de GPS, informações como: comentários, códigos, tipos, dentre vários outros, o que torna o trabalho mais complexo.

Como existe uma grande quantidade de informações no banco, as consultas tendem a retornar um grande volume de dados, tornando essa análise mais demorada. Sendo assim, são necessários **movimentadores de dados** para uma extração mais eficiente dessas informações.

No presente trabalho, foram utilizadas técnicas de otimização de consultas SQL, visando uma maior rapidez nas consultas, que pode ser obtido reduzindo-se as informações redundantes, e irrelevantes para a funcionalidade.

O **servidor de data warehouse**, conforme já foi explanado, não necessitou de implantação, tendo em vista que o mesmo já se encontrava em pleno funcionamento no CIOSP. A aplicação utilizada pelo Centro para gerenciar as ocorrências desempenha também essa funcionalidade e integra os dados em um repositório central.

Esse fator foi fundamental no desenvolvimento da aplicação, pois poupou o trabalho de realizar a integração dos dados, que possuem natureza heterogênea. Caso não houvesse algum mecanismo de integração das fontes de dados com os repositórios, esses deveriam ser implementados.

O servidor de **nível intermediário** utilizados no presente trabalho foi um servidor de relatórios, que teve a sua construção apoiada nas técnicas de integração de dados já implantadas no CIOSP.

O servidor baseia-se principalmente na sumarização dos dados. O servidor auxilia no processo de criação dos mesmos, agilizando assim a produção das saídas da ferramenta e consequentemente o tempo de resposta da aplicação.

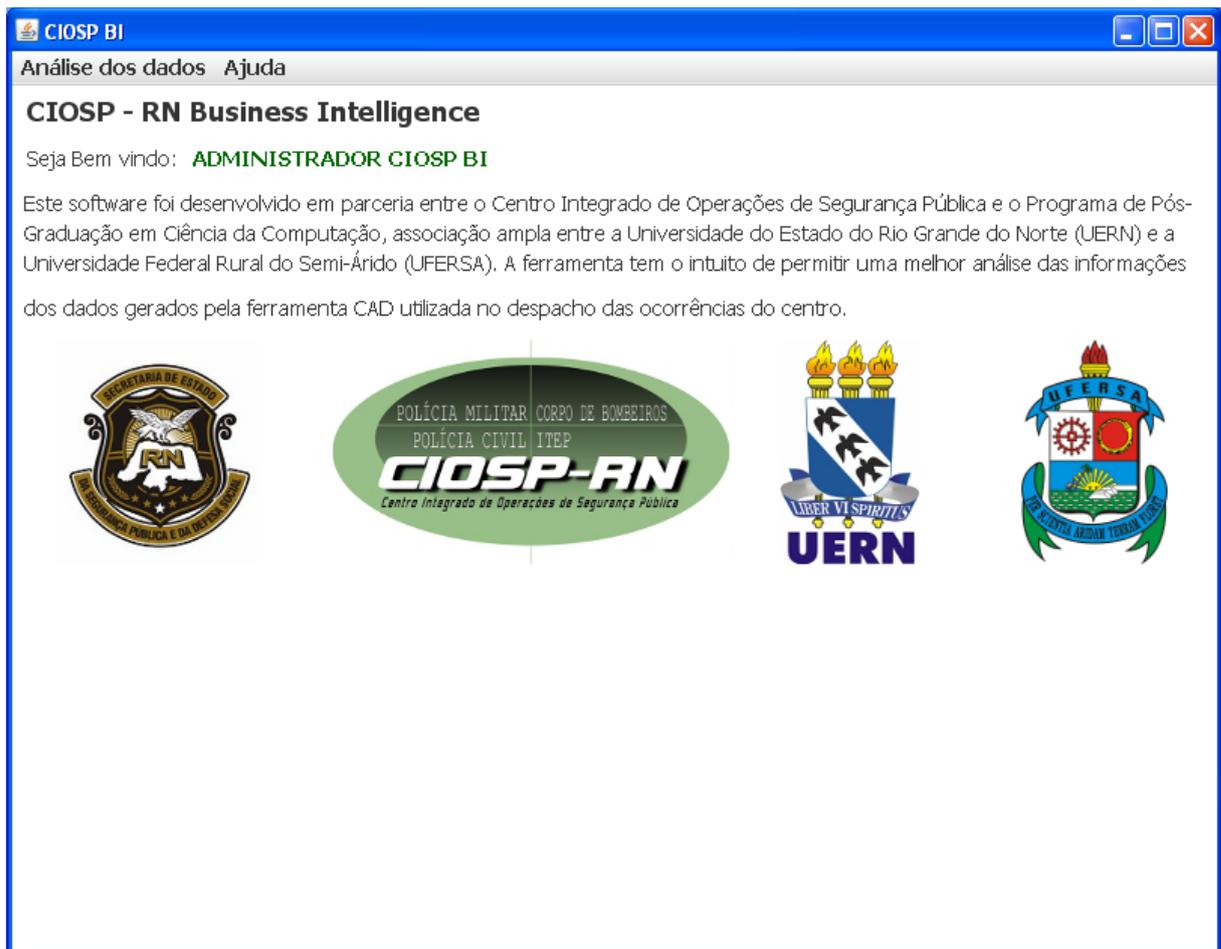
Por fim, a **aplicação front-end** desenvolvida foi a ferramenta CIOSP-BI, que possui uma interface simples e fácil de ser utilizada por usuários não técnicos. Detalhes da interface e dos relatórios gerados pela aplicação são apresentados no capítulo 5.

## 5 RESULTADOS E DISCUSSÕES

O presente trabalho foi desenvolvido com o intuito de validar a aplicação da técnica de BI em dados provenientes da ferramenta CAD utilizada pelo CIOSP-RN no município de Mossoró. Para isto, foi desenvolvida uma aplicação baseada na técnica, que teve suas funcionalidades apresentadas no capítulo 4.

A interface do software é simples, o acesso às funções pode ser realizado através de um *jMenu* ou utilizando os atalhos de teclado mencionados no capítulo 4. A tela inicial da aplicação pode ser visualizada na figura 8. A mesma é uma tela de boas vindas, que apresenta o nome do usuário que realizou *login* além de uma breve descrição da ferramenta. Na barra superior são exibidos os menus que levam o usuário para as funcionalidades.

Figura 8 – Tela inicial da aplicação



Fonte: Autoria própria

O presente capítulo apresenta os resultados de cada uma dessas funcionalidades, e discute como as mesmas podem ser úteis para o CIOSP-RN, no que diz respeito ao atendimento mais eficiente das ocorrências. Cada seção seguinte é referente a uma funcionalidade.

