

Walkir Alexandre Toscano de Brito

**ARCABOUÇO ANALÍTICO DE TEXTOS  
EM DENÚNCIAS: NOVAS TECNOLOGIAS  
NO TRATAMENTO E ANÁLISE DE  
DENÚNCIAS  
Estudo de Caso Milícia**



Instituto de Matemática



Instituto Tércio Pacitti de Aplicações  
e Pesquisas Computacionais

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
INSTITUTO DE MATEMÁTICA  
INSTITUTO TÉRCIO PACITTI DE APLICAÇÕES E PESQUISAS COMPUTACIONAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Walkir Alexandre Toscano de Brito

ARCABOUÇO ANALÍTICO DE TEXTOS EM DENÚNCIAS:  
NOVAS TECNOLOGIAS NO TRATAMENTO E ANÁLISE DE DENÚNCIAS  
ESTUDO DE CASO MILÍCIA

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Informática, Instituto de Matemática e Instituto Tércio Pacitti, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Informática.

Orientadora: Profa. Claudia Lage Rebello da Motta, D.Sc.

Rio de Janeiro  
2020

### CIP - Catalogação na Publicação

B862a Brito, Walkir Alexandre Toscano de  
Arcabouço Analítico de textos em denúncias: novas  
tecnologias no tratamento e análise de Denúncias -  
estudo de caso milícia / Walkir Alexandre Toscano de  
Brito. -- Rio de Janeiro, 2020.  
199 f.

Orientadora: Claudia Lage Rebello da Motta.  
Tese (doutorado) - Universidade Federal do Rio  
de Janeiro, Instituto Tércio Pacitti de Aplicações e  
Pesquisas Computacionais, Programa de Pós-Graduação  
em informática, 2020.

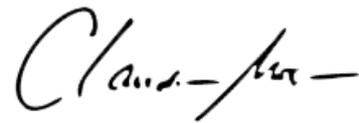
1. Mineração de dados. 2. Produção do  
conhecimento. 3. Arcabouço analítico. 4. Denúncias.  
I. Motta, Claudia Lage Rebello da , orient. II.  
Titulo.

Walkir Alexandre Toscano de Brito

**ARCABOUÇO ANALÍTICO DE TEXTOS  
EM DENÚNCIAS: NOVAS TECNOLOGIAS NO  
TRATAMENTO E ANÁLISE DE DENÚNCIAS  
ESTUDO DE CASO MILÍCIA**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Informática, Instituto de Matemática e Instituto Tércio Pacitti, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Informática.

Aprovada em 30/09/2020



---

Claudia Lage Rebello da Motta, D.Sc., PPGI/UFRJ (Orientadora)

participação por videoconferência

Claudio Miceli de Farias, D.Sc. – PPGI e NCE/UFRJ

participação por videoconferência

Daniel Serrão Schneider, D.Sc., PPGI e NCE/UFRJ

participação por videoconferência

Flávia Maria Santoro, D.Sc. - UERJ

participação por videoconferência

Joana da Costa Martins Monteiro, Ph.D. – FGV/IBRE

## AGRADECIMENTOS

Era início de 2016, quando recebi um e-mail avisando minha aprovação no Doutorado, uma mistura de emoção e medo em dar um passo tão relevante em minha vida, porém a certeza de um aprendizado inesquecível. Desde então, foi um mergulho profundo para entender um campo totalmente novo para mim e que escolhi como tema de meu caminho. O emprego de tecnologias a serviço da segurança pública.

Gostaria de começar agradecendo a todos que percorreram e trilharam esse caminho comigo, durante todo esse tempo, com momentos e objetivos diferentes, porém sempre juntos e compartilhando seus conhecimentos para a realização deste trabalho. Agradecer imensamente ao Zeca Borges, responsável pelo Disque Denúncia RJ, o famoso e carinhosamente chamado DD, por abrir as portas e dividir comigo toda a história do serviço e a sua própria história de vida, colocando-me em contato com todas as pessoas envolvidas na equipe e com contatos de pessoas ligadas a segurança do estado, uma rede de pessoas que pensa e age a cada dia na defesa e segurança de uma cidade melhor. Esse canal que tem como maior objetivo escutar a população através de suas denúncias e continuar a cobrar das autoridades que tenha responsabilidade e dê respostas ao clamor relacionado a violência de nosso estado. Reconhecer todo o esforço da equipe do DD, diariamente empenhada em manter o serviço funcionando apesar de todas as dificuldades enfrentadas em toda a sua história.

Aos meus amigos e companheiros da universidade, da escola e área de inteligência em segurança pública e de todos que aprenderam a conviver com minha ausência e minha necessidade em aprender o máximo possível para descobrir de que forma poderia contribuir, nessa área. Nesse momento, reconheço a importância de ter realizado o curso de inteligência em segurança pública que foi um diferencial necessário em minha busca e logo em seguida o curso de análise em segurança pública, junto a um mito da área de inteligência no Brasil, o coronel Romeu Ferreira. Obrigado por todos os ensinamentos doutrinários. Aproveito e agradeço a escola de segurança pública do Rio de Janeiro (ESISPERJ) e todos os seus membros em permitir meu acesso como aluno e agora como instrutor.

No entanto, o caminho é longo e perdemos amigos, familiares, tudo perde o sentido e não podemos perder nosso objetivo inicial, nosso foco diante das adversidades. Confesso que foi bastante difícil continuar nesse momento em que vivemos, a pandemia fez com que cada um reagisse de maneiras diferentes as perdas, aos medos as necessidades e principalmente aos nossos objetivos estabelecidos. Precisamos seguir pelos que se foram e pelos que continuam lutando por suas vidas ou para combater essa violência que enfrentamos diariamente. Deixo

aqui meus sentimentos aos familiares de todos aqueles que tive o prazer em conviver e aprender a cada dia, um pouco do que é preocupar-se e acolher o outro, com sua própria vida, pensando em uma sociedade melhor.

Meu agradecimento especial a minha orientadora e todos os Professores que tive até aqui, sem vocês não seria possível. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Gostaria de agradecer a todos os Professores da banca de defesa e os demais professores que tive pelo caminho, obrigado pelas possibilidades. Aos meus amigos que por algum motivo foram esquecidos neste momento, sintam-se abraçados e principalmente a minha família, a quem dedico esse trabalho, obrigado por todos os ensinamentos, todas as orações e pela fé nos momentos difíceis, e principalmente pela formação de caráter que carrego em mim para a vida. Sem vocês não seria possível, nem teria sentido.

Mãe, obrigado por me ensinar que tudo é possível, e que no final o importante é ser Feliz, Gratidão.

Sempre !!!!

## RESUMO

A cultura do valor da informação orientada a dados para a tomada de decisão, denominada *data driven*, vem sendo buscada por diversas organizações como suporte a tomada de decisão, na busca de eficácia, eficiência na finalização de seus processos de coleta de informações por órgãos e agências nas áreas de inteligência. No sentido de extrair dados, informações e produzir conhecimentos, a partir da integração de bases de dados, aliadas a mineração de dados textuais também conhecida como *text data mining*, e *knowledge discovery in texts*. Surge a partir da necessidade de descobrir de forma automática, informações (padrões e anomalias), na busca de encontrar conhecimentos. Abordagens para a denúncia de crimes têm sido amplamente discutidas, uma vez que grandes cidades têm enfrentado um aumento significativo de violência e de crimes relacionados a organizações criminosas, e que vem sendo vivido no Rio de Janeiro pelo crescimento do fenômeno chamado milícias.

Implementamos e utilizamos recursos tecnológicos aplicando mineração de dados textuais sensível ao contexto, para extrair conhecimentos de denúncias com a temática Milícias, buscando avanços e novas tecnologias em suas diferentes formas de atendimentos e canais de serviços, no Disque Denúncia (DD). Este estudo investiga perspectivas tecnológicas no atendimento para recepção, tratamento, análise e difusão, abrangendo novas formas de acolhimento dessas denúncias. Empregando inteligência estratégia em um arcabouço de mineração de dados textuais em textos de denúncias com a temática milícias em seu estudo de caso. Esse conjunto de ações e técnicas com tendências vindas da transformação digital consiste em converter dados brutos de denúncias em *insight*, auxiliando gestores e responsáveis na utilização fidedigna e específica dessas informações na tomada de decisão.

Este trabalho apresenta uma nova e ampla abordagem à análise de dados, incorporando o conceito de frações significativas ou pistas, utilizadas pela área de inteligência para a coleta de dados. Utilizamos técnicas de mineração de dados textuais – extração, stemming e clusterização – sobre o volume cada vez maior de dados textuais resultante dessas denúncias, para que o serviço consiga oferecer às autoridades policiais e as áreas de inteligência, informações úteis, bem encadeadas e com alto nível de confiança e aceitabilidade. Utilizamos conceitos e ferramentas de *data science*, *data driven*, *analytics*, *big data*, *BI*, *dashboard* e PLN na construção do modelo e em seus módulos aplicados neste estudo.

Propomos o uso de dados vinculados a uma abordagem que visa aprimorar o processo de criação de valores a partir da visualização de dados, fornecendo uma estrutura de avaliação

para analisar o impacto resultante baseado nas análises dessas informações e aplicamos em um estudo de caso abordando a aplicação destas ferramentas ao tema das milícias na base de dados do DD.

As tecnologias aplicadas ao estudo de caso, mostraram-se adequadas para a abordagem do problema apresentado, assim como para mineração de dados textuais contidos nas denúncias. Informações contidas nas denúncias por temas permitiu a criação de um dicionário específico e único de palavras, termos e semânticas utilizadas em diferentes tipos de crimes denunciados, nos permitindo empregá-los na construção das nuvens de palavras relacionadas as milícias.

**Palavras-chave:** Mineração de dados. Produção do conhecimento. Arcabouço analítico. Denúncias.

## ABSTRACT

The culture of the value of data-driven information for decision making, called data driven, has been sought by several organizations as support for decision making, in the search for effectiveness, efficiency in finalizing their processes of information collection by agencies and agencies in the areas of intelligence. In the sense of extracting data, information and producing knowledge, from the integration of databases, combined with textual data mining also known as Text Data Mining, and Knowledge Discovery in Texts. It arises from the need to automatically discover information (patterns and anomalies), in search of knowledge. Approaches to reporting crimes have been widely discussed, since large cities have faced a significant increase in violence and crime related to criminal organizations, and has been experienced in Rio de Janeiro by the growth of the phenomenon called Militia.

We implement and use technological resources applying context-sensitive textual data mining to extract knowledge of denunciations with the Militia theme, seeking advances and new technologies in its different forms of services and service channels, in the Disque Denúncia (DD). This study investigates technological perspectives in the reception, treatment, analysis and diffusion of these denunciations, covering new ways of receiving them. Employing strategic intelligence in a textual data mining framework in texts of denunciations with the militia theme in its case study. This set of actions and techniques with trends coming from digital transformation consists of converting raw data from denunciations into insight, assisting managers and decision-makers in the reliable and specific use of this information in decision making.

This work presents a new and broad approach to data analysis, incorporating the concept of significant fractions or clues, used by the intelligence area for data collection. We use textual data mining techniques - extraction, stemming, and clustering - on the increasing volume of textual data resulting from these reports, so that the service can provide law enforcement and intelligence agencies with useful, well-chained information with a high level of confidence and acceptability. We use concepts and tools of data science, data driven, analytics, big data, BI, dashboard and PLN in the construction of the model and its modules applied in this study.

We propose the use of data linked to an approach that aims to improve the process of value creation from data visualization, providing an evaluation structure to analyze the resulting impact based on the analysis of this information and we apply it in a case study approaching the application of these tools to the theme of militias in the DD database.

The technologies applied to the case study proved to be adequate for addressing the problem presented, as well as for mining textual data contained in the complaints. Information contained in the denunciations by theme allowed the creation of a specific and unique dictionary of words, terms and semantics used in different types of denounced crimes, allowing us to employ them in the construction of clouds of words related to militias.

**Keywords:** Text Data Mining. Knowledge Discovery in Texts. Analytical framework. Complaints.

## LISTAS DE FIGURAS

FIGURA 1: FLUXOGRAMA DE ORGANIZAÇÃO DO TEXTO.....	39
FIGURA 2: GESTÃO DO CONHECIMENTO E ATIVIDADE DE INTELIGÊNCIA.....	44
FIGURA 3: PRODUÇÃO DO CONHECIMENTO EM ATIVIDADES DE INTELIGÊNCIA.....	50
FIGURA 4: ANALYTICS E SISTEMA DE BI – INTEGRADO AO SISTEMA DE ATENDIMENTO PARA O PRODUTO – PROCURADOS.....	55
FIGURA 5: CONJUNTO DE FERRAMENTAS QUE PODERÃO SER UTILIZADAS NO BI.....	56
FIGURA 6: ARQUITETURA TÍPICA DE UMA SISTEMA BASEADO EM BUSINESS INTELLIGENCE...	57
FIGURA 7: RELAÇÕES E CATEGORIAS DE CLASSIFICAÇÃO.....	67
FIGURA 8: NUVEM DE PALAVRAS RELACIONADAS A CATEGORIA SUBSTÂNCIAS ENTORPECENTES.....	68
FIGURA 9: LIGAÇÃO TRÁFICO, PROCURADOS, MILÍCIA PARA ROUBO DE CARGAS EM 2017.....	70
FIGURA 10: CRIAÇÃO DAS ENTIDADES CAPTADAS E UTILIZADAS NAS DENÚNCIAS.....	76
FIGURA 11: INFORMAÇÕES E ENTIDADES QUE SÃO BUSCADAS NO ATENDIMENTO DAS DENÚNCIAS.....	77
FIGURA 12: DIAGRAMA DA ARRECADAÇÃO FINANCEIRA DA MILÍCIA.....	87
FIGURA 13: TERMOS CITADOS EM OUTROS ESTUDOS.....	92
FIGURA 14: LINHA TEMPORAL DO TOTAL DE DENÚNCIAS POR ANO E MÊS E A NUVEM DE PALAVRAS COM TEMAS ENCONTRADOS.....	93
FIGURA 15: CONTEXTO ATUAL DO DISQUE DENÚNCIA.....	100
FIGURA 16: SITUAÇÃO DE ATENDIMENTO ATUAL DO DD.....	101
FIGURA 17: DESENVOLVIMENTO DO SISTEMA BI INTEGRADO AO SISTEMA DE ATENDIMENTO .....	103
FIGURA 18: CENÁRIO PROPOSTO COM A INTEGRAÇÃO DE BASES EXTERNAS.....	104
FIGURA 19: METODOLOGIA DE DATA CLEANING EMPREGADA.....	111
FIGURA 20: MAIORES RELAÇÕES DE CLASSIFICAÇÃO DAS DENÚNCIAS.....	118
FIGURA 21: PALAVRAS, TERMOS E EXPRESSÕES UTILIZADOS NAS DENÚNCIAS POR LOCALIDADE.....	119
FIGURA 22: ENTIDADE PESSOA (PER).....	121
FIGURA 23: FLUXO DINÂMICO.....	122
FIGURA 24: PREPARAÇÃO BASE RASTRO A PARTIR DAS DENÚNCIAS (ENTIDADE PESSOA – PER) .....	122
FIGURA 25: RASTRO DOIS ÚLTIMOS 48 MESES DE UM ALVO.....	123
FIGURA 26: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES.....	127
FIGURA 27: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES PARA TIPOS DE CRIMES.....	128
FIGURA 28: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES.....	129
FIGURA 29: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES - TENDÊNCIAS.....	129
FIGURA 30: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES - MUNICÍPIOS	

.....	130
FIGURA 31: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES - ALVO.....	130
FIGURA 32: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES - CLASSIFICAÇÃO DE CRIMES .....	131
FIGURA 33: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES – MONITORAMENTO ANUAL DAS MILÍCIAS .....	132
FIGURA 34: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES – NARCO MILÍCIA.....	133
FIGURA 35: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES - FRANQUIAS .....	133
FIGURA 36: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES - DIFUSÃO	134
FIGURA 37: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES - ASSUNTOS .....	134
FIGURA 38: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES - QUEM.....	135
FIGURA 39: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES - NARCOTRAFICANTE.....	135
FIGURA 40: TELA DO SISTEMA AATDM E SUAS FUNCIONALIDADES DE ANÁLISES – CEMITÉRIO CLANDESTINO .....	136
FIGURA 41: TELA ONDE, NO SISTEMA DE INFORMAÇÃO DE MILÍCIAS - ONDE.....	136
FIGURA 42: TELA PARA QUEM FOI REALIZADA A DIFUSÃO DAS INFORMAÇÕES COLETADAS PELO SIM E QUAIS OS RESULTADOS OBTIDOS, RETROALIMENTANDO O SISTEMA DE INFORMAÇÃO.....	137
FIGURA 43: NUVEM DE PALAVRAS, EXPRESSÕES E TEMAS EMPREGADOS PELAS MILÍCIAS ...	138
FIGURA 44: ETAPAS DO PROCESSO DE MINERAÇÃO DE TEXTO.....	141
FIGURA 45: ETAPAS DO PIPELINE PLN REDUZIDO APLICADO AO ESTUDO DE CASO PARA O PROCESSO DE MINERAÇÃO DE TEXTO .....	142
FIGURA 46: ETAPAS DE MINERAÇÃO DE TEXTO .....	142
FIGURA 47: MODELO GENERALIZADO PROPOSTO .....	153

## LISTA DE QUADROS

QUADRO 1: COMPARAÇÃO ENTRE ESTADOS DA MENTE, TRABALHOS INTELECTUAIS E TIPOS DE CONHECIMENTO .....	52
QUADRO 2: COMPARAÇÃO ENTRE TIPOS DE CONHECIMENTO, ESTADOS DA MENTE, TRABALHOS INTELECTUAIS .....	53
QUADRO 3: ESTRUTURA DE UMA DENÚNCIA .....	65

## LISTA DE TABELAS

TABELA 1: TABELA DE STOPWORDS.....	144
TABELA 2: MATRIZ DE CLASSIFICAÇÃO GENÉRICA DE N CLASSES.....	149
TABELA 3: CONTRIBUIÇÕES DO TRABALHO....	159

## LISTA DE SIGLAS

ABIN	Agência Brasileira de Inteligência Nacional
AI	Agência de Inteligência
AISP	Agência de Inteligência em Segurança Pública
ALERJ	Assembleia Legislativa do Estado do Rio de Janeiro
AM	Aprendizagem de Máquina
APP	Aplicativo
APT	(Advanced Persistent Threat)
BI	Business Intelligence
BOPE	Batalhão de Operações Especiais
CC	Computação Cognitiva
CICC	Centro Integrado de Comando e Controle
CISP	Centro Integrado de Segurança Pública
CO	Crime Organizado
COAF	Controle de Atividades Financeiras
CPC	Ciclo Produção do Conhecimento
DD	Disque Denúncia
DDSYS	Sistema de denúncias do Disque Denúncia
DISPERJ	Doutrina de Inteligência de Segurança Pública do Rio de Janeiro
DHBF	Delegacia de Homicídios da Baixada Fluminense
DP	Delegacia Policial
DRACO	Delegacia Regional de Ações Criminosas Organizadas
ESISPERJ	Escola de Inteligência em Segurança Pública do Estado do Rio de Janeiro
ETL	Extração, transformação e carregamento
GAEGO	Grupo de Atuação Especial de Combate ao Crime Organizado
GIC	Gestão da Informação e do Conhecimento
GSI	Gabinete de Segurança Institucional da Presidência da República
IA	inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IBM	Internacional Business Machines
ISP	Instituto de Segurança Pública
LAI	Lei de acesso à Informação
MP	Ministério Público

MT	Mineração de Textos
ONG	Organização não-governamental
OPISP	Operações de Inteligência em Segurança Pública
ORCRIM	Organizações Criminosas
PLN	Processamento de Linguagem Natural
SC	Sistema Cognitivo
SESEG	Secretaria de Estado de Segurança
SSINTE	Subsecretaria de Inteligência em Segurança Pública
RI	Recuperação da Informação
RJ	Rio de Janeiro
TICs	Tecnologias da Informação e Comunicação
ZO	Zona Oeste

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	19
1.1 DISCUSSÃO DO PROBLEMA .....	30
1.2 MOTIVAÇÃO.....	31
1.3 OBJETIVO .....	32
1.3.1 OBJETIVO ESPECÍFICO.....	33
1.4 HIPÓTESE .....	33
1.5 TRABALHOS RELACIONADOS .....	34
1.6 CONTRIBUIÇÕES .....	36
1.7 ORGANIZAÇÃO DO TRABALHO .....	37
<b>2 DADOS DE INTELIGÊNCIA: COMPLEXIDADE E PRODUÇÃO DO CONHECIMENTO</b> .....	40
2.1 FORMAS DE CONHECIMENTOS PRODUZIDOS PELA ATIVIDADE DE INTELIGÊNCIA .....	41
2.2 PRODUÇÃO DO CONHECIMENTO E DADOS DE INTELIGÊNCIA.....	43
2.3 DADO, INFORMAÇÃO E CONHECIMENTO NA ÁREA DE INTELIGÊNCIA EM SEGURANÇA PÚBLICA.....	46
2.4 A PRODUÇÃO DO CONHECIMENTO NA ATIVIDADE DE INTELIGÊNCIA .....	48
2.5 VALOR DA INFORMAÇÃO ORIENTADA A DADOS PARA TOMADA DE DECISÃO .....	54
2.6 CONCLUSÕES DO CAPÍTULO .....	57
<b>3 ASPECTOS DA SEGURANÇA PÚBLICA NO RIO DE JANEIRO: O DISQUE DENÚNCIA E AS MILÍCIAS</b> .....	59
3 ASPECTOS DA SEGURANÇA PÚBLICA NO RIO DE JANEIRO: O DISQUE DENÚNCIA E AS MILÍCIAS .....	60
3.1 METODOLOGIA DE ATENDIMENTO E CANAIS DE ATENDIMENTO .....	62
3.2 ESTRUTURA DE UMA DENÚNCIA .....	64
3.3 SISTEMA DE CLASSIFICAÇÃO .....	65
3.4 TAXONOMIA DAS CLASSIFICAÇÕES .....	66
3.5 DESCRIÇÃO DA DENÚNCIA .....	68
3.6 TIPOS DE DENÚNCIAS.....	69
3.7 BASE DE DADOS DO DISQUE DENÚNCIA.....	71

3.8 ESTRUTURA DOS DADOS .....	72
3.9 CRIAÇÃO E IMPLEMENTAÇÃO DO CONCEITO DE ENTIDADES NO ROTEIRO DE ATENDIMENTO DAS DENÚNCIAS .....	74
3.10 O FENÔMENO MILÍCIAS .....	77
3.11 CPI DAS MILÍCIAS E SUA IMPORTÂNCIA EM ENTENDER O FENÔMENO.....	82
3.12 MODUS OPERANDI DAS MILÍCIAS NO RIO DE JANEIRO .....	85
3.13 ASSOCIAÇÕES E CORRELAÇÕES ENTRE OS ESTUDOS DE MILÍCIAS E A BASE DE DADOS DAS DENÚNCIAS .....	91
<b>4 PROPOSTA FRAMEWORK SENSÍVEL AO CONTEXTO EMPREGADO AO ESTUDO DE CASO MILÍCIAS .....</b>	<b>96</b>
4.1 CONTEXTUALIZAÇÃO DE CENÁRIOS .....	97
4.2 CENÁRIO REAL: ANTES DA IMPLEMENTAÇÃO DESTE TRABALHO .....	99
4.3 CENÁRIO PROPOSTO .....	101
4.4 MÓDULOS DO FRAMEWORK UTILIZADOS NO ESTUDO DE CASO .....	104
4.5 MÓDULO IDENTIFICAÇÃO:.....	105
<b>4.5.1 Ambiente seguro para carga e coleta de dados.....</b>	<b>106</b>
<b>4.5.2 Ações de coleta .....</b>	<b>107</b>
<b>4.5.3 Metodologia de data cleaning aplicada na base do disque denúncia .....</b>	<b>108</b>
4.5.3.1 Necessidade da limpeza de dados.....	108
4.5.3.2 Processo de limpeza dos dados (data cleaning).....	109
4.5.3.3 DATA CLEANING DESENVOLVIDA PARA O ESTUDO DE CASO MILÍCIAS .....	110
4.5.3.4 ESTRUTURA E ETAPAS: .....	113
4.5.3.4.1 Data cleaning - pré-processamento e geração de bases .....	113
4.5.3.4.2 Data cleaning: identificação e correções .....	114
4.5.3.4.3 Limpeza E Padronização De Logradouros Com Fuzzy Matching .....	115
4.5.3.4.4 Processo De Geocodificação Da Base Limpa .....	115
4.5.3.4.5 Utilizando O Serviço De Localização Here .....	116
4.5.3.4.6 Api Do Google Para Geração De Lat E Long .....	117
4.6 MÓDULO 2: INTERPOLAÇÃO .....	118
<b>4.6.1 Criação da base de personas e identificação dos XPTOS das palavras-chaves a partir das denúncias e a base de procurados .....</b>	<b>120</b>
<b>4.6.2 Criação da base de rastros a partir das denúncias.....</b>	<b>122</b>

4.7 CONCLUSÕES DO CAPÍTULO .....	124
<b>5 MÓDULOS: AGRUPAMENTO E ANÁLISES .....</b>	<b>125</b>
5.1 MÓDULO AGRUPAMENTO OU CLUSTERIG.....	126
5.2 MÓDULO ANÁLISE – PROCESSAMENTO DE TEXTOS DAS DENÚNCIAS .....	140
5.3 ETAPAS DO PROCESSO DE MINERAÇÃO DE TEXTOS .....	142
5.3.1 Coleta .....	143
5.3.2 Pré-processamento.....	143
5.3.3 Modelo .....	147
5.3.4 Métricas de avaliação .....	148
5.3.5 Análise dos resultados .....	149
5.3 CONCLUSÕES DO CAPÍTULO .....	150
<b>6 MODELO GENERALIZADO E CONSIDERAÇÕES FINAIS .....</b>	<b>151</b>
6 MODELO GENERALIZADO:.....	152
6.1 DESCRIÇÃO DO MODELO GENERALIZADO PROPOSTO .....	152
6.2 CONSIDERAÇÕES FINAIS E CONTRIBUIÇÕES.....	154
6.3 CONTRIBUIÇÕES .....	158
6.4 POSSIBILIDADES DE TRABALHOS FUTUROS: .....	166
<b>REFERÊNCIAS .....</b>	<b>170</b>
<b>APÊNDICE A - PILARES DO SERVIÇO DO DISQUE DENÚNCIA.....</b>	<b>180</b>
<b>APÊNDICE B - RELAÇÃO QUANTO A CLASSIFICAÇÃO DE DENÚNCIAS .....</b>	<b>181</b>
<b>APÊNDICE C – SISTEMA DE CLASSIFICAÇÃO DO DISQUE DENÚNCIA .....</b>	<b>182</b>
<b>APÊNDICE D - QUADRO DE DENÚNCIAS CADASTRADAS, QUE CONTÉM O XPTO “MILÍCIA” (OU A ESCRITA “MILÍCIA”) NO RELATO, DE ACORDO COM OS DEZ TIPOS DE ASSUNTO MAIS DENUNCIADOS, NOS ANOS DE 2003 A 2008 .....</b>	<b>183</b>
<b>APÊNDICE E - TIPO DE ERRO PARA OS OUTROS MUNICÍPIOS: ARMAÇÃO DOS BÚZIOS E CABO FRIO.....</b>	<b>185</b>
<b>APÊNDICE F - TIPO DE ERRO PARA O BAIRRO VILA KENNEDY .....</b>	<b>186</b>
<b>APÊNDICE G - ESTRUTURA E ETAPAS DE DATA CLEANING PARA PRÉ-PROCESSAMENTO E GERAÇÃO DE BASES UTILIZADAS NO MODELO .....</b>	<b>187</b>
<b>APÊNDICE H - DATA CLEANING: IDENTIFICAÇÃO E CORREÇÕES.....</b>	<b>188</b>
<b>APÊNDICE I - TRANSFORMAÇÃO DOS CARACTERES EM MAIÚSCULOS .....</b>	<b>189</b>
<b>APÊNDICE J - RETIRANDO ACENTUAÇÃO .....</b>	<b>190</b>

<b>APÊNDICE K - CORRIGINDO ERROS IDENTIFICADOS COM AS MAIORES FREQUÊNCIAS NA BASE.....</b>	<b>191</b>
<b>APÊNDICE L - IDENTIFICAÇÃO REALIZANDO CORREÇÃO DOS BAIRROS ..</b>	<b>192</b>
<b>APÊNDICE M - BASE NOVA ATUALIZADA E CORRIGIDA.....</b>	<b>193</b>
<b>APÊNDICE N - HERE CONFIGURAÇÃO ID DO APLICATIVO E A API KEY .....</b>	<b>194</b>
<b>APÊNDICE O - FLUXO CRIADO E UTILIZADO NO PROCESSO DO ALTERYX</b>	<b>195</b>
<b>APÊNDICE P - PADRONIZAÇÃO DA SAÍDA E VISUALIZAÇÃO DOS DADOS FINAIS.....</b>	<b>196</b>
<b>APÊNDICE Q - FLUXO DE TRABALHO CONFIGURAÇÃO DA APLICAÇÃO APP ID E O APIKEY .....</b>	<b>197</b>
<b>APÊNDICE R - API GOOGLE UTILIZADA NO PROCESSO .....</b>	<b>198</b>
<b>APÊNDICE S - JSON COM LATITUDE E LONGITUDE.....</b>	<b>199</b>

## 1 INTRODUÇÃO

Esta seção apresenta o tema desta pesquisa e as questões que a motivaram. Apresenta uma descrição sobre o problema, a metodologia utilizada neste estudo e seus objetivos. Por fim, aponta a importância deste trabalho para ações de inteligência estratégia em segurança pública.

Combater a criminalidade através de atividades de inteligência tem sido bandeira de campanha eleitoral em eleições municipais, estaduais e federais. Nas eleições de 2018, candidatos prometiam em seus discursos o endurecimento de leis penais, o fortalecimento do policiamento ostensivo e a revisão do Estatuto do Desarmamento, dentre outras medidas. A partir de 2019, ao Ministério da Justiça foi integrado o Ministério da Segurança Pública, Transparência e Controladoria Geral da União, além do COAF, tornando-se um superministério<sup>1</sup> – significativa mudança na segurança pública. O caso do Estado do Rio de Janeiro é ainda mais interessante e desafiador quando os aspectos socioeconômicos e geográficos do problema da violência são considerados. Esses aspectos permitem as aproximações, diferenciações e interações entre os espaços sociais estabelecidos na contemporaneidade e definidos através dos termos “favelas” e o “asfalto” (VENTURA, 1994). Pode-se dizer que uma parte da exacerbação da violência na atualidade decorre do convívio entre esses espaços preestabelecidos e dos conflitos oriundos dessas interações, demonstrando que questões ligadas à violência são problemas complexos. Historicamente, desde a década de 1990, observa-se uma ascensão da criminalidade no município e a crescente crise das instituições que tratam da segurança pública em todas as suas vertentes, somadas ao aumento da ligação entre tráfico de drogas e armas. E nesse contexto é que se desenvolvem as profundas e inevitáveis mudanças na dinâmica e caracterização da criminalidade e da violência na cidade (SILVA, 2013).

O caso especial de falência econômica enfrentado pelo estado, seu contexto político atual e suas especificidades humanas e de geografias física e urbana configuram grandes desafios para os agentes de segurança, assim como o cidadão morador desta cidade. O retorno da violência em áreas específicas e o aumento da criminalidade em algumas áreas e seus entornos são realidades enfrentadas no dia a dia da população e noticiadas com a mesma nos telejornais locais. As mudanças de planos e ações na política de seguranças acordadas pelo

---

<sup>1</sup> Notícia publicada em 01 de novembro de 2018. “‘É um superministério da Justiça’, diz Bolsonaro sobre pasta entregue a Moro “. Disponível em: <https://oglobo.globo.com/brasil/e-um-superministerio-da-justica-diz-bolsonaro-sobre-pasta-entregue-moro-23204742> . Acesso em: 19 fev. 2020.

estado e a não continuidade de ações bem-sucedidas na área de segurança são, sem dúvida, alguns dos maiores problemas enfrentados, bem como a falta de investimento em ciência, tecnologia, capacitação e treinamento. É inadiável o desenvolvimento de aparatos científicos e tecnológicos o que pressupõe o emprego de novas tecnologias para o tratamento e análise de denúncias, possibilitando aos tomadores de decisão ações específicas no combate à violência e ao crime organizado.

O setor de segurança pública busca novas tecnologias para dar suporte às suas atividades, estudos e operações. As instituições que atuam nessa área, com poucas exceções, possuem em sua estrutura uma ou mais unidades de inteligência, dedicadas ao tratamento e interpretação de grandes quantidades de dados. Atualmente, as atividades de investigação e os processos de análise criminal, denúncia e ocorrências envolvem receber, armazenar e processar dados das mais diversas fontes de informação, tais como: perfis em redes sociais, dados cadastrais de pessoa física e jurídica, informações sobre pessoas, fotos, veículos, áudios obtidos com processos de interceptação legal, informações via *web* e aplicativos móveis, dentre outras. Devido à diversidade da forma de coleta, os sistemas comumente empregados na área de inteligência em segurança pública possuem características tais como: bases disjuntas, dados heterogêneos e a dificuldade de guardar históricos longitudinais com características contínuas sobre o tema. A resolução desses problemas auxilia melhores análises para o desenvolvimento de investigações e de políticas públicas, bem como na geração e produção de conhecimentos aplicados a área em estudos necessários ao enfrentamento deste problema.

Nas análises mais profundas, bem como, no armazenamento, no tratamento e nos cruzamentos dos dados coletados surgem problemas e dificuldades de natureza operacional – derivadas da limitação tecnológica. A coleta, armazenamento, limpeza e cruzamento com outras bases de informações até a sua análise e produção de conhecimento para difusão aos órgãos competentes ficam comprometidos, devido as características das bases atuais. Problemas como bases disjuntas, não estruturadas e dados heterogêneos, como o dito anteriormente, são as principais características que atingem normalmente a realidade dessas bases que são utilizadas pelos sistemas empregados normalmente em órgãos e agências que trabalham com investigação, análises e dados de denúncias na composição de seus objetivos de investigação. Desse modo, faz-se necessária a busca por inovações tecnológicas que valorizem as informações que podem ser extraídas dos dados, bem como seus respectivos tratamentos, limpeza, e produção de conhecimento, possibilitando o monitoramento e visualização dessas informações. Assim, possibilita-se a criação de indicadores de performance que podem ser usados pela área de inteligência, bem como pelas áreas jurídicas e criminais.

É necessário e importante enfatizar que os relatos de denúncias podem e devem ser utilizados para esses fins e aproveitamos para explorar a possibilidade de questionar a forma em relação ao sigilo na geração do conhecimento, produzidas por essas áreas, que por questões culturais e doutrinários não comutam informações necessárias a resolução dos crimes ou delitos no decorrer de sua investigação, por exemplo. Em uma sociedade escassa em recursos, onde a produção do conhecimento torna-se um bem caro e necessário para produzir, carece em trocar e compartilhar informações, motivando a produção de conhecimento, torna-se urgente e necessária na resposta ao enfrentamento dessas organizações criminosas.

Diante desse panorama, setores de segurança ao redor do mundo vêm buscando tecnologias para dar suporte às suas operações em áreas de: inteligência, análise criminal, investigação, ciência criminal, produção do conhecimento, modelos preditivos, denúncias, redes sociais, monitoramento de ocorrências e alvos, visualização dos dados, terrorismo entre outros (MANSOUR, 2018).

Diversas iniciativas surgiram para auxiliar na resolução do problema da Segurança. Criado em 1995, o Disque Denúncia atua como um canal entre a população e os agentes de segurança pública no combate à criminalidade e solução de problemas que atingem comunidades, bairros e cidades do estado do Rio de Janeiro, recebendo as denúncias e coletando informações de importância fundamental na apuração de ações criminosas e, conseqüentemente contribuindo na redução da criminalidade.

O principal canal de comunicação da organização com a sociedade tem sido a linha telefônica, em um digitronco de número 2253-1177 que já recebeu mais de três milhões de denúncias ao longo dos últimos 25 anos. Mais recentemente, registra-se o incremento de denúncias realizadas através de mídias sociais, nas páginas de seus produtos nas redes, assim como a utilização de *WhatsApp*, e o Aplicativo do DD (App), dentre outras formas de coleta de informações, que já correspondem a cerca de 20% dos números de denúncias recebidas por meio desses canais digitais. As informações recebidas por estes canais têm como característica principal a forma não estruturada de dados, e principalmente a forma original de escrita de como essas denúncias são concretizadas, ou seja, uma escrita não formal e peculiar aos diversos grupos de denunciadores, exigindo uma ação de tratamento especial, assim como, a composição e compreensão simbólicas desses textos empregados pelos grupos, permitindo à construção de um dicionário de palavras, expressões e glossários utilizados em cada grupo, permitindo a classificação das denúncias por tipo e tema, respeitando-se sua escrita original e suas relações de entidades e relacionamentos, que serão empregados na produção do conhecimento resultante desse trabalho.

O fato da população com frequência demonstrar medo da exposição frente as incertezas de ameaças que sofrem ou pelo fato de muitas pessoas recusarem-se a testemunhar e exigirem garantias de sigilo para fazer é o que torna a pesquisa ainda mais desafiadora. Há vários fatores por trás disto, haja vista que a identidade dos membros das milícias, por exemplo, que precisa ser mantida em sigilo, diferentemente do narcotraficante tradicional em que sua fama é um componente primordial para sua dominação territorial. Basta considerar que os membros das milícias pareciam ser agentes do estado, basicamente policiais, o medo de que eles acabassem sabendo de qualquer depoimento era maior.

Conseqüentemente, é por isso que as denúncias tornam-se uma fonte especialmente importante na dinâmica, pois inúmeras categorias de crimes não são denunciadas diretamente ao Estado, seja por medo, ameaça ou até a própria descrença nas autoridades.

As denúncias ao DD podem ser realizadas porque seu anonimato é sempre garantido e precisa ser preservado para garantir sua efetividade. Quando falamos de crimes cometidos por funcionários públicos, como muitos membros das milícias, é normal que o cidadão denunciante não acredite que a polícia vá fazer qualquer tipo de investigação cabal na resolução destes crimes, permanecendo com medo de sofrer alguma consequência grave se for descoberto (LEMGRUBER, et al, 2003). Portanto a denúncia torna-se, a arma do cidadão, em conduzir o processo investigativo a seu favor (MENDONÇA, 2007).