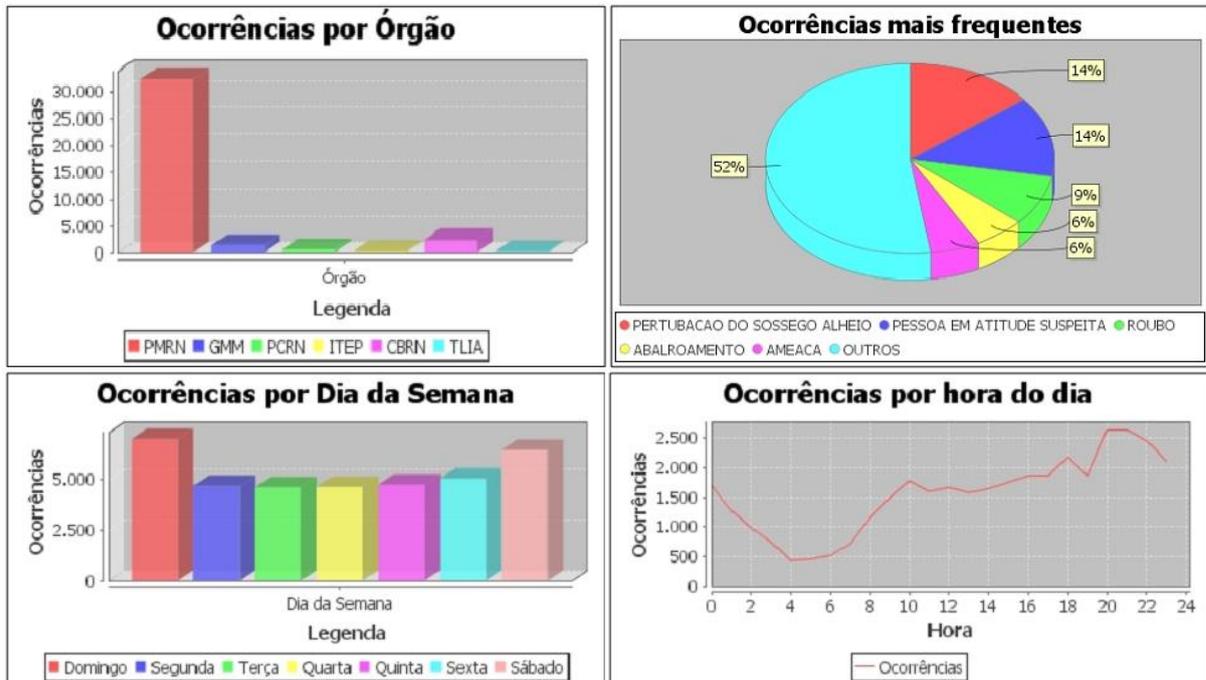
Os testes foram realizados em um computador portátil contendo uma parte do banco de dados, referentes às informações desde o início da operação do iCAD em Mossoró até julho de 2013, por isso, a maioria dos relatórios são referentes ao ano de 2012, que é o último com informações completas acerca de todo o ano.

## 5.1 RESUMO DAS OCORRÊNCIAS

A funcionalidade *Resumo das Ocorrências* cria um relatório resumido acerca de um determinado período de tempo definido pelo usuário. É possível também determinar o(s) órgão(s) que serão abordados no relatório bem como o(s) tipo(s) de crime apresentados no mesmo.

A figura 9 apresenta os gráficos contidos no relatório dessa funcionalidade. Os dados são referentes ao ano de 2012. Uma cópia completa desse relatório pode ser visualizada no anexo 1.

Figura 9 – Gráficos da funcionalidade: Resumo das Ocorrências



Fonte: Autoria própria

Para construir esse relatório é necessária a análise da tabela referente aos dados das ocorrências em si. Nesta tabela constam informações diversas acerca das ocorrências, como toda a composição temporal da mesma, informações sobre o despacho, encerramento, dentre outros.

No sistema utilizado pelo CIOSP, uma ocorrência se desmembra em vários eventos, sendo assim, para uma mesma ocorrência podem existir várias tuplas relacionadas a ela. Logo, são selecionadas apenas as tuplas de encerramento, pois são nelas que se encontram todas as informações sobre aquele evento.

A primeira informação presente no relatório é uma contagem das ocorrências no período indicado pelo usuário para aquele(s) órgão(s) que ele escolheu. Essa quantidade é seguida da média de ocorrências por dia, que demonstra uma quantidade aproximada das chamadas que são atendidas diariamente pelo CIOSP. Essa informação é relevante para o centro, pois a partir dela tem-se uma noção exata da quantidade de trabalho realizado pelo centro.

Como já foi abordado, o CIOSP é composto por diversos órgãos, e o usuário pode seleciona-los no momento de solicitar a construção dos relatórios. Caso seja selecionado mais de um órgão, é inserida no relatório uma contagem das ocorrências para cada um deles, seguido da percentagem que esse número representa dentro do período. Um gráfico de barras

com essa informação também é inserido no relatório visando facilitar a compreensão dessa informação.

Analisando os relatórios gerados percebe-se, por exemplo, que a maioria das ocorrências é atendida pelo órgão da Polícia Militar, normalmente com percentuais superiores a 85% do total de ocorrências nos períodos verificados. Isso é esperado devido ao fato da integração no município de Mossoró ainda não estar completamente concluída e a PM ser um dos órgãos mais solicitados no município.

A segunda informação apresentada nos relatórios é a distribuição temporal das ocorrências. Primeiramente é mostrada a distribuição das mesmas por dia da semana, sendo apresentada a quantidade seguida do percentual, no período selecionado pelo usuário. Posteriormente são apresentados dois gráficos.

O primeiro gráfico é de linhas e apresenta a distribuição das ocorrências por dia da semana. Nele, é possível perceber em que dias as ocorrências ocorrem mais frequentemente e com isso, torna-se possível pensar em mecanismos para coibir esses atos ou atende-los com maior rapidez.

Na maioria dos relatórios é notório o aumento das ocorrências nos fins de semana, principalmente das ocorrências do tipo perturbação do sossego alheio, o que pode ser explicado pelo consumo de álcool nesses períodos. Mas, como no momento de criação dos relatórios é possível filtrar o(s) tipo(s) de ocorrência, torna-se possível realizar o estudo individual de tipos específicos de ocorrência.

Outro gráfico apresentado nos relatórios é o de ocorrências por hora do dia. Esse gráfico de linha apresenta a distribuição por hora no período selecionado pelo usuário. Tornando possível analisar a que horas os crimes mais ocorrem.

Analisando alguns relatórios, percebe-se que o horário de pico das ocorrências encontra-se normalmente entre as 19 e 21 horas, provavelmente devido a este ser o horário de retorno do trabalho sendo assim um horário de grande circulação de pessoas.

Contrariando o que possa se pensar, os horários com menor incidência de ocorrências estão localizados normalmente entre as 03 e 05 horas da madrugada. Isso pode ser explicado devido ao baixo fluxo de pessoas nas ruas nesses horários.

De posse de informações como essa, o comando pode realizar um planejamento melhor de ações como a definição de rondas ostensivas nesses horários de pico de ocorrências, visando atende-las com maior rapidez, ou coibir a ocorrência do crime, tendo em vista que a presença do policial provavelmente intimida o criminoso evitando que o mesmo cometa o crime.

O último ponto abordado nesse relatório é o dos tipos de ocorrência mais comuns. Como existem diversos tipos cadastrados no banco, são mostrados apenas os mais frequentes naquele intervalo de tempo. Os tipos são seguidos das quantidades e das percentagens. Por fim é apresentado um gráfico de pizza contendo os cinco principais tipos de ocorrência no período, e o restante agrupado em um único campo denominado (OUTROS).