Desse modo com o aparecimento dos novos canais digitais com a sociedade, expandiu-se as possibilidades na contribuição do relacionamento com o denunciante, fomentando a etapa de empatia, primordial para a concretização da denúncia. É a fase da ciência sobre a denúncia, emprega-se a técnica de *rapport*, técnica usada para criar uma ligação de sintonia e empatia com outra pessoa, o denunciante. Essas ações e relações contém importância na percepção da(s) denúncia(s), sendo extremamente necessária a aplicação de novos recursos tecnológicos nesta atividade, conforme veremos adiante.

A concepção do que nomeamos neste trabalho como sendo arcabouço de mineração de dados textuais, sensível ao contexto em denúncias, trata da aplicação de novos recursos tecnológicos no tratamento das denúncias, cruzamentos e visualização de dados, complexos e relevantes, com outras bases de dados na extração e geração de informações. Trata-se de um grande projeto de ciência, análise e visualização de dados que busca entender e da mesma forma que entende essa extração de informações, modifica o processo de extrair esses dados das denúncias. Cabe ressaltar que, pode ser utilizado para diferentes cenários e contexto criminais, permitindo ser adaptado conforme seu novo contexto. Configura-se como um trabalho científico de caráter exploratório, em um contexto de contribuição social, utilizando-

se de base de dados com muitos dados significativos ao tema e concebendo uma base de conhecimento extremamente relevante. Faz parte de nosso objetivo final propor um modelo generalizado, que pode ser aplicado em outras áreas e contextos, de forma automatizada ou não, porém sensível a diversos contextos criminais, tipos, temas e classificações de denúncias, estabelecendo desta forma novas possibilidades para trabalhos futuros.

Empregamos ao estudo técnicas de mineração de dados textuais sensível ao contexto, *knowledge base*, análise, monitoramento e visualização dos dados para a aplicação do estudo de caso Milícias.

A análise a partir das denúncias identificadas como Milícia deu-se pela importância sociopolítica que os grupos milicianos ganharam no Estado do RJ ao longo dos anos. Pressupõem os conceitos aplicados pela inteligência, com o que é denominado frações significativas, que são as pistas, entidades, provas, vestígios, a busca pelo significado, é trazer o significado através dessas pistas. Porque uma peça de um todo, pode representar um significado completo por si (RIO DE JANEIRO, 2005).

Trata-se, portanto, da extração e inferência de dados ou informações, que podem ajudar a solucionar um delito ou crime referindo-se as pessoas e/ou organizações e entidades, ligadas as Milícias. Cabe ressaltar, a importância dessa fonte de dados histórica contida nas denúncias, sendo muitas vezes a única fonte de dados histórica ao tema milícia, e com escala de tempo e registros necessários aos estudos exploratórios por apresentar aspectos longitudinais representativos no decorrer dos anos.

Neste contexto, compreende-se que a expansão das milícias no Rio de Janeiro ocorre em dinâmica de “franquias”, onde a “matriz”, ou o modelo utilizado para a implantação dessas franquias, seria o que foi implementado em Rio das Pedras, primeira comunidade dominada por milicianos, conforme poderá ser verificado e observado ao longo desta pesquisa.

Os milicianos da capital criaram o que os promotores do MP chamam de "franquias" ao definirem regras, tais como:

- ✓ Política de não-agressão entre os grupos;
- ✓ Uso, inicialmente, do nome anterior da quadrilha;
- ✓ Envio de integrantes a essas regiões;
- ✓ Ações de tomada de territórios com fardas das polícias Civil e Militar.
- ✓ Pagamento à matriz de percentual dos lucros obtidos pela nova quadrilha. (GRANDIN et al, 2018).

Essas informações referentes as franquias, são informações altamente relevantes por

demonstrar as formas de atuação desses grupos e como são diversificadas as maneiras de expansão de seus negócios. Destacar as oportunidades e perspectivas de franquias empregadas pelas milícias, bem como as possibilidades em que operam suas franquias de forma diferente em seus territórios, e identificar quais os grupos milicianos utilizam o mesmo nome, porém operam de forma diferenciada em seus territórios, são algumas das possibilidades que o analista de inteligência pode enfatizar na produção do conhecimento de seus estudos, utilizando a ferramenta proposta por esse trabalho, assim poderá buscar as pistas que podem gerar sucesso em uma operação: O dado ou informação que faltava para a resolução do caso ou a identificação de um indivíduo, grupo ou franquia. O trabalho desenvolvido e os tipos de conhecimentos produzidos seguem um padrão e norma já estabelecido na Doutrina de Inteligência em Segurança Pública (RIO DE JANEIRO, 2005), seguindo suas normas e procedimentos na produção do conhecimento produzidos exigidos. A visualização dessas informações deve seguir protocolos de segurança e diferentes agências utilizam e reconhecem a necessidade, importância e o valor da informação.

Empregar inteligência estratégica em denúncias, assessorando a tomada de decisões com base na análise de dados e informações que visa embasar melhores decisões, identificar oportunidades, reduzir custos e gerar a produção de conhecimento através dessas denúncias, é o princípio deste trabalho. Para alcançar o escopo orientado por dados específicos na tomada de decisão urge a necessidade de alocar os dados no centro de todo o planejamento de estudo. Respeitando todos os seus requisitos que são: útil, verdadeiro, avaliado, significativo, preciso, imparcial, objetivo, simples, bem-apresentado, oportuno e seguro (RIO DE JANEIRO, 2005). Esses requisitos fazem parte da DISPERJ e precisam ser respeitados na concepção do arcabouço desenvolvido neste trabalho.

Estas informações podem ser coletadas e utilizadas para diversos propósitos como geração de *insights*, *leads* e engajamento, melhoria das denúncias, dentre outras, visto que tais organizações empregam diversas fontes de dados, de distintos tipos, diferentes banco de dados, utilizando-se com esse propósito, de conceitos ligados a *Big Data* e as diversas soluções para cada necessidade e cenários específicos são trazidas e oferecidas pela tecnologia.

Este trabalho apresenta uma nova e ampla abordagem à visualização e análise de dados, incorporando o conceito de frações significativas ou pistas, utilizadas pela área de inteligência para a coleta e extração de dados em denúncias, empregando a mineração de dados textuais para extrair e produzir conhecimentos das denúncias, em uma central de atendimento.

Este estudo investiga perspectivas tecnológicas no atendimento, recepção, tratamento, análise, visualização, extração e difusão desses dados, ou seja, todas as formas de acolhimento

dessas denúncias, empregando um arcabouço analítico de mineração de dados em textos de denúncias com a temática milícias, em um estudo de caso, que serviu para testar e executar todas essas possibilidades computacionais envolvidas, buscando um aval e prerrogativa da concepção de um modelo generalizado a partir dessa experiência que será detalhado no final deste trabalho.

O conjunto de ações e técnicas com tendências vindas da transformação digital consiste em converter dados brutos de denúncias em *insight*, auxiliando gestores e responsáveis na utilização fidedigna e específica dessas informações na tomada de decisão. Utilizamos técnicas de mineração de dados textuais sensível ao contexto, *knowledge base* – extração, *stemming* e clusterização – sobre o volume cada vez maior de dados textuais resultante dessas denúncias, para que o serviço consiga oferecer às autoridades policiais e as áreas de inteligência, informações úteis, bem encadeada, com alto nível de confiança e aceitabilidade. Utilizamos conceitos e ferramentas de *data science*, *data driven*, *analytics*, *big data*, BI, *dashboard* e PLN (processamento de linguagem natural) na construção do modelo e em seus módulos aplicados neste estudo. Propusemos o uso de dados vinculados a uma abordagem que visa aprimorar o processo de criação de valores a partir da visualização de dados, fornecendo uma estrutura de avaliação para analisar o impacto resultante baseado nas análises dessas informações.

A implementação de ferramentas tecnológicas e análise de textos contribuem para uma economia de tempo, recursos financeiros e humanos. Permite corrigir problemas de escritas e limpezas dos dados existentes na estrutura atual, validando a importância deste trabalho. Ao trabalharmos com o armazenamento, tratamento e associação em grandes bases de dados, precisamos considerar também que esses dados precisam ser compreendidos a partir da natureza de suas fontes, de suas implicações, de seus relacionamentos e entidades. Assim tornou-se possível a transformação de informações relevantes em uma nova estrutura denominada ambiente analítico, empregado neste trabalho ao estudo de caso.

A complexidade para lidar com vários motivos, causas, fatores e dados quando tratamos desses assuntos, exige um esforço tecnológico muito grande de todos os envolvidos, bem como o emprego de novas ferramentas informatizadas que represente um conjunto de processos, procedimentos e tecnologia úteis, comportando e potencializando os recursos existentes. Deste modo, surgem novas possibilidades de tratamento, análises e difusão das informações produzidas entre os órgãos e agências integradas ao sistema de segurança do estado (WARDEN, 2010). Ansiamos por contribuir para melhores práticas nas políticas públicas de segurança, aliando eficiência, eficácia e efetividade aos processos e seus respectivos resultados.

Todo o armazenamento e diagnóstico diferenciado dessas denúncias tem por finalidade

a integração com outras bases de dados para a produção do conhecimento ao estudo de caso Milícia. Cabe ressaltar que todo o processo de produção do conhecimento utilizado por órgãos que tratam com informações de inteligência foram respeitados e aplicados ao estudo, assim como suas regras, de classificação referentes a fonte e conteúdo, preservando sempre o anonimato dos denunciante, preservando o pressuposto de um dos pilares do DD, a garantia do anonimato. Buscamos nos relatos de denúncia utilizados nesta pesquisa, as frações significativas. Isto é, entidades e informações relevantes para a composição dos cenários, envolvidos, vulgos<sup>2</sup>, carros, características, personas, rastros entre outras. Convém verificar a pertinência e credibilidade dos dados recebidos, bem como a relevância dessas frações significativas, que normalmente tratam-se de elementos com grau de classificação por fonte e conteúdo relevantes a investigação criminal ou delito, a(s) prova(s) de envolvimento(s), a(s) pista(s) de elemento(s), as redes de conhecimentos, a hierarquia da facção, a formação de grupos e envolvidos entre outros. São elementos utilizados neste estudo para compor, comparar e correlacionar com os aspectos essenciais dos assuntos ligados a Milícias e que não poderiam ser desconsiderados na criação do ambiente analítico empregado por conter toda a complexidade do tema milícias e suas relações na ciência do crime processada neste trabalho, como forma de produção do conhecimento para apoiar aos tomadores de decisão quanto ao assunto milícias. Neste sentido, levando em consideração as vertentes, as várias incógnitas e o envolvimento de todos os atores, precisamos recorrer aos sistemas complexos, a fim de abarcar a complexidade tanto das informações e dados quanto da interação entre os diferentes órgãos responsáveis pela inteligência e segurança (FURTADO, SAKOWSKI e TÓVOLI, 2015).

Apesar da contribuição para a resolução de crimes que o Disque Denúncia vem trazendo, o serviço ainda encontra alguns obstáculos que limitam o alcance pleno de seu objetivo principal, que é ser um agente facilitador para a resolução rápida e eficiente de crimes e violações aos direitos dos cidadãos. A primeira dificuldade diz respeito ao volume de informações recebidas pela central de atendimento diariamente – devido a uma recorrente divulgação do serviço pela mídia – somando-se ao aumento da criminalidade na cidade. Isto faz com que os atendentes recebam aproximadamente 380 denúncias por dia. O grande número de casos recebidos faz com que as áreas de análise tenham um gasto excessivo de tempo para organizar, classificar e analisar as informações recebidas. Há um trabalho ainda maior em

---

<sup>2</sup>Vulgo aqui é entendido como apelido e/ou codinome e não como os sentidos do dicionário que definem como o povo ou a maior parte das pessoas. Optou-se por manter a expressão desta maneira porque é termo recorrente na Segurança Pública.

associá-las aos respectivos sistemas classificatórios, para que elas sejam separadas e divulgadas às autoridades policiais responsáveis pelas denúncias individualmente. Cabe ressaltar também que há policiais, delegacias e/ou batalhões especializados para cada tipo de crime e essas informações geradas – que são utilizadas na concepção das cenas ou criação dos cenários investigáveis da ocorrência - servem para compor o que é denominado mapa de inteligência (MARTINS, 2009), que deve ser gerados de maneira mais organizada e eficiente possível e levantadas todas as informações necessárias a cada caso.

Além disso, cada palavra tem sua importância nessa cadência da denúncia, cada expressão e/ou termo utilizado possui um significado único e expressivo a cada grupo analisado e classificado. Trata-se muitas vezes de sentimentos de apelo, desespero, gritos de um abandono pelas autoridades competentes e suas responsabilidades sobre a população e território abandonado e sem qualquer representatividade e voz. Um idioma local, nativo, cheio de expressões e significados, uma linguagem não formal representada por expressões e gírias que precisam ser compreendidas nesse abismo social e racial em que se encontra o Estado. Ressalta-se que a desigualdade social deve ser considerada quando tratamos de segurança pública na compreensão do dia a dia para entendermos a complexidade dessas denúncias e o apelo do cidadão que busca ajuda. Assim, desenvolver um glossário de palavras, expressões, termos semânticos a cada tipo ou classe de denúncia no decorrer desses 25 anos de base, torna presumível um potencial estudo sobre o tema, mesmos com todos os desafios de se trabalhar em língua portuguesa e respeitando todas as características específicas da própria linguagem utilizada nas denúncias.

A denúncia produzida pela central e seus dispositivos digitais (Aplicativo, *messenger face*, *Whatsapp*), configura-se como uma ferramenta dinâmica que não se esgota no texto produzido pela correspondência e troca de informações com os atendentes. Depende de outros fatores e processos para consolidar sua legitimidade como instrumento de produção de conhecimento e informação, sobretudo quando caracterizada como delito ou ato criminal. Por isso, a importância de contribuições de estudo ligado a essa temática na consolidação e entendimento dessas questões e seus atores perante a área de inteligência. É importante considerar que uma denúncia anônima, mesmo que seja uma única denúncia, é um instrumento poderoso. Ela alcança todo seu potencial quando chega a tempo às mãos certas e ao agente certo – a quem compete fiscalizar ou investigar os fatos que ela aponta. A experiência de mais de duas décadas de atuação do DD mostra que o correto tratamento e utilização da informação favorecem o melhor desempenho das autoridades. Quanto maior a adesão da sociedade a esse canal de informação, aliada a uma maior e mais inteligente capacidade de gestão de dados,

maior se demonstra a eficácia, a eficiência e a efetividade de políticas públicas para a área. As denúncias recebidas todos os dias pelo serviço são avaliadas por analistas e encaminhadas às autoridades e aos sistemas de inteligência das polícias e agências para serem investigadas.

Algumas ideias de automatização desse processo de classificação vêm sendo testadas (KY; LEROY, 2014), no intuito de reduzir esse tempo e permitir que a atividade de inteligência em segurança amplie os conteúdos informacionais que são facilitadores das atividades de prevenção e da resolução dos problemas vinculados à violência e à criminalidade.

A pressão imposta pelas novas tecnologias e sua respectiva utilização sobre a sociedade faz com que o DD precise se adaptar à nova realidade e ao formato digital dessas denúncias. Estas novas formas de comunicação fazem com que seja necessário o desenvolvimento de um modelo analítico que combine tecnologias, processos de visualização de dados, técnicas de processamento de linguagem natural, *data mining* e mineração de textos em denúncias, que possibilitam estudos e análises de tais informações, permitindo uma melhor disponibilidade dessas informações para as áreas de inteligência e difusão, visando a contribuir de forma eficiente, aos estudos e aportando a tecnologia como suporte a essas transformações. Motivando e transformando a forma de como visualizar esses dados, permitindo traçar objetivos, medir resultados, planejar monitoramento, incorporar novas bases, desenvolver novas visões, entender seus significados, gerar conhecimento e usá-los na melhoria dessas ações e na tomada de decisões sobre o tema.

O acesso às informações, a geração de conhecimentos territoriais, a integração dos dados, a visualização desses dados, a análise textual empregada e a produção de conhecimento com as denúncias trabalhadas é o propósito fundamental que orienta este trabalho. A iniciativa é pioneira e está voltada ao desenvolvimento, capacitação e implantação de novos processos, tanto no que se refere ao processo de coleta da denúncia e seu tratamento textual quanto à entrega de novas respostas à sociedade. Neste modelo, são incorporados todos os temas já analisados e classificados pela instituição, tais como: violência doméstica, meio ambiente, tráfico de drogas, foragidos da justiça, organizações criminosas entre outros. A estratégia é preparar-se para atender às novas demandas presentes na sociedade e dessa maneira melhor cooperar com o denunciante a despeito dos diversos temas tratados em suas denúncias, fornecendo o encaminhamento das denúncias e cobrando das autoridades responsáveis as respostas para essas demandas. Contudo, a aplicação do modelo proposto neste trabalho utilizou como estudo de caso as Milícias.

Mais especificamente, no contexto dessa base de dados, torna-se necessário encontrar os vínculos entre esses dados, as relações de entidades existentes e principalmente entender

como estão geograficamente distribuídos. É preciso correlacionar a abordagem dos textos das denúncias e analisar as formas de escritas, bem como seus sentidos diferentes a cada grupo, permitindo a classificação e criação de glossário de sentidos e expressões empregadas nas denúncias e garimpar informações relevantes entre centenas ou milhares desses textos, características, dados, fotos, vídeos e objetos. Para isso, partimos de uma abordagem quantitativa e qualitativa que possibilitou a compreensão desse processo como captação de dados que geram informações (valores), e, posteriormente, conhecimento. Para obtenção de resultados dinâmicos e eficientes focamos na utilização dessas tecnologias para capacitar recursos humanos, agentes de segurança e inteligência em ferramentas de visualização e análise de informações em inteligência. Consideramos ainda que essa nova abordagem pode ser útil a potenciais pesquisadores interessados ao tema bem como uma nova fonte de informação a estudos concernentes as áreas afins.

Podemos citar como exemplo existente o sistema utilizado pela Secretaria de Estado da Segurança Pública de Santa Catarina, que adquiriu um produto de tratamento de *Big Data* para apoio à inteligência chamado “Guardião”<sup>3</sup>. Tal sistema é parte de uma plataforma que aplica vários conceitos com a capacidade de tratar de maneira eficiente um número muito maior de dados, onde agentes de segurança podem responder de forma mais efetiva à resolução e investigação de denúncias e crimes e, principalmente, também preveni-los e (ou) antecipá-los. Visualizando-se pela perspectiva da segurança pública, como no exemplo, fica fácil observar que *big data* não é algo que pode ser considerado uma moda momentânea das empresas de tecnologia, mas sim um conceito que já está sendo aplicado em benefício de toda a sociedade.

A partir da necessidade de integração de bases e produção do conhecimento em sistemas complexos usando denúncias com o tema Milícia, conjuntamente com o tratamento, limpeza, carga e agregação em grandes bases de dados, este estudo desenvolveu o que intitulamos arcabouço analítico para análise de textos, empregando ferramentas tecnológicas a um modelo proposto neste trabalho na base de denúncias do Disque Denúncia, e possibilitou o cruzamento de outras bases relacionadas ao tema, com o objetivo intrínseco de visualização de informações para estudos e análises. Novas formas e emprego de tecnologia em visualização de dados para a tomada de decisão, possibilita que informações contidas em denúncias e muitas vezes ignoradas devido a toda complexidade e a grande quantidade de dados recebidas muitas vezes

---

<sup>3</sup> Mais a respeito do software guardião pode ser encontrado nos links <http://claudiotognolli.com.br/saiba-tudo-sobre-o-guardiao-o-software-da-pf-que-grampeou-todos-os-corruptos/e> <https://securityinformationnews.com/2014/02/01/plataforma-guardiao-de-escutas-ganha-selo-feito-no-brasil/>

não sejam observadas, devido a questões como o volume de denúncia recebida, variedade ou velocidade na composição desses dados e/ou informações ocultas nesse contexto. Deste modo, foi criado um ambiente analítico experimental que estimula, apoia e registra novas formas de coleta, tratamento, análise e difusão de denúncias e informações de interesse dos órgãos de segurança e atividades de Inteligência.

### ***1.1 DISCUSSÃO DO PROBLEMA***

A necessidade de extrair e produzir conhecimentos no excessivo volume de denúncias, é hoje o maior problema enfrentado, seja pela grande quantidade de dados recebidas pelos canais existentes das denúncias ou o pouco poder de processamento existente na extração e mineração dos dados em suas bases de conhecimentos, que possam ser utilizados pelas áreas de inteligência, na produção e geração de conhecimento.

É inegável a importância dessas denúncias no decorrer dos anos, bem como sua escala representativa ao tema milícias, no entanto, a falta de aparatos tecnológicos para receber, limpar, trabalhar esses dados para futuras análises em o grande e diversos tipos e classificações número de denúncias recebidas, mostra-se como um grande e complexo desafio de projeto de ciência de dados neste trabalho.

Apoiar, construindo um instrumento que contribui socialmente, no enfrentamento a criminalidade através de seu uso pelas forças de segurança configura-se como parte deste problema, na extração de conhecimento, sensível ao contexto e na composição da base de conhecimento necessária para a extração das informações relevantes. Faz-se necessário a aplicação de um estudo de caso, para limpeza, mineração de dados, extração e geração de conhecimento, decorrer de especialista ao tema para sua validação, para desta forma entender e modificar ou ressignificar o processo de extração de conhecimento.

Trata-se, portanto, de um problema computacional complexo, que transforma a extração de dados em uma base de conhecimento sensível ao contexto. Podendo ser automatizada ou não, porém precisa ser independente de qualquer cenário ou tema, e pode ser empregado e aplicado a diversos temas, por exemplo, a violência a mulher, ou homicídios, ou milícias, entre outros.

O processo de entender o contexto e extrair dele conhecimento, sendo empregado a qualquer cenário, requer sensibilidade apurada e a aplicação de um estudo de caso, utilizando metodologia exploratória ao tema milícias. Fazer a PLN sensível ao contexto que é o objetivo, não o processamento ou o uso da PLN em si, mas a extração e inferência do conhecimento produzido para a base de conhecimento sobre milícias (*knowledge base*).

A ideia da construção do arcabouço conceitual sobre como fazer um projeto de análise e ciência de dados, sensível ao contexto e independente de um padrão ontológico a temática milícias é empregada a um estudo de caso.

Parte-se dos problemas analíticos enfrentados a seguir : a capacidade de extrair e produzir conhecimento através de denúncias e operacionalização do uso dessas informações; problema da produção de provas e material investigativo para a resolução de delitos e crimes, até hoje pouco utilizado como ferramental em denúncias. E, o problema de inteligência estratégica para a extração e produção de conhecimento a partir de dados para a área de inteligência com a temática milícias.

## **1.2 MOTIVAÇÃO**

As áreas de inteligência vêm buscando o emprego cada vez maior de tecnologias em seus serviços de coleta e análise de dados. Isso deve-se às novas formas de crimes, na sua maioria ligados a crimes cibernéticos que alcançam um número elevado de ocorrências nos últimos anos pelo mundo, principalmente com o desenvolvimento e utilização das redes sociais com seus alcances cada dia mais expressivo e globalizado. As instituições que atuam nessa área, com poucas exceções, possuem em sua estrutura uma ou mais unidades de inteligência dedicadas a coleta, tratamento e interpretação de grandes quantidades de dados. Normalmente, contam com poucos recursos financeiros, tecnológicos e principalmente pessoal técnico habilitado e treinado com as novas habilidades e competências necessárias para a função. Atualmente, as atividades de inteligência, investigação, análise criminal, denúncias e registros de ocorrências envolvem receber, armazenar e processar dados das mais diversas fontes de informação, tais como: perfis em redes sociais, dados cadastrais de pessoa física e jurídica, informações sobre pessoas, fotos, veículos, áudios obtidos com processos de interceptação legal, informações via *web* e aplicativos móveis, dentre outras. Devido à diversidade da forma de coleta, os sistemas comumente empregados na área de inteligência possuem as características ditas anteriormente como: bases disjuntas, dados heterogêneos, dificuldade de guarda e acesso de grandes históricos, executar análise utilizando a integração de várias fontes de dados de coleta, cruzamentos entre dados de denúncias e ocorrências de ligação ao serviço de emergência policial, entre outros.

Portanto, quando se tratam de coleta de dados, armazenamento, tratamento, cruzamentos dos dados coletados e análises mais profundas, surgem diversos problemas e dificuldades de natureza operacional – derivadas da limitação tecnológica.

Um ambiente seguro de coleta de dados, onde sejam feitas cargas de bases (ETL),

necessárias a estudos específicos costumam ser objetos distantes as áreas de inteligência por limitações estruturais, operacionais e/ou técnica.

A coleta, armazenamento, limpeza e cruzamento com outras bases de informações até a sua análise e produção de conhecimento para difusão aos órgãos competentes ficam comprometidos, conforme dito anteriormente, devido as características das bases atuais.

Problemas como bases disjuntas, não estruturadas e dados heterogêneos, são as principais características que abordam normalmente a realidade dessas bases que são utilizadas pelos sistemas empregados normalmente na área de inteligência, órgãos e agências, que trabalham com investigação, análises e dados, seja policiais, de denúncias, ou análise de redes sociais empregando a coleta na composição de seus objetivos de investigação. Desse modo, faz-se necessária a busca por inovações tecnológicas que valorizem as informações que podem ser extraídas dos dados, bem como o tratamento, limpeza, e aplicação de produção de conhecimento, possibilitando o monitoramento e visualização dessas informações e criando indicadores de performance, *dashboard* entre outros na atualização dessa realidade.

### **1.3 OBJETIVO**

Deste modo, este trabalho objetiva analisar o desenvolvimento de um ambiente analítico de coleta e extração de dados sensíveis ao contexto, em um ambiente seguro, que empregado na área de inteligência, contemplando textos de denúncias. Utilizamos conceitos e ferramentas de *data science*, *data driven*, *analytics*, *Big Data*, *BI*, *knowledge base*, *dashboard* e PLN.

A construção de uma proposta de um *framework*, com o estudo de caso contendo denúncias relacionadas a temática milícias, tendo como foco um projeto de ciência e análises de dados sensível ao contexto de forma que seus módulos tratem de questões referentes a carga, limpeza dos dados, extração de informações e mineração de dados significantes para a extração e produção de conhecimento sensível ao contexto, produzindo uma base de conhecimento, e permitindo a criação de uma proposta de um modelo generalizado empregado em diferentes cenários, contextos e composições na extração de conhecimentos específicos sensíveis ao contexto como contribuição deste trabalho.

Propomos o uso de dados vinculados a uma abordagem que visa aprimorar o processo de criação de valores e extração, geração e produção de conhecimento a partir da visualização desses dados pela implementação do arcabouço de mineração de dados textuais, fornecendo uma estrutura de avaliação das informações pelos *dashboards* criados para analisar o impacto e monitoramento da extração e geração do conhecimento produzido resultante baseado nas análises e visualização dessas informações.

### *1.3.1 OBJETIVO ESPECÍFICO*

Os objetivos específicos deste estudo são:

- ✓ Aplicar técnicas de análise textuais em relatos de denúncias e ilícitos de crimes na obtenção de dados e informações relevantes a temática milícias;
- ✓ Validar a aplicação de técnicas de mineração de dados, e análise textuais, sensível ao contexto como ferramenta de captação de informações e visualização de informações em denúncias;
- ✓ Contribuir na melhoria do atendimento para: qualidade da captação, número de denúncias relevantes, capacidade de armazenamento e tratamento dessas informações;
- ✓ Desenvolver uma proposta de um modelo de aplicação que pode ser utilizado por outros temas e contextos, a partir do estudo de caso, utilizando a base de dados do DD, e integrando as bases de dados aos órgãos e áreas de inteligência;
- ✓ Identificar padrões de ocorrências nas localidades estudadas para melhor difundir essas informações as autoridades competentes;
- ✓ Ampliar as possibilidades de análises, estudo e visualização de dados por áreas, temas e indivíduos, estabelecer perímetro de atuação e identificar redes de conhecimento ao tema milícias.

### *1.4 HIPÓTESE*

Esta pesquisa objetiva examinar a criação de um arcabouço analítico aplicando mineração de dados textuais sensível ao contexto, em denúncias, que permita a identificação de informações, bem como sua extração para a geração e produção do conhecimento, a partir dos relatos contidos nas denúncias com o tema milícias. Busca-se para isso, ferramentas de análise de textos, aplicando novas tecnologias em seu tratamento, correção e limpeza dos dados, permitindo a mineração de dados com o intuito de explicitar e gerar informações para produção de pesquisas, análises, base de conhecimento e visualização de dados importantes para a área de inteligência referente a temática Milícias.

Duas hipóteses conduzem o presente trabalho.

Hipótese 1 – (H1): O uso de um ambiente analítico com novas tecnologias, aplicando mineração de dados textuais sensível ao contexto, em denúncias, configura-se como um instrumento de coleta e extração de informações eficientes na produção do conhecimento sobre milícias para a atividade de Inteligência.

Hipótese 2 – (H2): A utilização de um arcabouço de mineração de dados textuais, sensível ao contexto em denúncias, e a cultura do valor da informação, orientada a dados para a visualização e extração de informações, que auxilie a tomada de decisão, agrega valor à produção do conhecimento a temática milícias.

### **1.5 TRABALHOS RELACIONADOS**

Na revisão de literatura a respeito de trabalhos relacionados, percebemos que são utilizadas algumas técnicas de *machine learning* (WANG *et al*, 2006), processamento de linguagem natural, relacionamentos de entidades nominais entre outros. Entretanto, há poucos estudos em língua portuguesa sobre linguagem natural aplicados quando em comparação ao número de estudos em língua inglesa, o que torna o presente trabalho bastante significativo ao tema. Não foram encontrados trabalhos com referências a denúncias na literatura com o foco em transformar dados em informações que devem ser usadas pelos analistas de inteligência na composição de um estudo de caso como, por exemplo, o caso das milícias abordado neste estudo. Foram encontradas pesquisas que auxiliam análises criminais, modelos preditivos de rotas de patrulhamento e monitoramento de áreas policiais. Além de pesquisas que envolvem *Big Data*, dados sobre terrorismo, análise de risco, maximização de efetivo de patrulhamento em uma determinada área entre outros (TOURE; GANGOPADHYAY, 2016; DAS *et al*, 2019). Estes trabalhos extraíam seus dados dos registros ou boletins de ocorrência. Não foram encontrados estudos que utilizassem dados como os do Disque Denúncia em que o processo de recebimento e análise de denúncias conta com a participação efetiva da sociedade civil na busca e auxílio de resoluções de problemas de segurança que assolam suas localidades, trazendo informações e ao mesmo tempo auxiliando aos órgãos de segurança com informações. Há também poucos estudos aplicados utilizando dados de segurança pública em língua portuguesa. Não foram encontrados até o momento, tratamentos de denúncias empregando o idioma português.

Objetivando apontar medidas que possam prevenir e minimizar os efeitos de conflitos e padrões espaciais de crimes (QUICK, 2018), possibilitando aos gestores e tomadores de decisões em políticas públicas e segurança por meio da informação gerada pela tecnologia sobre cada expressão, frase, contexto (LASLEY *et al*. 1995) em circunstâncias ou similaridades em textos das denúncias (FUENTES *et al*. 2012), tracem suas ações de inteligência, montem suas linhas de análise, investiguem a diferença entre as novas e as antigas expressões empregadas em cada delito e traçar formas e o *modus operandi* referente a cada grupo criminal envolvido e que possa assim investigar e diminuir o efeito da violência e territórios em conflitos, (TSENG

et al., 2012).

Consideramos que este estudo está relacionado a trabalhos que apontam para a compreensão da complexidade envolvida na problemática da segurança pública e seus campos de estudo. Aqui podemos abordá-los do ponto de vista social, tecnológico, através do emprego de centrais telefônicas no recebimento de denúncia, da utilização de tecnologias analíticas que auxiliem as operações policiais e atividade de inteligência de órgãos de segurança pública. O cruzamento das bases de dados desses órgãos sobre esses temas é algo tão raro e complexo de ser realizado, devido a poucas bases existentes na área. Há trabalhos que abordam o uso da tecnologia na gestão do conhecimento e integração entre as instituições da segurança pública, no contexto de suas bases de dados, aplicada à atividade de inteligência (SHEPTYCKI, 2017).

Procuramos nas publicações e nos periódicos científicos artigos que pudessem transpor essas relações, no emprego de tecnologia, algoritmos e aspectos tecnológicos envolvidos, nos atendimentos e centrais telefônicas relacionadas ao tema, na coleta, agrupamento e relação de disseminação da informação em inteligência, na prevenção e redução de crimes em determinadas áreas, em modelos computacionais para medir atratores de crimes, na detecção e mapeamento de pontos críticos, em algoritmos e clusterização para encontrar comportamentos semelhantes e relacionados em crimes, no uso de BI para melhor alocação de recursos para combate ao crime (NETO CASTRO, 2014), no vocabulário e temas comuns ou tendências cenários de relatórios de crimes, na rede de termos utilizados na mineração de textos, em crimes econômicos, na mineração de dados de crimes, na integração de redes sociais, em locais perigosos para crimes sexuais e na atividade de inteligência policial, integradas a previsão de crimes e cybercrimes entre outros (BOWEN, 1994; HEWITT *et al.* 2018; WANG *et al.* 2006).

Há estudos que avaliam a tecnologia para coletar, agrupar relatar e disseminar a inteligência de maneira oportuna e precisa para apoiar a tomada de decisão o emprego de *business intelligence* no policiamento de forma a prevenir ou reduzir crimes utilizando ferramentas de BI, integração de bases e *analytics* na previsão de crimes utilizando redes sociais e crimes cibernéticos (VOMFELL, HÄRDLE, LESSMANN, 2018). Encontramos trabalhos brasileiros em que os dados analisados seguem a mesma abordagem de levantamento operacional entre áreas geográficas de policiamento *versus* índices de criminalidades e emprego de tecnologia no monitoramento e redução desses índices.

Podem também ser vistos um conjunto de artigos que representa o conhecimento compartilhado tendo como exemplo encontrado, de mineração de dados criminais auxiliando o trabalho policial, aumentando a eficiência na coleta de informações de dados de suspeitos. Neste caso foram bastante utilizados dados abertos e de vital importância no campo

investigativo e de atividade de inteligência, análise de redes criminais e análise de vínculo de organizações criminosas e suspeitos de atividades ilícitas em grandes conjuntos de dados, análise de *links* e navegabilidade para crimes cibernéticos e violação de narcótico, terrorismo e sequestro podem facilitar o trabalho policial, permitindo que os investigadores encontrem maiores detalhes (KU *et al* 2011, CHEN *et al.* 2004, XU; J.J., 2004, 2006; WARDEN, 2010).

O tema da violência compreende a participação de diversos atores e espaços de análise e de reflexão. Assim, incorporam-se novas e diferentes formas de enfrentamento ao fenômeno da insegurança, da violência e da criminalidade ao contexto. O problema da (in)segurança vem causando preocupação recorrente aos governos, ensejando que a atividade como esta não esteja restrita a questões que dizem respeito apenas marcação das áreas de conflito no Estado, mas também da possibilidade de uma sociedade mais digna nesses territórios.

Diferentes campos de estudo empregam a busca de um conjunto de diagnósticos e possibilidades de projeções, no sentido de identificar cenários de risco aos atores envolvidos nessa temática e mitigar situações de conflitos, oferecendo, por meio da tecnologia, atividades de diagnóstico e prevenção a um tema tão importante para o Estado, e podendo, assim, agir em prol da defesa da sociedade e do cidadão. A busca e análise realizada permeou diversos campos de estudos e saberes nas publicações de periódicos referentes ao tema. Mostrando-se bastante eficaz para esse levantamento e a integração de bases de informação nessas diversas áreas e saberes.

As contribuições teóricas alcançadas com esta pesquisa dizem respeito às possibilidades que as atividades e o uso de tecnologia podem trazer para o campo da segurança pública e da defesa do cidadão. Pode tornar-se um aliado a descoberta e prevenção de áreas e contextos violentos territoriais (LOECK *et al.* 2016), e serve para auxiliar a tomada de decisões em políticas públicas e a inteligência estratégica em segurança.

## **1.6 CONTRIBUIÇÕES**

Este trabalho favorece a melhoria no processamento, análise e encaminhamento das denúncias recebidas pela central de atendimento, produzindo uma base de dados que pode ser compartilhada e transmitida aos órgãos de inteligência, bem como às autoridades responsáveis. Para tanto, torna-se necessário classificar e relacionar uma grande quantidade de dados e informação de maneira que seus sentidos e valores sejam claros e úteis, e ainda, classificar suas fontes, permitindo desse modo, aumentar a eficiência das informações recebidas. Esta pesquisa também subsidia:

- ✓ o aumento da qualidade na captação de informações;
- ✓ o aumento do número de denúncias relevantes por atendentes;
- ✓ o aumento da capacidade de armazenamento das informações;
- ✓ o aumento da capacidade de tratamento das informações, produção de análises customizáveis, análises e relatórios;
- ✓ o aumento da capacidade de otimização e envio de informações aos órgãos competentes;
- ✓ o a aumento da integração das bases e automatizar os processos de carga, limpeza e troca de informações (ETL).

Nosso estudo auxilia a conversão de dados brutos de denúncias em *insight*, auxiliando gestores e responsáveis na utilização fidedigna e específica dessas informações na tomada de decisão. Essas técnicas são largamente utilizadas no mundo dos negócios, mas são escassos os trabalhos que aplicam efetivamente essas técnicas em dados de segurança pública.

De tal modo, este trabalho possibilita a aplicação dessas técnicas com base na análise de informações recebidas pela central de atendimento, realizando uma análise direta e melhor difusão dessas denúncias, visando a incentivar a utilização dessas tecnologias na área de inteligência em segurança pública e entre os órgãos e áreas de inteligência.

### **1.7 ORGANIZAÇÃO DO TRABALHO**

Este trabalho está dividido da seguinte maneira: parte introdutória – apresentando a discussão do problema, a motivação, o objetivo, a justificativa que levaram ao início desta pesquisa, a hipótese e a metodologia proposta para responder as questões levantadas na primeira e segunda parte da hipótese.

No capítulo 2, apresentamos as definições contendo todos os conceitos necessários ao entendimento do problema, suas características, complexidade e cuidados em trabalhar com dados sigilosos e de segurança, os chamados dados sensíveis. Mostraremos a normativas e a doutrina necessária para trabalhar com esses dados sigilosos para produção do conhecimento pela área de inteligência. Como são as relações de troca de informações entre os órgãos, agências e áreas. Todas essas questões são bastante relevantes para a criação e atualizações do arcabouço de texto empregado no estudo de caso.

No capítulo 3, é apresentado o Disque Denúncia, sua metodologia de atendimento em todos os canais, sua história e pilares, suas premissas em sua metodologia e técnicas de acolher uma denúncia, a estrutura de uma denúncia, sua classificação, o que caracteriza uma denúncia ou atendimento, a base de dados geradas por essas denúncias, suas características e a implementação das entidades nomeadas em sua metodologia de atendimento para incrementar

a produção do conhecimento gerado por essas denúncias ao arcabouço textual das milícias, que foi empregado em nosso estudo de caso neste trabalho. Este capítulo também apresenta o problema das milícias no estado do Rio de Janeiro. São apresentadas as bases históricas para seu surgimento e consolidação ao longo do tempo, seu *modus operandi* e o modo como aparecem na base de dados, bem como a relação da base de dados com estudos relevantes ao tema milícias, demonstrando dessa forma todo o potencial da base de dados e aspectos e informações que não foram abordados por esses estudos no decorrer do tempo.

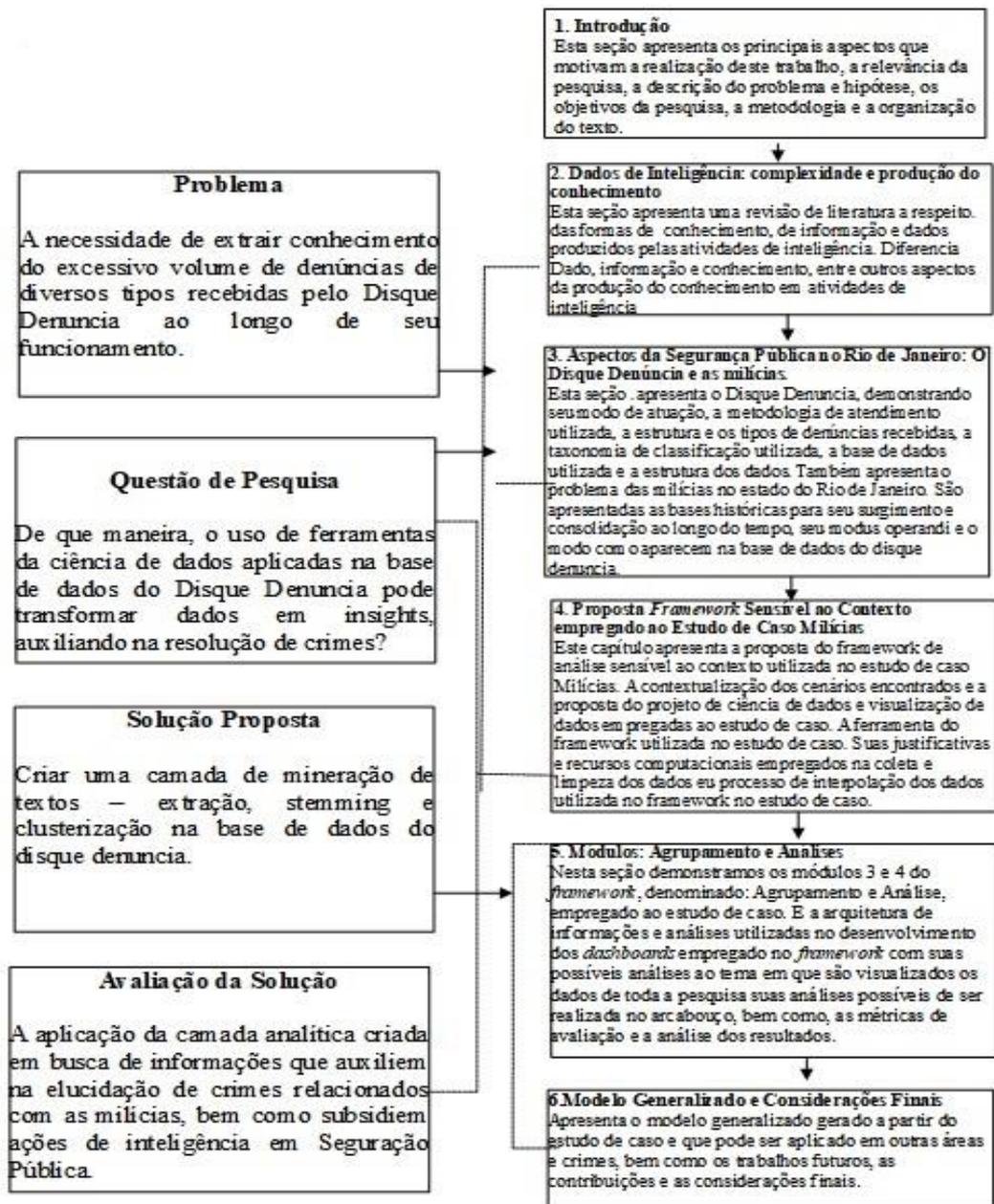
O capítulo 4 apresenta a proposta do *framework* de análise sensível ao contexto utilizada no estudo de caso Milícias. A contextualização dos cenários encontrados e a proposta do projeto de ciência de dados e visualização de dados empregadas ao estudo de caso. A ferramenta do *framework* utilizada no estudo de caso. Suas justificativas e recursos computacionais empregados na coleta e limpeza dos dados e processo de interpolação dos dados utilizada no *framework* no estudo de caso.

No capítulo 5 demonstramos os módulos 3 e 4 do *framework*, denominado: Agrupamento e Análise, empregado ao estudo de caso. E a arquitetura de informações e análises utilizadas no desenvolvimento dos *dashboards* empregado no *framework* com suas possíveis análises ao tema. Onde são visualizados os dados de toda a pesquisa suas análises possíveis de ser realizada no arcabouço, bem como, as métricas de avaliação e a análise dos resultados.

Por último, no capítulo 6, é apresentado o modelo generalizado gerado a partir do estudo de caso e que pode ser aplicado em outras áreas e crimes, bem como os trabalhos futuros, as contribuições e as considerações finais.

A seguir mostraremos as relações dos capítulos e suas seções deste trabalho e de que maneira elas interagem para a concepção e realização deste trabalho: arcabouço Analítico de textos em Denúncias: Novas Tecnologias no Tratamento e Análise de Denúncias – Estudo de Caso Milícias. Este trabalho foi estruturado em seis capítulos. Abaixo apresentamos graficamente a organização do texto.

Figura 1: Fluxograma de organização do texto



## **2 DADOS DE INTELIGÊNCIA: COMPLEXIDADE E PRODUÇÃO DO CONHECIMENTO**

Neste capítulo, apresentamos os conceitos necessários ao entendimento do problema e suas características. Mostra a complexidade e a segurança necessária para trabalhar com dados sigilosos, as relações de troca de informações entre os órgãos e agências de Inteligência. E discute o papel do Disque Denúncia neste contexto, com seus dados e informações recebidas pelos denunciante.

Neste capítulo, objetivamos demonstrar os fundamentos teóricos e metodológicos para a produção de conhecimento em atividades de inteligência. Em particular, descrever ideias de método e metodologia aplicada na produção do conhecimento em atividades de inteligência, reconhecer a importância do seu uso cotidiano através de sua metodologia chamada de ciclo de produção do conhecimento utilizada para essa atividade e, principalmente como e quais são os modos de produção do conhecimento utilizado na área de inteligência, seus princípios e os métodos para compreender a produção da verdade e fontes de informação em atividades de inteligência. Faz-se necessário apresentar as características, terminologia e conceitos da informação e do conhecimento no campo e área de domínio da atividade de inteligência para compreender esses recursos no estudo de caso que foi aplicado a milícia, foco deste trabalho, e encaminhar conceitos da área de inteligência, presente na Doutrina de Inteligência (2005), para explicar como as informações e o conhecimento são produzidos nesta área.

Isto posto, torna-se notório que o problema de milícia, do crime organizado, do tráfico de drogas e de armas e da segurança pública, por envolver diversas variáveis que se influenciam mutuamente (pobreza urbana, desemprego, crime, violência, homicídio, armas, tráfico de drogas, crescimento econômico entre outros) possui características bastante marcantes com indicações e contextualizações complexas para a produção do conhecimento na atividade de inteligência. Isso é dificultado ainda por não haver um centro integrado de estudos com informações e dados para suporte a área de inteligência que possa contribuir para a produção de material investigativo, normalmente necessário nas relações complexas entre as diversas variáveis envolvidas neste contexto, não deixando de fora a distribuição geográfica desses acontecimentos.

## ***2.1 FORMAS DE CONHECIMENTOS PRODUZIDOS PELA ATIVIDADE DE INTELIGÊNCIA***

São considerados sigilosos e/ou estratégicos, os conhecimentos que possuem como características alto valor agregado e cujo acesso não autorizado pode comprometer a consecução dos objetivos nacionais, estaduais e/ou municipais, e que possam resultar em algum tipo de prejuízo ou dano ao País. São também chamados de conhecimentos sensíveis e necessitam de medidas especiais de proteção. A produção de conhecimento da atividade de inteligência define característica em sua Doutrina como:

[...] por meio de metodologia específica, denominada de Ciclo de Produção

de Conhecimento (CPC), reúne dados e conhecimentos e transforma-os em novos conhecimentos, com a finalidade de assessorar os usuários no processo decisório. (RIO DE JANEIRO, 2005, p. 10)

Cabe ressaltar que neste trabalho, estamos sempre direcionando ao tipo de conhecimento, que é produzido e trabalhado pelos órgãos de inteligência, ou atividade de inteligência, ou inteligência de Estado, ou defesa, e/ou segurança pública. Todo esse conjunto de nomes e normas são procedimentos adotados para a produção de conhecimentos considerados sensíveis pelos canais de inteligência, e aqui estão sendo considerados e respeitados garantindo o padrão estabelecido pela DISPERJ (RIO DE JANEIRO, 2005) e pelo Programa Nacional de Proteção ao Conhecimento estabelecido pela ABIN, órgão competente, segundo:

O Art. 4 da Lei nº 9.883, de 7 de dezembro de 1999, estabelece a competência da ABIN para planejar e executar a proteção de conhecimentos sensíveis, relativos aos interesses e à segurança do Estado e da sociedade. Assim, a ABIN, como órgão responsável por planejar e executar a proteção dos conhecimentos sensíveis, que são “aqueles que, por sua natureza e potencial necessitam de medidas especiais de proteção, tendo em vista sua importância estratégica para a defesa dos interesses nacionais e a segurança do Estado e da sociedade” (BRASIL, 1999b), criou o Programa Nacional de Proteção ao Conhecimento (PNPC), cuja responsabilidade de execução está no Departamento de Contra-Inteligência da ABIN. (ABIN, 2015)

A atividade de inteligência é o exercício permanente de ações especializadas orientadas para a obtenção de dados, produção e difusão de conhecimentos, com vistas ao assessoramento de autoridades governamentais nos respectivos níveis e áreas de atribuição para o planejamento, a execução e o acompanhamento das políticas de Estado. Engloba também a salvaguarda de dados, conhecimentos, áreas, pessoas e meios de interesse da sociedade e do Estado. Diante disso, fica evidente a necessidade de consolidação de uma cultura de proteção do conhecimento considerado sensível.

Para esse fim, é necessária a associação entre as instituições públicas e privadas, os estados, as agências e os órgãos e as atividades de inteligência que produzem, manipulam, processam e difundem esses conhecimentos no alcance, identificação a qualquer antecipação de identificação de ameaças e oportunidades, principalmente, para o apoio do processo decisório, em nível estratégico, das diversas esferas governamentais, relacionadas à atividade de inteligência de Estado, e especialmente a contrainteligência, na proteção e gestão do conhecimento produzido. Veremos em nosso estudo, uma proposta de ambiente seguro para coleta de dados e informações utilizadas e uma sugestão de boas práticas que deve ser seguido e compartilhada junto ao CICC e seu órgão de inteligência.

## **2.2 PRODUÇÃO DO CONHECIMENTO E DADOS DE INTELIGÊNCIA**

A produção do conhecimento e a atividade de inteligência trabalham com uma metodologia própria e clássica na produção e salvaguarda de seus conhecimentos, que são utilizados para assessorar a tomada de decisão de interesse da segurança e em proveito dos órgãos de segurança. O uso adequado de procedimentos específicos e de técnicas de assessoria voltadas para a produção de conhecimento excluem práticas de ações intuitivas e a adoção de procedimentos sem orientação racional (RIO DE JANEIRO, 2005).

A produção do conhecimento compreende um adequado tratamento pelos profissionais de inteligência de dados, informações e conhecimentos. Esses profissionais centram-se no que é denominado ciclo de produção do conhecimento, para tomada de decisão, utilizando-se desses dados, informações e conhecimentos na produção e difusão desses novos conhecimentos, seus monitoramentos, análises de vínculos, construção de cenários e perspectivas de curto, médio e longo prazo.

Para que a atividade de inteligência apoie o processo decisório, a partir da produção de conhecimentos, deve ser orientada através de um planejamento estratégico, onde sua atuação deve ser em colaboração com a gestão e envolve ações de curto, médio e longo prazos na criação e construção de cenários que devem ser continuamente acompanhados no intuito de preservar a ordem pública.

Autores como Drucker (1992; 1993), Stewart (1998), Sveiby (1998) e Davenport (1998) afirmam que estamos no limiar de uma nova era na qual o conhecimento é reconhecido como o principal ativo das organizações e a chave para uma vantagem competitiva sustentável.

O conhecimento e sua gestão são produzidos pela Agência de Inteligência de Segurança Pública e deve sobrepujar os requisitos básicos dos dados e/ou conhecimentos, em outros conhecimentos, com essas características: avaliado, verdadeiro, significativo, preciso, imparcial, objetivo, útil, simples, bem-apresentado, oportuno e seguro nas seguintes situações:

- ✓ De acordo com um Plano de Inteligência;
- ✓ Em atendimento à solicitação de uma Agência de Inteligência congênere;
- ✓ Em atendimento à determinação da autoridade competente;
- ✓ Por iniciativa própria.

A produção do conhecimento em Inteligência de Segurança Pública deve ser executada e gerenciada por pessoas treinadas e capacitadas diante das necessidades e pré-requisitos produzidos pelos órgãos de segurança pública pois abordam questões sensíveis a investigação

e seus ilícitos. Tratam-se normalmente de informações do tipo estratégicas, gerenciais e operacionais e estão ligadas a decisão, gestão e operação e seus tomadores de decisão, Figura 1. Essas informações são importantes na resolução de casos e refletem normalmente todas as formas de violências aplicadas à sociedade, sendo de vital importância aos agentes de inteligência do estado e suas linhas e pastas de investigação.

**Figura 2: Gestão do Conhecimento e Atividade de Inteligência.**



Fonte: elaboração própria.

No Brasil, utilizamos a definição de Inteligência do Subsistema de Inteligência de Segurança Pública (SISP), criado pelo Decreto n° 3.695, em 21 de dezembro de 2000, que consiste em:

Identificar, acompanhar e avaliar ameaças reais ou potenciais de segurança pública e produzir conhecimentos e informações que subsidiem ações para neutralizar, coibir e reprimir atos criminosos de qualquer natureza<sup>1</sup> (BRASIL, 2002).

Nesse contexto, e para que se compreenda a dimensão da atividade de inteligência, ocorre-se ao conceito disposto no artigo 1º, § 2º, do Decreto n° 4.376 de 2002, que regulamentou a Lei n° 9.883 de 1999, onde:

inteligência é a atividade de obtenção e análise de dados e informações e de produção e difusão de conhecimentos, dentro e fora do território nacional, relativos a fatos e situações de imediata ou potencial influência sobre o processo decisório, ação governamental, a salvaguarda e a segurança da sociedade e do Estado. (BRASIL, 2002)

O estado do Rio de Janeiro, foi o primeiro a criar, em 01 de janeiro de 1995, com base

na doutrina da inteligência militar, uma Agência de Inteligência, diretamente ligada ao Secretário de Segurança, chamado Centro de Inteligência de Segurança Pública, até 2019, denominado Subsecretaria de Inteligência. Criada com o objetivo de: planejar, dirigir e executar a atividade de inteligência em segurança pública, assim como regular, coordenar e supervisionar o Sistema de Inteligência de Segurança Pública do Estado do Rio de Janeiro (RIO DE JANEIRO, 2005).

A concepção da atividade de inteligência tem sofrido um processo de transformação. Na atualidade, aponta-se para uma perspectiva na qual a inteligência não tem poder de polícia, mas usa-se o trabalho intelectual para avaliar as informações. Estas informações podem ser classificadas de diversas maneiras, tais como: informação militar, tática, geral, diplomática, política, econômica, social, biográfica, científica, tecnológica e informação sobre comunicações e transportes. Seu processo envolve as seguintes fases: necessidade de conhecimento; coleta de dados na imprensa ou outros similares, incluindo coleta de dados não disponíveis; processamento dos dados; e disseminação do conhecimento ao usuário, para a tomada de decisão. A atividade deve ser centralizada e seu quadro de profissionais deve ser preenchido por pessoas íntegras e com bons propósitos (RIO DE JANEIRO, 2005).

Tradicionalmente, o exercício da atividade de inteligência tem despertado com frequência debates sobre as exigências de sigilo em defesa dos interesses do Estado e sobre as necessidades de transparência dos atos governamentais, uma das características da democracia. Temos vistos diversas reportagens onde a transparência dos dados e a Lei de Acesso à Informação são empregadas na expressão do discurso de uma camada da sociedade. Como resultado desses debates, fica o entendimento de que os serviços de Inteligência necessitam ser submetidos a oportunos controles, como forma de garantir que as ações governamentais não se sobreponham aos interesses da sociedade e do Estado, sejam por objetivos políticos, ou de manejo em massa sobre algum assunto. O emprego de ferramentas de inteligência artificial, reconhecimento facial, *big data* entre outras tecnologias de coleta de dados empregados na coleta de dados por órgãos de inteligência, necessitam ter em suas prerrogativas seus marcos regulatórios previstos sobre o ponto de vista ético, jurídico, legal e de governança dos dados e informações estabelecidos sobre o ponto de vista ético prescrito.

Além destes controles, a ética é imprescindível no rumo dessas ações e intrínseca ao funcionamento da atividade de inteligência, desde a seleção ou recrutamento rígido de seus profissionais, que, a partir de suas habilidades e competências e de seus atributos intelectuais e morais, precisam ter a plena consciência e treinamentos para a realização de suas atividades, tal como saber que seu trabalho é destinado à consecução e execução dos maiores interesses da

nação (SANTOS; BARBOSA, 2016). Destaca-se a extrema importância deste ponto para todos os profissionais que trabalham com dados sigilosos e que tratam de segurança e inteligência, ou aqueles que venham a ter contato com dados sigilosos como, por exemplo, a realização deste trabalho.

Todo o processo de recrutamento, treinamento e dispensa deve seguir normas e procedimentos estabelecidos pela DISPERJ e direcionar a execução dos profissionais durante sua função ativa de atividade.

### ***2.3 DADO, INFORMAÇÃO E CONHECIMENTO NA ÁREA DE INTELIGÊNCIA EM SEGURANÇA PÚBLICA***

A partir de nossas reflexões, definimos dado como toda e qualquer representação de foto, situação, comunicação, notícia, documento, extrato de documento, fotografia, gravação, relato, denúncia, dentre outros, ainda não submetida a um profissional ou agente de inteligência de segurança pública e à metodologia de produção e gestão de conhecimento. O conhecimento é produzido pela agência de inteligência de segurança pública e consiste na transformação dos dados de acordo com a metodologia própria e específica esses dados em informações segundo um determinado grau de classificação. A Lei de Acesso à Informação, 12527/2011, em seu art.24 classifica como sigilosas essas informações porque tratam de matéria de caráter restrito a agências e órgãos de inteligência de segurança pública. O prazo de restrição de acesso é de 15 anos sem possibilidade de prorrogação, conforme os prazos constam abaixo:

Os graus de classificação de informação e seus respectivos prazos máximos de restrição de acesso são os seguintes:

Reservado: 5 anos, sem possibilidade de prorrogação.

Secreto: 15 anos, sem possibilidade de prorrogação.

Ultrassegredo: 25 anos, podendo ser prorrogado uma única vez por igual período, apenas pela Comissão Mista de Reavaliação de Informações.

Importante: o prazo de sigilo é contado a partir da data de produção da informação, e não de sua classificação. (BRASIL, 2011)

O início do século XXI foi revolucionário quanto ao uso de dados. Percebeu-se que os dados eram fonte de informação importante, desde que uma grande quantidade destas informações seja efetivamente gerenciada e análise de padrões seja realizada. Para esse fim, considerou-se útil uma quantidade volumosa de dados coletados de várias fontes e em diferentes formatos.

A presença de novas ferramentas baseadas em tecnologia apresentou muitas mudanças na coleta, processamento e análise de dados. Foi percebido que a aplicação de ferramentas

tecnológicas de comunicação e informação aumenta o poder de coleta e a capacidade de processamento que transforma dado bruto em informações. Esta transformação tem o objetivo de extrair das informações produzidas um alto valor agregado, anteriormente impossível de ser processado devido ao seu grande volume de processamento. O uso de *big data* é movido por vários processos de tomada de decisão em diversas áreas do conhecimento. Deste modo, o impacto do *big data* também pode ser percebido nas ações de inteligência.

São distintas as fontes de entrada de dados, abertas e/ou fechadas, tais como: perfis em redes sociais, dados cadastrais de pessoas físicas e jurídicas, informações sobre pessoas, fotos, veículos, áudios obtidos mediante processos de interceptação legal, informações via *web*, aplicativos móveis, dentre outras tantas fontes disponíveis. São basicamente orientadas hoje na produção de conhecimentos e necessários para subsidiar os estudos realizados por analistas na construção e execução de estudos na área de inteligência. Dispor acesso e treinamento a essas ferramentas aos profissionais da área de inteligência são fundamentais para a melhoria da produção do conhecimento produzido, bem como, da coleta de melhores e maiores informações sobre possíveis alvos ou assuntos demandados em seu dia a dia.

A rápida massificação das tecnologias, o acesso contínuo à internet e redes sociais amplificaram a interatividade entre indivíduos, organizações e dispositivos, gerando lapsos de vulnerabilidade da informação. No Brasil, tal convergência gerou um aumento exponencial de armazenamento, transferências de dados e interatividade entre informações, indivíduos, dispositivos e sistemas. Segundo Sêmola (2014), assim como o *big data* tornou-se um meio de melhorar a capacidade dos sistemas de segurança na detecção de atividades maliciosas, ele é uma ferramenta ilimitada de mineração de conhecimento. Neste trabalho, o *big data* é entendido como um grande volume de dados, estruturados e não estruturados utilizados na área de inteligência previamente coletados.

Para que os dados sejam transformados em informação é fundamental que as correlações entre os vários fatos e suas implicações para os indivíduos e para a organização sejam evidenciadas, tornando-se visíveis e explícitas. No entendimento de Drucker (1992), para que dados sejam convertidos em informações é preciso perguntar do que se necessita, de quem, quando e de que forma, além da certeza de que aqueles que podem prover as informações conhecem e compreendem suas responsabilidades. Segundo Nonaka e Takeuchi (1997), a informação é um fluxo de mensagens e por meio dela não só se extrai, como também se constrói o conhecimento. Alerta Sveiby (1998) que o valor não está na informação armazenada, mas na criação de conhecimento a partir dela. Informação é a disposição dos dados de tal forma que possam fazer sentido, criando padrões e ativando significados na mente das pessoas. Entende

Fialho (2001) que um conjunto de dados analisados e organizados sob um determinado contexto e que satisfaçam um objetivo específico recebe o nome de informação.

#### **2.4 A PRODUÇÃO DO CONHECIMENTO NA ATIVIDADE DE INTELIGÊNCIA**

Para Paula et al. (2012), a atividade de inteligência deve ser compreendida como um sistema complexo adaptativo em que os processos de construção, produção e gestão do conhecimento e da informação são utilizados na composição ou ação, e que sejam capazes de otimizar a sua organização e utilização na segurança pública para defesa do Estado, da sociedade e do cidadão. Deste modo, trataremos a informação e o conhecimento produzido, gerado e difundido entre as atividades de inteligência com características de um sistema complexo. A maior parte das cidades, a sociedade civil, os agentes econômicos e as organizações políticas enfrentam problemas muitas vezes com características incompreensíveis.

O ciclo de produção do conhecimento é a metodologia conduzida pelo profissional da área de inteligência na produção do conhecimento. Trata-se de uma atividade ou exercício, racional composto por fases, disposta de forma ordenada. A metodologia empregada é aplicável ao todo ou em partes aos quatro tipos de conhecimento produzidos pela área de inteligência. São o informe, a informação, a apreciação e a estimativa. Sendo:

- Informe: conhecimento resultante de juízo formulado pelo analista de inteligência sobre a narração de fato ou situação passada ou presente. É a narração de um fato ou situação à qual foi aplicada uma técnica de avaliação de dados, tendo recebido um juízo de valor quanto à sua credibilidade.
- Informação: conhecimento resultante de raciocínio elaborado pelo analista de Inteligência que expressa sua certeza sobre situação ou fato passado ou presente.
- Apreciação: conhecimento resultante de raciocínio elaborado pelo analista de Inteligência que expressa sua opinião sobre situação ou fato passado, presente ou futuro imediato.
- Estimativa: conhecimento resultante de raciocínio elaborado pelo analista de Inteligência que expressa a sua opinião sobre a evolução futura de um fato ou de uma situação. (RIO DE JANEIRO, 2005).

O CPC envolve a sequência ordenada de procedimentos executados pelo analista, de forma racional e lógica, para dar credibilidade ao conhecimento produzido. Os atributos pessoais do analista, sua experiência e seu embasamento cultural favorecem o sucesso da atividade. O emprego deste método não deve ser entendido como condição suficiente para o êxito de um trabalho de inteligência, mas, como condição necessária. Para o Coronel Romeu Ferreira – figura importante da atividade de inteligência e um dos percursores da

implementação da atividade de inteligência no Brasil – comenta em suas aulas do Curso de Inteligência em Segurança Pública (BRITO, 2018) que o analista precisa ter habilidade e competências desenvolvidas pelos anos de atividade, para possuir experiência necessária com o intuito de exercer bem suas atividades de análise.

Além disso, o CPC compreende quatro fases lógicas, cronológicas ou não, que são respectivamente: planejamento, reunião dos Dados, processamento e utilização. A área de Inteligência utiliza a técnica de avaliação de dados, compreendendo as fases de julgamento da fonte; julgamento do conteúdo; e a determinação do grau de credibilidade do dado.

Para o julgamento das fontes são observados aspectos a respeito da autenticidade, confiança e competência, que permitem aferir à fonte o grau de: idônea; regularmente idônea; regularmente inidônea; e não avaliada. Mostraremos no estudo de casos essas avaliações aplicadas em denúncias e a metodologia que foi construída que vem sendo aplicada com fontes primárias e secundárias no atendimento na central de atendimento para algumas classificações importantes e, principalmente, para o estudo de caso da milícia.

O julgamento do conteúdo objetiva determinar o grau de veracidade do teor do dado, de acordo com sua semelhança, coerência e compatibilidade, bem como a determinação do grau de credibilidade do dado é determinada pelo analista que produziu o conhecimento, transformando o dado em informe, é expresso por código alfanumérico.

O emprego do método contribui para garantir que todos os aspectos do problema sejam considerados, produzindo conhecimento com base científica, uniformizando procedimentos e assegurando que o usuário confira credibilidade ao conhecimento produzido. As etapas de produção são: planejamento; reunião de dados; análise e síntese; interpretação; e formalização e difusão. A reunião dos dados a serem analisados e a produção de informação contextualizada a ser interpretada gera conhecimento de inteligência, apoiando o processo de tomada de decisão estratégica do usuário da informação. Adequa-se a orientar o planejamento da política de segurança do Estado e possibilita as atividades de acompanhamento, estudo, produção e difusão dos conhecimentos produzidos através desse método.

**Figura 3: Produção do Conhecimento em atividades de inteligência**



Fonte: elaboração própria.

Desta maneira, o sistema contribui diretamente com as iniciativas de produção do conhecimento, objetivando a adoção de medidas, procedimentos e comportamentos adequados, considerada a natureza sensível dos assuntos aos quais tenham acesso, tendo em vista que o usuário da informação muitas vezes é o tomador de decisões em relação a produção do conhecimento, conforme a Figura 3 acima.

A definição de que dados deverão ser protegidos e os que deverão ser compartilhados seguem regras preestabelecidas de segurança, de maneira sistêmica, integrada e abrangente aos órgãos competentes de difusão.

Por estar diretamente relacionada a tomada de decisão estratégica necessita que toda sua comunicação produzida, seja clara, concisa e direta, na intenção de minimizar distorções e incompreensões.

Neste sentido, a atividade de inteligência utiliza-se de uma linguagem específica que foi respeitada na construção deste trabalho com foco em milícia para que possa utilizada como fonte de atividades de inteligência ao tema. A construção e produção do conhecimento necessitam alcançar essas etapas, e são fundamentadas em três noções, como descrito abaixo:

A verdade consiste na perfeita concordância do conteúdo do pensamento (sujeito) com o fato (objeto).

A verdade - significativa, imparcial, oportuna e bem apresentada - é a aspiração que norteia o exercício da atividade de ISP. Todos os profissionais que a exercem devem acautelá-la contra a mera ilusão da verdade, contra o erro.

(...)

A mente humana pode encontrar-se, em relação à verdade, em quatro diferentes estados: certeza, opinião, dúvida, ignorância.

1) Certeza: consiste no acatamento integral, pela mente, da imagem por ela mesma formada, como correspondente a determinado objeto. A certeza é, dessa forma, o estado em que a mente adere à imagem de um objeto, por ela mesma formada, sem temor de enganar-se.

2) Opinião: é o estado em que a mente acata, porém com receio de enganar-se, a imagem, por ela mesma formada, como correspondente a determinado objeto. A opinião é, portanto, um estado no qual a mente se define por um objeto, porém com receio de equívoco. Por isso, o valor do estado de opinião expressa-se por meio de indicadores de probabilidades.

3) Dúvida: é o estado em que a mente encontra, metodicamente, em situação de equilíbrio, razões para aceitar e razões para negar que a imagem, por ela mesma formada, está em conformidade com determinado objeto.

4) Ignorância: é o estado da mente que se caracteriza pela inexistência de qualquer imagem de determinado objeto. A ignorância é um estado puramente nulo da mente. Nesta situação, o espírito encontra-se privado de qualquer imagem sobre uma realidade específica. (RIO DE JANEIRO, 2005, p. 12-13).

Neste sentido, conceber ideias é a projeção mental “da imagem de determinado objeto, sem adjetivá-lo” (ESG apud SANTOS, 2016, p. 85). Já a formulação de juízos é entendida como “a atividade pela qual a mente estabelece uma relação entre ideias” (ESG apud SANTOS, 2016, p. 86). Por último, temos a elaboração de raciocínios que vem a ser “a atividade pela qual a mente, a partir de dois ou mais juízos conhecidos, alcança outro que deles decorre logicamente” (ESG apud SANTOS, 2016, p. 86).

Os conhecimentos também são classificados quanto ao nível de utilização e validade no tempo, como descrito a seguir (SANTOS, 2016). Quanto ao nível de utilização podem ser estratégicos, operacionais ou táticos. A saber, o conhecimento estratégico é aquele conhecimento necessário para a formulação das avaliações estratégicas dos órgãos componentes da Defesa, de planos e de políticas nacional ou internacionalmente, relativos à Defesa Nacional. O conhecimento operacional é o tipo de conhecimento utilizado para o planejamento, condução e sustento das operações militares de grande envergadura, com a finalidade de alcançar objetivos estratégicos inserido em um Teatro de Operações ou Zona de Operações. Por último, o conhecimento tático que é aquele que é requerido para a condução de operações de combate no nível tático (SANTOS, 2016), conforme quadro 1 abaixo.

**Quadro 1: Comparação entre estados da Mente, Trabalhos Intelectuais e Tipos de Conhecimento**

ESTADOS DA MENTE (EM)	TIPO DE CONHECIMENTO			
	INFORME	INFORMAÇÃO	APRECIACÃO	ESTIMATIVA
CERTEZA	X	X	-	-
OPINIÃO	X	-	X	X
DÚVIDA	X	-	-	-
IGNORÂNCIA	-	-	-	-
<b>TRABALHO INTELECTUAL (Relint)</b>				
IDEIA	-	-	-	-

Fonte: BRANCO (2017), a partir da DISPERJ/ 2005.

Quanto à validade no tempo, o conhecimento pode ser classificado em básico e corrente. O conhecimento básico é “pouco sensível à ação do tempo”, compreendendo conhecimentos geográficos, demográficos, políticos, biográficos, e de “estrutura geral das forças e das características operacionais de áreas do país ou do exterior”. O conhecimento corrente é aquele identificado como “altamente sensível à ação do tempo e que trata de assuntos e atividades em andamento, ou recentemente concluídos, apresentando reflexos para a conjuntura atual” (Quadro 2, ESG, 2005).

Com o aumento do uso das TICs, inevitavelmente crescem as vulnerabilidades e ameaças em nível digital para as organizações brasileiras que precisam lidar com episódios de vazamento de dados, espionagem e sabotagem virtual. São desafios para a área de inteligência o gerenciamento de riscos na segurança cibernética, e de seus dados produzidos, impostos pelo *big data*.

**Quadro 2: Comparação entre Tipos de Conhecimento, Estados da Mente, Trabalhos Intelectuais**

TIPO DE CONHECIMENTO	ESTADOS DA MENTE	TRABALHOS INTELECTUAIS	TRABALHO VOLTADO PARA :
INFORME	CERTEZA	JUÍZO	PASSADO
	OPINIÃO		e
	DÚVIDA		PRESENTE
INFORMAÇÃO	CERTEZA	RACIOCÍNIO	PASSADO
			e
			PRESENTE
APRECIÇÃO	OPINIÃO	RACIOCÍNIO	PASSADO,
			PRESENTE
			e
			FUTURO (FAZER O ESTUDO SEM TÉCNICAS AUXILIARES, SERÃO PROJEÇÕES SOBRE PERCEPÇÕES E DESDOBRAMENTOS PELO PROFISSIONAL DE ISP)
ESTIMATIVA	OPINIÃO	RACIOCÍNIO	FUTURO

Fonte: BRANCO (2017), a partir do conteúdo da DISPERJ/2005.

Tem sido crescente o uso de técnicas de *machine learning* para que, sem intervenção humana, seja possível classificar qualquer denúncia anônima utilizando apenas sua natureza textual (PINHO, 2017). Quando os dados são textuais e, portanto, dados não estruturados, pode-se dispor de métodos e modelos de mineração de textos que têm sido bastante explorados em muitos problemas em diversas áreas (classificação de documentos, extração de informação, sumarização, análise de sentimentos, dentre outros). Torna-se necessário e importante lembrar a diferença entre sistemas transacionais e sistema de apoio à decisão.

Nos sistemas transacionais ocorrem transações de negócio, através dos quais se realizam

compras, vendas entre outros. As características de sistema necessárias são alta velocidade no tempo de resposta, baixo volume de dados envolvidos nas transações e onde informações históricas não são tratadas como relevantes. Podemos citar como exemplo, um cliente fazendo um saque bancário, em um caixa automático. O cliente necessita que, as respostas sejam rápidas. O volume de informações envolvidas é baixo na operação de saque. São necessárias informações como banco, agência, conta-corrente que constam em seu cartão de saque e as demais senhas, valor requisitado, saldo disponível e limite de cheque especial, por exemplo, estão cadastradas no sistema.

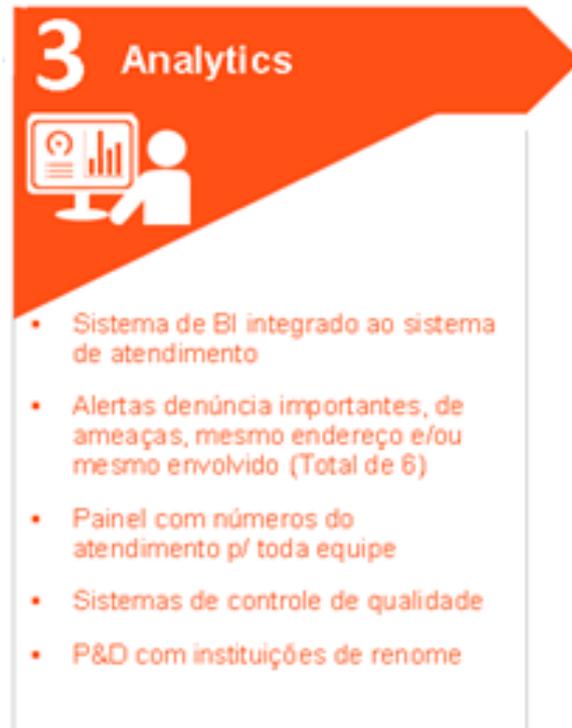
Já sistemas de apoio à decisão possuem características completamente distintas por serem construídos com base em uma grande quantidade de dados e possuem uma arquitetura típica para *data warehouse* em que estão presentes os dados operacionais de uma entidade, modelados de uma maneira diferenciada e denominada modelagem dimensional. Possui características voltadas para análises de dados e inteligência de negócios, sendo suporte e apoio para tomada de decisão. O *data warehouse* oferece uma solução de armazenamento de dados centralizados, contendo um grande volume de informações históricas, otimizados para operações de leitura e confiáveis para a tomada de decisão.

## **2.5 VALOR DA INFORMAÇÃO ORIENTADA A DADOS PARA TOMADA DE DECISÃO**

A cultura do valor da informação orientada a dados para a tomada de decisão – denominada *data driven* – vem sendo buscada por diversas organizações como suporte a tomada de decisão em busca de eficácia, eficiência e efetividade na finalização e finalidade de seus processos.

Usualmente, os dados e suas fontes, são extraídos dos sistemas transacionais para serem carregados e armazenados em servidores de *data warehouse* através de ferramentas de ETL (extração, transformação e leitura) que representam a extração dos dados e suas fontes de dados, normalmente de vários sistemas transacionais e banco de dados, passando pelo módulo de transformação que normalmente são padronizações necessárias para a carga dos dados no servidor de *data warehouse*.