Como exemplo, tem-se que no relatório sobre o ano de 2012, as ocorrências de perturbação do sossego alheio, pessoa em atitude suspeita, roubo, alborroamento e ameaça, correspondem a quase 50% das ocorrências no período.

Desse modo, de posse de um quadro que denote o cenário das ocorrências, é possível a criação de campanhas de conscientização, políticas de prevenção e combate a essas ocorrências, dentre várias outras atitudes que podem ser tomadas pelo comando no intuito de reduzir estas situações.

Vale salientar que as informações são referentes ao município de Mossoró e cada localidade possui características próprias, não sendo possível, a generalização dessas informações.

Outro ponto relevante é a velocidade com que os relatórios são gerados. O relatório apresentado no anexo 1, referente a todos os tipos de crime e todos os órgãos, para o ano de 2012 demorou pouco mais de 2,5 segundos para ser gerado. Esse tempo foi reduzido devido às otimizações realizadas nas consultas aos dados.

## 5.2 COMPARAR INTERVALOS

A funcionalidade de comparar intervalos é semelhante à de resumo das ocorrências, até mesmo nos relatórios produzidos por elas. A diferença principal está no fato de que na funcionalidade de comparar intervalos, as informações serem referentes a dois intervalos distintos de tempo e não a um só, como ocorre no resumo das ocorrências.

Esse tipo de comparação é rotineiramente utilizado pelo CIOSP e pelos centros integrados de maneira geral, como uma forma de identificar aumentos ou diminuições nas ocorrências entre dois intervalos de tempo distintos.

Um exemplo prático do seu uso seria a verificação da eficácia de determinada política de prevenção ou combate a determinado tipo de ocorrência. Isso pode ser feito comparando

um período de tempo sem o uso dessa política e outro período, onde esta fosse adotada, visando encontrar alterações significativas que viessem a validar a eficiência da política.

A metodologia de criação dos relatórios é semelhante à utilizada na funcionalidade resumo das ocorrências, porém nesta, os dados são exibidos lado a lado para possibilitar uma comparação direta entre os intervalos. Os percentuais de ambos os intervalos também são exibidos, facilitando a comparação entre períodos de tempo com durações distintas.

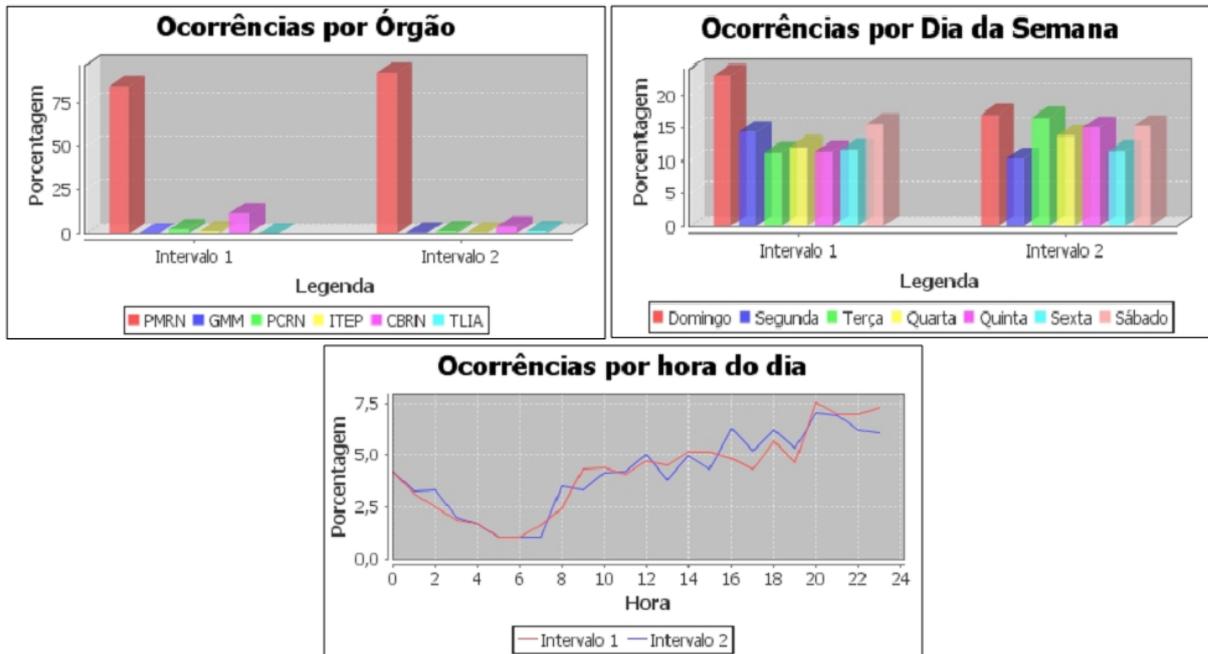
A tabela analisada é a mesma que no *Resumo das Ocorrências*, porém na funcionalidade de *Comparar Intervalos* são realizadas duas consultas ao banco e não somente uma, nesse caso as consultas devem ser otimizadas visando agilizar esse processo.

Quanto aos gráficos, ao invés de receber os valores absolutos das ocorrências recebem os valores relativos. Isso ocorre porque em alguns casos pode ser necessária a comparação entre períodos de tempo com durações distintas, e o uso de valores absolutos seria errôneo, visto que comumente um intervalo de dias maior tendem à apresentar um maior número de ocorrências.

A ferramenta se mostrou bastante eficiente, criando o relatório comparativo dos meses de janeiro de 2012 e janeiro de 2013 em 2,28 segundos. O que agiliza uma comparação dessa natureza, tendo em vista que caso a mesma fosse realizada manualmente poderia demorar horas para ficar concluída.

A figura 10 apresenta os gráficos resultantes da comparação entre os meses de janeiro de 2012 e janeiro de 2013. Nela pode-se notar a utilização dos valores relativos, conforme foi ressaltado. Uma cópia completa desse relatório pode ser visualizada no anexo 2 apresentado ao fim do presente documento.

Figura 10 – Gráficos da funcionalidade: Comparar Intervalos



Fonte: Autoria própria

### 5.3 ANÁLISE DO ATENDIMENTO

A funcionalidade *Análise do Atendimento* foi criada no intuito de auxiliar o comando com mecanismos de avaliar o atendimento à população realizado pela corporação. Os relatórios gerados pela ferramenta podem levar o comando a identificar deficiências e assim agilizar o atendimento.

O atendimento às ocorrências é um processo que depende de diversos fatores para acontecer. Inicialmente tem-se o atendimento à ligação. Quando um informante disca para algum dos números atendidos pelo centro, é feita inicialmente a coleta das informações para caracterizar a ocorrência como: tipo de ocorrência, localidade, suspeitos (se houver), dentre outros.

Em seguida, caso seja necessária alguma viatura se dirigir ao ocorrido, essa ocorrência entra em uma fila para despacho. Com o apoio de softwares que denotam a localização das viaturas em tempo real, ou através do rádio (no caso de Mossoró), o despachante localiza a viatura disponível mais próxima do ocorrido e a direciona para o mesmo.

Após esse ponto, a viatura se dirige à localidade e dá prosseguimento aos procedimentos padrão, que variam para cada tipo de ocorrência. A partir desse momento, a atividade da base do CIOSP é de monitorar as ocorrências e as suas viaturas em campo, para em caso de uma eventual complicação, solicitar reforços, montar barreiras, dentre outras atividades de apoio.

Um fato ainda não mencionado é que além dos códigos das ocorrências, a ferramenta utilizada pelo centro possui códigos de finalização para as ocorrências. Estes são necessários para indicar algumas situações, como: trotes, desistência por parte do solicitante, indisponibilidade de viaturas, ofensor não localizado, dentre outros.

Para a construção desse relatório é necessária a análise de três tabelas distintas: a das ocorrências em si, a das viaturas e a dos eventos. Pois cada uma delas apresentam informações diferentes. Para evitar a realização de várias consultas, todas as informações são requeridas em uma única consulta.

O Anexo 3 apresenta o relatório completo gerado pela funcionalidade referente ao ano de 2012. O mesmo demorou cerca de 3,5 segundos para ser construído. Algumas situações devem ser explicadas para melhor compreensão do mesmo.

Inicialmente o relatório apresenta informações referentes ao despacho das ocorrências. Não é sempre necessário haver o despacho das ocorrências, muitas ligações são somente de informações, ou ocorrências duplicada, ou mesmo referentes a situações em que não é necessário o envio de uma viatura ao local.

Então, cabe aos atendentes analisar cada caso e verificar a necessidade do despacho dessa ocorrência. O primeiro gráfico do relatório apresenta as ocorrências despachadas e não despachadas no período solicitado, bem como o tempo médio para a realização do mesmo.

O segundo tópico do relatório refere-se à saída das viaturas. Como se sabe, veículos em órgãos de segurança pública são recursos escassos. Sendo assim, existem situações em que mesmo com a ocorrência despachada, a viatura não se dirige ao local da ocorrência.

No relatório apresentado no anexo 3 é mostrado também o gráfico das ocorrências em que alguma viatura se dirigiu à ocorrência e as que não se dirigiu, além do tempo médio desde a ligação até o momento em que ela foi ao ocorrido.

Analisando esse relatório percebe-se que em alguns casos o tempo médio da viatura se dirigir até o local é menor que o tempo de despacho. Esse fato pode ser explicado pelo fato de muitas vezes a comunicação com as viaturas se darem antes da realização do despacho no sistema.

O terceiro ponto abordado no relatório é a chegada das viaturas ao local dos ocorridos. Além do fator trânsito, existem outros que interferem diretamente nesse componente. Outro ponto a ser ressaltado é que em certos casos, mesmo com a ocorrência despachada e com o veículo se dirigindo à localidade do ocorrido, a viatura não chega ao local. Isso pode se dar devido: desistências, informações erradas, dentre outros.

No relatório é apresentado um gráfico de pizza comparando as ocorrências nas quais as viaturas chegam ou não ao local, além do tempo médio gasto pelas viaturas desde o telefonema até a chegada.

O último ponto refere-se ao encerramento das ocorrências. Em alguns casos, devido a falhas ou por motivos diversos, as ocorrências não são encerradas. Isso pode também auxiliar os gestores a identificar problemas no atendimento. É apresentado assim, um gráfico comparativo das ocorrências encerradas e das não encerradas, bem como o tempo médio gasto desde a chamada até o encerramento.

Percebe-se que essa componente temporal é a mais dispendiosa, mas isso é esperado devido a ser nesse momento, entre a chegada da viatura ao local até o encerramento, em que a maior parte do tempo é gasta, tendo em vista que deve haver todos os procedimentos padrão para cada ocorrência, além de muitas vezes haverem complicações, o que torna a ocorrência mais demorada.

#### 5.4 ANÁLISE ESPACIAL DAS OCORRÊNCIAS

Esta funcionalidade é utilizada para realizar a análise espacial das ocorrências do CIOSP em um dado período informado pelo usuário. A saída da aplicação serve como entrada para o software QGIS (QGIS, 2013) que permite a visualização das ocorrências além da criação de mapas de calor.

Segundo Gonzales *et al.* (2005), *Hot Spots* ou sua tradução literal, pontos quentes, ocorrem em regiões onde há uma alta concentração de ocorrências. Nessas regiões onde existem altos índices de ocorrências, há também uma alta tendência à vitimização e assim, merecem uma atenção especial por parte das autoridades.

Gonzales *et al.* (2005) destaca que a análise dos *Hot Spots* podem auxiliar a polícia a identificar áreas onde os tipos de crime estão ocorrendo e assim criar mecanismos de combate aos mesmos nessas localidades.

Existem diversas maneiras de realizar a análise e classificação de pontos quentes. Gonzales *et al.* (2005) apresenta algumas dessas maneiras focando principalmente na componente prática. Mas diversos trabalhos, dentre eles Yu *et al.* (2011) e Zhang *et al.* (2010) que foram apresentados na seção 2.5 apresentam formas variadas de realizar essa análise.

No Quantum GIS existem diversos *plugins* gratuitos, dentre eles o “Mapa de Calor” (versão em português) que cria um mapa de calor como uma camada *raster* para pontos vetoriais de entrada, analisando a concentração de pontos nas regiões do mapa. Não há uma definição precisa de qual algoritmo o *plugin* utiliza para realizar a criação desses mapas.

O método de interpolação utilizado, no presente trabalho, para gerar os mapas de calor e identificar os *Hot Spots* foi o Mapa de *Kernel* ou a interpolação por densidade de *Kernel*, que permite a análise espacial de um grande volume de dados e é indicado para posições de dados individuais. (Bailey e Gatrell, 1995).

O software utilizado pelo CIOSP para gerenciamento das ocorrências armazena informações referentes à localização geográfica das mesmas em uma tabela especial relacionada aos eventos. As mesmas são expressas em coordenadas X e Y.

Diferentemente do que ocorre em Natal, onde todas as viaturas possuem GPS e a localização geográfica das ocorrências é dada em coordenadas de GPS, Mossoró ainda não possui esse sistema em pleno funcionamento e as localizações são informadas ao sistema pelos atendentes e despachantes a partir das informações dadas pelos informantes. Sendo assim, a localização não é tão precisa quanto ocorre na capital.

Por outro lado, mesmo tratando-se de localizações aproximadas, as mesmas podem auxiliar na análise espacial das ocorrências. No banco, essas coordenadas são dadas em UTM com base em um mapa georreferenciado pela prefeitura municipal.

O QGIS permite a importação de camadas a partir de textos delimitados como, por exemplo, o formato CSV. A ferramenta CIOSP-BI aplica os devidos filtros solicitados pelo usuário como: período temporal, tipo de ocorrência e órgão, gerando como saída um arquivo em formato CSV que pode ser automaticamente importado pelo QGIS.

Após a importação, o usuário necessita clicar no botão que chama a o *plugin* e após informar o raio (na maioria dos testes foi usado o valor 150 unidades do mapa) e a imagem de saída, o mapa de calor é criado. Por padrão, o mapa é gerado em escala de cinza, mas para facilitar a visualização, pode-se aplicar uma classificação espectral entre pseudocores.