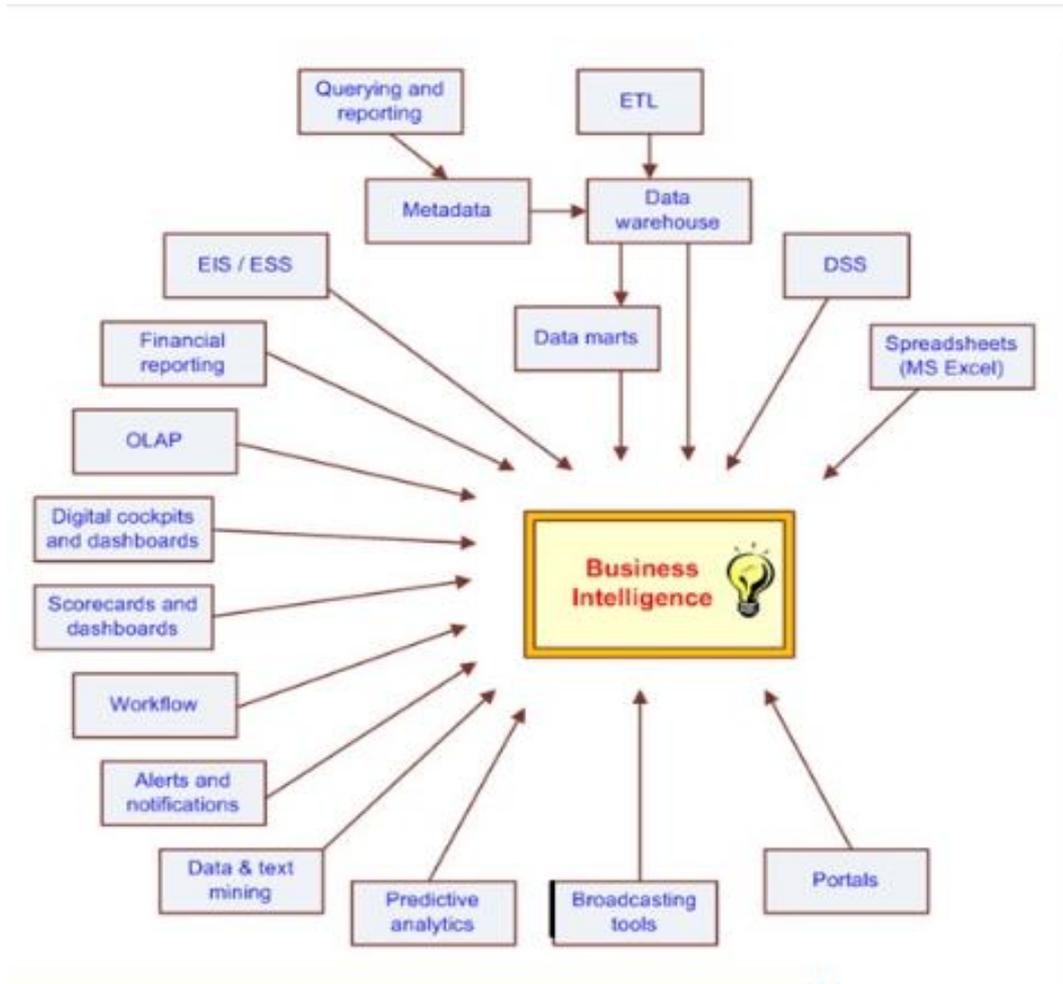
**Figura 4: Analytics e Sistema de BI – Integrado ao Sistema d**



Fonte: Figura adaptado pelo autor

Utiliza-se de servidores analíticos e intermediários para processos e ferramentas de BI, Figura 4 acima, e estudos analíticos, assim como a retroalimentação de informações históricas na linha do tempo e geram informações, que alimentarão aplicações *front end*.

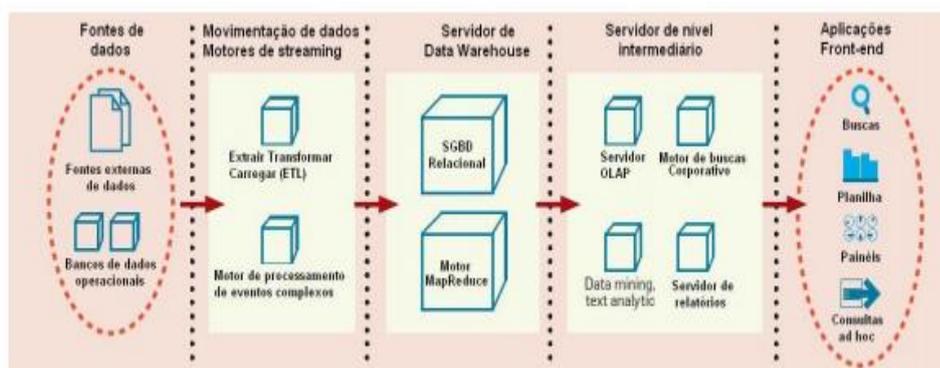
**Figura 5: Conjunto de Ferramentas que poderão ser utilizadas no BI**



Fonte: Figura adaptada pelo autor

Espera-se com a aplicação do BI auxiliie em algumas questões, tais como a padronizar a gestão e o planejamento em órgãos de segurança pública para garantir a ordem pública, aumentar a sensação de segurança em determinados locais, minimizar a criminalidade, aumentar as informações sobre suspeitos e foragidos da justiça, combater a criminalidade, aumentar a gestão da informação através do acompanhamento de seus indicadores de violência, dentre outros, os quais serão analisados com esse estudo, em parceria com órgãos de inteligência em segurança pública.

**Figura 6: Arquitetura típica de uma Sistema baseado em Business Intelligence**



Fonte: Chaudhuri *et al.* (2011)

Fonte: Chaudhuri *et al.* (2011)

Trata-se de um conjunto de ferramentas que auxiliam o trabalho da atividade de inteligência voltada a organizações, órgãos e agências de segurança pública do Estado (Figura 5 acima), auxiliando e tornando possível a cultura orientada a dados no processo e organização, auxiliando a tomada de decisão, bem como a utilização e obtenção dessa cultura em órgãos e organizações de segurança pública, com um sistema baseado em BI (Figura 5 acima).

No sentido de extrair conhecimento e informações, a partir da integração de bases de inteligência em órgãos de segurança pública, aliadas a construção e gestão de conhecimento em denúncias por meio da criação de glossários de palavras e expressões, utilizadas em relatos de denúncias no decorrer do tempo e classificadas em sistema de denúncias.

Nesta seção foram apresentadas as etapas que envolvem o desenvolvimento de uma arquitetura de *business intelligence* com o sistema de atendimento das denúncias e a integração com bases de analítica utilizada no estudo de caso. Assim como o modelo, a criação dos glossários e textos, que reflitam as palavras, expressões e semânticas utilizadas no decorrer do tempo por tipo de denúncia.

## 2.6 CONCLUSÕES DO CAPÍTULO

Este capítulo serve como base teórica para a compreensão das demais seções deste trabalho. A partir dela podemos apreender noções a respeito das atividades de inteligência, e a metodologia aplicada na área de inteligência, chamada de ciclo de produção do conhecimento. Também abordamos o conceito de dados, informação e conhecimento; e a complexidade da

questão informação e a produção da informação nas atividades de inteligência; e essas relações com sistemas complexos. Nosso objetivo foi contribuir com iniciativas de produção do conhecimento, a partir da análise das áreas de interação existentes entre a ciência da informação e aplicação de tecnologias na atividade de inteligência. Também foram apresentados a ferramenta de *business intelligence* dirigida ao nosso estudo de caso e o valor da informação como processos de decisão organizacional. O propósito dessa abordagem foi tornar a metodologia proposta de fácil compreensão e implementação para a área de inteligência que, independentemente de seu porte, área de atuação, ou recursos, seja capaz de se relacionar e utilizar a teoria apresentada. Além da síntese conceitual proposta, o presente trabalho estabelece um referencial teórico que pode ser útil para pesquisas futuras ligadas ao tema como a implementação de sistemas de informação integrados nas organizações de atividades de inteligência, que pode ser aplicado a trabalho futuros nos ramos de proteção como as ameaças cibernéticas, hackerativismo, *big data* sobre terrorismo, espionagem digital, corrupção, lavagem de dinheiro e evasão de divisas, interferência externa, crime organizado, roubo e manipulação ilegal de bens sensíveis entre outros.

### **3 ASPECTOS DA SEGURANÇA PÚBLICA NO RIO DE JANEIRO: O DISQUE DENÚNCIA E AS MILÍCIAS**

Neste capítulo é apresentado o Disque Denúncia e todas as suas Premissas em relação aos canais de atendimento, e a metodologia de atendimento, seus canais, estrutura de uma denúncia e sua classificação, bem como a base de dados que será utilizada no estudo. Este capítulo também apresenta a questão das milícias, seu modo de operação e de que maneira é possível perceber sua ocorrência na base de dados do DD.

### 3 ASPECTOS DA SEGURANÇA PÚBLICA NO RIO DE JANEIRO: O DISQUE DENÚNCIA E AS MILÍCIAS

O Rio de Janeiro conta com o Disque Denúncia<sup>4</sup>, uma central de atendimentos, criada inicialmente para ajudar as polícias Federal, Civil e Militar no esclarecimento de sequestros, crimes e delitos, por intermédio do recebimento de denúncias feitas por meio de ligações anônimas da população. Este instrumento tem sido caracterizado, ao longo dos seus 25 anos de funcionamento, por receber as mais variadas demandas que, apesar de não serem apenas criminais, guardam relação direta com os sentimentos de temor, insegurança e medo da população carioca ao longo dos anos. Criado como uma alternativa para contribuir para a diminuição da violência na cidade, o Disque Denúncia começou a funcionar em 1995, com o objetivo de receber informações da população a respeito de atividades criminosas e casos de violação aos direitos humanos, para que fossem transmitidas aos órgãos competentes. O serviço, que tem como gênese a importância da garantia do anonimato de quem fornece as informações através da denúncia atende desde reclamações sobre a conduta de um vizinho até dados completos sobre quadrilhas e crimes ligados ao tráfico de drogas, armas, atividades criminosas, milícia e sobre o modo de agir de determinadas facções, entre outras denúncias realizadas no seu dia a dia.

O acesso da sociedade às novas tecnologias têm alterado a rotina e o processo de captura, recepção e formas de atendimentos dessas denúncias, um desafio cada dia maior nessa comunicação direta com o denunciante. Manter vários canais de comunicação com a mesma metodologia de atendimento já empregada com sucesso também se torna um grande desafio ao DD. Por isso, buscar novas tecnologias que auxiliem nessa comunicação na busca de informações pertinentes a cada assunto torna-se uma necessidade diária de adaptação à nova realidade de comunicação tecnologia exercida pelos diferentes públicos denunciante. Ter um mesmo formato de comunicação e respeitar as formas de comunicação desses canais através da aplicação de sua metodologia de atendimento e levar a mesma régua de atendimento para todos os canais são alguns dos pontos que enfatizamos neste trabalho. A aplicação do conceito de *Omnichannel*, uma estratégia de conteúdo multicanal usada pelas organizações para melhorar sua experiência do usuário e promover melhores relacionamentos com seu público através de pontos de contato. Isso requer um conjunto de habilidades e competências dos atendentes que

---

<sup>4</sup> Disponível em: <<http://www.disquedenuncia.org.br/>>.

não apenas têm a função de receber essa denúncia, mas também analisar as necessidades relacionadas a cada tema e o canal de contato do denunciante. Para isso, precisam ser treinados e orientados a se comportar com um perfil de trabalho diferenciado de um atendente normal de teleatendimento, é necessário ter uma postura de analista de informações para a coleta de dados. E só conseguimos esses resultados aliando um bom treinamento de simulações de denúncias, treinamentos e simulando o atendimento na identificação dos dados, informações e entidades necessárias a cada tipo de denúncia, trataremos dessas entidades mais à frente, no item 3.9 desta seção.

O principal canal de comunicação da organização com a sociedade continua sendo a linha telefônica, em um digitronco de número 2253-1177. No entanto, mais recentemente nos últimos 6 anos, registra-se o incremento de denúncias realizadas através de mídias sociais e aplicativos. O crescimento dessa forma de comunicação entre as pessoas também reflete na utilização do serviço. Através do Aplicativo e do crescente uso das mídias sociais, como *Facebook* e *WhatsApp*, entre outras, já correspondem a cerca de 35% do número de atendimentos de denúncias recebidas por esses canais. Deste modo, no estado do Rio de Janeiro, o Disque Denúncia atua como um canal entre a população e os agentes de segurança pública no combate à criminalidade e na solução de problemas que atingem comunidades, favelas, bairros e cidades. Ao receber denúncias anônimas e coletar informações de importância fundamental na apuração de ações criminosas e delitos, contribui conseqüentemente para a redução da criminalidade. Recebendo mais de 100 mil relatos anônimos de crimes e/ou delitos por ano, este é um programa bastante semelhante ao *American Crime Stoppers*<sup>5</sup>, por exemplo. A informação desses relatórios de crimes torna-se extremamente importante para reduzir os níveis de criminalidade, ajudando a polícia e a área de inteligência e outras autoridades, a obter informações mais seguras podem assim realizar operações mais eficientes contra criminosos em uma determinada área, por exemplo, exercendo o papel de servir como um termômetro de atenção e monitoramento do que, onde, como e o que está ocorrendo na cidade. O DD possui como Pilares do serviço, a relação entre os agentes de segurança pública, a população e a Imprensa, Apêndice A. Através dos seus canais de recebimento de denúncias com a participação da população, e sempre com a preservação e a garantia do anonimato em seus canais, o serviço é o elemento de operação entre os órgãos de segurança, através da integração dos dados das denúncias difundidas, e estabelecendo pauta de notícias para a imprensa na

---

<sup>5</sup> Disponível em: <http://www.crimestoppersusa.org/>>. Acesso em 09 abr. 2020.

resolução e cobrança dessas denúncias. A representação simbólica dessa metodologia de emprego e uso de metodologia própria na captação da informação na denúncia, a garantia de seu anonimato ao denunciante e a difusão aos corretos órgãos competentes e resolver essas questões, colabora para o trabalho das autoridades policiais na busca de diminuição dos índices de violência urbana.

### 3.1 METODOLOGIA DE ATENDIMENTO E CANAIS DE ATENDIMENTO

Com o aparecimento desses novos canais digitais com a sociedade amplifica as capacidades no aporte em se transformar em um canal único com o denunciante, fomentando a etapa de empatia, primordial para a concretização da denúncia e na aplicação da metodologia de atendimento utilizada pela Central de Atendimento. Trata-se da fase da ciência sobre a denúncia. Nesta etapa, emprega-se a técnica de *rapport*, técnica usada para criar uma ligação de sintonia e empatia com outra pessoa, um espelhamento com o denunciante, algo semelhante a uma entrevista. Essas ações e relações contém importância na percepção da(s) denúncia(s), sendo extremamente essencial a aplicação de novos recursos tecnológicos nesta atividade, auxiliando esse analista com essa habilidade.

Essas informações podem ser coletadas e utilizadas para diversos propósitos como geração de *insights*, *leads* e engajamento, melhoria das denúncias, complemento de algo denunciado trazendo novas informações, dentre outras.

A técnica utilizada para criar uma conexão entre o denunciante e o atendente chama-se *rapport*, conforme o dito anteriormente. Esta palavra tem origem no termo em francês *rapporter* que significa "trazer de volta". O *rapport*, aplicado ao denunciante através do atendente, ocorre quando existe uma sensação de sincronização entre duas ou mais pessoas, possibilitando que elas se relacionam de forma agradável, diminuindo a ansiedade e o medo inicial que normalmente acontece. Esta técnica é muito útil porque cria laços de compreensão entre dois ou mais indivíduos envolvidos, e onde se busca sua necessidade, uma visão geral do(s) assunto(s) abordado(s), e ainda, apreender se como o denunciante: colocando-se em seu lugar, procurando enxergar o que, como e quando os fatos ocorrem, os envolvidos, sua ordem cronológica, frequência, características relevantes, dentre outros. A interação é o foco na busca do maior número possível de dados e informações para sua plausível descrição, gerando rastros e pistas indispensáveis a uma boa denúncia. Deste modo, fonte primária é quando o denunciante é um agente que viu, presenciou, participou ou foi envolvido. Isto implica, no relato da denúncia, que os verbos estejam no presente do indicativo. Por exemplo, uso de expressões

como “é”, “está nesse momento” etc. Conseqüentemente, os profissionais de inteligência classificam essas fontes de maneira diferenciada as fontes secundárias. Quando o denunciante não fez parte ou não está envolvido diretamente ao ocorrido, apenas foi informado sobre todo o relato deverá ser escrito no futuro do pretérito. Por exemplo, “estaria”, “é provável que” etc. Estas regras são aplicadas a pessoas e organizações pela doutrina de inteligência em segurança pública e são relevantes na produção dos conhecimentos produzidos pelos órgãos que produzem e que foi implementado ao estudo.

Vale lembrar que, são assuntos ligados a criminalidade, violência, milícias, facções criminosas, enfim temas que trazem uma carga emocional pesada. Por isso, o denunciante precisa sentir-se confortável e seguro quanto ao anonimato para que todos os elementos, frações significativas, entidades referentes a classificação da denúncia possam ser coletados. Isso exige técnica, treinamento, destreza por parte do atendente em saber ouvir e se comportar junto ao denunciante. Existe uma metodologia aplicada nessa fase, utilizando o *script* de atendimento. Componentes comportamentais são essenciais nessa fase de coleta da denúncia, tais como: atenção mútua, positividade mútua e coordenação pelo atendente de uma sequência lógica e cronológica do cenário e seus elementos importantes para uma ação futura (MORAES, 2006).

Torna se necessário observar a situação denunciada como se fosse uma grande novidade, sem fazer suposições nem julgamentos, encarando todos os fatos como primordiais à composição da denúncia e assim, classificá-los. Ainda é preciso identificar do que se trata o assunto ou assuntos, seus envolvidos, sua periodicidade, suas principais características quanto a pessoas, locais, veículos, ou aos elementos necessários a cada tipo de denúncia. É necessário entender todo o relacionamento necessário no contexto da denúncia, suas características, assim como no modelo de classificação que ela será registrada. O relacionamento em relação a denúncia e sua classificação pode ser vista no Apêndice B.

Sem dúvida, com o surgimento desses novos canais digitais com a sociedade amplia possibilidades para a contribuição do DD com os dados coletados, assim como testar novas formas de atendimento específicas a cada canal, respeitando suas características e formas diferenciadas de comunicação e o público característico de cada um deles.

Diante da complexidade na captação e coleta de dados e com o propósito de aliar conceitos empregados na área de inteligência a esses dados, considerávamos necessários o emprego do conceito de frações significativas, já citado anteriormente, com a intenção de ter uma classificação da fonte anônima bem como do relato, portanto a utilização de fontes primárias e secundárias foram aplicadas. Cabe ressaltar que, através do estudo de automatização as denunciadas aplicada no trabalho (PINHO *et al*, 2017) podemos observar que a quantidade

de palavras que compõem uma denúncia caiu de 38 palavras em média em 2015, para cerca de 18 palavras em média em 2018. Propor a implementação dessas frações significativas que tratam de pistas importantes relacionadas aos temas específicos tornava-se uma necessidade imprescindível a melhoria do serviço. A utilização desta técnica não é para aumentar o número médio de palavras utilizado nas denúncias, mas para captar as melhores pistas que podem compor aquela denúncia. A partir desse momento chamaremos essas pistas de entidades que podem representar pessoas, objetos, carros, facções, vulgos, classes, foragidos entre outras. Tudo e qualquer informação que possa ser relacionado na composição de informações sobre algo ou alguém. Portanto entender o sentido da palavra, o termo, a frase, a expressão, a terminologia utilizada na denúncia, nos traz elementos simbólicos sobre uma representação territorial, o comportamento de uma facção, a análise de sentimento que aquela denúncia expressa como medo, apelo, urgência. Cooperando na produção do conhecimento dos temas estudados e na produção dos dicionários semânticos existentes nessas denúncias. É sobre isso que trata esse arcabouço analítico de texto em denúncia deste trabalho, que veremos de uma melhor condição nas próximas seções.

Desta forma, urge a necessidade em empregar diversas fontes de dados, bases com características abertas e fechadas, textos originais e pré-definidos, distintos tipos de expressões e significados, diferentes bancos de dados, com tamanhos, volumes, velocidades etc. Para esse fim, utilizou se o propósito, de conceitos ligados ao *Big Data* e as diversas soluções para cada necessidade e obstáculos que foram surgindo no caminho na realização desse trabalho. Muita mais do que qual a tecnologia escolhida e porque a escolha dessa tecnologia, trataremos neste estudo de caso as necessidades, dificuldades e como foi realizado e proposto a solução, não queremos ficar presos a alguma condição tecnológica visto que, as mesmas mudam em uma velocidade assombrosa nos últimos tempos. Trataremos de qual maneira essas dificuldades nesses cenários específicos são trazidas e de qual maneira foram aplicadas e oferecidas pela tecnologia e trataremos da criação da captação dessas entidades nas próximas seções.

### **3.2 ESTRUTURA DE UMA DENÚNCIA**

Entender como funciona a estrutura de uma denúncia anônima é essencial para a correta compreensão deste trabalho. Uma denúncia anônima basicamente é composta por três partes fundamentais. São elas a "descrição da denúncia", a "classificação principal" e as "classificações secundárias". Os dados recebidos das denúncias e os relatos com as palavras, termos, expressões, significados e sentimento do objeto denunciado compõem a parte principal deste trabalho na construção do arcabouço e análise de textos. Quadro 3 abaixo, ilustramos a

estrutura de uma denúncia e as características citadas.

**Quadro 3: Estrutura de uma denúncia**

<p><u>Classificação Principal</u></p> <p>TRÁFICO DE DROGAS</p> <p><u>Descrição</u></p> <p>NA RUA XPTO, PRÓXIMO A UM COLÉGIO E A UMA PRAÇA, SOB UMA AMENDOEIRA, DIARIAMENTE, À PARTIR DAS 18H, INDIVÍDUOS DO SEXO MASCULINO (NÃO IDENTIFICADOS),          ALGUNS USANDO TORNOZELEIRA ELETRÔNICA, ARMADOS, COMERCIALIZAM ENTORPECENTES,          MENORES CIRCULAM JUNTO AOS INDIVÍDUOS, CARROS COM SOM MUITO ALTO TOCAM FUNKS.</p> <p><u>Classificações Secundárias</u></p> <p>LOCALIZAÇÃO DE FORAGIDO          POSSE ILÍCITA DE ARMAS DE FOGO          CONSUMO DE DROGAS          CORRUPÇÃO DE MENORES</p>
---

Fonte: Adaptada pelo autor

### 3.3 SISTEMA DE CLASSIFICAÇÃO

Atualmente, as informações passadas à central DD por meio de ligações anônimas e suas fontes de entradas são filtradas e classificadas entre 190 assuntos, divididos em 19 classes, Apêndice C. Realizamos um teste para este estudo – considerando os dados desde 2015 até abril 2018 – baseado no sistema de classificação de denúncias desenvolvido e utilizado pelos gestores do DD.

Os temas ligados ao combate à criminalidade e ao crime organizado, cujos principais ilegítimos são o tráfico de drogas, o contrabando de armas, a corrupção institucional e a lavagem de dinheiro, por tratar dos temas de maiores ocorrências ligados ao estudo de caso “milícia”. Aplicamos um ambiente analítico com a finalidade de testar os objetivos propostos neste trabalho, e assim poder verificar na prática todas as dificuldades inerentes ao procedimento real e solucionar os problemas existentes.

Com este corte, a partir do exame da base de Foragidos da Justiça – denominada “Procurados” – foi possível aperfeiçoar e nomear os vínculos existentes e incrementar novas informações a essa base, evento esse, um tanto significativo e necessário as futuras análises, visto que se trata da base com os principais suspeitos desses delitos. Empregamos para isto consulta interna à base de classificação do Disque Denúncia e relacionamos os principais temas de denúncia análogas a este estudo para traçar as entidades nome, vulgo e as pistas, para cada envolvido, permitindo assim uma hierarquização e a criação da Base de *Personas*, aplicada ao

estudo de casos, alvos identificados e monitorados a partir das denúncias.

Especificamente, no caso das principais denúncias recebidas, é possível perceber que há relações e padrões que podem ser observados, analisados e que relacionam mutuamente com denúncias de crime organizado e às demais classificações das denúncias, chamadas de XPTO, são marcações (etiquetas) de interesse imputadas no sistema, Figura 6 acima. Ainda, é possível perceber como tais denúncias refletem as dinâmicas sociais associadas aos diferentes crimes, bem como, seus impactos na sociedade, através do uso das classificações de crimes e delitos utilizadas nas denúncias através do tempo.

### **3.4 TAXONOMIA DAS CLASSIFICAÇÕES**

As plausíveis classificações para crimes apresentadas ao longo deste trabalho seguem estritamente a taxonomia legal, adotada pelo código penal vigente e pelos órgãos de segurança pública do Estado do Rio de Janeiro e, conseqüentemente, adotada por todas as forças policiais e órgãos competentes da região. Seguir exatamente a taxonomia adotada no Estado é uma premissa fundamental deste trabalho para que o mesmo seja compatível e útil para adoção em soluções que contribuam para a segurança pública do Rio de Janeiro. No entanto, vimos nas seções anteriores a tabela de classificação adotada ao DD desde sua criação e a explicação das formas e códigos adotados. Porém para a realização deste estudo foi realizada uma revisão de todas as classificações, estruturas e atualizações de códigos penais vigentes para a validação dos especialistas, sua modificação no histórico da base de dados do DD no decorrer dos anos, bem como a atualização da base com seus respectivos novos tipos e classes que foram usadas no *dashboard*.

Um trabalho árduo realizado para este trabalho, em várias etapas e com uma equipe multidisciplinar, foi a reclassificação da tabela com todas as novas classificações e realocando toda sua série histórica, adequando aos códigos vigentes e para a validação dos novos tipos de classes que foram surgindo diante as modificações do código penal vigente, sendo aprovada pela direção do DD, e implementada a partir deste trabalho.

No primeiro momento, não foi nosso foco analisada a complexidade das redes gerada pelas conexões entre as categorias de denúncias, apenas identificar quais as classes que estão sendo denunciadas em conjunto e desse modo inferir suas relações criminais, por exemplo, para denúncias com Milícias quais os outros tipos de crimes que estão associadas a ela, assim como para todas as outras denúncias foram levantadas todas as classes correlacionadas. Todavia e exclusivamente, neste momento analisamos a semântica formada pelas palavras mais frequentes nessas denúncias, desde o início da análise, para a geração das nuvens através do





hierárquicas, locais de palavras, termos e expressões são caracterizadas diante ao fato denunciado e sendo de vital importância ao contexto de conhecimento produzido. Essas expressões semânticas e um glossário dinâmico dessa identificação das expressões são utilizados no arcabouço de texto contido neste trabalho e empregados em todos os canais de atendimento, assim como as entidades criadas nesse reconhecimento das pistas, ou frações significativas como são reconhecidas pela área de inteligência.

A descrição é a denúncia propriamente dita, isto é, o relato recebido e transcrito pelo analista de atendimento, digitada e salva no sistema de denúncia do DD. Os analistas de atendimento responsáveis pela escuta e coleta dessas denúncias conversam com o denunciante e tentam coletar o máximo de informações possíveis sobre o cenário e composições dessas denúncias. Através do *script* de conversa, conduz o denunciante a descrever a cena observada com atenção a detalhes importantes para cada tipo de crime. Ao longo da conversa, é transcrito de forma mais resumida o que foi dito pelo denunciante sem perder, porém, a essência da informação transmitida. Ao término do atendimento são revistos e complementados, a sequência e fatos. O texto da descrição começa geralmente pela descrição da localidade do crime com referências precisas sobre o local, o relato do ato criminoso que ocorreu ou está ocorrendo e a descrição do suposto criminoso contendo detalhes como um possível porte de arma ou drogas por exemplo.

### **3.6 TIPOS DE DENÚNCIAS**

A classificação principal define qual é o delito central da denúncia, servindo como um dos mais importantes meios de organização da denúncia e posterior difusão as autoridades competentes pela investigação. A organização das denúncias é fundamental para o sistema interno do Disque Denúncia bem como para órgãos governamentais que processam as mesmas posteriormente.

No processo atual, a classificação principal é escolhida através de métodos manuais, onde o atendente interpreta o conteúdo da denúncia e atribui uma classificação baseado em seu treinamento e sua experiência de trabalho cadastrando as denúncias.

Posteriormente, essa classificação pode ser revisada por um superior na hierarquia de trabalho, denominado monitor, que tem o papel de ler todas as denúncias, verificar caso ocorra as correções e alertas a difusão sua relevância e importância. É uma espécie de monitoramento de ocorrências real das denúncias recebidas e o que está acontecendo no exato momento. Por exemplo, caso o crime denunciado seja classificado como, em um primeiro momento, como algo não muito grave, como barulho e no decorrer do tempo identifique-se que, na realidade,

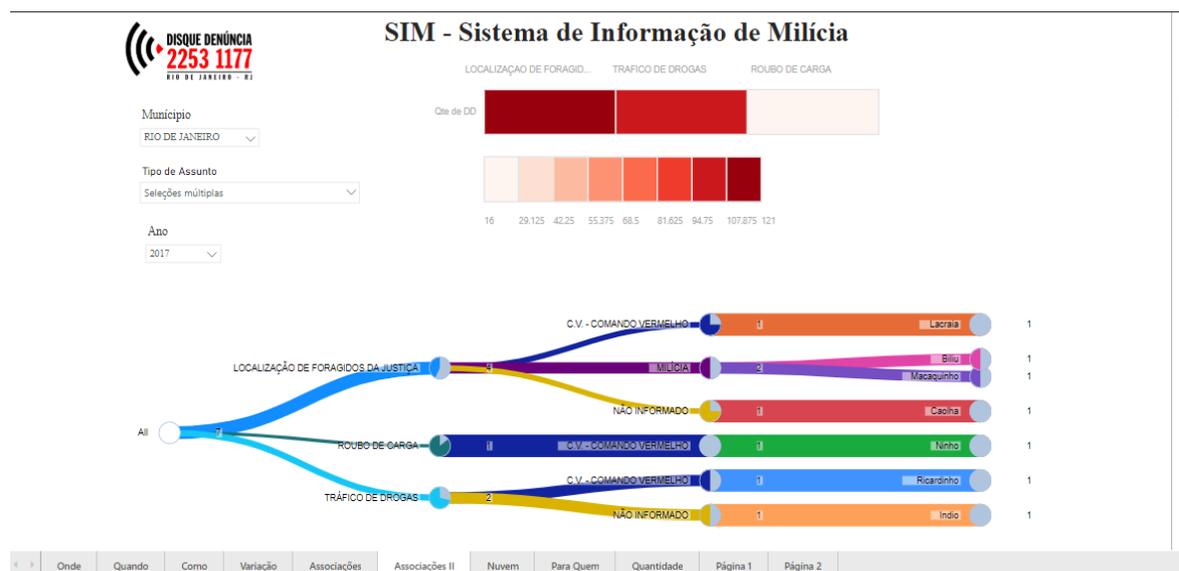
trata-se de barulho de um baile funk organizado por milicianos, identifica-se um crime mais grave que requer uma difusão diferenciada ou um alvo investigado, por exemplo.

Sem a classificação correta e apropriada das denúncias, crimes graves podem ficar sem resolução por estarem mascarados como delitos menores ou mesmo serem encaminhados para autoridades erradas na difusão, que não deveriam tratar daquele crime ou objeto investigativo, fazendo com que toda investigação torne-se ainda mais lenta. A experiência do monitor em conduzir e alertar essas ocorrências na difusão e resultados são importantes nesta fase de captação da denúncia. Sem uma correta priorização decorrente da urgência de cada denúncia, crimes de maior gravidade poderiam ficar renegados frente a imensa quantidade de crimes a serem investigados no Estado.

A última parte da estrutura de uma denúncia é a de Classificações Secundárias. Essa é uma estrutura que permite a atribuição de outras classificações de crimes que também foram observadas na descrição além da principal. É utilizada quando são relatados mais de um assunto na mesma denúncia, uma forma de desmembrar os delitos ou crimes contidos na mesma denúncia.

Assim como a classificação principal, as classificações secundárias são atualmente determinadas pelo operador que recebe a denúncia anônima e podem posteriormente ser revisadas por um supervisor. Ao ler a descrição da denúncia, o responsável pela classificação determina o crime principal em curso na denúncia e caso encontre outros delitos na mesma descrição, estes são atribuídos nas classificações secundárias.

**Figura 9: ligação Tráfico, Procurados, Milícia para Roubo de cargas em 2017**



Fonte: Figura produzida pelo Arcabouço

Sem as classificações secundárias seria mais difícil fazer essa associação direta entre traficantes, o roubo de cargas e Milícia, uma vez que o rótulo de "tráfico de drogas", aparece apenas nas Classificações Secundárias quando o denunciante descreve ao operador que são traficantes que estão roubando caminhões de cargas ou quando citam a participação de Milicianos dando apoio e proteção de alguma forma, Figura 8 acima.

As classificações secundárias são importantes pois ajudam as autoridades e os investigadores a determinarem com mais precisão o nível de urgência da denúncia em vez de usar simplesmente a classificação principal para isso. Assim podem determinar melhor as prioridades de investigação e quais tipos de ações policiais devem ser executadas. Essa estrutura de classificação é de extrema importância para a associação entre crimes, um exemplo relevante são crimes de roubo de carga que assolaram o estado do Rio de Janeiro, principalmente no ano de 2017. Esse tipo de crime que antigamente era executado principalmente por ladrões sem maiores ligações com outros tipos de delitos passou a ser executado principalmente por indivíduos ligados ao tráfico de drogas e, posteriormente, por grupos de milicianos.

### 3.7 BASE DE DADOS DO DISQUE DENÚNCIA

O primeiro passo, antes de se pensar nas estratégias de relacionamento entre as bases de dados, é conhecer sua estrutura e avaliar a qualidade dos dados. O objetivo nesta etapa é detectar anomalias, verificar se os dados estão íntegros e avaliar a qualidade do preenchimento de todos os atributos para, finalmente, conseguir uma visão geral sobre quais informações de fato são possíveis de ser extraídas e quais atributos podem ser utilizados com maior eficiência.

Antes da utilização dos relatos de textos das denúncias, diversas transformações foram necessárias para que o arcabouço pudesse trabalhar com o texto de forma coerente. Sem as transformações, limpezas e padronizações todo o resultado do arcabouço ficaria comprometido. Mostraremos a seguir, toda a estrutura dos dados, formatos, distribuição e conteúdo da base de dados. Em seguida, discorreremos sobre o processo de limpeza de dados necessária a base de dados e as ferramentas e *softwares* necessários para esses procedimentos, assim como os problemas e soluções encontradas em sua implementação do estudo de caso.

### **3.8 ESTRUTURA DOS DADOS**

A estrutura original da base de dados possui um formato relacional e é composta de 13 colunas. São elas:

1. Mês da denúncia

2. Número da denúncia Número sequencial que serve como identificador único da denúncia no mês.

3. Classificação: Tipo do delito ou crime, ex.: tráfico de drogas, Roubo de Carga.

4. Tipo da classificação: Marcação binária que demonstra se a classificação dada é principal ou secundária. Uma denúncia pode conter mais de um tipo de crime, sendo decomposta em outra denúncia, o que é chamada de principal, a ocorrência prioritária e normalmente mais grave. E secundária a outra(s) citações contidas na denúncia a outros crimes. Serve para um acompanhamento das correlações conjuntas características a ocorrências criminais. Ex: barulho de som, porém é um baile funk com vendas de droga e prostituição.

5. U.F. do delito: Unidade Federativa onde ocorreu o crime.

6. Cidade do delito: cidade onde ocorre o relato da denúncia

7. Tipo do logradouro: Rua, Beco, Viela, Avenida, Praça, etc.

8. Logradouro do delito: Tipo do logradouro e suas características. Ex: Casa verde de portão preto

9. Número do logradouro: Número do logradouro referente a denúncia. Ex: 13 B

10. Complemento do logradouro: Complemento quando ocorre do logradouro. Ex: casa 02

11. Bairro do delito: Bairro onde ocorreu os fatos relatados na denúncia. Ex: Boa Esperança

12. Localidade do delito: Ponto de referência ou nome popular do local.

13. Texto da denúncia: Relato da denúncia com todos os fatos narrados e com todas as características sobre o que está sendo denunciado, suas entidades e informações necessárias para compor a denúncia. Cabe ressaltar que, quando não temos todas as informações para compor uma denúncia essa é cadastrada como um atendimento, sendo salva no sistema para

posterior análise. Nesta análise busca-se verificar se existem dados ou informações que possam complementar esses dados com outras denúncias ou se existe correlação com o histórico de denúncia do banco de dados para produzir conhecimento.

A base de dados original foi recebida em formato texto, mas precisamente no formato CSV, ou seja, texto delimitado por vírgulas, sendo cada linha uma entrada diferente da base de dados e cada conteúdo separado por vírgula uma coluna diferente da base.

As colunas mais relevantes para o desenvolvimento desse trabalho foram:

**Mês da denúncia e Número da denúncia:** compondo juntos a chave única de identificação de cada denúncia, desta forma foi possível associar cada linha da base a uma determinada denúncia.

**Classificação:** como essa informação foi atribuída individualmente para cada denúncia.

**Tipo da classificação:** o conteúdo deste campo foi utilizado para determinar se a classificação observada era principal ou secundária.

**Texto da denúncia:** representando o conteúdo escrito da denúncia em si com todas as características já citadas anteriormente. Este campo refere-se ao maior objeto de estudo deste trabalho e de vital importância ao arcabouço de texto de denúncias e todas as suas relações envolvidas neste estudo.

O arquivo original da base possui mais linhas do que denúncias únicas. Isso deve-se ao fato que uma denúncia pode conter uma classificação principal e diversas classificações secundárias, e cada uma delas é apresentada em uma linha separada, ou seja, uma denúncia que tenha duas classificações secundárias ocupará três linhas do arquivo, sendo uma linha com a classificação principal e outras duas linhas contendo uma classificação secundária em cada.

A diferença entre as linhas deste exemplo estará apenas no campo *Classificação* onde cada linha terá um valor diferente da outra, representando a classificação dada, e o campo *Tipo da classificação* onde a linha que representa a classificação principal terá o número 1 e as classificações secundárias terão o número 0.

A qualidade de dados é um fator importante para determinar a precisão das informações geradas para a tomada de decisão quando se utiliza informações provenientes de bancos de dados. Buscando-se sempre a menor quantidade possível de irregularidades e anomalias, procura-se corrigir os padrões de erro encontrado. A qualidade dos dados não é prerrogativa de perfeição dos dados, mas permite a definição de alguns conceitos como a garantia da informação correta e confiável (acuracidade), espelha se a sua fonte ou conteúdo (reputabilidade) é verdadeira (credibilidade), se é imparcial (objetividade), se a informação está disponível de forma fácil (acessibilidade), se a informação é útil para a tarefa em questão

(contextualidade), se a informação é restrita para resguarda a sua segurança (Segurança da informação ou salvaguarda) entre outros. Dessa forma o nível de qualidade de um banco de dados pode ser medido de forma específica (PIPINO; LEE; WANG, 2002). Cabe ressaltar que, nenhum sistema de informação trabalha com qualidade de dados em 100% ORR (1998).

A qualidade do banco de dados vai depender dos conceitos citados e do processo de limpeza utilizado. Criamos uma metodologia de processo de limpeza dos dados que considerou as necessidades técnicas necessárias, além de considerar os procedimentos para o trabalho com dados da área de segurança pública e seus protocolos e normas, bem como a proteção e salvaguarda desses dados e serão detalhados em partes deste trabalho. O nível de qualidade dos dados encontrado no estudo garante um nível de qualidade onde foi possível tomar decisões razoáveis, de forma rápida, e com garantia de segurança do arcabouço de texto empregado ao estudo de caso.

### ***3.9 CRIAÇÃO E IMPLEMENTAÇÃO DO CONCEITO DE ENTIDADES NO ROTEIRO DE ATENDIMENTO DAS DENÚNCIAS***

Propor a criação de entidades a metodologia de atendimento implementada ao *script* dos atendentes do DD, tornou se uma necessidade para captura, extração e obtenção de informações relevantes nas denúncias. Pressupõem conceitos aplicados a área de inteligência, denominando frações significativas, que são as pistas encontradas para composição de: entidades, provas, vestígios, fragmentos de informações entre outros. Composta, de dados ou informações, que podem ajudar a solucionar um delito ou crime referindo se as pessoas e/ou organizações, fazendo um elo na produção de conhecimento. São informações altamente relevantes aplicadas a área de inteligência para compor mapas de inteligências que são montados para a resolução de um crime normalmente mostradas em filmes investigativos, onde em uma parede são locadas fotos, recorte de jornais, pistas na composição dos cenários, tempo e fatos. Onde o analista de inteligência utiliza desses recursos para a produção do conhecimento. Estamos implementando o mesmo conceito aplicado em denúncias, quando o denunciante liga para realizar sua denúncia nossos atendentes são treinados na captação dessas pistas ou frações que quando estão relacionadas a alguém ou alguma coisa tornam se entidades, entidades relacionadas a essa pessoa ou objeto. Portanto são essas pistas que podem gerar um sucesso ou fracasso de uma operação por exemplo. Trataremos nesse trabalho como: O dado ou informação que faltava para a composição da informação investigada ou a identificação e características necessárias na identificação de um indivíduo denunciado, por exemplo etc. O trabalho desenvolvido dessas

pistas, em como os tipos de conhecimentos produzidos por elas, seguem um padrão e norma já estabelecido em uma doutrina de inteligência em segurança pública (RIO DE JANEIRO, 2005), para a sua aplicação. Precisando para sua implementação seguir normas e procedimentos padrões de inteligência para a produção do conhecimento com dados sigilosos e a visualização e acesso do conhecimento produzido dessas informações deve seguir esses protocolos de segurança. Sendo empregada por todos os órgãos e agência da área de inteligência, que utilizam e reconhecem sua necessidade, importância e o valor da informação produzida.

Desta maneira implementamos esses princípios buscando nos relatos de denúncias, as frações significativas e suas entidades, as informações relevantes para a composição do conteúdo dessas denúncias, na esperança em trazer maiores informações as policiais e na aplicação de estudos de cenários, envolvidos, vulgos<sup>6</sup>, carros, características, personas, rastros entre outras. Convém verificar a pertinência e credibilidade dos dados recebidos, bem como a relevância dessas frações significativas, que normalmente tratam se de elementos com grau de classificação por fonte e conteúdo relevantes a investigação criminal ou delito, a(s) prova(s) de envolvimento(s), a(s) pista(s) de elemento(s), as redes de conhecimentos, a hierarquia da facção, a formação de grupos e envolvidos entre outros. Nesta metodologia desenvolvida e aplicada neste estudo, são para compor, comparar e correlacionar com os aspectos essenciais dos assuntos ligados as Milícias e contém toda a complexidade do tema, suas relações criminais, seus elementos, sua forma de obtenção de lucro entre outros. Permitindo a produção de conhecimento desses temas para apoiar aos tomadores de decisão quanto ao assunto milícias, seu combate e sua repressão. Implementamos esses conceitos ao arcabouço de texto, não apenas para tentar entender essas questões e problema, mas principalmente para coletar as palavras, texto e semânticas direcionados a geração dessas informações que foram usadas no arcabouço na composição das nuvens de palavras resultante ao estudo milícias.

---

<sup>6</sup>Vulgo aqui é entendido como apelido e/ou codinome e não como os sentidos do dicionário que definem como o povo ou a maior parte das pessoas. Optou-se por manter a expressão desta maneira porque é termo recorrente na Segurança Pública.

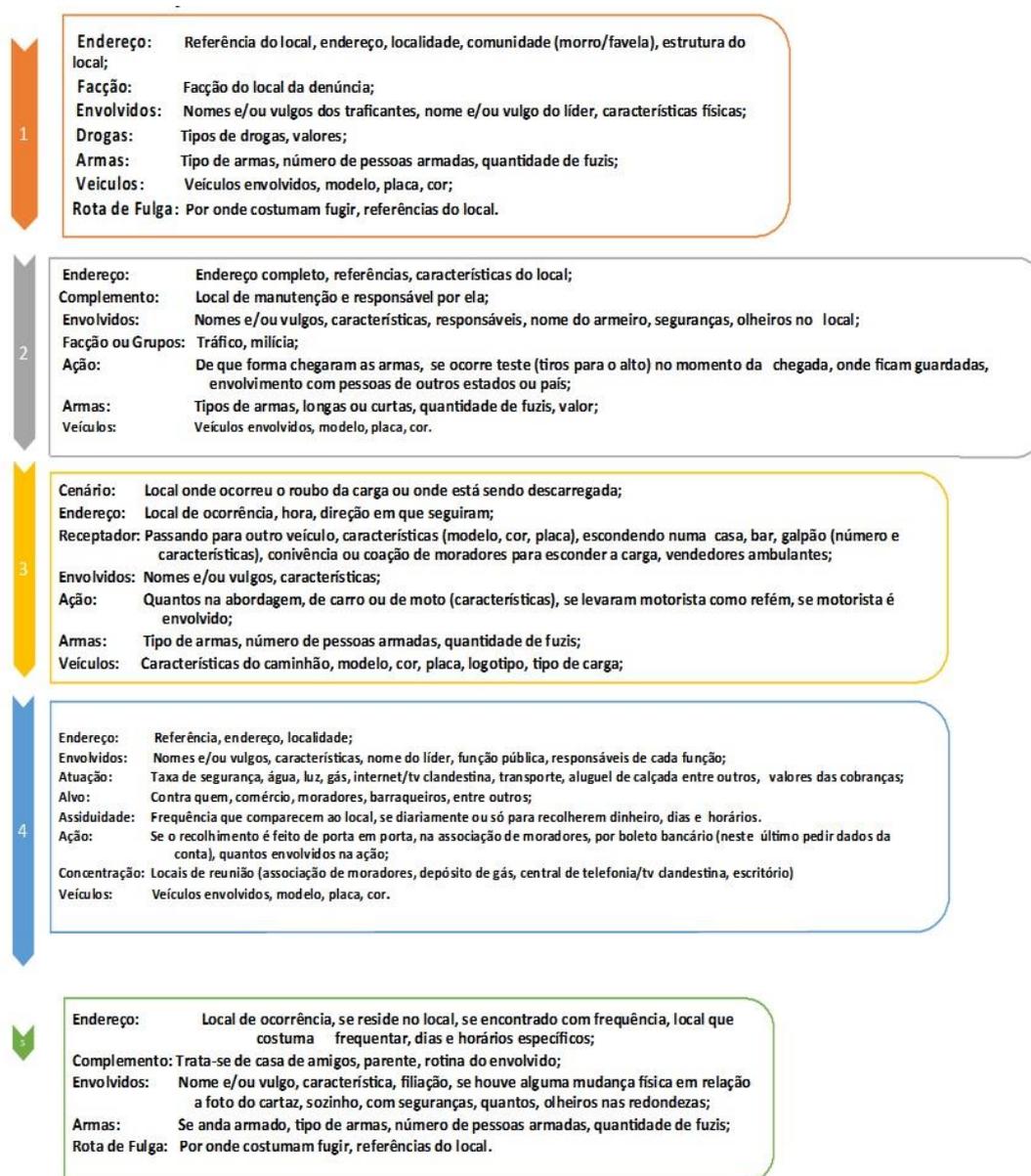
**Figura 10: Criação das Entidades captadas e utilizadas nas Denúncias**



Fonte: criação e implementação do autor

Consideramos ser uma boa base na obtenção de pistas relacionados as organizações criminosas necessárias ao estudo. E a seguir, as informações e entidades que são buscadas nas denúncias no atendimento e captação com os denunciante, Figura 10 abaixo. E que foram utilizadas em conjunto com o modelo proposto neste estudo, para o tema Milícias, e que serão mostrados nas próximas seções deste trabalho.

**Figura 11: Informações e entidades que são buscadas no atendimento das denúncias**



Fonte: Elaborada pelo autor.

Na seção a seguir, discorreremos sobre o problema do fenômeno das milícias no estado do Rio de Janeiro e sua relação com o Disque Denúncia.

### 3.10 O FENÔMENO MILÍCIAS

A milícia é uma forma de poder paralelo ao estado que vem ganhando espaços principalmente em favelas no Rio de Janeiro. Os primeiros indícios destes grupos na base de dados do Disque Denúncia surgem a partir de 2002. Inicialmente, os membros que constituíam

esses grupos começaram a ser citados com outros nomes, tais como: justiceiros, grupos de extermínio, polícia mineira, paramilitares dentre outros, apêndice D. Esse tipo de atuação ganhou grande proporção de denúncias a partir de 2006, quando houve maior engajamento popular acerca da atuação dos recém-denominados grupos de “milícias”, Apêndice D. No seu início a atuação de milícias possuiu formação, denominações e características específicas em seu modus operandi.

Alba Zaluar e Isabel Siqueira da Conceição (2007) definem as milícias como:

organizações formadas primordialmente por policiais e bombeiros militares, além de guardas penitenciários – ativos ou aposentados –, que garantem a segurança de moradores de algumas vizinhanças em troca de uma taxa mensal (ZALUAR; CONCEIÇÃO, 2007, p. 90).

Para o antropólogo e ex-comandante do BOPE, Paulo Storani (FONSECA, 2019), o surgimento das milícias cariocas ocorreu em três fases distintas. A primeira, no início dos anos 80, quando organizações que atuavam nas comunidades ofereciam proteção empregando trabalhadores e moradores da própria comunidade nesta dinâmica. A segunda fase configura-se quando agentes do Estado receberam suas moradias em conjuntos habitacionais principalmente na zona oeste do estado, e passaram a se organizar de modo a combater a entrada de traficantes nessas áreas. E, por último, a atuação da milícia nos modelos atuais, onde esses grupos avançaram para as organizações criminosas que conhecemos. Isto ocorreu quando policiais enxergaram a oportunidade de exploração de atividades econômicas nessas comunidades, principalmente após a expulsão dos traficantes e a proteção exercida através de pagamentos mensais.

Contudo, o uso do termo milícia ao longo da história mundial não tem o mesmo significado ao empregado a esses grupos paramilitares (ou, no caso do Rio de Janeiro, efetivamente militares). Ao realizar uma análise epistemológica do termo Milícia, originário da junção entre as palavras latinas *miles* (“soldado”) e *itia* (“estado”, “condição” ou “atividade”), dando sentido, quando juntas, de “serviço militar” ou “atividade militar”. Identifica-se também que este termo nem sempre foi utilizado para referenciar atuação de organizações criminosas. Historicamente também encontrado para simbolizar um poder paralelo, que não integra as forças armadas ou de polícia de um país. É frequentemente usado para indicar uma força militar composta de cidadãos ou civis que pegam em armas para garantir sua defesa e o cumprimento da lei, bem como o serviço paramilitar em situações de emergência, e uma das características é a de seus integrantes não receberem salários e não possuírem funções específicas, ou mesmo

não serem regidos por normas institucionais (ZALUAR; CONCEIÇÃO, 2007, p. 90).

Atualmente no Brasil, o termo *ilícia* usualmente refere-se a policiais e ex-policiais (principalmente militares), bombeiros e agentes penitenciários, todos com treinamento militar e pertencentes às fileiras profissionais do Estado, que tomaram para si a função de dar proteção e “segurança” em áreas supostamente ameaçadas por traficantes, além de civis armados. De acordo com a Anistia Internacional (2015), as milícias utilizam da força para extorquir a população em determinados territórios urbanos.

Portanto, a milícia que historicamente define-se como um poder paralelo que não integra as forças armadas ou de polícias de um país ganhou nova definição no contexto brasileiro, a partir do momento em que essas organizações passaram a ocupar os centros urbanos, espaços e territórios nos quais a presença do Estado é quase nula. Espaços esses considerados aglomerados subnormais, denominado comunidades ou favelas.

O estudo do arcabouço de texto com a temática milícia em denúncias recebidas pelo DD torna-se extremamente necessário para o entendimento, construção e compreensão desses cenários através do tempo, permitindo aos diversos campos de estudo e da área de Inteligência buscar soluções e compreensões a esse tema e suas representações simbólicas através dos anos. Dispor de uma ferramenta que trabalhe as análises de textos das denúncias, a possibilidade de identificação das entidades relacionadas a ela, a construção dos dicionários de palavras, expressões e temas envolvidos na temática milícia, é uma alternativa relevante para a atuação daqueles que integram as instituições que formam e trabalham com o tema Milícia.

Devido ao problema da violência que assola o RJ, faz-se necessário utilizar o Rio de Janeiro como estudo de caso para melhor compreensão da origem e da expansão desses grupos de milícia, que atuam na capital e em outros municípios do Estado.

Segundo abordado em matéria publicada pelo portal G1 de Notícias<sup>7</sup>, em 1998 a única comunidade que sofria com a ação de milicianos era a de Rio das Pedras, localizada na Zona Oeste do Rio de Janeiro. Dez anos depois, em 2008, o número de comunidades dominadas por milicianos era de 161. Em 2018, este número aumentou para 165 comunidades localizadas em 37 bairros de 11 municípios do Estado, demonstrando o crescimento e a força que as milícias ganharam ao longo de 20 anos. Neste contexto, compreende-se que a expansão das milícias no

---

<sup>7</sup> A matéria publicada, que faz parte da coluna “Franquia do Crime” produz uma análise acerca da atuação desses grupos criminosos no Estado do Rio de Janeiro com base em fontes oficiais (Ministério Público, Polícia Civil, Secretaria de Segurança e IBGE). Disponível em: <<https://g1.globo.com/rj/rio-de-janeiro/noticia/franquia-do-crime-2-milhoes-de-pessoas-no-rj-estao-em-areas-sob-influencia-de-milicias.ghtml>>. Acesso em 13 jun. 2020.

Rio de Janeiro ocorre em dinâmica de “franquias”, onde a “matriz”, ou o modelo utilizado para a implantação dessas franquias, seria o que foi implementado em Rio das Pedras, primeira comunidade dominada por milicianos, conforme poderá ser observado ao longo desta pesquisa.

Os milicianos da capital criaram o que os promotores do MP chamam de "franquias" ao definirem regras, tais como:

- ✓ Política de não-agressão entre os grupos;
- ✓ Uso, inicialmente, do nome anterior da quadrilha;
- ✓ Envio de integrantes a essas regiões;
- ✓ Ações de tomada de territórios com fardas das polícias Civil e Militar.
- ✓ Pagamento à matriz de percentual dos lucros obtidos pela nova quadrilha. (GRANDIN et al, 2018).

Conforme abordado pelo promotor Jorge Luís Furquim, do Ministério Público do Rio de Janeiro, em trecho divulgado na mesma matéria: “Eles estão se expandindo por meio de acordos nos municípios limítrofes, com operações que parecem muito com uma franquia” (GRANDIN *et al*, 2018). A seguir, abordaremos a história das milícias no Rio de Janeiro.

O surgimento de grupos de ex-policiais ocorreu da mesma forma que outros grupos militares vistos em nossa história, como, por exemplo, os grupos de extermínio que tiveram sua atuação marcante nas décadas de 60, 70 e 80 principalmente na Zona Oeste da cidade, bem como na Baixada Fluminense. Na Baixada Fluminense e Grande Rio, desde o início da década de 60, pessoas ofereciam meios ilegítimos na resolução de conflitos, afastando inimigos políticos e fornecendo algum tipo de proteção à população. Na década de 70, em algumas comunidades da zona oeste da cidade, já existiam grupos de extermínios que ofereciam proteção e investiam seus lucros em negócios imobiliários, em locais como Rio das Pedras por exemplo, uma comunidade com intensa migração nordestina e que se organizavam para impedir a entrada de traficantes.

O grande diferencial da região de Rio das Pedras é que esses grupos de extermínio já exerciam um controle sobre o território de maneira diferenciada e que o número de policiais e bombeiros pertencente a esse grupo era desde seu início bastante significativo. A principal diferença entre o fenômeno das milícias e o de grupos de extermínio está basicamente na ampliação da atuação dos negócios gerados com a chamada venda de produtos e serviços prestados. Por meio da extorsão e o envolvimento em atividades comerciais que são oferecidos inicialmente como proteção ou segurança contra assaltos e o distanciamento de traficantes desses locais, as milícias mantêm essa população refém, impondo seus serviços aos moradores.

Nos momentos iniciais da articulação desses grupos, buscava-se o domínio do território com a prerrogativa de fornecer segurança à população e uma suposta moralidade contra o tráfico de drogas, fazendo com que estes grupos tivessem suas ações de certa forma legitimadas pelos moradores da região dominada e, até mesmo, em alguma medida, pelo poder público. Além da própria segurança privada, exploravam serviços como: venda de botijões de gás, água, sinal pirata de TV a cabo, a compra de mercadorias mais caras, o sinal ilegal de TV – o famoso “gatonet”, a cobrança de taxa as cooperativas de transporte alternativo que circulam no território, a cobrança de pedágios para entregas nesses territórios, o pagamento de percentuais para compra, venda e aluguel de imóveis, o que garante o domínio dos novos moradores desde o primeiro momento.

Estas são algumas das características iniciais desses grupos. Podemos observar claramente um cunho comercial de opções de serviços a essa população e o controle desse território que passa ser dominado militarmente como características desses milicianos e uma das formas de distinção quando comparados aos grupos de extermínios, encontrados normalmente na região Metropolitana da cidade e que cobravam dos comerciantes locais e matavam por encomenda utilizando para isso pistoleiros nordestinos em sua maioria (BURGOS, 2002). A seguir, demonstraremos a diversidade entre os grupos milicianos.

Os “casos de repercussão” (KANT DE LIMA; EILBAUM; MEDEIROS, 2017), tais como, a tortura sofrida pela equipe<sup>8</sup> do Jornal O Dia, ocorrido em 2008 e o homicídio da Juíza Patrícia Acioli<sup>9</sup>, ocorrido em 2011, além da CPI das Milícias, realizada no ano de 2008, são eventos que deram maior visibilidade à atuação das milícias por meio da veiculação de reportagens sobre o tema. Esta visibilidade midiática influenciou a ação das milícias que passaram a funcionar de forma mais discreta (CANO; DUARTE, 2012). Como o dito anteriormente, segundo matéria publicada pelo portal g1 de notícias no dia 14 de março de 2018<sup>10</sup>: anos após o início da milícia em Rio das Pedras, de acordo com um relatório elaborado pela Subsecretaria de Inteligência, à época ligada a Secretaria de Estado de Segurança do Rio de Janeiro, em 2008 foram listadas 171 comunidades dominadas por milicianos, sendo 161 na região metropolitana do Rio de Janeiro. É inegável pensar, portanto, que neste processo histórico

---

<sup>8</sup> Para saber mais sobre a tortura sofridas por jornalistas do veículo O Dia, consultar <https://piaui.folha.uol.com.br/materia/minha-dor-nao-sai-no-jornal/>. Acesso em: 13 jul. 2020.

<sup>9</sup> Para saber mais sobre o assassinato da Juíza Patrícia Accioli, consultar: <http://g1.globo.com/rio-de-janeiro/noticia/2011/08/juiza-assassinada-sofreu-emboscada-e-levou-21-tiros-diz-delegado-no-rio.html>. Acesso em: 13 jul. 2020.

<sup>10</sup> Matéria intitulada “Franquia do Crime”, disponível em <<https://g1.globo.com/rj/rio-de-janeiro/noticia/franquia-do-crime-2-milhoes-de-pessoas-no-rj-estao-em-areas-sob-influencia-de-milicias.ghtml>> acesso em 23 de março de 2020.

que permeia os primórdios dos movimentos, ainda nos anos 60 e 70, que posteriormente viriam a ser solidificados e dariam origem às milícias na forma como conhecemos hoje, não trata-se de uma milícia unicamente, mas sim de vários grupos diferentes que agem de maneira semelhante. Estes grupos não são homogêneos. Eles respeitam as especificidades de suas lideranças, dos locais onde se instalam e, ainda, do perfil dos moradores e comerciantes de cada uma dessas localidades. Ou seja, apesar dessa expansão em larga escala e da ideia de franquia implantada pelos milicianos de Rio das Pedras, é importante ressaltar que esses grupos não atuam de forma única e padronizada nas diversas regiões onde se fazem presentes. Por exemplo: um grupo de milicianos que atua em Austin, Nova Iguaçu, pode possuir formas específicas de administrar os conflitos locais e não necessariamente adotam as mesmas medidas utilizadas por um grupo de milicianos que atua na região da Taquara, na Zona Oeste do Rio de Janeiro.

Observando as formas de atuação desses grupos, é interessante pensar que todos eles começaram a modificar suas formas de atuação em meados de 2006, até o ano de 2008, a partir de então nota-se que os milicianos passam a atuar de forma mais "discreta" (CANO; DUARTE, 2012).

Isso ocorre devido a alguns fatores determinantes, como o caso dos homicídios dos jornalistas do jornal o dia, ocorridos em 2008, fato que será melhor abordado no tópico adiante, e a repercussão (EILBAUM; MEDEIROS; KANT DE LIMA, 2016) que envolve esse caso, que acaba gerando uma demanda por respostas e ações do poder público. Com a postura de enfrentamento mais direta do Estado às Milícias, já sendo vista em seu plural e com suas características específicas em cada área dominada, esses grupos milicianos passam a agir de forma menos enérgica, conforme abordam os autores Inácio Cano e Thais Duarte (2012).

### ***3.11 CPI DAS MILÍCIAS E SUA IMPORTÂNCIA EM ENTENDER O FENÔMENO***

No início do ano de 2008, jornalistas do jornal O Dia são torturados por milicianos. enquanto buscavam produzir matérias em área dominada pela milícia. O fato gera uma comoção pública, repercutindo em mídia nacional e internacional, despertando muito interesse pelo tema.

Esse foi o marco para que a Assembleia Legislativa do Estado do Rio de Janeiro aprovasse a criação da CPI das Milícias. Realizada entre os meses de junho e novembro de 2008, sendo composta por reuniões com estudiosos, profissionais de Segurança, delegados e membros do Ministério Público. Foram realizados encontros para discutir desde as definições do termo Milícia, seu *modus operandi*, sua atuação e estrutura, formas de comércio e produtos, taxas cobradas, envolvimento com políticos e candidatos a cargo público, a franquia de

expansão de território, as formas de atuação entre outros fatores durante o período. O resultado produzido um relatório final encaminhado a todos os órgãos competentes os Resultados CPI das Milícias no final dos trabalhos.

A primeira curiosidade trata que nenhum dos membros participantes desse grupo foram unânimes quanto a uma definição do termo milícia, demonstrando a dificuldade em definir este fenômeno complexo (ALERJ, 2008).

Os atores envolvidos, suas dinâmicas, atividades econômicas, entidades de relacionamentos, assim como o processo semântico utilizado em suas definições serão utilizadas na comparação e construção de entendimento do tema.

Algumas perguntas precisavam ser respondidas, tais como: O que são as milícias? O que elas representam para a segurança pública e para a cidade? Que condições e consequências têm para moradores dos locais por elas dominados? Que diferenças existem entre a dominação das milícias e de outros grupos armados como o narcotráfico? Quais são os cenários futuros para a segurança pública da cidade em função do surgimento das milícias?

José Mariano Beltrame, ex-Secretário de Estado de Segurança pública, defende que “as milícias constituem, uma das maiores ameaças concretas ao Estado”. Destaca-se que “no México, por exemplo, grupos milicianos evoluíram para o que lá já é chamado de Narco milícia, ou seja, a milícia atuando no narcotráfico.”(ALERJ, 2008, p. 43).

Um estudo de destaque foi apresentado a CPI das Milícias e ajudou a compreender algumas das características do fenômeno milícia, *Seis por Meia Dúzia? Um estudo Exploratório do fenômeno das chamadas “Milícias”, no Rio de Janeiro*. Trata-se de pesquisa pioneira ligada a milícia, realizada pelo sociólogo Ignácio Cano, a partir de matérias jornalísticas dos jornais O Globo e O Dia, entre janeiro de 2005 e setembro de 2007, que continham a palavra: Milícia ou Polícia Mineira, e a utilização dos registros de relatos das denúncias do Disque Denúncia, entre janeiro de 2006 e abril de 2008.

Analisando seus dados procurou responder questões-chave, como, por exemplo, o que são as milícias, realçando as dificuldades de um único conceito sobre o termo, como vimos acima, principalmente por abordar de uma realidade extremamente dinâmica e complexa os eventos que fizeram com as milícias alterassem significativamente seu comportamento nos últimos anos. Diversos mecanismos e marcos foram apontados neste estudo como fatores de reposicionamento do comportamento adotado pela milícia, seja com o intuito de ludibriar ações de investigação ou dificultar os levantamentos de informações relevantes ao seu modo de operação.

O fato de a população com frequência demonstrar medo de se expor frente as incertezas

de ameaças que sofrem ou pelo fato de muitas pessoas se recusarem a testemunhar e exigirem garantias de sigilo para fazê-lo é o que torna a pesquisa ainda mais desafiadora nesta área. Há vários fatores por trás disto, haja vista que a identidade dos membros da milícia precisa ser mantida em sigilo, diferentemente do narcotraficante tradicional em que sua fama é um componente primordial para sua dominação territorial. Basta considerar que os membros da milícia pareciam ser agentes do estado, basicamente policiais, o medo de que eles acabassem sabendo de qualquer depoimento era maior.

Consequentemente, é por isso que as denúncias recebidas pelo DD tornam-se uma fonte especialmente importante nesta dinâmica, pois inúmeras categorias de crimes não são denunciadas diretamente ao estado, seja por medo, ameaça ou até a própria descrença nas autoridades.

Para Cano, a definição de Milícia tem partida em função de cinco eixos que devem ocorrer simultaneamente. Sendo:

- ✓ O controle de um território e da população que nele habita por parte (ou em seu todo) de um grupo armado irregular (não oficial);
- ✓ O caráter coativo exercido desse controle;
- ✓ O impulso de lucro individual como motivação central;
- ✓ O discurso de legitimação alegado à proteção desses moradores e a criação de uma ordem (imposta);
- ✓ A participação ativa e reconhecida dos agentes do Estado.

Podemos observar que os três primeiros eixos não são características apenas de grupos milicianos. Encontramos os mesmos elementos como o controle territorial, a coação e o lucro também em outros grupos armados como o narcotráfico. Os traços mais marcantes encontrados na Milícia portanto são: o discurso de legitimação e a participação dos agentes públicos. Esses são os grandes diferenciais.

No discurso de legitimação, que é o ponto central dos milicianos, trata da “proteção” dos habitantes naquele determinado território dominado. É apresentado como uma alternativa ao narcotráfico, como proteção contra esses grupos e contra todo tipo de ameaça que eles poderiam exercer. Desta forma, tentando se legitimar, tornam-se um “mal necessário” e apresentam-se como a solução ao problema do tráfico de drogas. Segundo Jaqueline Muniz, estudiosa de segurança pública,

o medo também é origem da legitimidade informal dos grupos de controle. Ele instrumentaliza de tal forma a insegurança que torna viável a crença e a aceitação de justiceiros, salvadores ou qualquer sorte de libertador que ofereça

segurança.

É nesse sentido que comércios ilegais e informais de segurança se expandem e dão origem, na prática, à privatização informal da segurança, diante de uma opção política que se absteve, nos últimos anos, de oferecer uma alternativa de segurança pública para a população. (ALERJ, 2008, p. 39)

Quanto ao quinto ponto e não menos importante, a participação de agentes do Estado, funciona como uma divulgação desses grupos por ser um grande diferencial, uma espécie de certificação da Ordem Pública, o “lado bom da força”. São agentes que fazem parte do Estado, portanto, qualquer problema no território, poderão acionar usar sua rede oficial de segurança (as polícias) para ajudar na proteção territorial. Diversas denúncias demonstram ações quase que conjuntas entre forças policiais e milicianos em entradas de territórios ou abrindo caminho para milicianos em territórios dominados por narcotraficantes, ou a expulsão dos narcotraficantes para, em seguida, os milicianos entrarem.

Para os estudiosos Jaqueline Muniz e Domício Proença o termo milícia é inadequado, pois, para eles, tratam-se de gangues formadas por policiais e ex-policiais que vendem segurança contra eles próprios. Para Domício Proença, o conceito de milícia é: — arranjo de gente armada querendo prover segurança fora da lei. Todo e qualquer grupo que age de forma ilegal. (ALERJ, 2008, p. 35).

### **3.12 MODUS OPERANDI DAS MILÍCIAS NO RIO DE JANEIRO**

A expansão e criação de novos grupos de milícias na cidade demonstrou diversos avanços e involução, decorrentes das dinâmicas locais desses territórios (Cano, 2006). Em algumas comunidades da ZO, esse domínio demonstra ter sido muito mais sutil e progressivo que em outros territórios onde foi feito por invasão armada. Isso deve-se principalmente ao fato de que não havia tráfico, nem crime organizado antes da entrada dos milicianos. Depois que as comunidades são ocupadas pela milícia, as associações de moradores passam a ser controladas por milicianos que escolhem um representante para essa função – assim como era feito pelos narcotraficantes anteriormente nos locais onde havia tráfico de drogas. É implementado um forte esquema de vigilância a essas comunidades, típico de locais de expansão de milícia.

Implementa-se uma organização e controle através de cadastros, reuniões, recibos, patrulhamento entre outros. Em depoimento à CPI, o delegado Claudio Ferraz narrou:

A situação de Campo Grande, na Zona Oeste, onde há patrulhamento de viaturas com 25, 30 homens uniformizados, de fuzil, com sistema de comunicação, sob o comando de milicianos denunciados. Há registros de práticas semelhantes em Volta Redonda, na Região dos Lagos, em Macaé, em diversas comunidades grandes, pequenas, desde que tenham condições de

produção de impostos informais, porque já existem casos de comunidades que foram tomadas e depois abandonadas. (ALERJ, 2008, p. 44)

O *modus operandi* de um grupo na ocupação de uma comunidade é variável em relação a cada local. E vai depender da presença de alguns dos seguintes elementos:

- ✓ Existe tráfico na área pretendida, então emprega o uso da força;
- ✓ Não há tráfico e a população resiste, milicianos passam a assaltar as casas e o comércio;
- ✓ Moradores e comerciantes intimidados e acuados pela situação passam a contribuir financeiramente com valores mensais estipulados pelos milicianos;
- ✓ Comunidades menores aceitam sem resistência a presença dos milicianos;
  - ✓ Para manter o domínio conquistado, é costume expulsar pessoas ligadas ao crime, às facções criminosas e familiares de traficantes;
  - ✓ Agir de forma violenta na área conquistada, espancando, torturando e matando viciados em drogas e criminosos que resistem a obedecer a suas ordens;
  - ✓ Estabelece o toque de recolher;
  - ✓ Quem se recusa a pagar sofrem represálias: assaltos, ameaças, agressões, espancamento, tortura, expulsão da comunidade e até mesmo a morte;
  - ✓ Mortos têm seus corpos largados em cemitérios clandestinos para não serem encontrados.

O controle econômico exercido pelas milícias dá-se pela imposição da cobrança de taxas aos moradores. Algumas dessas taxas foram levantadas e precificadas em algumas comunidades sendo o conjunto de todas as opções em outras apenas uma parte delas, gerando dessa forma um fluxo de arrecadação financeira que pode ser visto na Figura 12 abaixo.

De acordo com denúncias recebidas pelo Disque Milícias na época das CPI das Milícias e de dados do DD, os valores cobrados oscilam conforme a área dominada e o grupo dominante: Segurança de moradores de R\$ 15,00 a R\$ 70,00 Segurança do comércio de R\$ 30,00 a R\$ 300,00 Instalação de TV a cabo de R\$ 50,00 a R\$ 60,00 Sinal de TV a cabo de R\$ 20,00 a R\$ 40,00 Internet de R\$ 10,00 a R\$ 35,00.

**Figura 12: Diagrama da arrecadação Financeira da Milícia**



Fonte: Relatório CPI das Milícias (ALERJ, 2008)

A cobrança da arrecadação dessas taxas pode ser oriunda de:

- ✓ Taxa de segurança pessoal;
- ✓ Taxa para moradores que possuem veículos;
- ✓ Taxa de instalação e mensalidade dos serviços de sinal de TV a cabo e Internet;
- ✓ Controle e ágio na venda de gás e garrafão de água;
- ✓ Cobrança de alimentos para composição da cesta básica;
- ✓ Taxa do valor da venda de imóveis (que varia de 10 a 50%);
- ✓ Taxa para legalização de imóveis;
- ✓ Taxa para permitir construções na comunidade;
- ✓ Controle e cobrança de taxas do transporte alternativo: mototáxi, vans e kombis;

- ✓ Cobrança de taxa para funcionamento de barracas de camelô, festas e pipoca e demais eventos;
- ✓ Cobrança de instalação de portões e guaritas nas comunidades.

O documento apresentado à Comissão Parlamentar de Inquéritos das Milícias, em 2008, da Assembleia Legislativa do RJ, apresentava um total de 161 comunidades sob o comando das Milícias. Estima-se que em 2018, uma década depois, cerca de 2 milhões de pessoas no Rio de Janeiro estejam em área sob influência da Milícia.

Em 37 bairros e 165 comunidades da região metropolitana do RJ – o que equivale a ¼ um quarto da cidade do Rio em área de atuação, estas áreas de influência desses grupos equivalem a 348 KM<sup>2</sup> <sup>11</sup>. Essa população permanece coagida a cobranças e taxas como as citadas anteriormente e sem grandes mudanças em relação ao quadro apresentado anteriormente.

Esses dados foram produzidos e publicados por informações do Ministério Público Estadual, da Polícia Civil e da Secretaria de Estado de Segurança em conjunto com o Instituto Brasileiro de Geografia e Estatística (IBGE)<sup>12</sup>. A principal diferença em relação ao ano de 2008 foi a expansão para bairros inteiros, principalmente na região da Zona Oeste, Baixada Fluminense e em outros municípios como Itaguaí. A forma de entrada também não se modificou, sendo em locais sem atenção do poder público e que serviços básicos como transporte e segurança foram ignorados há muito tempo. O apoio de grupos anteriores continua sendo uma porta de entrada para esses grupos como o que pode ser denominado “Franquias”. Segundo o Promotor do Ministério Público do Rio, Jorge Luís Furquim, a expansão das milícias ocorre por meio “de acordos nos municípios limítrofes, com operações que parecem muito com uma franquia.” O promotor prossegue em seu relato apontando que:

---

<sup>11</sup><https://g1.globo.com/rj/rio-de-janeiro/noticia/franquia-do-crime-2-milhoes-de-pessoas-no-rj-estao-em-areas-sob-influencia-de-milicias.ghtml> estudo acesso em: 25/03/2020.

<sup>12</sup> <https://epoca.globo.com/como-rio-das-pedras-virou-sede-do-escritorio-do-crime-23398750> estudo acesso em: 25/03/2020.

A partir de informações ou de relações com pessoas desses municípios, eles chegam fazendo acordos com moradores, grupos de milicianos ou grupos de extermínio. Empréstam o 'know-how', deixam que se use o nome intimidador. Agem fardados com roupas da Polícia Civil ou Militar nas ações de tomada de territórios. (GRANDIN et al, 2018, s.p.)

Um ator novo neste processo e que tem como função ser, o ponto em comum, o ponto de ligação entre os grupos da rede, fazendo cumprir as determinações sobre as funções e gerindo as brigas internas dessas células que são criadas, é o Franqueador, que é o responsável pelas marcas, dizendo a forma de atuação dos grupos. Para Furquim, este foi o modo utilizado em Itaguaí e Seropédica, apesar de serem grupos diferentes (GRANDIN et al, 2018)

Já quando o assunto é o combate a essas organizações, na última década, 1.403 pessoas foram levadas a cadeia. Mas entre 1997 e 2006, apenas seis integrantes destes grupos haviam sido presos pelas polícias. Essas informações que foram coletadas de uma reportagem do G1, chamada a Franquia do Crime<sup>13</sup>.

O crescimento das quadrilhas na baixada fluminense, ocorreu de maneira muito acelerada e ficou conhecida como a nova fronteira. Para Alves, em seu livro, Uma guerra pela região geográfica do Rio de Janeiro, a reorganização da estrutura do crime. Esta dimensão pode ser expressa em números: nove municípios, 43 localidades. Essa é a abrangência das atividades de milícias na Baixada Fluminense, que tem 13 municípios. Já há grupos de milicianos em: Belford Roxo, Duque de Caxias, Itaguaí, Mesquita, Nova Iguaçu, Queimados, São João de Meriti, Seropédica e Magé.

Os dados foram obtidos junto ao núcleo de inteligência da Delegacia de Homicídios da Baixada Fluminense, documentos do Ministério Público e investigações da Polícia Civil. Uma das vertentes de expansão das milícias saiu de Campo Grande, na Zona Oeste do Rio, berço dessas organizações, e partiu em direção à Baixada. Esse deslocamento fica explícito no depoimento do miliciano Ricardo Teixeira da Cruz, conhecido como "Batman", um dos maiores milicianos, preso em 2016. O criminoso contou para a polícia que, a milícia denominada Liga da Justiça, teve sua origem nos bairros de Santa Cruz, Campo Grande, Cosmos, Inhoaíba e Paciência.