A figura 11 ilustra as etapas do processo informado. O primeiro quadro mostra o mapa do município de Mossoró utilizado pelo CIOSP para referenciar suas ocorrências; no segundo são exibidas as ocorrências do ano de 2012 importadas de um arquivo CSV gerado pela

ferramenta CIOSP-BI, cada ponto em vermelho representa uma ocorrência; e o terceiro, exibe o mapa de calor gerado pelo QGIS após a aplicação da classificação por pseudocor. Todas as imagens foram obtidas com ajuda da ferramenta Quantum GIS.

Figura 11 – Exemplo de uso da funcionalidade Análise Espacial das Ocorrências



Fonte: Autoria própria

Por questões éticas, os nomes dos bairros e ruas foram ocultados. Deve-se salientar que o relatório é referente a todo o ano de 2012, todos os órgãos que integram o CIOSP e a todo tipo de ocorrência, não só as criminais. Pode ser observado também que como as

localizações são aproximadas, elas encontram-se normalmente nas ruas da cidade e não sobre residências.

Os gestores do CIOSP podem criar outros mapas focados em determinados tipos de ocorrência para auxiliar no planejamento das suas decisões, como posicionamento das viaturas, por exemplo.

A ferramenta CIOSP-BI demonstrou um desempenho satisfatório, criando o relatório do ano de 2012 em 0,62 segundos. O processo completo de criação do mapa de calor durou aproximadamente 2 minutos.

## 6 CONSIDERAÇÕES FINAIS

O presente trabalho foi desenvolvido com o objetivo de aplicar a técnica de *Business Intelligence* em dados oriundos da ferramenta CAD utilizada pelo CIOSP-RN. Como estudo de caso, foi desenvolvida uma aplicação utilizando a técnica de BI para o banco de dados do município de Mossoró-RN.

Percebeu-se que apesar de ainda longe do ideal, os órgãos de segurança pública, estão passando a investir também em tecnologia da informação para apoiar as suas atividades. Esse investimento é útil e pode vir a melhorar o atendimento desses órgãos à população e com isso, tentar reduzir também o número de ocorrências.

Com o uso de ferramentas computacionais, que atualmente já estão em pleno funcionamento na maioria dos centros, são gerados dados de granularidade fina, que podem ser analisados com o apoio de ferramentas de análise de dados que podem ser baseadas em diversas técnicas como: Mineração de Dados, *Business Intelligence*, dentre outras.

Essas técnicas, empregadas com muita frequência no mundo corporativo, mostram-se promissoras também para órgãos públicos, em especial órgãos de segurança pública, tendo em vista que permitem aos gestores tomarem decisões melhores em um curto espaço de tempo.

Além do trabalho em si, existe também o fato deste estudo, juntamente com outros que estão sendo desenvolvidos paralelamente, marcarem o início de uma parceria entre PPgCC e CIOSP, visando a criação de mecanismos computacionais que possam ser úteis para o Centro, ao mesmo tempo que possibilitam aos pós-graduandos o contato com problemas sociais reais vivenciados por toda a sociedade brasileira.

As técnicas de aperfeiçoamento das consultas ao banco mostraram-se eficientes, criando relatórios anuais completos em poucos segundos. Assim, percebe-se que os resultados foram satisfatórios, tendo em vista que minimizou o tempo de produção de relatórios estatísticos do CIOSP-RN/Mossoró. E o que muitas vezes custava várias horas de trabalho de alguns funcionários, pode agora ser feito em segundos, e o tempo economizado pode ser aplicado em outras atividades.

Outro trabalho futuro que pode ser desenvolvido é um estudo dos benefícios da implantação da ferramenta no CIOSP, tendo em vista que ainda não foi possível concluir com precisão os benefícios reais que uma ferramenta desse tipo vai oferecer no Centro, tendo em vista que estes se darão em longo prazo.

Analisando-se o histórico de versões do BI apresentado na seção 2.3.1 percebe-se que o presente trabalho apesar de não utilizar a plataforma WEB, enquadra-se na versão 2.0, pois

abstrai as funcionalidades do usuário e pode ser usada facilmente até por pessoas sem conhecimentos de banco de dados, e das técnicas de análise de dados, diferentemente do que ocorre na versão 1.0. Essa foi uma das principais preocupações durante o desenvolvimento, tendo em vista que a maioria dos funcionários do CIOSP não possui conhecimento técnico acerca de banco de dados.

Um possível trabalho futuro seria a migração completa da ferramenta para alguma plataforma WEB, bem como a adaptação para dispositivos móveis, que é justificável pelo fato da recente aquisição de *tablets* para apoiar o serviço do CIOSP.

Essa modificação é totalmente viável, e apoiada no fato do Governo do Estado do RN estar adquirindo *tablets* para serem utilizados nas viaturas do CIOSP, e que podem também serem utilizados pelo comando para apoiá-los na tomada das suas decisões através da análise dos dados do CIOSP.

Uma modificação dessa natureza auxiliaria o Centro na realização das funções, pois permitiria aos comandantes analisarem informações acerca do CIOSP até mesmo nas suas residências ou quando estivessem em viagens, não sendo necessário estarem presentes fisicamente no Centro.

A limitação maior nesse caso seria relativa à segurança da informação, pois com uma ferramenta em rede aumentar-se-ia o número de ataques. Outro fator está relacionado ao baixo poder de processamento desses dispositivos, o que necessitaria ainda mais de melhorias visando à economia de recursos.

As principais dificuldades na construção do trabalho foram referentes à compreender o funcionamento do CIOSP como um todo, sendo necessárias diversas visitas e conversas com os operadores do Centro. Assim, os funcionários do CIOSP foram fundamentais para o desenvolvimento da aplicação,

Por fim, tem-se que, apesar do trabalho ter sido desenvolvido para ser aplicado em todo o Centro Integrado, a problemática principal atacada foi a violência. A situação nessa área encontra-se crítica como foi apresentado no capítulo 1, e necessita do apoio de todos os setores da sociedade para enfrentar os problemas com maior eficiência.