No depoimento, podemos perceber a narrativa do mesmo diagrama de arrecadação financeira apresentado na Figura 14 acima, sendo agora acrescentado maiores requintes de

---

<sup>13</sup> <https://g1.globo.com/rj/rio-de-janeiro/noticia/franquia-do-crime-2-milhoes-de-pessoas-no-rj-estao-em-areas-sob-influencia-de-milicias.shtml> acessado em 25/03/2020.

ameaças e crueldades. Segundo o criminoso, muitos crimes envolvem a ocultação de cadáveres, ameaças de morte, extorsões financeiras a comerciantes e empresários, esbulho de imóveis, além de uso de veículos clonados.

As principais atividades econômicas da milícia na Baixada, estão:

- Segurança privada;
- Extração clandestina de areia e saibro (Itaguaí e Seropédica);
- Exploração de transporte alternativo;
- Comércio e venda de água mineral e botijões de gás;
- Venda de cestas básicas
- Comércio ilegal de sinal de TV a cabo.

O tripé Polícia, Milícia e Poder<sup>14</sup>, estudado e discutidos por alguns autores entre eles Ignácio Cano (2019), estuda o avanço da milícia no Rio de Janeiro, apresenta que a ligação entre a milícia e a política é historicamente relacionada a atuação dos grupos de extermínio na Baixada Fluminense, fato já citado anteriormente. Como a região fica em território com baixo policiamento e distante do centro de poder do RJ, torna-se espaço propício para a atuação desses grupos. Os grupos de extermínios já estavam infiltrados na política (CANO, 2019). Portanto, foi fácil para que as milícias fizessem o mesmo. Buscava-se a terminologia que era empregada e compreendida nesses territórios, seus atores e propostas, as palavras empregadas nessas denúncias sempre demonstravam importância nesses estudos relacionados ao tema Milícia.

Neste processo, para Cano:

As denúncias sobre milícias realizadas ao Disque-Denúncia são casos em que as pessoas se sentem, presumivelmente, vitimizadas pelas ações desses grupos até o ponto de ligar para denunciar. De uma forma indireta, esses registros devem permitir acompanhar a extensão das atividades das milícias e a natureza das mesmas. (Relatório CPI das Milícias)

O autor continua

Nesse sentido, mesmo que cada denúncia individual não possa ser comprovada, o conjunto delas representa uma fonte significativa de informação criminal. O Disque-Denúncia é uma fonte particularmente importante sobre tipos de crimes que não costumam ser denunciados oficialmente, seja por medo dos denunciadores ou por descrença no

---

<sup>14</sup> <https://noticias.uol.com.br/reportagens-especiais/milicia-do-rio-se-infiltra-na-policia-e-na-politica-para-expulsar-rivais-e-expandir-territorio/#tematico-2 estudo> acessado em 25/03/2020.

sistema.(Relatório CPI das Milícias)

Cano também destaca a importância do anonimato pois, segundo ele, “é bem mais provável que os cidadãos se animem a realizar denúncias anônimas, razão pela qual os dados do Disque-Denúncia representam uma fonte de valor inestimável.” (Cano, 2008)

Por outro lado, é preciso destacar a importância do serviço prestado no acolhimento dessas denúncias e encaminhamento as autoridades competentes, bem como todos os dados contidos nas semânticas dessas denúncias, foco deste trabalho, que permite estudos de acompanhamento da expansão desses grupos em outras áreas e de toda as palavras, termos e expressões semânticas utilizadas nas denúncias.

Toda a semântica textual envolvida as denúncias em relação as palavras, expressões, frases e seus significados são objetos importantes de atenção neste trabalho por trazerem representações simbólicas referentes às dinâmicas e suas relações com o tema da milícia dentro dos territórios estudados, assim como de seus atores e entidades envolvidos, servindo como uma rede de entidades relacionadas nas denúncias que podem compor pistas de dados e informações sobre os envolvidos. Desta forma, contribuem para sua utilização em estudos na área de inteligência sobre esse contexto.

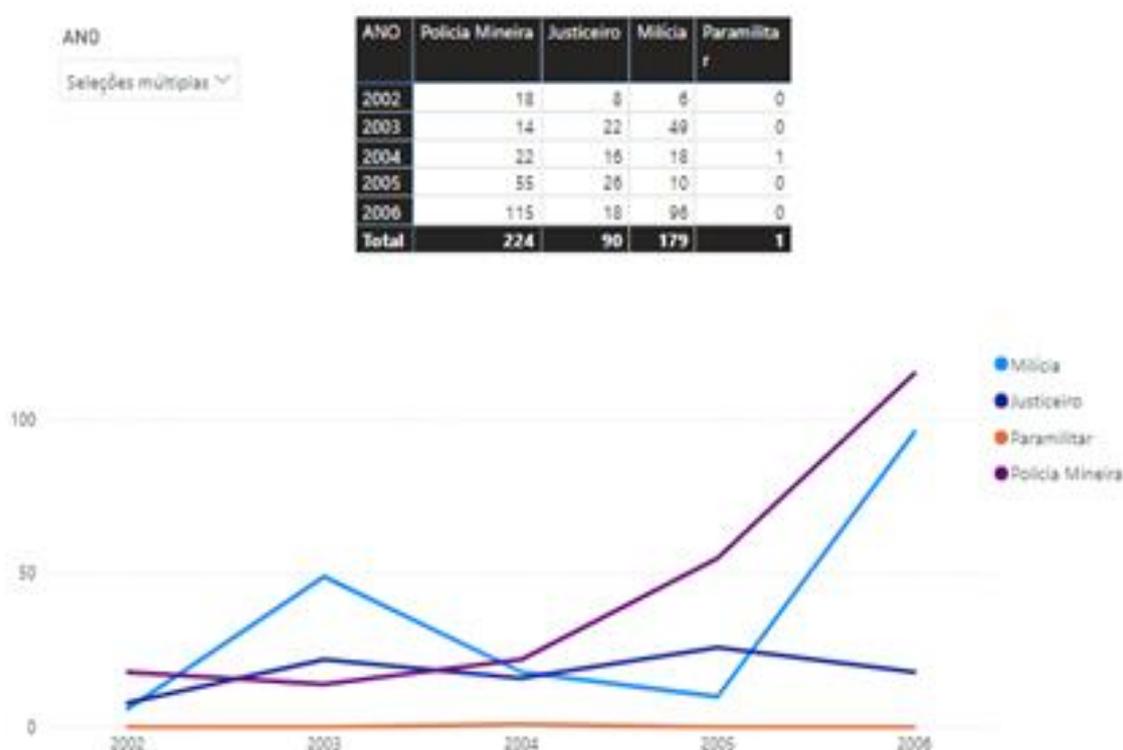
### ***3.13 ASSOCIAÇÕES E CORRELAÇÕES ENTRE OS ESTUDOS DE MILÍCIAS E A BASE DE DADOS DAS DENÚNCIAS***

Observando os dados quantitativos e qualitativos de denúncias cadastradas no sistema do Disque Denúncia, foi possível identificar os momentos iniciais, a conseqüente expansão das atividades das milícias no Estado do Rio de Janeiro, sobretudo na capital e em alguns municípios da baixada fluminense. Relacionamos os dados referentes as milícias e cruzamos com os pontos relevantes encontrados no decorrer do tempo, permitindo desta forma, aprofundar alguns pontos e neste caso apontar algumas generalizações e curiosidades em seu comportamento e forma de agir dessas milícias, para aplicar esse conhecimento produzido ao estudo de caso, objeto prático deste trabalho.

Neste aspecto, cabe ressaltar que as denúncias anônimas apuradas e analisadas para a realização deste trabalho estão compreendidas no intervalo temporal entre: 01 de maio de 2002 a 31 de dezembro de 2019. Portanto, elas refletem determinados momentos específicos da atuação de grupos milicianos no Estado do Rio de Janeiro.

No primeiro momento, as denúncias recebidas sobre o tema trazem em seus relatos os mesmos termos citados no estudo de Cano (2008). Alguns exemplos são: “Grupo Paramilitar”, “Justiceiros” e “Polícia Mineira”, como são chamados os grupos responsáveis pelo controle de Rio das Pedras até hoje, independentemente do surgimento e da ampla divulgação do termo milícia a partir de 2006 (Figura 13, abaixo).

**Figura 13: Termos citados em outros estudos**

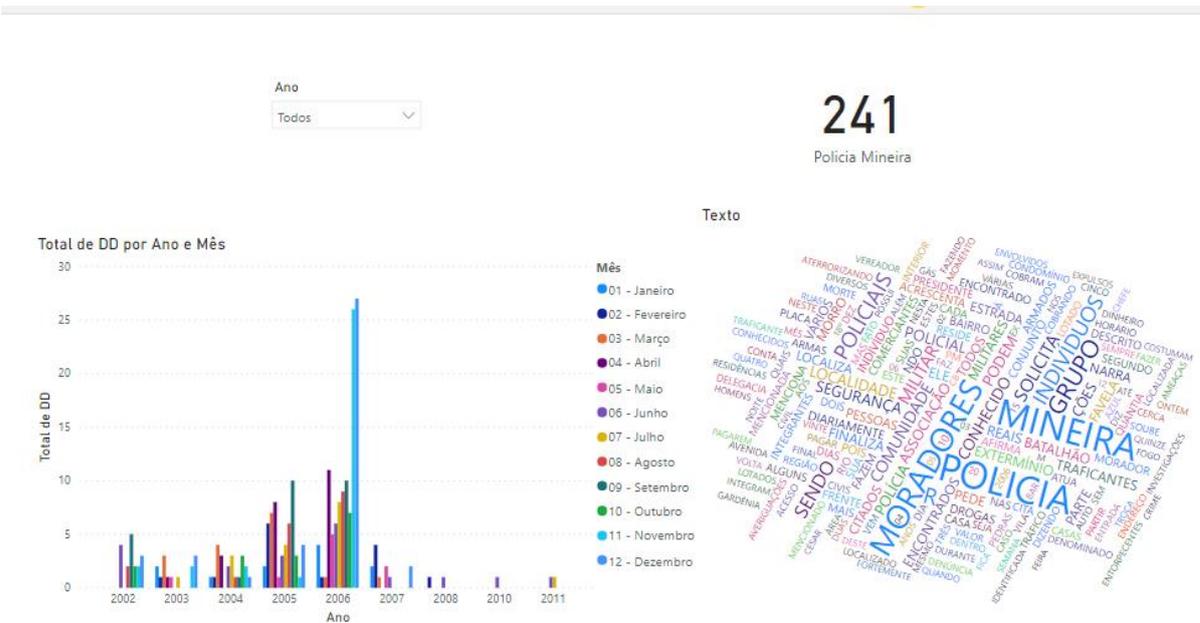


Fonte: Figura Produzida pelo Arcabouço

No entanto, podemos verificar na linha temporal, Figura 14 abaixo que, determinadas localidades já utilizavam alguns desses termos desde 2002 e em alguns casos continuam usando esses termos em suas denúncias até hoje, e em alguns casos utilizam por um longo período de tempo e depois alguns caíram em desuso, dependendo das localidades dessas denúncias. Algo que até então não havia sido citado em nenhum dos estudos encontrados e mostrando desse modo o potencial de produção de conhecimento da base de denúncias. Podemos perceber que, em geral, a expressão desses termos não é utilizada apenas como uma expressão local vinculada ao tema. E sim uma representação simbólica desses grupos e como são chamados nessas

localidades.

**Figura 14: Linha Temporal do Total de Denúncias por ano e mês e a nuvem de palavras com temas encontrados**



Fonte: Figura Produzida pelo Arcabouço.

Isto ocorre porque Rio das Pedras era dominada por um grupo conhecido como Polícia Mineira na década de 70 e 80. Este grupo supostamente protegia a comunidade utilizando-se da mesma ordem utilizada pela milícia atualmente. A Polícia Mineira utilizava o argumento de manter a ordem e empregava o código de conduta de acertos de contas, violando os direitos de cidadania da população local, fazendo com que as pessoas, residentes, comerciantes ou frequentadoras daquela comunidade, se sentissem intimidadas. Conseguimos perceber, nos cinco primeiros anos de análise, o uso do termo em denúncias sobre essa localidade e outras localidades próximas como apontado no estudo de Cano. As primeiras denúncias encontradas foram cadastradas no ano de 2002, e dão conta de informar sobre o *modus operandi* de alguns desses grupos.

Para melhor compreensão da representatividade das denúncias anônimas cadastradas no Disque Denúncia acerca da atuação desses grupos, é importante compreender o papel do Disque Denúncia como um refúgio, uma válvula de escape, uma “arma do cidadão” (MORAES, 2006 p.33) no combate às práticas criminosas. Sob esta ótica, os dados dispostos no banco de dados

do DD podem ser utilizados como termômetros que geram alertas sobre como determinadas dinâmicas criminais que causam reflexos no sentimento de insegurança da população fluminense. Com base no referencial teórico selecionado para a realização deste trabalho, bem como pode ser compreendido ao analisar as denúncias contidas no banco de dados do Disque Denúncia, é possível compreender como essa garantia de segurança local e o início do fenômeno da privatização da segurança de áreas “de risco”, juntamente a imposição da rede de produtos oferecida a comunidade já eram características desses grupos milicianos.

A maior modificação, no início desse processo de enraização e posterior expansão das milícias, foi a aproximação da população a associação de moradores, despertando o interesse eleitoral de eleger candidatos que representasse a comunidade no governo e assim serem ouvidos através de suas reivindicações e carências do território.

### **3.14 CONCLUSÕES DO CAPÍTULO**

Este capítulo apresentou a organização não-governamental Disque Denúncia, responsável pela base de dados aplicada a este estudo, demonstrando todo seu processo de trabalho e metodologia aplicada nos seus 25 anos de serviços e sua relevância no combate ao crime no estado do Rio de Janeiro. Da mesma forma, abordamos os procedimentos e métodos de trabalho adotados pela instituição em todos os seus canais de atendimento e a metodologia adotada nestes canais. A seção demonstra a classificação utilizada pelo DD e a importância que esta classificação tem na atenção dada pelas autoridades policiais ao crime identificado na difusão dessas denúncias. A seção discorre sobre a constituição das denúncias e suas relações com as classes, considerando que um mesmo relato pode apresentar mais de uma denúncia. Esta seção apresenta também a implementação do conceito de entidades no roteiro de atendimento da central e a importância e necessidade desta criação para melhores informações e denúncias na captura e obtenção de informações relevantes para melhores resultados e pesquisas.

Além do papel do DD na segurança pública no RJ, foi possível compreender a questão da milícia no contexto do Rio de Janeiro e sua relação na base de dados do DD. A milícia não é um grupo coeso e único, portanto, tratam-se de milícias, que atuam de modos diferentes, apesar de suas semelhanças. Milícias são grupos paramilitares que exercem um poder paralelo ao do estado em diversas zonas do estado. Estas organizações são fundamentalmente formadas por policiais, bombeiros e outros agentes da lei, mas também possuem em seus quadros outros cidadãos civis armados. As milícias impõem medo e violência nos locais onde atuam porque

muitos de seus agentes são também agentes do estado, as milícias contaram com a conivência do estado por muitos anos. Expandiram suas atividades e passaram a atuar também na prestação de serviços como o gato net, "proteção" e extorsão de comerciantes e moradores, entre outras atividades ilegais. Na base de dados do DD, as milícias começam a aparecer a partir de 2002. Isto não quer dizer que elas não atuassem anteriormente, mas as denúncias não eram relacionadas tão facilmente a esse nome, podemos identificar outros termos utilizados para se referenciá-la nas denúncias, algo não relatado em outros estudos. Procuramos explicitar nesta seção o *modus operandi* central de atuação destes grupos paramilitares, sua relação com a política e a tentativa do poder público de identificar estes grupos por meio da Comissão de Inquérito Parlamentar, a CPI das Milícias, terminado em 2008 e as informações encontradas na base de dados do DD.

#### **4 PROPOSTA *FRAMEWORK* SENSÍVEL AO CONTEXTO EMPREGADO AO ESTUDO DE CASO MILÍCIAS**

Neste capítulo apresenta a proposta do *framework* de análise sensível ao contexto utilizada no estudo de caso Milícias.

#### 4.1 CONTEXTUALIZAÇÃO DE CENÁRIOS

Com o objetivo de trabalhar com coleta de dados, integração de bases, extração e produção de conhecimento através de denúncias com a temática milícias, e simultaneamente com a reflexão do tratamento, limpeza, carga, agregação em grandes bases de dados, esta proposta tem por finalidade preparar o que intitulamos arcabouço de mineração de dados sensíveis ao contexto em denúncias, utilizado com um conjunto de ferramentas tecnológicas e utilizando o modelo proposto por este trabalho em um estudo de caso em milícias. Dispõe-se de um projeto de ciência de dados, complexo e relevante que tem como objetivo intrínseco a visualização dessas informações bem como extração e inferências de informações produzidas pelas denúncias, podendo ser visualizadas e utilizadas através dos *dashboard*<sup>15</sup> pelas áreas de inteligências, e implementando todos os pressupostos doutrinários empregados na construção do conhecimento pela área de inteligência, propiciando como resultados estudos, análises e textuais e indícios de pistas contidas em denúncias na produção de conhecimentos relacionados as milícias. Para isso, recorreremos às novas formas e atualizações do emprego de tecnologia na análise de textos dessas informações e frações de pistas muitas vezes escondidas, porém presentes nas denúncias, devido a toda complexidade inerente aos dados já discutidas anteriormente e que muitas vezes não foram observadas. Empregamos conceitos de *big data*, *data science* aliados a programação de linguagem natural na criação do ambiente analítico de coleta dos dados bem como, na construção do ambiente de análise dos dados e informações disponíveis em toda a base de dados do DD, considerando a metodologia de atendimento e a utilização das entidades aplicadas na metodologia – já mostrada anteriormente –, aumentando desta forma o volume de informações possíveis de conexões por meio de pistas ou fragmentos dispostos nas denúncias e muitas vezes ignorados ou não percebidos. Quer pelo volume de dados e informações, variedade ou velocidade da composição no recebimento e processamento desses dados, bem como a importância semântica neste assunto, mostramos todas as dificuldades encontradas em cada etapa do modelo proposto, assim como toda a teoria e prática aplicada na construção desse arcabouço e de que maneira resolvemos cada problema enfrentado ao longo do desenvolvimento.

O novo tipo de abordagem no tratamento dos dados do DD, bem como seus conhecimentos e valores foram obtidos não somente a partir do rigor metodológico no tratamento dos dados apurados, mas também da observação, da leitura e do compartilhamento de ideias entre os principais atores envolvidos na utilização dessas análises. Neste caso, são os

---

<sup>15</sup> Ferramenta para visualização dos dados e informações.

analistas da área de inteligência que utilizam essas informações na produção de conhecimento para suas agências e órgãos. O anonimato dos denunciante foi mantido e sempre será preservado, por ser um dos principais pilares estratégicos que sustentam a existência do DD e o que garante o reconhecimento por parte dos denunciante ao longo dos anos de atuação da ONG.

O método proposto nesta pesquisa permitiu a visualização de indícios ocultos pela complexidade das informações textuais que envolvem a denúncia em seus diferentes canais e propiciou a descoberta de conhecimento entre associações de palavras, termos e expressões ao aplicar, para esta finalidade, um modelo de mineração de texto composto de algoritmos híbridos e processamento de linguagem natural na obtenção de pistas e informações importantes ao contexto milícia.. Outro ponto que merece destaque é que todos os pressupostos aplicados a Doutrina de Inteligência em Segurança Pública (RIO DE JANEIRO, 2005), seus métodos e princípios de coleta de dados, bem como sua metodologia (o que, onde, quando, como e de qual forma) foram respeitados na aplicação deste método desde a construção dos ambientes (necessariamente seguros e sigilosos), até a confecção dos *dashboards* com suas informações pertinentes e seus resultados testados por pessoas ligadas a análise e estudo referentes ao tema milícia. Este estudo foi apresentado a alguns agentes de inteligência estaduais e federais durante a implementação e execução deste trabalho.

É necessário criar um ambiente tecnológico propício às interações que favoreçam a inovação, acompanhem, validem e monitorem os dados recebidos, não bastando apenas a estruturação de um sistema de informações através de tecnologias de *big data* e *data science*. O retorno por parte daqueles utilizam esse sistema em seu dia a dia ajudam a captar a real necessidade, aplicabilidade da ferramenta e o melhoramento contínuo a essas novas necessidades devem estar previstas. A troca de informações entre os agentes e instituições é vital para o bom andamento dessas necessidades e recursos aplicados e que durante todo esse trabalho contou com uma parceria frutífera com diversos atores envolvidos.

A metodologia proposta envolveu os ciclos de tratamentos históricos de denúncias recebidas pelo serviço do DD, de uma forma ampla, com toda a classificação de denúncias cadastradas em seu sistema – chamado de DDSYS. Utilizamos a totalidade de denúncias e dados cadastrados desde maio de 2001. A partir disso, foram aplicados algoritmos de extração de radicais temáticos para a construção de glossários e dicionários especializados a cada classificação de denúncia, em que se sobrepuje todas as principais palavras utilizadas em cada classificação de denúncia no decorrer dos anos. Semelhantemente, foi utilizada a extração dos relacionamentos de entidades contidas nas denúncias para identificar os temas, palavras e

expressões envolvidas nas narrativas denunciadas, promovendo o desenvolvimento de possíveis pistas, vínculos e cenários que compõem os crimes em formato de uma matriz de relacionamentos, com as principais palavras utilizadas para cada denúncia e sub-denúncia. Cabe ressaltar que todo o modelo de classificação das denúncias foi revisto e será explicado adiante neste trabalho. Todo o processo da construção dessas classificações foi longo e contou com a colaboração diversos especialistas ao tema e, por si só, já demonstra um resultado bastante positivo as análises que serão vistas nos *dashboards* apresentados. Por meio das ferramentas criadas neste estudo, tornou-se possível gerar nuvens de palavras por linha de tempo (período de tempo que deseja analisar) e por tipo de classificação e temas empregados na concepção de cenários. Os cenários são, então, convertidos em mapas de inteligência – metodologia empregada por agentes e área de inteligência, destinada ao diagnóstico de vínculos. Estes vínculos poderão ser de pessoas, locais, organizações, dentre outros, acoplados as pastas de investigações pertencentes e de responsabilidade de cada área. Podemos citar, como exemplo, a pasta Foragidos da Justiça (Procurados) que são informações sobre pessoas identificadas como foragidas da justiça e, hoje em dia é um dos principais produtos do DD.

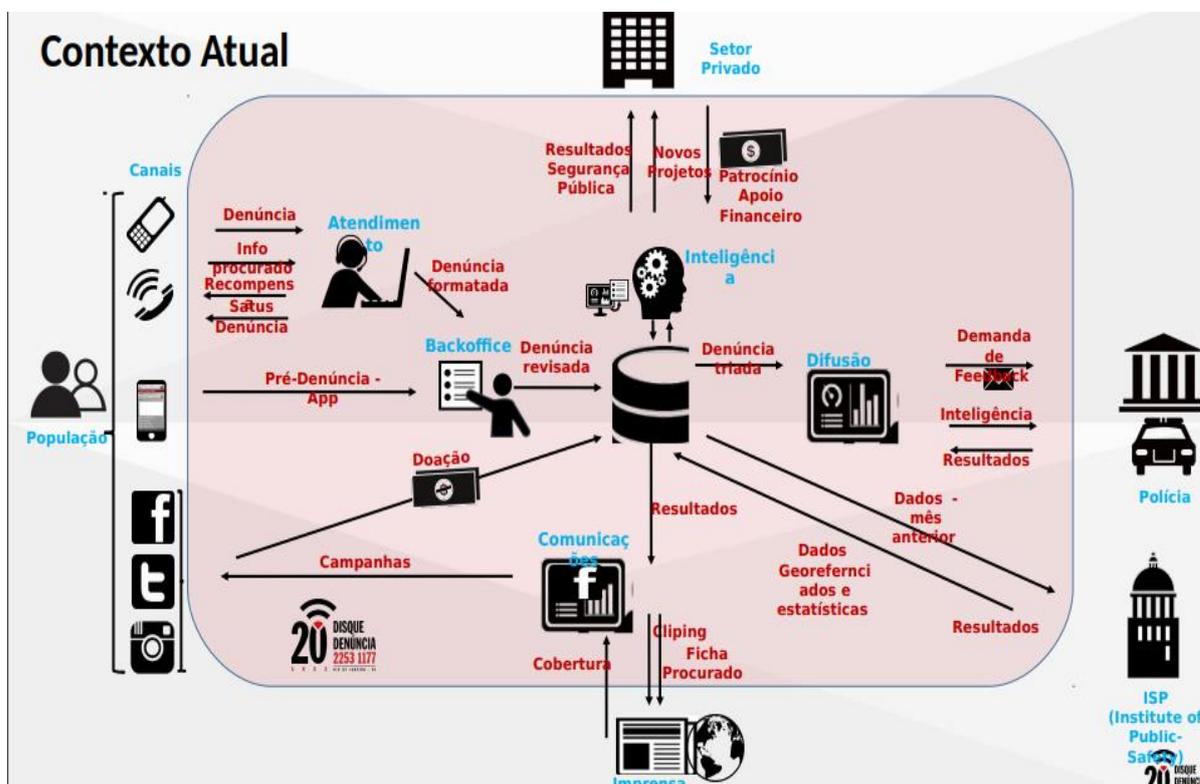
A visualização de indícios ocultos pela complexidade das informações textuais que envolvem o texto da denúncia deve-se ao fato da transcrição entre o texto original escrito ou narrado no idioma “nativo” e o texto estrutural composto na denúncia e enviado aos órgãos competentes. Para que se propicie a descoberta de conhecimento entre associações de palavras e expressões aplicadas a essa realidade, um modelo de mineração de texto composto de algoritmos híbridos aplicando um conjunto de metodologia próprio foi criado e empregado ao estudo de caso. A composição e criação dos módulos, as soluções empregadas e a tecnologia utilizada para a resolução dos problemas foram pensadas a partir dos recursos disponíveis na realidade das organizações voltadas para a Segurança Pública no estado do Rio de Janeiro. Por isso, era necessário considerar a costumeira falta de recursos de ponta, aparato tecnológico e pessoal técnico capacitado com as devidas habilidades e competências necessárias ao trabalho.

#### 4.2 CENÁRIO REAL: ANTES DA IMPLEMENTAÇÃO DESTE TRABALHO

No cenário atual, as informações passadas à Central DD através de ligações anônimas ao sistema são filtradas e classificadas por analistas de informações, visando uma melhor compreensão da mecânica criminal. O contexto atual, conforme a Figura 15 abaixo, demonstra os canais e como é a relação com os demais órgãos. Portanto, hoje as informações recebidas de todos os canais, são registradas de forma individual (telefone, aplicativo e mídias sociais) em

software *in-house* para posterior tratamento e, através de um *software* proprietário a fim de serem extraídas para difusão, junto às instituições responsáveis pela apuração.

Figura 15: Contexto Atual do Disque Denúncia



Fonte: Figura adaptada pelo Autor.

A ausência de *softwares* ou mecanismos de descoberta de informações em grandes bases de dados, mineração de dados, e/ou outros de desempenho e análises correlatas, limitava as possibilidades de interpretação das informações coletadas no banco de dados existente no DD. Conseqüentemente, a capacidade de atuação e difusão tornava-se comprometida e a produção do conhecimento e estudos de análises não potencializavam descobertas e monitoramento dos dados de entrada dessas denúncias, nem a aplicação da metodologia proposta neste estudo. A situação atual até este momento era bastante simples tecnologicamente falando, como podemos observar sua representação na Figura 16.

Figura 16: Situação de Atendimento Atual do DD



Fonte: Figura adaptada pelo Autor.

#### 4.3 CENÁRIO PROPOSTO

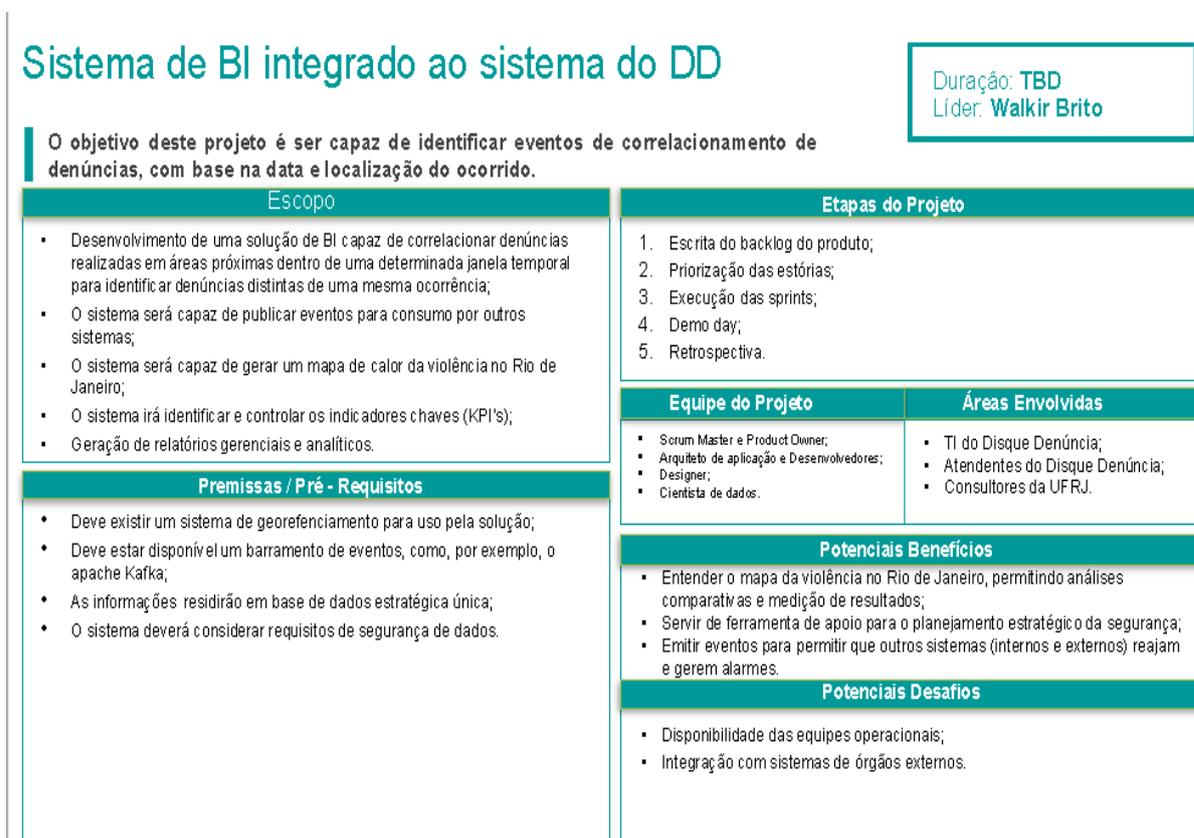
O novo cenário proposto é alocado em um serviço de distribuição de chamadas (URA) para garantir agilidade no processo de denúncias e facilitar o treinamento dos atendentes na condução do *script* de recepção e atendimento utilizando a metodologia de atendimento, já descritos anteriormente, além do padrão de captação das denúncias em seus diferentes canais (*App*, *Whatsapp*, Mídias sociais e *Widget* no site do Disque Denúncia e na página dos procurados). As informações recebidas permaneceram registradas no *software in-house*, entretanto, a utilização de um *software open source* altamente customizável permite que todas as informações sejam depositadas em um ambiente seguro na nuvem e com os lagos de dados necessários para o desenvolvimento de todas as etapas do estudo de caso milícia. Trata-se de um grande repositório seguro de armazenamento e coleta de dados para estudos e análises aplicados à área de inteligência, baseado em objetos que armazenam dados em seu formato original, até que seja necessário usá-los e encaminhá-los para serem tratados por uma série de

técnicas e *scripts* de tratamentos, que serão detalhados nas próximas etapas deste trabalho. Foram aplicadas técnicas de mineração de textos nas denúncias – as quais pudemos ler uma grande quantidade de dados, e correlacioná-los as variáveis necessárias ao estudo na produção de vínculos, tendências e indicadores.

Foram utilizados, como princípio, não somente o conjunto de dados armazenados a esse ambiente analítico, como também os princípios da metodologia *analytics* de visualização de dados para obter melhores resultados. Tornou-se necessário que o conjunto de dados fosse limpo, extraído, transformado e carregado, utilizando novas tecnologias de armazenamentos, tratamento, cruzamento e limpeza, capazes de receber e lidar com dados estruturados e não estruturados de diferentes fontes e origens, em que a aplicação dessa metodologia não interferisse na dinâmica do serviço.

Portanto, foram incorporados tais conjuntos de dados, por conseguinte, ao que denominamos ambiente analítico (*analytics*), permitindo a visualização de informações para a tomada de decisões com base nas análises e informações disponíveis, abordando não somente as integrações de bases, produção do conhecimento e ferramentas de *business intelligence* – como ferramenta de integração com o sistema de atendimento do DD, Figura 17.

**Figura 17: Desenvolvimento do Sistema BI integrado ao sistema de atendimento**



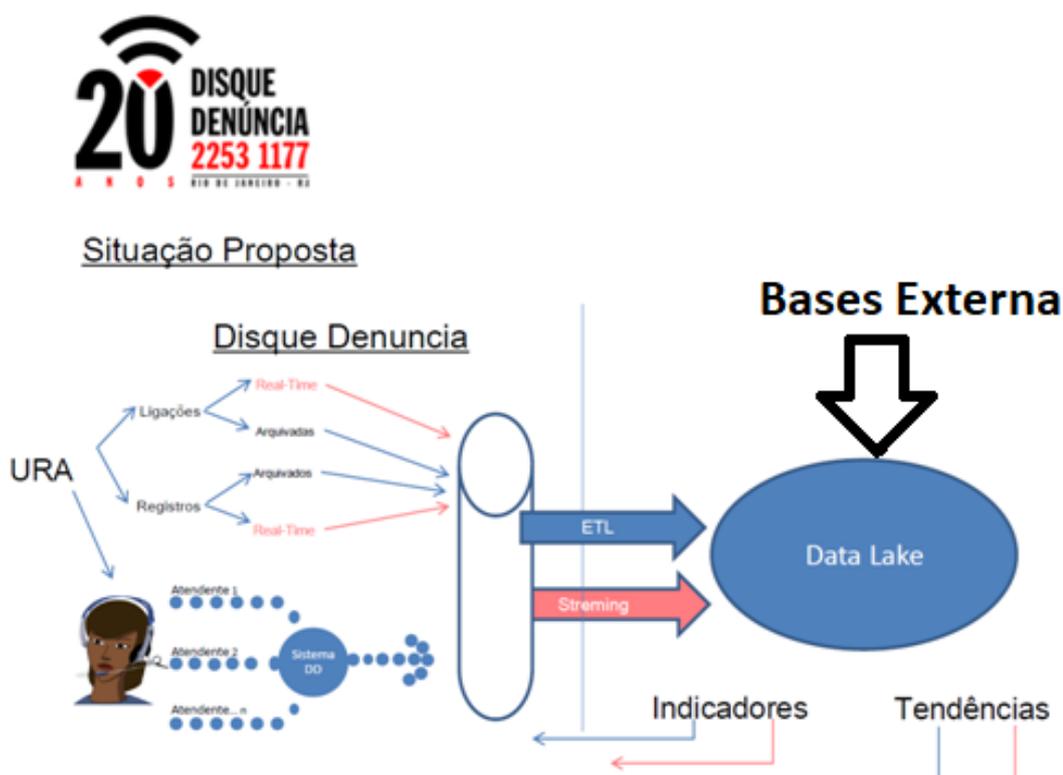
Fonte: Figura adaptada pelo Autor

O principal objetivo da implementação do projeto de *business intelligence* integrado ao sistema de atendimento, utilizando dados do Disque Denúncia pode ser visto no seu escopo, tratando-se basicamente do monitoramento, integração e a execução do georeferenciamento das denúncias, principais fontes nessa etapa do projeto. Incluiu ainda a elaboração de requisitos de negócio, requisitos técnicos e desenvolvimento da solução e permite a análise dos dados de denúncias, bem como, dados de importância na identificação de pessoas, locais, objetos, organizações, entre outros (entidades).

Fizemos o levantamento de metodologias e melhores práticas adotadas por referências na área de inteligência. Foram utilizadas ferramentas e técnicas que incluíam processamento de linguagem natural e mineração de textos para as etapas de limpeza e extração. A criação de um novo modelo de classificação das denúncias foi realizada, análise de todas as informações recebidas, relacionamento das entidades nomeadas nas denúncias e glossários padrões e específicos foram criados no decorrer do tempo. Todas as etapas de processamento e associação

dessas informações e entidades foram georreferenciadas para localização e descrição dos envolvidos (*persons*) nas respectivas denúncias ligadas aos possíveis suspeitos, delitos e/ou crimes, e no cruzamento dessas bases com bases de parceiras envolvidas nesse projeto (base de inteligência, serviço 190 e 181, entre outras).

**Figura 18: Cenário Proposto com a integração de bases externas**



Fonte: Figura adaptada pelo Autor.

Acima para demonstração, Figura 18, segue a descrição do lago de dados (*analytics*), criados e incorporados neste estudo do arcabouço textual. A base trata-se da parte analítica e foi utilizada como a fonte de dados de entrada e análise do estudo de casos. Toda a parte de limpeza e tratamento será detalhada nas etapas do modelo.

#### 4.4 MÓDULOS DO FRAMEWORK UTILIZADOS NO ESTUDO DE CASO

Nesta sessão são apresentados os módulos 1 e 2, do *framework* do arcabouço de mineração de dados textuais sensível ao contexto utilizada no estudo de caso. Respeitamos todas as suas complexidades para a metodologia de captação, coleta, correção, geração de dados e informações, aplicados a área de inteligência, mantendo o pressuposto do anonimato e na

extração do conhecimento produzidas pelas denúncias. A seguir, detalharemos os módulos utilizados no *framework*.

#### 4.5 MÓDULO IDENTIFICAÇÃO:

Apresentaremos o módulo 1 do *framework*. O módulo 1 (Identificação) que compõe parte do módulo do arcabouço analítico de texto em denúncias. No primeiro módulo, Identificação, foi produzida toda a lista de palavras-chaves (*stems*) utilizada no estudo, através da utilização de um algoritmo de *stemming* empregado com o intuito de transformar cada palavra em sua forma raiz, permitindo reduzir a palavra ao seu menor radical significativo, chamado neste texto palavra-chave (raiz) produzindo a base de conhecimento em milícias. Esta ação serve para realçar as palavras, os termos, os assuntos, os temas, as frases e expressões que são utilizadas em denúncias com a temática Milícia. Dessa maneira, construímos um dicionário específico com suas representações e significados textuais das denúncias, permitindo simplificar consultas aos temas e relacioná-los a outros temas, preparando a base para análises específicas e delimitando para cruzamentos com outros estudos.