## REFERÊNCIAS

- BAILEY, T.C., GATRELL, A. C. *Interactive Spatial Data Analysis*. Essex: Longman; 1995
- BEZERRA, E. *Princípios de análise e projeto de sistemas com UML*. Rio de Janeiro: Editora Campus, 2006.
- BRASIL. Constituição (1988). *Constituição [da] República Federativa do Brasil*. Brasília: Senado Federal, 1988.
- BRASIL. *Plano Diretor de Tecnologia da Informação 2012–2014*. Brasília: Ministério do Planejamento Orçamento e Gestão, 2011.
- CERQUEIRA, D. R. C; CARVALHO, A. X. Y; LOBÃO, W. J. A.; RODRIGUES, R. I. *Análise dos custos e consequências da violência no Brasil*. Brasília: Instituto de Pesquisa Econômica Aplicada, 2007.
- CHAN, J.; BRETON, D.; LEGOSZ, M.; DORAN, S. *E-policing: The impact of Information Technology on police practices*. Queensland: Criminal Justice Commission, 2001.
- CHAUDHURI, S., DAYAL, U., e NARASAYYA, V. An overview of business intelligence technology. *Communications of ACM*, v. 54 n. 8, p. 88–98, 2011.
- CHEN, H., CHIANG, R. H. L., STOREY, V. C. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, v. 36, n. 4, p. 1165–1188, 2012.
- COHEN, M. A.; MILLER, T. R.; ROSSMAN, S. B. *The Costs and Consequence of Violent Behavior in The United States*. Edited by Albert J. Reiss, Jr. and Jeffrey A. Roth. Understanding and Preventing Violence. Washington, D. C.: National Academy Press. 1994.
- COVER, T., HART, P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, v. 13, n. 1, p. 21-27, 1967.
- CZERNICKI, B. *Next Generation Business Intelligence Software with Silverlight 3*. New York: Apress, 2009.
- ECONOMIST, T. Beyond the PC, 2011. Disponível em: <<http://www.economist.com/node/21531109>> Acesso em 27 de Novembro de 2013.
- ENASP, *Meta 2: A Impunidade como alvo*. Brasília: Conselho Nacional do Ministério Público, 2012.
- GONZALES, A. R., SCHOFIELD, R. B., HART, S. V. *Mapping Crime: Understanding Hot Spots*. Washington: US Department of Justice, 2005.
- GOPALKRISHNAN, V., STEIER, D., Lewis, H., GUSZCZA, J. Big data, big business: Bridging the gap. *Preceedings of the International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, p. 7–11, 2012.

HARDOIM, H. O CIOPS como fator da melhoria da operacionalidade na polícia militar do Ceará: uma análise crítica de sua atuação – 2007. Disponível em: <<http://www.sspds.ce.gov.br/canalDetalhado.do?tipoPortal=1&codCanal=179&titulo=Informa%E7%F5es&action=detail>> Acesso em 19 de outubro de 2013.

HEARST, M. A., DUMAIS, S. T., OSMAN, E., PLATT, J., SCHOLKOPF, B. Support vector machines. *IEEE Intelligent Systems and their Applications*. v. 13, n. 4, p. 18-28, 1998.

IBGE. Projeção da População do Brasil por sexo e idade 1980-2050 – Revisão 2008. Disponível em: <[http://www.ibge.gov.br/home/estatistica/populacao/projecao\\_da\\_populacao/2008/default.shtm](http://www.ibge.gov.br/home/estatistica/populacao/projecao_da_populacao/2008/default.shtm)> Acesso em 15 de outubro de 2013.

IBGE. IBGE Cidades@ – Mossoró RN. – 2010a. Disponível em: <<http://www.ibge.gov.br/estadosat/perfil.php?sigla=rn>> Acesso em 16 de outubro de 2013.

IBGE. IBGE Estados@ – Rio Grande do Norte. – 2010b. Disponível em: <<http://www.ibge.gov.br/estadosat/perfil.php?sigla=rn>> Acesso em 16 de outubro de 2013.

IBM. Fast track to the future: The 2012 IBM tech trends report. – 2012. Disponível em: <<http://ibm.com/developerworks/techtrendsreport>> Acesso em 27 de novembro de 2013.

IWANSKI, N., FRANK, R., DABBAGHIAN, V., REID, A., BRANTINGHAM, P. Analyzing an offender as journey to crime: A criminal movement model (CRIMM). *Proceedings of the European Intelligence and Security Informatics Conference*. p. 70–77, 2011.

IWANSKI, N., FRANK, R., REID, A., DABBAGHIAN, V. A computational model for predicting the location of crime attractors on a road network. *Proceedings of the European Intelligence and Security Informatics Conference*. p. 60–67, 2012.

ITEXT. iText – Programmable PDF Software. Disponível em: <<http://itextpdf.com/>> Acesso em 08 de dezembro de 2013.

JFREE. Welcome To JFreeChart. Disponível em: <<http://www.jfree.org/jfreechart/>> Acesso em 08 de dezembro de 2013.

KANTARDZIC, M. *Data Mining: Concepts, Models, Methods and Algorithms*. New Jersey: IEEE Press, 2011.

KOZIELSKI, S., WREMBEL, R. *New Trends in Data Warehousing and Data Analysis*. New York: Springer, 2009,

MA, L., CHEN, Y., HUANG, H. Ak-modes: A weighted clustering algorithm for finding similar case subsets. *Proceedings of the International Conference on Intelligent Systems and Knowledge Engineering*, p. 218–223, 2010.

MCAFEE, A., BRYNJOLFSSON, E. Big data: The management revolution. *Havard Business Review*, p. 3-9, 2012.

MENDES, Claudio A. *O processo de informatização dos registros de eventos de defesa social e a sua utilização como instrumento de gestão*. Monografia (conclusão de curso) – Fundação João Pinheiro, CEGESP, 2007.

MILLER, H. J., HAN, J. *Geographic Data Mining and Knowledge Discovery*. Segunda Edição. New York: CRC Press, 2009.

MONTEIRO, J. M. *Relatório e parecer prévio sobre as contas do Governo da República – Exercício de 2011*. Brasília: Tribunal de Contas da União, 2012.

NAGABHUSHANA, S. *Data Warehousing – OLAP and Data Mining*. New Delhi: New Age, 2006.

ORACLE. Oracle Database 10g. Disponível em: <<http://www.oracle.com/technetwork/database/database10g/overview/index.html>> Acesso em 08 de dezembro de 2013.

OUF, S., NASR, M. Business intelligence on the cloud. *Proceedings of the International Conference on Communication Software and Network*. p. 650–655, 2011.

QGIS. QGIS – Um sistema de informação geográfica livre e aberto. Disponível em: <[http://www.qgis.org/pt\\_PT/site/](http://www.qgis.org/pt_PT/site/)> Acesso em 08 de dezembro de 2013.

RIO GRANDE DO NORTE (Estado). *Lei complementar nº 442*. Diário Oficial do Estado do Rio Grande do Norte, Natal, 2010.

SAPORI, L. F.; ANDRADE, S. C. P. Integração Policial em Minas Gerais – Desafios da governança da política de segurança pública. *Civitas Portalegre*, v.8, n.3, p.428-453, 2008.

SERRANO-BERTHET, R.; LOPEZ, H. *Crime and Violence in Central America: A Development Challenge*. Washington: World Bank, 2011.

SILBERSCHATZ, A., KORTH, H. F., SUDARSHAN, S. *Sistema de Banco de Dados – Tradução da 5ª edição*. Rio de Janeiro: Elsevier, 2006.

SILVA, J. A. S. *Centro Integrado de Operações de Segurança (CIOPS) – Uma análise da tecnologia a serviço da comunidade da área operacional integrada nº 3*. 2004, 134 f. Trabalho de Conclusão de Curso (Aperfeiçoamento de Oficiais), Academia de Polícia Militar, Natal, 2004.

SOUZA NETO, C. P. A. Segurança pública na constituição de 1988: conceituação constitucionalmente adequada, competências federativas e órgãos de execução das políticas, *Revista de Direito do Estado*, Vol. 8, 2007.