Criamos e definimos desta maneira o número de palavras, termos e expressões utilizados para as milícias, destacando todos os assuntos relacionados, seus temas e termos encontrados, bem como a categorização das entidades encontradas. Buscamos na base de dados a listagem de denúncias com a mesma representatividade de temas, e criamos dessa maneira a relação de documentos (denúncias) representativa por assunto. Originamos desse modo uma base de dados específica para o estudo, com suas palavras, termos, assuntos e temas. Escolhemos para o estudo as palavras, termos e expressões, suas relações de classificação nas denúncias, envolvendo o contexto milícias, milícia, miliciano, venda, confronto, operação, tiroteio, disparo de armas, violência, entre outras, Apêndice D. Destacamos que o mesmo processo pode ser aplicado a diversas temáticas futuras relativas à criminalidade, por exemplo, violência a mulher, homicídios entre outros. Portanto, todos os procedimentos e fluxos criados podem ser replicados a quaisquer outros temas se necessário em estudos futuros.

No entanto, para que todas essas ações sejam realizadas, descreveremos a seguir todos os processos necessários e a metodologia desenvolvida neste estudo de caso para alcançar estes resultados esperados.

Primeiramente, descreveremos a criação de um ambiente seguro de coleta e tratamentos dos dados, seguindo as normas e recomendações de OPISP, conforme proposto pela DISPERJ (RIO DE JANEIRO, 2005). Estes são procedimentos necessários e utilizados pelas áreas de inteligência para carga de dados, ações de coleta, reunião dos dados, natureza das fontes de

dados e processamento das informações. Descreveremos brevemente abaixo suas etapas necessárias ao ambiente de coleta segundo seus aspectos doutrinários, e no final posicionamos nossa referência de boas práticas para futuros trabalhos com dados sigilosos, seguindo os preceitos e rigor empregado na Doutrina (RIO DE JANEIRO, 2005).

#### 4.5.1 AMBIENTE SEGURO PARA CARGA E COLETA DE DADOS

Foi criado um ambiente para ações de coleta de dados, baseado nos procedimentos recomendados pela DISPERJ (RIO DE JANEIRO, 2005), para que todas as normas, estruturas e ferramentas pudessem ser usadas neste trabalho. Denominado Arcabouço Analítico de Textos em Denúncias – Milícias (AATDM). Baseia-se em toda a estrutura de segurança necessária a um ambiente seguro para carga, coleta, pesquisa e análise de dados oriundos do contexto de segurança. Envolve desde a troca de arquivo entre os órgãos envolvidos, suas questões de sigilo e cooperação entre os órgãos por exemplo: PMERJ, CICC, CEG, TJ.

O ambiente de coleta estabelece protocolos de segurança não apenas para as cargas e suas composições de dados, importados ou exportados, como para: regras de senhas e permissões de acessos, origem e fontes de dados, entre outros. É orientada a toda parte trabalhada na geração das informações, e na reunião dessas fontes para a produção do conhecimento, seja com fontes internas e externas de dados. Serve para os procedimentos de geração de *dataset* e consultas envolvidas em temas de interesse da área de inteligência. Neste estudo, o enfoque central será no relato das denúncias, seu processo semântico, todos os temas e procedimentos relacionados a temática milícia.

O procedimento para obter dados necessários e suficientes para a produção do conhecimento utilizado pela área de inteligência em uma temática de pesquisa, análise e/ou investigação denomina-se ação de busca em inteligência e faz parte do ciclo de produção de conhecimento para a área. Ela pode ocorrer de duas formas: ações de coletas e as ações de busca. As ações de busca tratam de operações exercidas pelas forças policiais na obtenção dos dados – costumam ser sigilosas e não são tratadas neste estudo. Cabe apenas ressaltar que, os resultados dessas ações de buscas também foram incluídos no ambiente seguro de coleta. Abaixo descreveremos alguns dos procedimentos adotados desde a coleta dos dados até a produção do conhecimento exercida pela área de inteligência, para que fique evidente que todas essas ações necessitam estar presentes neste ou em qualquer outro sistema desenvolvido e utilizado pela área de inteligência.

A DISPERJ, estabelece o uso do CPC, Ciclo de Produção do Conhecimento, na

produção de conhecimento produzido pelas áreas de inteligência e descrevemos suas etapas adiante (RIO DE JANEIRO, 2005).

#### 4.5.2 AÇÕES DE COLETA

Ações de coleta, ou simplesmente, coleta, são todos os procedimentos realizados pela área de inteligência, ostensiva ou sigilosamente, a fim de reunir dados cadastrados e/ou catalogados em fontes abertas disponíveis. Podem ser oriundos de indivíduos, órgãos públicos ou privados, denúncias ou material investigativo. Essas ações, quando as fontes estão protegidas, caracterizam-se, particularmente, pelo acesso, credenciado aos órgãos que dispõem desses dados. E são solicitados por ordem judiciais e todo o aparato legal necessário. De posse da coleta dos dados, o segundo passo é a reunião desses dados. Reunião dos dados (REUDA) é a fase do CPC na qual o analista da área de inteligência procura obter os dados necessários para produzir o conhecimento. Sejam em fontes abertas ou fechadas, em dados protegidos ou negados. Para obtê-los e produzir o conhecimento, os profissionais necessitam planejar e executar essas normas e procedimentos sequencialmente. Alguns dados podem ser obtidos com uma simples coleta nos sistemas de segurança disponíveis. Outros, ditos negados, necessitam ações e autorizações específicas envolvendo autorizações judiciais. (RIO DE JANEIRO, 200, p. 14)

Quanto à natureza das fontes de dados, as fontes podem ser abertas ou fechadas. Fontes abertas são aquelas cujos dados são de livre acesso. Fontes fechadas são aquelas cujo os dados são protegidos ou negados. Dado protegido é aquele que necessita de credenciamento para acesso. Já o dado negado é aquele que necessita de ações policiais e autorizações para sua obtenção (RIO DE JANEIRO, 2005).

Normalmente, a coleta é executada por analistas que podem ser pessoas civis ou militares e as ações de busca, apenas por agentes e policiais treinados e autorizados nessa função. Desta forma, a reunião de dados pode ser assim esquematizada:

- Consulta aos arquivos existentes;
- Pesquisa (coleta);
- Acionamento do Elemento de Operações para dados protegidos;
- Ligações entre os órgãos e agências para obtenção e troca de dados (RIO DE JANEIRO, 2005).

O processamento é a fase de produção do conhecimento (RIO DE JANEIRO, 2005). É a fase dita "intelectual" do ciclo, na qual o analista da área de inteligência utiliza toda sua

capacidade e experiência, percorrendo as etapas lógicas de avaliação, análise, integração e interpretação para produzir o conhecimento necessário. Esta é a parte analítica do processo. Por isso, o resultado precisa ser sintético, direto, objetivo e deve atender a demanda inicial solicitada atendendo as expectativas necessárias com esse conhecimento. Baseia-se no princípio da oportunidade, que é produzir a melhor informação possível no tempo estabelecido para não perder a oportunidade de êxito de uma investigação ou operação. Em nosso estudo de caso, trata-se dos *dashboard* produzidos para auxiliar a produção de análise e conhecimento por parte desse analista. Portanto, inserir conhecimentos, métodos, metodologias e normas faz parte do desenvolvimento do sistema que precisa refletir a metodologia já utilizada pela área de inteligência e precisava ser descrita neste trabalho.

Pontuamos a integração entre os órgãos de inteligência e seus recursos, bem como o que consideramos boas práticas adotadas para a criação de um ambiente seguro para carga e coleta de dados empregada no estudo de caso da Milícia, totalmente validada, vinculada e utilizada com outros parceiros para o projeto, baseado nas doutrinas vigentes de inteligência em segurança pública, apêndice 6.2.1.

#### 4.5.3 METODOLOGIA DE DATA CLEANING APLICADA NA BASE DO DISQUE DENÚNCIA

*Data cleaning*, ou limpeza de dados – também conhecida como *scrubbing*, são conjuntos de técnicas utilizadas para detectar, remover e corrigir anomalias em base de dados.

Essa fase de limpeza dos dados torna-se essencial e requer a abordagem de uma metodologia clara com passos contínuos necessários para a obtenção de um bom resultado, (RAHM & DO, 2000). Abordaremos a seguir algumas das etapas de limpeza construída e utilizada neste trabalho, considerando todas as normas estabelecidas a dados sensíveis empregada na área de inteligência, proporcionando as transformações necessárias aos dados originais e corrigindo o necessário para um bom resultado das informações geradas.

##### 4.5.3.1 Necessidade da limpeza de dados

A necessidade de dispor de diversas possibilidades de software na construção do processo partiu dos recursos existentes nas diversas áreas de inteligência. Detectamos que, infelizmente, não existe um padrão tecnológico adotado, dificultando a troca de bases e procedimentos de carga, por exemplo. Além disso, as áreas tecnológicas nem sempre dispõem

de recursos financeiros ou pessoal capacitado e treinado para os procedimentos necessários. Para evitar esse problema e ter um ambiente único de carga, coleta e tratamento dos dados, optamos por utilizar os recursos tecnológicos disponíveis, de fonte aberta e também alguns já existiam anteriormente com suas licenças pagas e adaptá-los conforme as necessidades encontradas na base. Dessa forma, buscou-se um melhor desempenho com os recursos atuais sem a necessidade de investimentos altos ao projeto. Descreveremos abaixo alguns dos procedimentos abordados na metodologia criada para limpeza e tratamento dos dados. Em anexo, colocaremos alguns *scripts* e códigos criados nesse procedimento.

A limpeza de dados envolveu não apenas os bairros da cidade do Rio de Janeiro, como correções ortográficas e a identificação de bairros em outros municípios do estado do RJ, distinguindo que bairros que pertenciam a que municípios. Além de identificar os tipos de erro encontrados para poder corrigi-los.

Por exemplo, na Apêndice E, podemos observar a qualidade dos dados para a Localidade de Armação de Búzios e Cabo Frio e seus bairros correspondentes. Ficou evidente a necessidade da limpeza dos dados e a padronização dos logradouros. Mostraremos algumas das técnicas empregadas e ferramentas utilizadas no processo de limpeza adiante.

#### 4.5.3.2 Processo de limpeza dos dados (data cleaning)

Trata-se de uma etapa trabalhosa e que exigiu bastante tempo de testes e comparações até chegar a uma metodologia eficaz que pudesse ser replicada automaticamente a cada carga do sistema.

Apresentamos em seguida alguns problemas enfrentados no desenvolvimento e criação da metodologia e de que maneira estes problemas foram resolvidos para o processo de limpeza dos dados. Assim, explicitamos quais as transformações executadas e incorporada a metodologia e como foi realizado o trabalho. As técnicas de limpeza de base de dados também devem ser empregadas para identificar partes dos dados que são sujas, incompletas, incorretas ou irrelevantes e, então, substituí-las ou removê-las. O emprego destas técnicas é útil no sentido de tornar a execução mais rápida e os resultados mais preciso e correto. Os autores RANDAL, FERRANTE e SEMMENS (2013), demonstraram em sua pesquisa que as etapas de preparação das bases de dados são de irrefutável importância para aumentar a qualidade dos registros.

Estes problemas podem ser provenientes da falta de padrão/validação na entrada de dados, da agregação de sistemas legados e da ruptura na coleta de informação em determinados períodos, entre outras razões. Em nosso caso o sistema de captura das denúncias utilizado

apresenta campos abertos o que facilita o erro ao digitar, causando esse tipo de inconsistência. Problemas deste tipo representam entre 75% a 85% do esforço total de tempo para a correção e limpeza consistente. A literatura indica que não existe um percentual único para esse esforço de correção em diversos casos, estima-se que o processo de limpeza é um dos principais passos no processo de correlação podendo representar 75% do esforço para sua correção (GILL; STATISTICS, 2001).

Damasceno et al (2010) relaciona cerca de 80% do tempo total e defende que são fases cruciais para o sucesso do processo por ser onde os dados são tratados e, posteriormente, possam ser submetidos às técnicas de inteligência computacional, gerando resultados satisfatórios. Essas irregularidades existentes na base podem ser de sintaxe, contendo erros lexicais, erros no formato do domínio e anomalias, ou semântica que consistem ao não entendimento do dado registrado, podendo conter violações das restrições de integridade, duplicidade de registro, registros inválidos e contradições.

A seguir, mostraremos alguns dos problemas encontrados referente a qualidade dos dados da base de dados do DD e como foram trabalhados para sua correção.

Foram identificados alguns tipos de erros ortográficos encontrados na base e foi necessário realizar a limpeza dos dados. São estes: Bairros (DenLogrBairro) foi realizado a identificação de quais e quais formas de erro, para posterior limpeza dos dados e suas padronizações desses bairros (DenLogrBairro), para serem georreferenciados corretamente.

No Bairro VILA KENNEDY, por exemplo, fizemos uma busca e encontramos diversas formas de escrita para o mesmo local Apêndice F. Isso ocorre basicamente pelo campo ser aberto no sistema de coleta de denúncias e os analistas de atendimento escrevem o mesmo bairro de maneiras diferentes, mostrando que a qualidade dos dados não se apresentava boa e necessitava da realização do processo de limpeza dos dados, ou *data cleaning*, para esse campo. Foram identificados todos os bairros e tipos de erros existentes, e realizado o *data cleaning*, em toda a base do estado do Rio de Janeiro.

#### 4.5.3.3 DATA CLEANING DESENVOLVIDA PARA O ESTUDO DE CASO MILÍCIAS

Foi construído e montado todo um conjunto de *scripts* para limpeza e georreferenciamento dos dados, que vem sendo atualizado constantemente e utilizado em todos os trabalhos que utilizam a base, trata-se de um esforço contínuo em melhorar a cada dia esse processo.

Considerando-se todas as premissas já citadas no texto sobre o trabalho com dados sigilosos, foi necessário a criação de uma metodologia de limpeza dos dados criada e adotada neste estudo. A seguir, descreveremos esta metodologia e seus principais desafios encontrados no caminho.

A limpeza de dados tem como objetivo conseguir uma base isenta de erros, porém nem sempre é possível. Devido a esse motivo é benéfico e eficaz a adoção de medidas preventivas nesse processo de limpeza, tornando o custo e o tempo menor para sua correção. Isto permite uma base mais saudável, confiável e com um melhor poder analítico.

A utilização de *data cleaning* foi realizada mediante um processo controlado e mensurado após a atualização do banco de dados. Mostramos a metodologia adotada neste estudo e as etapas necessárias a limpeza de dados na Figura 19, além da execução das atividades de limpeza como e com quais *softwares* foram automatizados.

**Figura 19: Metodologia de Data Cleaning empregada**



Fonte: Adaptado pelo autor

Empregamos a abordagem de correção para erros voltados a erros lexicais, erros de formato de domínio, irregularidades nos dados, violações de integridade, correção de tipo de logradouro e bairros, eliminação de duplicidade de registros, valores omissos, registros omissos e *missing*.

O processo de limpeza dos dados não ficou restrito apenas a uma abordagem tecnológica, nem a uma única linguagem para o processo. Criamos e aplicamos a metodologia que pode ser rodada em diversos softwares como: R, Python, PowerBI, Tableau e Alteryx.

Adiante, demonstraremos algumas vantagens e desvantagens a escolha de cada um. Todavia optou-se por poder ter escolhas devido à falta de recursos existentes em cada órgão ou agência da área de inteligência, bem como a capacitação do pessoal operacional e analista, conforme já citado anteriormente. Dessa maneira, proporcionamos a opção de estruturar os procedimentos necessários utilizando softwares livres ou a cada tipo de necessidade necessária.

A execução das atividades e a construção do modelo utilizada neste trabalho, foi parcialmente automatizada com o apoio dos seguintes softwares e linguagens:

Python: Utilizando a versão 3.7;

R : é um software estatístico gratuito e *open-source*. Foi criado em 1993, por ROBERT GENTLEMAN e ROSS IHAKA. Em 2014, contava com mais de 2 milhões de usuários (RAJESWARI, BASU E MAURYA, 2017).

Durante a execução de tarefas intensivas, códigos programados com C, C ++ e Fortran podem ser chamados no ambiente R. Além disso, usuários experientes podem chamar diretamente objetos R em C. Na verdade, R é uma realização da linguagem S, que é uma linguagem interpretada desenvolvida pela AT&T Bell Labs e usada para exploração de dados, análise estatística e gráficos de desenho. Comparado ao S, o R é mais popular, já que é *open-source*. Devido à popularidade do R, os fabricantes de bancos de dados, como Teradata e Oracle, lançaram produtos que suportam R. RAJESWARI, BASU e MAURYA (2017) definem algumas vantagens e desvantagens do uso do R:

#### Vantagens:

- ✓ Utilizado para identificação e correção das falhas identificadas
- ✓ As habilidades gráficas de R são extraordinárias, dando um design completamente programável que supera a maioria dos outros pacotes gráficos;
- ✓ R não tem restrições de licença. É possível executá-lo em qualquer lugar e a qualquer momento, e até vendê-lo sob as condições da licença;

#### Desvantagens:

- ✓ É necessário aprender muito bem a ferramenta antes de usar. Caso contrário, não poderá ser usado de forma eficaz.
- ✓ Todos os pacotes usados em R nem sempre dão resultado perfeito (RAJESWARI, BASU E MAURYA, 2017).

*Tableau*: é uma das ferramentas de BI mais rápidas. É rápido de implantar, fácil de aprender e muito útil para o cliente. O Tableau possui cinco produtos principais que facilitam as diversas necessidades de profissionais e organizações. Sendo:

- ✓ *Tableau desktop*: para uso individual;
- ✓ *Servidor tableau*: colaboração para qualquer organização;
- ✓ *Tableau online*: BI em caso de nuvem;
- ✓ *Tableau reader*: usado para ler arquivos salvos no *desktop tableau*;
- ✓ *Tableau public*: para publicar dados interativos online. Apesar das muitas vantagens do *Tableau*, como a ótima visualização de dados, soluções de baixo custo para atualização e excelente suporte móvel, há muitas desvantagens, como falta de capacidade de previsão, segurança arriscada e problemas de gerenciamento de mudanças.

Rajeswari, Basu e Maurya (2017), definem algumas vantagens e desvantagens do uso do Tableau, sendo:

Vantagens:

- ✓ Tem uma excelente interface de usuário.
- ✓ Bom recurso de integração, pode se integrar a outras plataformas de *big data*, como o *Hadoop*;
- ✓ Compatível com dispositivos móveis;
- ✓ Software de baixo custo, fácil de atualizar e consome menos espaço de memória.

Desvantagens:

- ✓ Necessário processamento inicial de dados. E isso deve ser feito por especialista;
- ✓ Não suporta todas as estatísticas.

#### 4.5.3.4 ESTRUTURA E ETAPAS:

##### 4.5.3.4.1 *Data cleaning* - pré-processamento e geração de bases

A seguir apresentaremos a esquematização e etapas de *data cleaning* referente a Identificação de falhas e correção das falhas, mostrada na segunda etapa da metodologia empregada no modelo, Apêndice G, empregadas no pré-processamento e na separação das bases de textuais seguindo as técnicas utilizadas, originando-se na estrutura e etapas de *data cleaning* para pré-processamento e geração de bases empregadas ao modelo.

Para a Base 1, empregamos as seguintes técnicas: Remoção de acentos, *Case Folding*,

Remoção de dígitos, Remoção de pontuação, Tokenização, Substituição de acrônimos e Remoção de *stopwords*.

Para a etapa de *case folding*, optou-se em transformar os caracteres para maiúsculos. A Base 2 foi criada a partir da Base 1, sendo adicionada apenas etapa de *stemming*. A Base 3 também foi criada a partir da Base 1, e contém a etapa de correção ortográfica, ou seja, ela não contém a etapa de *stemming*. Já a Base 4 foi criada a partir da Base 3, contendo, portanto, todas as etapas de pré-processamento e técnicas mencionadas.

A remoção de pronomes oblíquos e placas, apesar de ser uma prática de remoção de *stopwords*, está sendo mostrada de forma distinta por ser aplicada sobre os dados não tokenizados, visto que, a exclusão de *stopwords* convencionalmente é feita sobre os *tokens*. Cabe ressaltar que, a etapa de *stemming* aparece duplicada na figura, por ser aplicada em circunstâncias distintas e em bases distintas, recebendo uma base que sofreu correção ortográfica ou não. Faz-se necessário explicar que precisamos guardar todos os processos realizados em bases distintas por questões de segurança dos dados e princípios doutrinários. Nas etapas de pré-processamento de remoção de *stopwords*, *stemming* e tokenização foram utilizados métodos da biblioteca NLTK (*Natural Language Toolkit*), que é um conjunto de bibliotecas do *Python* para processamento de linguagem natural. A remoção de acentos utilizou a biblioteca Unidecode e as demais etapas utilizaram funções nativas da linguagem e da sua biblioteca de *strings*.

#### 4.5.3.4.2 *Data cleaning: identificação e correções*

O primeiro passo abordado na metodologia do processo de limpeza criada foi transformar todos os caracteres em maiúsculo, eliminar todos os acentos ortográficos e outros caracteres especiais através do código. Utilizamos o R para esses *scripts* porque permite rodar dentro ou fora do *PoweBi*, gerando como resultado a Apêndice H. Após a primeira limpeza básica, de forma a tornar os endereços mais precisos, gerou-se uma lista com nomes únicos de bairros. Com isso, observamos uma grande quantidade de erros ortográficos e digitação, o que demonstra que os dados podem ser digitados de maneira imprecisa e de diferentes maneiras para a mesma informação. Assim, tivemos que identificar esses padrões de erros e corrigir cada padrão.

Adiante, mostraremos algumas telas do processo de identificação e correção utilizado na limpeza de dados. Apêndice I, J, K, ilustram alguns dos processos de transformações dos dados realizados para a identificação e correção das falhas existentes nos dados durante o processo de limpeza. Ao fim, a base deixou de ser como na Apêndice L, mostrada

anteriormente, passando a ficar como na Apêndice M, totalmente corrigida, testada e atualizada para o processo de geoprocessamento que foi realizado posteriormente.

#### *4.5.3.4.3 Limpeza E Padronização De Logradouros Com Fuzzy Matching*

Em um primeiro momento, usamos o software *Alteryx* para a limpeza e padronização de logradouros e seu modo de *Fuzzy Matching*, uma técnica reunida que produz como resultado a latitude e longitude através de correspondência parcial. Nele podemos controlar a acurácia dos resultados e delimitar a linha de corte de aceitação. No entanto, para não ficar preso a um *software* e a necessidade de renovações de licenças anuais necessárias, estabelecemos a necessidade de utilizar um conjunto de diferentes ferramentas para limpeza de dados e atribuição de endereços, para a melhor correspondência de localização física e geográfica das denúncias. Utilizamos o conjunto de código gerado nessa ação adaptando essa codificação nessas ferramentas utilizadas no estudo.

Nessa etapa, foi feita a limpeza da base de dados do Disque Denúncia para associar com os *shapes* dos bairros. O formulário usado para registrar as denúncias têm campos abertos para texto, então muitas vezes os nomes dos bairros e outros detalhes são escritos incorretamente, como apontado anteriormente. Foi realizado o *script* de limpeza e o processo que antes era manual – no qual se perdia muito tempo – foi automatizado em toda a base de dados. O que mostrou ser um procedimento bastante satisfatório para o estudo, o procedimento e realizados nas atualizações e cargas da base realizadas diariamente. Cabe ressaltar apenas que não se trata de uma limpeza total, persistindo inconsistências que podem ser trabalhadas em estudos futuros contribuindo dessa forma para uma maior eficácia do *script*.

#### *4.5.3.4.4 Processo De Geocodificação Da Base Limpa*

A partir da limpeza dos dados do campo endereços foi necessário à sua geocodificação, que consiste no processo de conversão de endereços em coordenadas de latitude e longitude que foram utilizadas para criar pontos em um mapa e a localização geográfica dos locais constantes nas denúncias e suas localidades.

Precisamos ter os dados de latitude e longitude de cada local referente a cada denúncia, tornando a base totalmente geocodificada. Isto permite melhores análises e informações. Normalmente não temos todos esses detalhes na base de dados. Em geral, temos o endereço da denúncia, os pontos de referência e as entidades coletadas pelo analista de atendimento. Com

isso, podemos obter os dados necessários para iniciar a análise geoespacial. Para isso utilizamos dois serviços de localização, através de API:

- ✓ *Here serviço de localização;*
- ✓ *API do Google*

Investigamos qual serviço seria melhor para a geocodificação com as características da base de dados do Disque Denúncia, resguardando os procedimentos de anonimato necessário, descrevemos essas duas etapas a seguir, pois este foi um dos grandes problemas iniciais enfrentados. Não pela questão da geocodificação da base, mas sim para garantir a premissa de não identificação do denunciante.

#### 4.5.3.4.5 Utilizando O Serviço De Localização Here

A *Here*<sup>16</sup> é uma empresa que fornece soluções de geolocalização. Disponibiliza uma API gratuitamente para realizar a geocodificação de endereços, que pode ser usada para endereços em todo o mundo. Os serviços de geolocalização e pesquisa oferecem 250 mil endereços por mês, sem nenhum custo. O primeiro passo foi registrar-se como desenvolvedor e usuário da plataforma para criar um projeto<sup>17</sup> Em seguida, dentro da seção projetos selecionamos o projeto que criamos, copiamos o ID do aplicativo e a *API KEY* e pudemos começar, conforme Apêndice N. Em um primeiro momento, utilizamos o *Alteryx* e o seu módulo de geocodificação com *Fuzzy Matching*, com todos os requisitos de anonimato imputados em suas regras de localização. O desafio foi coletar as informações utilizando as ferramentas geográficas do *Alteryx*, para criar um centroide da área que precisamos geocodificar, devolvendo a latitude e longitude desejada. Trabalhando com a área do centroide, preservamos o anonimato e a precisão dos dados, conforme Apêndice O. Os endereços são utilizados na entrada de texto em um único campo e o endereço da API onde ele se relaciona é capturado pela fórmula, permitindo sua conexão, Apêndice P. Para isso, é necessário modificar o *AppID* e o *ApiKey* provenientes do *Here*, sendo codificado como uma constante que o sistema forneceu. Isso é codificado como uma constante do sistema. O valor é atualizado, conforme Apêndice Q. Finalmente, executamos o fluxo criado, permitindo ver os dados e sua localização aproximada no mapa.

---

<sup>16</sup> <https://www.here.com/>, acessado em 26/05/2020

<sup>17</sup> [https://developer.here.com/documentation/authentication/dev\\_guide/topics/api-key-credentials.html](https://developer.here.com/documentation/authentication/dev_guide/topics/api-key-credentials.html), acessado em 26/05/2020

No entanto, como sabíamos que não teríamos a disponibilidade da licença de uso prorrogada por muito tempo e por questões financeiras, resolvemos examinar o código gerado e replicar o código em R, permitindo dessa maneira rodar em outras ferramentas, e tomar posse do conhecimento desenvolvido. Dessa maneira, escolhemos os resultados com maior acurácia e levamos para o *script* escrito em R, mantendo o resultado encontrado e melhorando outros padrões.

#### 4.5.3.4.6 Api Do Google Para Geração De Lat E Long

Conforme já citado anteriormente, tivemos a oportunidade de testar a geração da *Lat e Long* por outra ferramenta de localização além do *Here*. Isto foi de extrema importância para comparar os resultados entre os dois e escolher quando e de que forma usá-los. É importante lembrar que este estudo utiliza base em dados de endereços complexos, em sua maioria não pertencentes aos ditos, asfalto, zonas de localização da cidade. Estamos pesquisando ruas, travessas, becos e vielas com características específicas e conglomerados habitacionais em localidades complexas onde os complementos dessas denúncias muitas vezes apontam sua localização ou área dessas comunidades. Tentamos verificar de que forma ferramentas de localização comportam-se em ambientes como esses e o que tivemos que fazer para melhorar a metodologia empregada no estudo para resolver essas situações de endereços não encontrados. Por isso trabalhar com duas opções e modificar para uma maior precisão foi extremamente relevante ao trabalho e na obtenção desses dados. A seguir, descreveremos como foi feito através da *API* do *Google*.

Para realizar a inserção da Latitude e Longitude nas denúncias, o sistema envia a *API* do *Google* através do *Script* abaixo, Apêndice R, o endereço cadastrado pelo operador e que passou pelo processo de limpeza de dados. O *Google* então retorna um *Json* com a latitude e longitude da denúncia desejada, conforme Apêndice S.

Com base no *json* retornado, foi realizada a chamada para atualizar a tabela de denúncia e em seguida atualizamos os campos: *den\_gps\_lat* e *den\_gps\_long*, compondo dessa forma a latitude e longitude necessária a base de denúncias. Para os endereços não encontrados foi incorporado ao *Data Cleaning*, um procedimento de verificação dentro das denúncias em busca das entidades relacionadas a locais. Do mesmo modo, buscamos a ocorrência denúncias dos nomes das comunidades denunciadas através do uso de *string* a essa informação. Desta maneira, associamos essas informações aos resultados ligados aos endereços, realizando uma nova pesquisa nas ferramentas de localização. Assim, foi possível classificar um número maior de

casos e foi possível associar a *Lat e Long* a Base.

#### 4.6 MÓDULO 2: INTERPOLAÇÃO

No segundo módulo, interpolação, a partir da lista de *stems* de cada denúncia, busca-se a lista de significados mapeados com suas respectivas estatísticas de frequência do banco, i.e., o número de ocorrência de palavras relacionadas a cada classificação das denúncias, Figura 20, abaixo.

**Figura 20: Maiores Relações de Classificação das Denúncias**



Fonte: Disque Denúncia.

Com o seu resultado, geramos a nuvem de palavras e criamos o dicionário relacionando as relações de classificação utilizados nas denúncias, por tipo e classe de categoria utilizada no DD. Por fim, realizamos as geolocalizações dessas denúncias, preservando seu anonimato, Figura 21 a seguir. Assim, é possível relacionar as palavras encontradas as suas áreas específicas e a intensidade de frequência dessas palavras.



principais variáveis do estudo ao nível de setor censitário em mapas e, posteriormente, utilizando os fatores resultantes da combinação linear destas variáveis, comparando com as informações sociodemográficas para os municípios, por exemplo.

Com a necessidade de formas mais sofisticadas de análise espaciais, assim como a incorporação de procedimentos que permitam uma avaliação da confiabilidade e segurança dos resultados obtidos, empregamos o método de interpolação, a partir da lista de *stems* de cada denúncia. Buscamos gerar uma lista de significados mapeados por geolocalização, importante no contexto aplicado as milícias, tema central do nosso estudo. Desta forma, o significado das palavras, seus termos e expressões já bem enfatizado anteriormente na construção deste estudo, podem ser geolocalizados, permitindo a criação de um dicionário estabelecido pelas *stem* e sua geolocalização para a construção das nuvens de palavras utilizadas no arcabouço de texto de denúncias. O arcabouço cria a lista de *stems* com suas respectivas estatísticas de frequência do banco, i.e., o número de ocorrência de palavras relacionadas a cada categoria. Foi utilizada neste estudo para o resultado das nuvens de palavras e a criação do dicionário específico por categoria e tipo de classificações empregadas nas denúncias e empregue nas análises do DD.

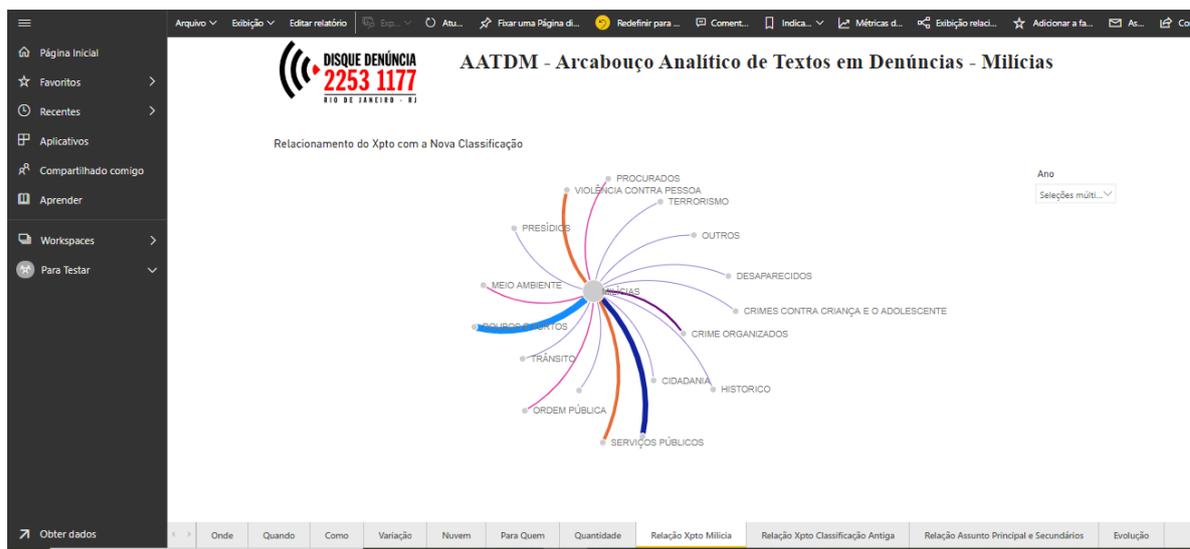
Adiante, descreveremos alguns desses procedimentos empregados na base, bem como a criação da base de personas associadas a essas palavras, as palavras-chave (*Xptos*) e seu processo de identificação dinâmica empregada ao estudo na atualização da base. Foi criada a base de rastros, que considera de maneira longitudinal na linha de tempo da base as identificações dos alvos e sua movimentação, apontada pelas denúncias com sua geolocalização e a suas denúncias respectivamente, permitindo ao analista da área de inteligência produzir conhecimento e identificar movimentações de envolvidos em grupos milicianos. As ações de processamento de textos, processamento de linguagem natural, *stem* e dicionários que foram utilizados para a construção das nuvens de palavras estarão descrita nos módulos 3 e 4 deste *framework*.

#### ***4.6.1 CRIAÇÃO DA BASE DE PERSONAS E IDENTIFICAÇÃO DOS XPTOS DAS PALAVRAS-CHAVES A PARTIR DAS DENÚNCIAS E A BASE DE PROCURADOS***

Baseado no texto de cada denúncia e utilizando a base de dados desde 2002, identificamos que palavras estão associadas a que etiquetas (*XPTOs*) ou *hashtags*, para compor a criação da base de pessoas (entidade pessoa – *PER*), que contempla não somente a base de foragidos da justiça, a "Procurados", Figura 22 abaixo, mas também todos os nomes de pessoas encontradas no cruzamento.



**Figura 23: Fluxo Dinâmico**

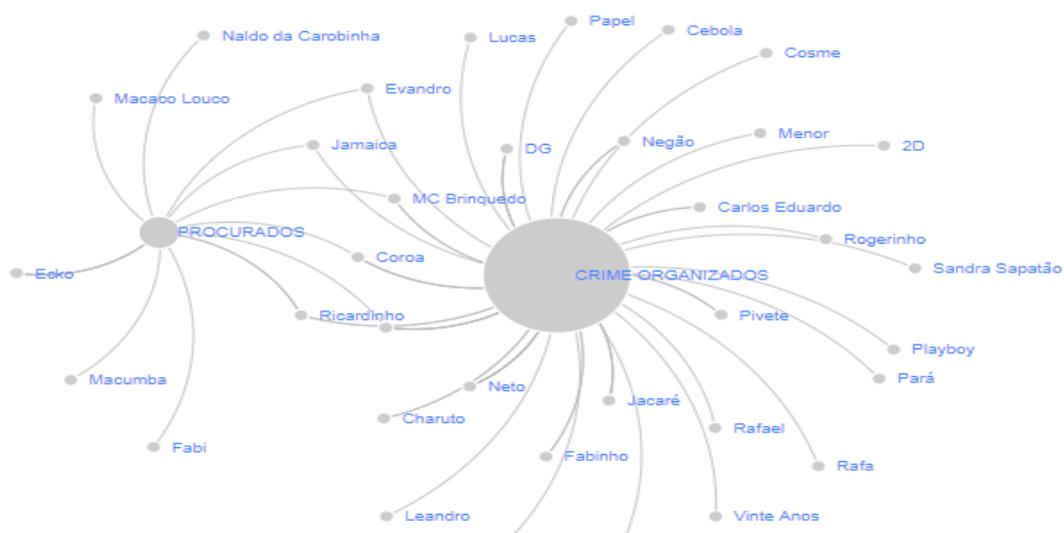


Fonte: Figura Adaptada pelo Autor.

#### 4.6.2 CRIAÇÃO DA BASE DE RASTROS A PARTIR DAS DENÚNCIAS

Com a criação da base de *Personas*, buscou-se identificar todas as denúncias associadas às personas em um período longitudinal da base de procurados para criar o rastro das pessoas identificadas como procuradas, Figura 23 abaixo.

**Figura 24: Preparação Base Rastro a partir das Denúncias (entidade pessoa – PER)**



Fonte: Figura Adaptada pelo Autor

Esse procedimento possibilita rastrear o alvo através das denúncias no decorrer do tempo, ou seja, a base de rastro mostra ao longo do tempo por onde a pessoa tem sido vista (utilizando o *Lat* e *Long* encontrado) e quais as denúncias que já foram associadas a ela. Deste modo, gerou-se um padrão de movimentação dentro do mapa que mostra a trajetória da entidade pessoa – Per denunciadas. Podemos visualizar um exemplo do rastro de um alvo fictício na Figura 25 abaixo.

A imagem mostra a trajetória de denúncia sobre a entidade pessoa em relação aos locais de denúncias ligados a ela, tornando-se uma importante fonte de informação extraída das denúncias. Isto permite que a área de inteligência monitore esse alvo e perceba locais de movimentação, domínio, expansão, novos territórios, entre outros.

**Figura 25: Rastro dois últimos 48 meses de um alvo**



Fonte: Figura Adaptada pelo Autor

Fonte: extraída do arcabouço.

#### **4.7 CONCLUSÕES DO CAPÍTULO**

Neste capítulo, apresentamos os módulos 1: identificação e 2: interpolação, que compõem o arcabouço analítico de denúncias. Esses módulos são compostos por toda a parte de carga, limpeza e tratamento de dados que foram usados nos módulos seguintes. Conforme o exposto nesta seção, apresentamos a criação e implementação um ambiente seguro para a coleta de dados. Apresentamos a sugestão das boas práticas que devem ser adotadas baseadas na DISPERJ (RIO DE JANEIRO, 2005) e de extrema necessidade na área de inteligência envolvendo máquinas, agentes, analistas, órgãos e instituições. Relatamos o processo de *data cleaning* utilizado na base de dados que possibilitou a correção de falhas. Apresentamos também a metodologia de correção dos logradouros com *fuzzy matching* e a geocodificação da base. Nesta seção foi demonstrada a criação da base de *personas* a partir da base dos Procurados e das denúncias, a identificação das dinâmicas de *XPTO* baseada em palavras-chave para a temática a milícia, a criação da base de rastros a partir das denúncias e a atualização da base e sua medição de qualidade. Em resumo, todos os recursos necessários para os próximos módulos do arcabouço, que veremos no próximo capítulo.