WARDEN, J. BI – Intelligence for the business of crime fighting. *Proceedings of the IEEE International Conference on Intelligence and Security Informatics*, p. 159–161, 2010.

WAISELFISZ, J. J. *Mapa da violência 2012 – Os novos padrões da violência homicida no Brasil*. São Paulo: Instituto Sangari, 2011.

WASELFISZ, J. J. *Mapa da violência 2013 – Homicídios e juventude no Brasil*. Rio de Janeiro: CEBELA, 2013.

WIN, W. Y., HTUN, H. H. A Simple and Efficient Framework for Detection of SQL Injection Attack. *International Journal of Computer & Communication Engineering Research (IJCCER)*. v. 1, p. 26-29, 2013.

WITTEN, I. H., FRANK, E., HALL, M. A. *Data Mining – Practical Machine Learning Tools and Techniques*. 3ª Edição. Burlington: Elsevier, 2011.

WORTLEY, R., MAZEROLLE, L. *Environmental criminology and crime analysis*. Portland: Willian Publishing, 2008.

YU, C. H., WARD, M. W., MORABITO, M., DING, W. Crime forecasting using data mining techniques. *Proceedings of the IEEE International Conference on Data Mining Workshops*. p. 779–786, 2011.

ZHANG, X., HU, Z., LI, R., ZHENG, Z. Detecting and mapping crime hotspots based on improved attribute oriented induce clustering. *Proceedings of the International Conference on Geoinformatics*, p. 1–5, 2010.

## **ANEXOS**

## Anexo 1 – Relatório do ano de 2012 gerado pela funcionalidade Resumo das Ocorrências

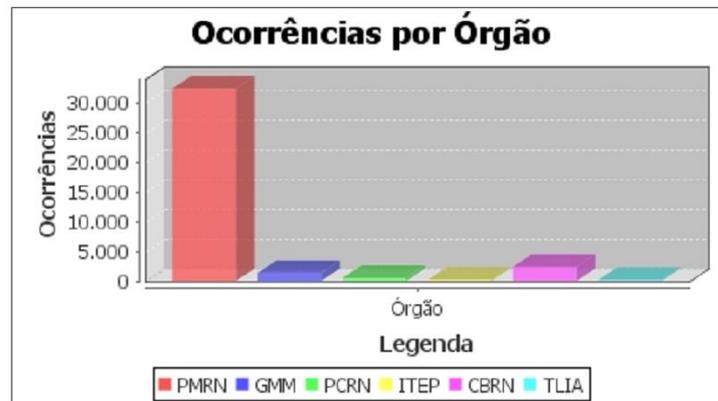
**Relatório - CIOSP BI**

Início: 01/01/2012 - 00:00:00 Fim: 31/12/2012 - 23:59:59

Total de ocorrências: 37065  
Quantidade de dias: 365  
Média: 101.54794 ocorrências por dia

**OCORRÊNCIAS POR ÓRGÃO**

PMRN: 32435 (87.50843 %)  
GMM: 1362 (3.6746256 %)  
PCRN: 525 (1.4164306 %)  
ITEP: 187 (0.5045191 %)  
CBRN: 2398 (6.4697156 %)  
TLIA: 158 (0.42627817 %)



**OCORRÊNCIAS POR DIA DA SEMANA**

Domingo: 6975 (18.818293%)

Segunda: 4672 (12.604883%)

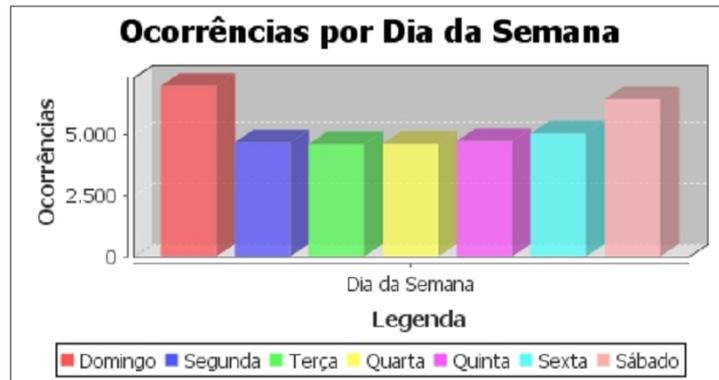
Terça: 4601 (12.413328%)

Quarta: 4621 (12.467287%)

Quinta: 4720 (12.734385%)

Sexta: 5033 (13.578849%)

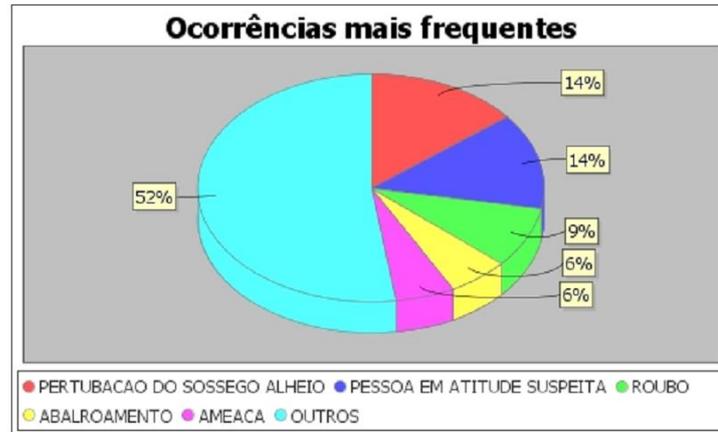
Sábado: 6443 (17.382977%)





## OCORRÊNCIAS MAIS FREQUENTES

PERTUBACAO DO SOSSEGO ALHEIO:5314 (14.336975%)  
PESSOA EM ATITUDE SUSPEITA:5018 (13.538379%)  
ROUBO:3176 (8.568731%)  
ABALROAMENTO:2113 (5.7007957%)  
AMEACA:2062 (5.5632%)  
LESAO CORPORAL:1913 (5.1612034%)  
ESTACIONAMENTO IRREGULAR:1190 (3.2105758%)  
FURTO:1060 (2.8598406%)  
SOLICITACAO DE POLICIAMENTO:1016 (2.7411304%)  
VIAS DE FATO:965 (2.6035345%)  
ENTORPECENTES E DROGAS AFINS:954 (2.5738568%)  
DISPARO DE ARMA DE FOGO:820 (2.2123299%)  
CHOQUE DE VEICULO:693 (1.8696884%)  
ARROMBAMENTO:691 (1.8642925%)  
DESORDEM:661 (1.7833537%)  
INVASAO :552 (1.4892756%)  
DANO / DEPREDAÇÃO:540 (1.4569%)  
APOIO A ÓRGÃO PÚBLICO:488 (1.3166059%)  
INFRACAO AMBIENTAL:471 (1.2707406%)  
INCENDIO:469 (1.2653446%)  
OUTROS: 6896 (18.605154%)



Anexo 2 – Comparação entre os meses de Janeiro de 2012 e Janeiro de 2013, criada pela funcionalidade Comparar Intervalos

## Relatório - CIOSP BI

Intervalo 1 - Início: 01/01/2012 - 00:00:00 Fim: 31/01/2012 - 23:59:59

Intervalo 2 - Início: 01/01/2013 - 00:00:00 Fim: 31/01/2013 - 23:59:59

Total de ocorrências: 2538 || 2800

Quantidade de dias: 30 || 30 ocorrências

Média: 84.6 || 93.333336 ocorrências por dia

### OCORRÊNCIAS POR ÓRGÃO

PMRN: 2148 (84.63357%) || 2584 (92.28571 %)

GMM: 0 (0.0%) || 13 (0.46428573 %)

PCRN: 65 (2.5610716%) || 36 (1.2857143 %)

ITEP: 31 (1.2214342%) || 14 (0.5 %)

CBRN: 293 (11.544523%) || 114 (4.071429 %)

TLIA: 1 (0.039401103%) || 39 (1.3928572 %)



**OCORRÊNCIAS POR DIA DA SEMANA**

Domingo: 586 (23.089046%) || 478 (17.071428%)

Segunda: 365 (14.381403%) || 292 (10.428572%)

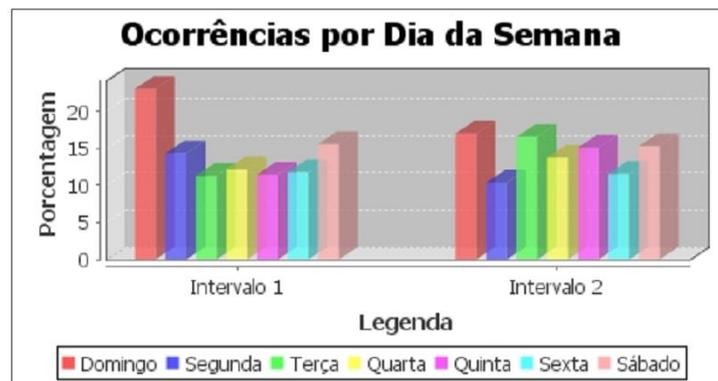
Terça: 288 (11.347517%) || 465 (16.607143%)

Quarta: 310 (12.214342%) || 386 (13.785714%)

Quinta: 292 (11.505122%) || 424 (15.142857%)

Sexta: 300 (11.820332%) || 325 (11.607142%)

Sábado: 397 (15.642238%) || 430 (15.357142%)





Anexo 3 – Análise do atendimento no ano de 2012, criada pela funcionalidade Análise do Atendimento

## Relatório - CIOSP BI

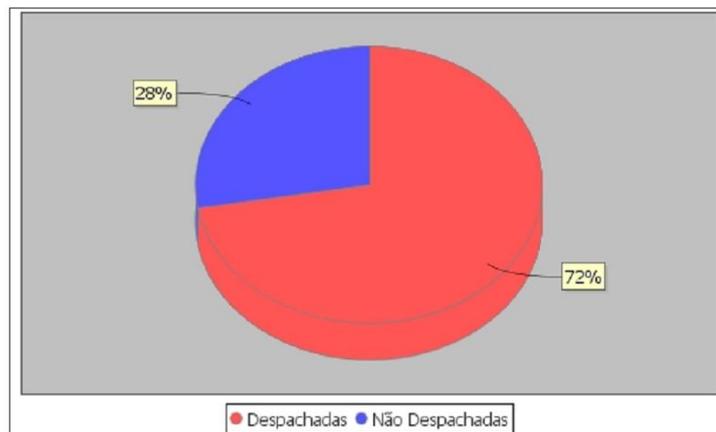
Início: 01/01/2012 - 00:00:00 Fim: 31/12/2012 - 23:59:59

### Despacho das ocorrências

Ocorrências despachadas: 26734 (72.30095%)

Ocorrências não despachadas: 10242 (27.699047%)

Tempo médio de despacho: 0 hora(s) 16 minuto(s) 49 segundo(s)



## Saída das viaturas

Alguma viatura se dirigiu à ocorrência: 23509 (63.579075%)

Nenhuma viatura se dirigiu à ocorrência: 13467 (36.42092%)

Tempo médio para alguma viatura se dirigir à ocorrência: 0 hora(s) 16 minuto(s) 17 segundo(s)

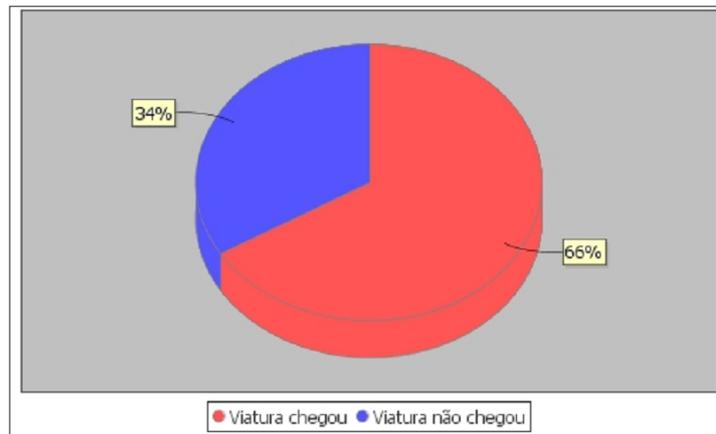


## Chegada das viaturas

Alguma viatura chegou ao local da ocorrência: 24535 (66.35385%)

Nenhuma viatura chegou ao local da ocorrência: 12441 (33.64615%)

Tempo médio para a viatura chegar ao local da ocorrência: 0 hora(s) 27 minuto(s) 48 segundo(s)

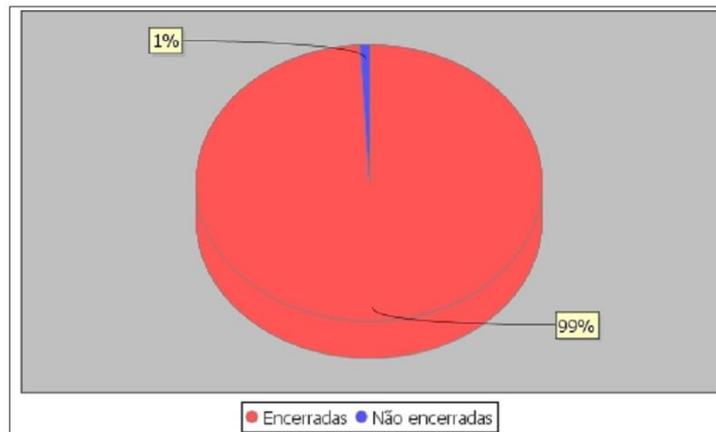


## Encerramento das ocorrências

Ocorrências encerradas: 36664 (99.156204%)

Ocorrências não encerradas: 312 (0.84379053%)

Tempo médio para encerramento: 1 hora(s) 17 minuto(s) 50 segundo(s)



## **Códigos de Finalização das ocorrências**

(ELOC) Endereço não localizado: 507  
(NLOC) Interessado não localizado: 1470  
(NR) Não resolvido: 657  
(OFR) Ofensor deixou a cena: 7243  
(RES) Resolvido no local: 7324  
(SAL) Sem alteração no local: 8555  
(SOL) Solicitante desistiu: 666  
(VTN) Viatura não disponível: 375  
(ORG) Encaminhado ao órgão competente: 1529  
(TRT) Trote: 292