## 5 MÓDULOS: AGRUPAMENTO E ANÁLISES

Neste capítulo, são apresentados os módulos 3 e 4 do *framework*, denominado: Agrupamento e Análise, empregado ao estudo de caso, e que descreveremos a seguir.

## 5.1 MÓDULO AGRUPAMENTO OU CLUSTERING:

O ambiente analítico denominado Arcabouço Analítico de Textos em Denúncias – Milícias (AATDM) – é um arcabouço de mineração de dados textuais, sensível ao contexto em denúncias, e foi criado como um ambiente seguro na nuvem. As denúncias são *clusterizadas* a partir dos significados extraídos de cada uma delas. Primeiro, agrupa-se as denúncias similares e adicionam-se informações estatísticas aos temas classificados. Cria-se um sumário executivo para posterior validação por um especialista, gerando a base de conhecimento que será utilizada no estudo.

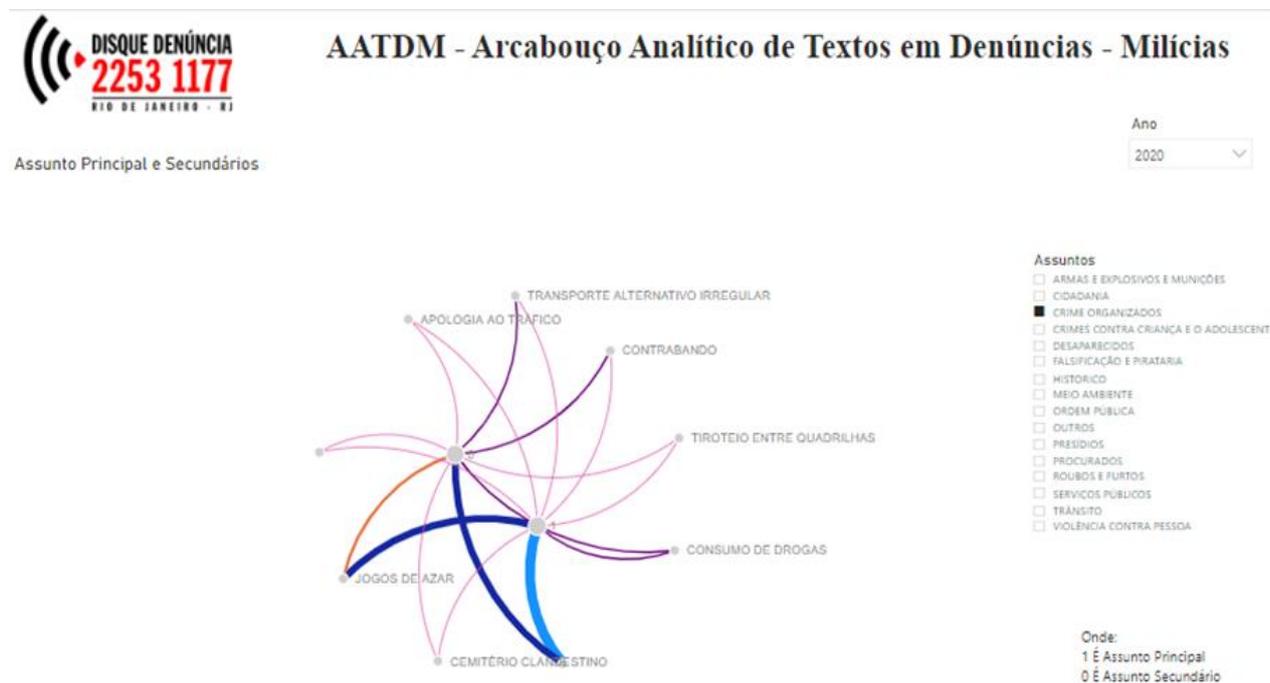
A partir disso, e feita a criação de seus *dashboards*, que configuram a parte analítica que pode ser utilizada pelo analista de inteligência na busca de informações pertinentes as suas pesquisas e análises. Respeitou-se todas as premissas doutrinárias em sua produção, contribuindo desta forma para a construção do conhecimento empregado na área de inteligência, baseamos nos aspectos doutrinários da DISPERJ para sua criação e implementação.

O arcabouço possibilita funcionalidades de pesquisas e análises, segundo o formato utilizados pelos órgãos de inteligência seguindo a tipologia do ciclo de produção do conhecimento (CPC): “Tendências”, “Onde?”, “Quando?”, “Para Quem”, “Nuvem de palavras” e temas e classificações de crimes específicos interligados ao contexto, não apenas as milícias e serão apresentados seus resultados na sessão sobre as análises. Utilizamos toda a classificação de crimes das denúncias utilizada pelo DD por assuntos.

A Figura 26 adiante, serve como uma ferramenta de monitoramento em tempo real das denúncias recebidas, comparadas com seu histórico de anos, meses e ciclos anteriores, tornando-se viável escolher a classificação por tema e assuntos para realização das análises, possibilitando observar seus comportamentos e resultados ao longo dos períodos escolhidos, por áreas, temas ou indivíduos.

Trata-se de um ambiente com características de visualização de dados e *data science*, e com a composição da nuvem de palavras com as maiores frequências das palavras trabalhadas na mineração textual utilizada pelas bases do arcabouço. Por questões de segurança, usabilidade e acessibilidade das informações, não podemos detalhar os processos empregados em cada uma das bases.

**Figura 26: Tela do Sistema AATDM e suas funcionalidades de análises**

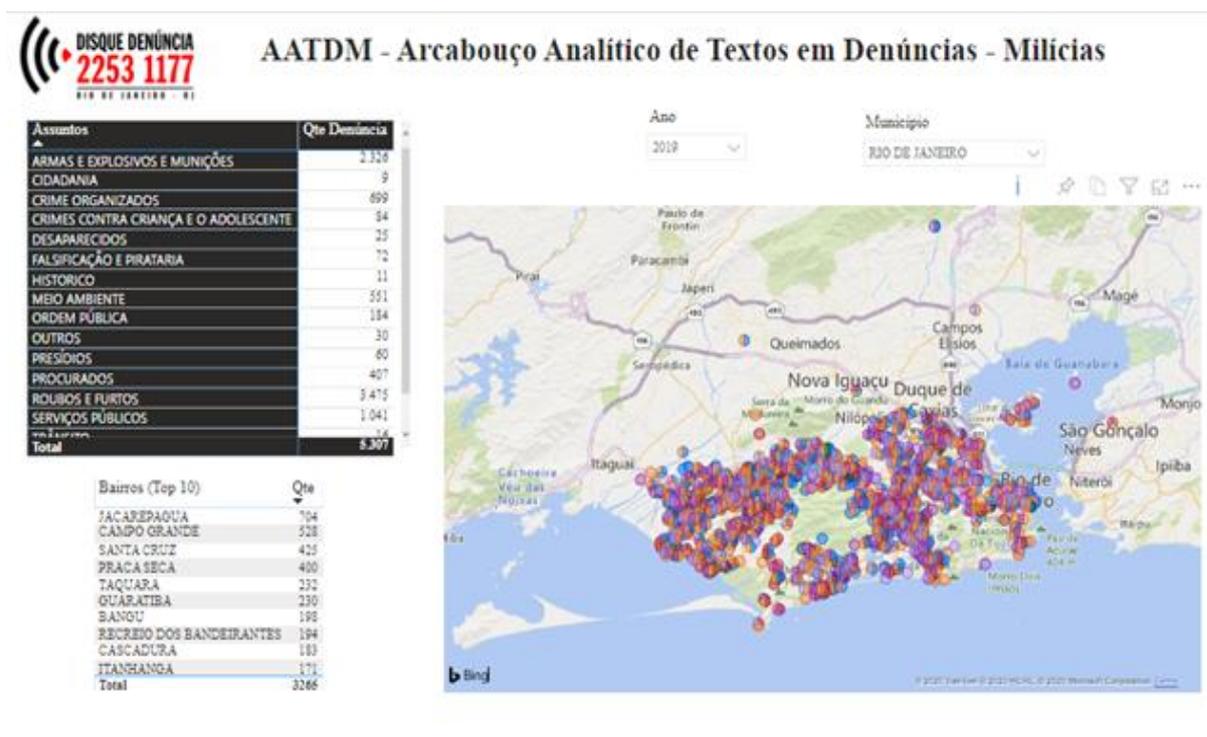


Fonte: extraída do Arcabouço.

Trata-se de um processo inovador no âmbito de mineração de dados textuais e tratamento de dados para produção de conhecimento, bem como de relacionamento de dados de denúncias, na produção de informações com o foco para área de estudos e pesquisas empregados a área de inteligência com o objetivo de mitigar riscos e tentar melhorar os recursos aplicados a segurança. Trata-se de uma estrutura analítica de mineração de textos, relevantes para a produção do conhecimento utilizando denúncias advindas da população, e transformando seus dados de relatados em informações relevantes a tomada de decisão, envolvendo a temática milícias e todos os assuntos associados ao tema. Seja na descoberta das frações significativas (pistas), para a composição e contribuição na linha de investigação, ou para traçar perfis e modos operacionais desses grupos em seus territórios.

É uma ferramenta que objetiva contribuir para os aspectos de descoberta e produção de conhecimento, utilizando a mineração de textos e palavras importantes a esses fatos. Do lado direito, encontramos sempre o mapa do estado do RJ contendo a quantidade de ocorrências e do lado esquerdo, outras informações relevantes e os filtros que podem ser mudados nas análises.

Figura 27: Tela do Sistema AATDM e suas funcionalidades de análises para tipos de crimes



Fonte: extraídas do arcabouço.

Os pontos coloridos no mapa refletem a localização geográfica aproximada de onde, quando estão ocorrendo essas denúncias e quais os conteúdos relacionados à milícia, lembrando que preservamos o anonimato da mesma forma nos algoritmos. Serve para entender quando essas ocorrências aconteceram, suas frequências por dia e mês, possibilitando desta forma analisar sua sazonalidade, ciclos, dias da semana, horários entre outros.

As figuras a seguir apresentam as visualizações dos dados do arcabouço no formato de telas do *dashboard* utilizadas no AATDM (Figura 28-43):

A Figura 28: Distribuição das Denúncias no decorrer dos anos e suas nuvens de palavras a Difusão.



Figura 30: Tela do Sistema AATDM e suas funcionalidades de análises - Municípios



F

Fonte: extraída do arcabouço

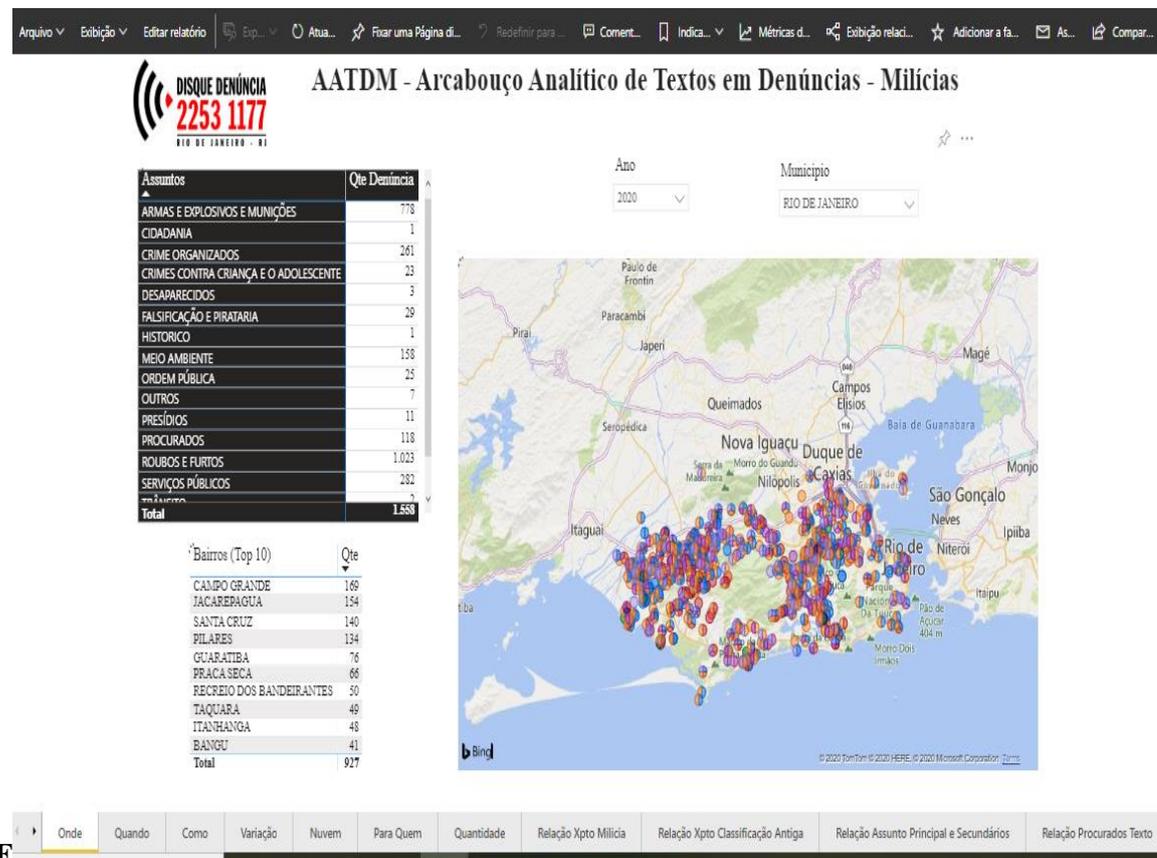
A Figura 30, a distribuição das ocorrências por municípios, e suas respectivas Nuvens de Palavras. A Figura 31, representa a escolha por um alvo.

Figura 31: Tela do Sistema AATDM e suas funcionalidades de análises - Alvo



Fonte: extraída do arcabouço

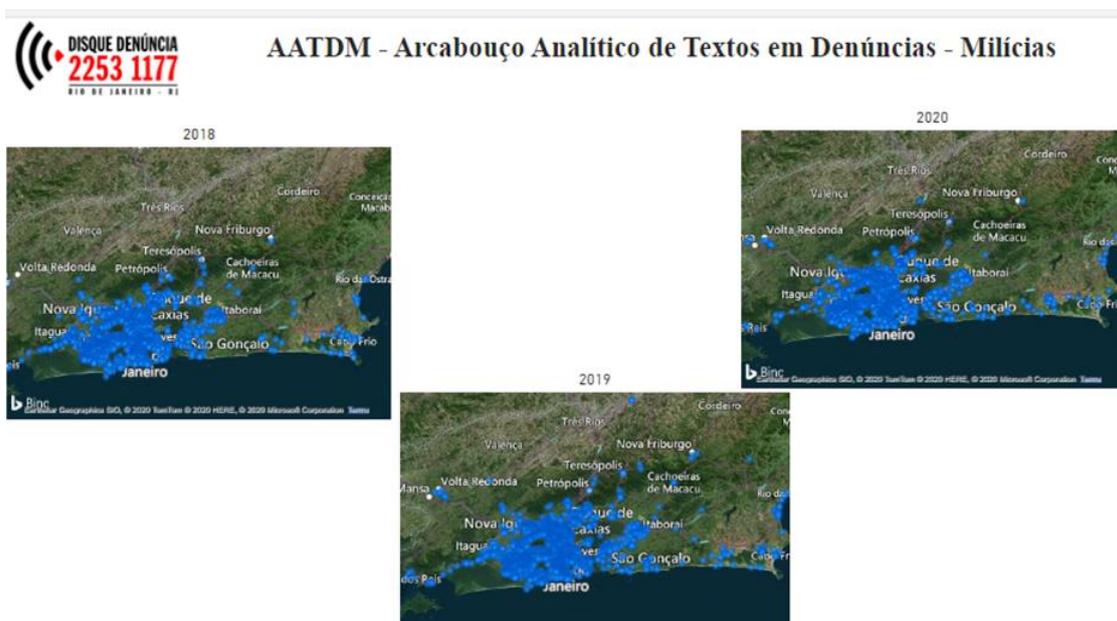
**Figura 32: Tela do Sistema AATDM e suas funcionalidades de análises - Classificação de Crimes**



Fonte: extraída do arcabouço

A Figura 32 por classificação de crimes. A Figura 33, o monitoramento anual a temática Milícias.

**Figura 33: Tela do Sistema AATDM e suas funcionalidades de análises – Monitoramento anual das Milícias**



Fonte: extraída do arcabouço

A Figura 34, a representação da Narco Milícia.



Na Figura 36, a seguir, a relação da difusão das denúncias e seu acompanhamento e resultados.

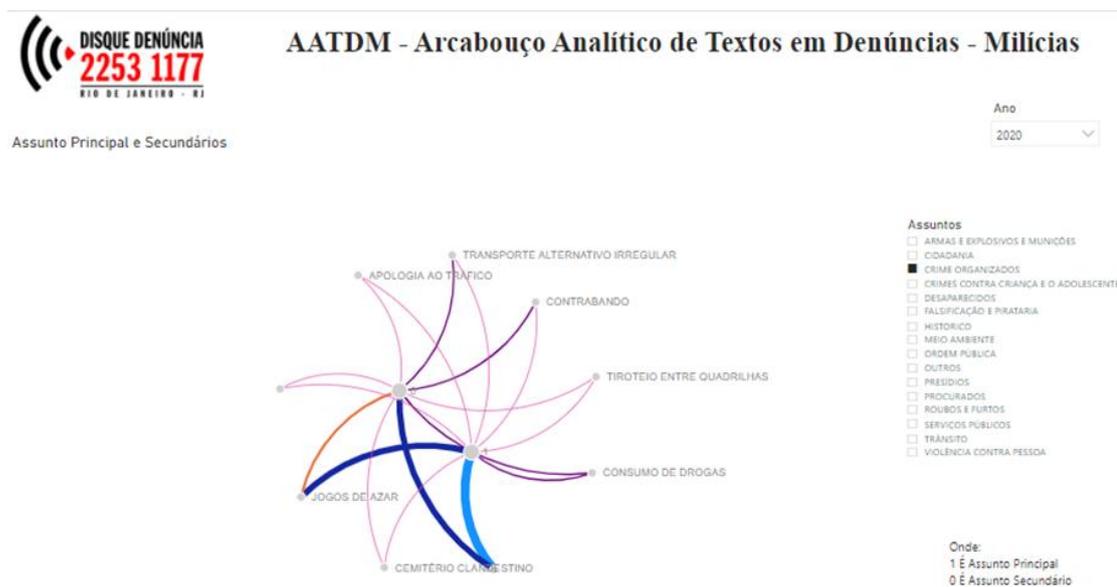
**Figura 36: Tela do Sistema AATDM e suas funcionalidades de análises - Difusão**



Fonte: extraída do arcabouço

Na Figura 37, os assuntos de análises.

**Figura 37: Tela do Sistema AATDM e suas funcionalidades de análises - Assuntos**



Fonte: extraída do arcabouço



Figura 40: Tela do Sistema AATDM e suas funcionalidades de análises – Cemitério Clandestino



Fonte: extraída do arcabouço

Na Figura 41, a opção Onde. Cabe Ressaltar que, todas essas análises podem ser combinadas em novos *dashboards* para a formação e seleção de novos temas desejados, que por questões de sigilos não demonstraremos, No entanto, podemos ver a potencialidade da visualização e extração de dados a respeito do tema ampliando o poder de análise aos órgãos de inteligência.

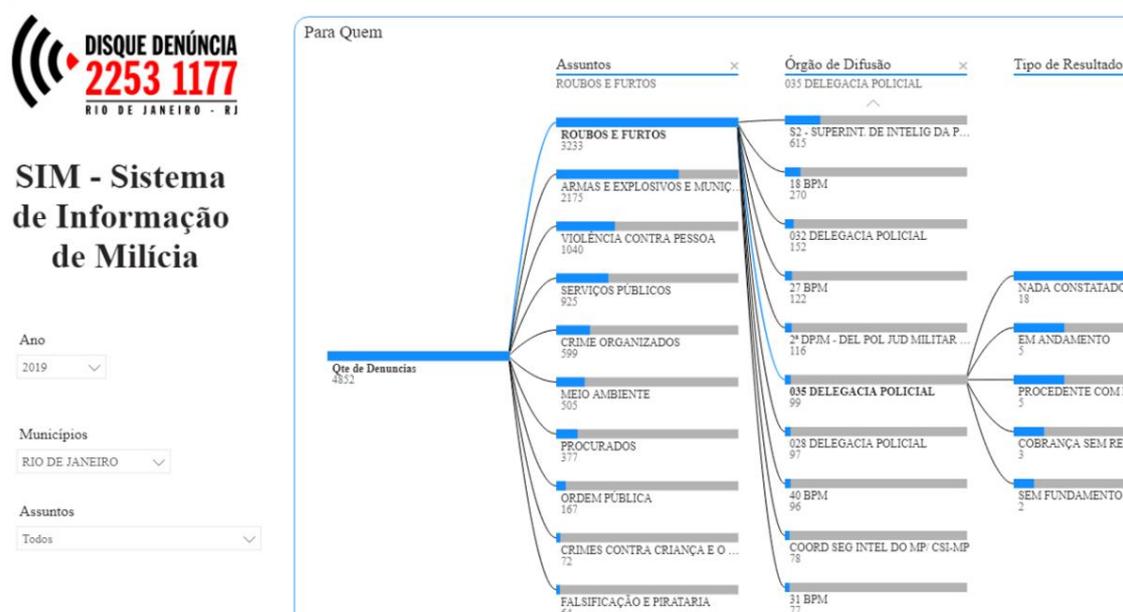
Figura 41: Tela Onde, no Sistema de Informação de Milícias - Onde



Fonte: extraída do arcabouço

Na Figura 42 a seguir, apresentamos quais os assuntos e para quais órgãos de difusão foram enviadas a denúncia e quais os resultados obtidos por esses órgãos. Este tipo de conhecimento permite realizar um melhor planejamento das ações de inteligência e mensurar a eficácia dos resultados de cada órgão. Cabe ressaltar que, nessa etapa de difusão não detalharemos totalmente as especificações adotadas no trabalho por questões de sigilo das informações, porém podemos dizer que há órgãos, delegacias e batalhões especializados para cada tipo de ocorrência criminal e que foram todos adequados, contemplados e validados no sistema de informação.

**Figura 42: Tela Para Quem foi realizada a Difusão das informações coletadas pelo SIM e quais os resultados obtidos, retroalimentando o sistema de informação**



Fonte: extraída do arcabouço

Na Figura 42, adiante, apresentamos uma demonstração das nuvens de palavras extraída seguindo todos os critérios de mineração de dados textuais mencionados neste trabalho que foram incorporados ao sistema originário das denúncias e a localização geográfica dessas ocorrências em seus distintos contextos de crimes. É importante dizer que são locais com organizações criminosas diferentes. Por isso, são também diferentes os estilos de expressão semântica e verbal encontrados na análise de texto das denúncias, mostrando diferentes representações simbólicas para o analista de inteligência, o que permite verificar o modo de agir dessas milícias e sua movimentação no espaço geográfico ocupado.

O objetivo inicial no desenvolvimento do AATDM e a expectativa para a área de



adequadas para a abordagem do problema apresentado, assim como a mineração de dados textuais, em que classificou-se e relacionou-se um grande volume de dados textuais de forma que fossem fornecidas informações confiáveis e que tivessem sentido no contexto. Conforme percebemos, o impulso analítico de visualização dos dados abarca todo o ciclo de tratamento dos históricos de denúncias, desde a extração de radicais temáticos, formados pelas principais palavras relacionadas às diferentes categorias de denúncia, passando pela construção dos dicionários especializados para apoio à extração de entidades, sua localização e formação das redes relacionais e temas que foram construídos a partir das denúncias. Os cenários foram convertidos em *dashboards*, utilizando técnicas de visualização de dados *self-service*, permitindo ao analista da área de inteligência pesquisar e visualizar toda a análise de vínculos, além de aplicar a sua utilização no ciclo de produção do conhecimento, reconhecido e identificados nas denúncias ao mesmo tempo em que há descoberta de novos conhecimentos e descobertas sobre o tema milícias.

Como exemplo do que foi descoberto em nossas análises, temos a representação das palavras dentro de seus respectivos territórios. Assim, é possível extrair mensagens subliminares a partir de suas representações de palavras dentro da denúncia, identificando os significados e as ações que são capazes de gerar. A partir disso, é possível elaborar um glossário e dicionário de termos que atenda a complexidade sintática e gramatical na base do DD em territórios determinados.

A conversão dos elementos de tendência utilizados na área de inteligência foi respeitada e aplicada em nosso trabalho. Sendo eles: onde, quando, comparativo, variação, e nuvem de palavras encontrando-se apta ao sistema de informação. Desta maneira, foi possível uma visão mais profunda da complexidade relacionada ao tema milícias, possibilitando ao analista da área de inteligência ampliar sua visão e possibilidades de análises. Esta é a grande contribuição desse sistema de informação. Um ambiente de coleta e análises de informações, em que os resultados obtidos até o momento são bastante expressivos, pois aumentou de maneira significativa a utilização do sistema e um melhor aproveitamento dos recursos tecnológicos ligados ao seu desenvolvimento aplicado. Desta maneira, a metodologia aplicada ao estudo mostra-se relevante e promissora aos objetivos estabelecidos, demonstrando sua aplicação em um mundo real e cheio de desafios para a sociedade e auxiliando órgãos e áreas de inteligência no enfrentamento das milícias e da criminalidade como um todo.

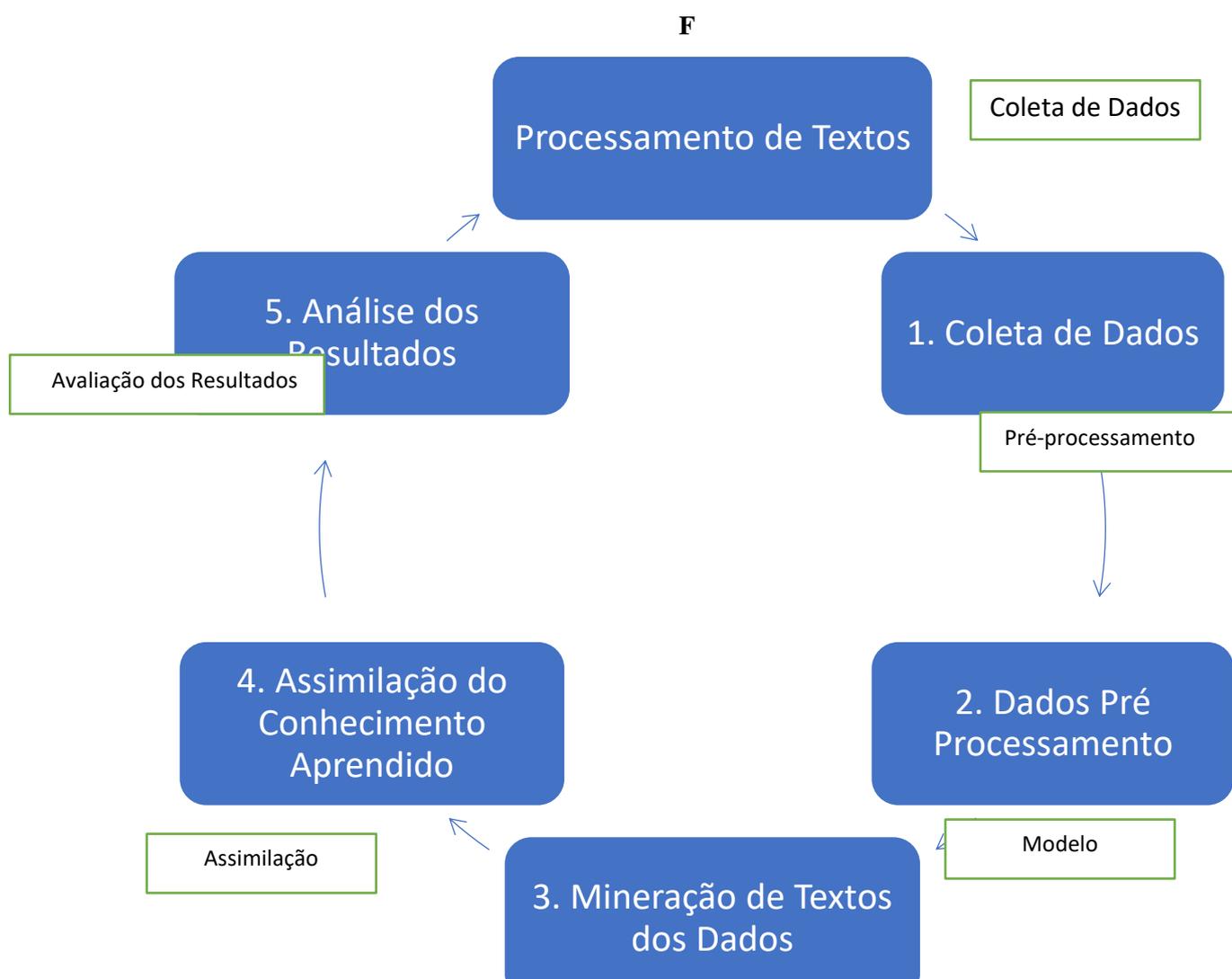
A partir das análises no decorrer do tempo, foi possível examinar as principais modificações e seus contextos, os temas de atuação e área geográficas apontadas, ocasionando a construção de conhecimento para ao tema milícias. A seguir, abordaremos as etapas

necessárias para esses resultados bem como a metodologia adotada para este contexto.

## **5.2 MÓDULO ANÁLISE – PROCESSAMENTO DE TEXTOS DAS DENÚNCIAS**

O processamento de textos aplicado ao AATDM para a mineração de dados de textos sensível ao contexto em denúncias, ocorreu em cinco etapas: coleta de dados, pré-processamento, mineração de textos dos dados, assimilação do conhecimento aprendido e análise dos resultados, conforme exibido na Figura 44 abaixo.

Foi aplicado o *pipeline* PLN reduzido nesta etapa para a formação dos *clusters* necessária às análises produzidas bem como as nuvens de palavras geradas, conforme Figura 45 abaixo, com o propósito de agrupamento automático de instâncias similares, isto é, para a formação dos *clusters*. O intuito da clusterização neste momento é uma classificação não-supervisionada dos dados. Ou seja, um algoritmo que *clusteriza* os dados e classifica em conjuntos de dados que “se assemelham” de alguma forma – independentemente de classes pré-definidas. Os grupos gerados por essa classificação são chamados *clusters*. Foi utilizado para o processamento de texto, descrito adiante, assim como na extração de atributos encontrados nas denúncias. Também foi usado para a modelagem necessária as análises ao estudo de caso e a própria mineração de textos empregada.



Fonte: Adaptada pelo autor.

**Figura 45: Etapas do Pipeline PLN Reduzido aplicado ao estudo de caso para o Processo de Mineração de Texto**

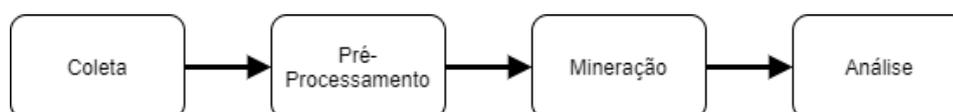


Fonte: Adaptada pelo Autor

### **5.3 ETAPAS DO PROCESSO DE MINERAÇÃO DE TEXTOS**

O processo de mineração de texto exige algumas etapas. Podemos visualizá-las de forma macro como a coleta dos dados, a mineração em si, que engloba todo o processo, seguida da análise humana dos resultados. No entanto, por tratar-se em sua grande maioria de textos em linguagem natural e com pouca estruturação de seu conteúdo, a mineração requer etapas de preparação dos dados trabalhados, como formatação, aumento da qualidade, redução de termos insignificantes para o propósito buscado e transformação do dado textual em dado numérico para que modelos matemáticos possam ser aplicados. Desta forma, a mineração de texto pode ser desmembrada nas seguintes etapas, Figura 46 abaixo:

**Figura 46: Etapas de Mineração de Texto**



Fonte: Adaptada pelo Autor

### 5.3.1 COLETA

A primeira etapa do processo de mineração de texto utilizada ao AADTM é a coleta. Esta etapa consiste na procura e seleção de documentos relevantes ao domínio que irá compor a base de dados, e onde serão analisados as pesquisas e busca do conhecimento requerido nos textos das denúncias. Para isso, precisamos reunir todos os documentos de texto que são relevantes a esse domínio, para que possamos extrair o conhecimento desejado, isto é, a extração do conhecimento, formando o chamado *corpus*.

### 5.3.2 PRÉ-PROCESSAMENTO

A segunda etapa do processo é o pré-processamento. Nesta etapa, os dados coletados são preparados para serem utilizados pelas técnicas de pré-processamento empregadas no AATDM. Essa etapa consiste em uma série de ações realizadas com o objetivo de transformar o texto em linguagem natural em uma série de termos úteis, geralmente representada por uma estrutura de tabela atributo-valor.

O pré-processamento teve como objetivo melhorar a qualidade e uniformizar do texto, para que as bases fossem melhor utilizadas nos modelos de modo que utilizássemos os termos mais importantes na base para haver uma melhor classificação dos modelos selecionados. Foram aplicadas diferentes sub-etapas de pré-processamento para as bases de dados descritas no projeto e feitas diferentes avaliações dos modelos conforme se obteve o avanço deste processo:

- Tokenização
- *Case Folding*
- Remoção de Acrônimos
- Remoção de Pontuação
- Remoção de *Stopwords*
- Correção Ortográfica
- Remoção de Acentos
- *Stemming*

Cabe ressaltar que utilizaremos como exemplo frases genéricas apenas para demonstrar as técnicas utilizadas e não as frases encontradas nas denúncias – por uma questão de sigilo, referente aos dados sensíveis encontrados nas denúncias, explicado anteriormente neste estudo. Não se trata apenas da técnica executada, mas do conhecimento que dela pode ser produzido. Portanto, toda essa composição do conhecimento descoberto a partir das atividades executadas

neste estudo que comprometam investigações em curso foram mantidas sob sigilo.

Dentre as técnicas de pré-processamento (PERKINS, 2014), as seguintes foram utilizadas em nosso estudo:

- ✓ *Tokenização*: É o processo de separar as frases de um texto em palavras.

Portanto, as demais técnicas de pré-processamento serão aplicadas a cada palavra. Por exemplo, a frase: “Se perdermos outro jogo seremos eliminados e isso me deixa com pavor.”. Após o processo de *tokenização*, a frase seria representada como: “[Se] [perdermos] [outro] [jogo] [seremos] [eliminados] [e] [isso] [me] [deixa] [com] [pavor]”

- ✓ *Case Folding* : Processo de converter todo o texto para letras maiúsculas ou minúsculas visando desfazer a capitalização de palavras e descaracterizar nomes próprios.

Por exemplo, para a frase: “Se o Flamengo perder outro jogo seremos eliminados e isso me deixa com pavor”. Após o processo de *case folding*, a frase ficaria: “se o flamengo perder outro jogo seremos eliminados e isso me deixa com pavor.”

- ✓ *Remoção da pontuação*: Apesar de alterar o significado da frase, a pontuação não altera o significado da palavra, desta forma ela não agrega nenhum valor significativo à base de dados, podendo ser removida para diminuir a quantidade de elementos a serem analisados.
- ✓ *Remoção de stopwords*: remoção de palavras que não trazem nenhum significado à frase. Ex.: artigos, conjunções, preposições, etc.

Abaixo na tabela 1 a seguir, a construção e utilização da tabela de *stopwords* empregada para remoção das palavras nas denúncias exatamente por não agregar nenhum valor significativo.

Tabela 1: Tabela de STOPWORDS

STOPWORDS						
a	eles	estivesse	houver	no	serei	terão
à	em	estivessem	houvera	nos	seríamos	teríamos
ao	entre	estivéramos	houveram	nossa	seria	teu
aquela	eram	estou	houverem	nossas	seriam	teus
aquelas	éramos	está	houvesse	nosso	será	teve
aquele	essa	estávamos	houvessem	nossos	serão	tinha
aqueles	essas	estão	houverámos	num	seríamos	tinham
aquilo	esse	eu	houvéssemos	numa	seu	tive
as	esses	foi	há	não	seus	tivemos
às	esta	fomos	hão	nós	somos	tiver

até	estamos	for	isso	o	sou	tivera
com	estas	for a	isto	os	sua	tiveram
como	estava	foram	já	ou	suas	tiverem
da	estavam	forem	lhe	para	são	tivemos
das	este	fomos	lhes	pela	só	tivesse
de	esteja	fosse	mais	pelas	também	tivessem
delas	estejamos	fui	me	pelo	te	tivéramos
dele	estes	fôramos	mesmo	pelos	tem	tivéssemos
deles	estive	fossemos	meu	por	temos	tu
depois	estive	haja	meus	qual	tenha	tua
do	estivemos	hajam	minhas	quando	tenham	tuas
dos	estiver	hajamos	minhas	que	tenhamos	tínhamos
e	estivera	havemos	muito	seja	teremos	você
ela	estiveram	hei	na	sejam	teria	vocês
elas	estiverem	houve	nas	sejamos	teriam	vos
ele	estivermos	houvemos	nem	sem	terá	obs

Fonte: Adaptada Pelo Autor

Por exemplo, na frase: “Se perdermos outro jogo seremos eliminados e isso me deixa com pavor.”. Após a remoção das *stopwords*, teremos: “perdermos outro jogo eliminados deixa pavor”.

- ✓ Remoção de acentos: Consiste na remoção dos acentos das palavras para evitar repetição de termos com significados semelhantes.

Na língua portuguesa, a acentuação pode mudar a classe gramatical da palavra. Exemplo: “sábua” – adjetivo e “sabia” – verbo. É importante que esse passo seja executado por último, pois na língua portuguesa, a presença de acentos é significativa em algumas regras gramaticais. Por exemplo, o singular de “sóis” é “sol”. Se nessa etapa as palavras já estivessem sem acento, a palavra “dois”, por exemplo, poderia ser entendida como plural de “dol”.

- ✓ *Stemming*: Processo de retirada dos prefixos e sufixos das palavras para deixar apenas o radical, que é a parte da palavra que contém seu significado.

Além dos afixos, também são eliminadas características de gênero, número e grau. O objetivo é diminuir a quantidade de dados a serem analisadas pelo algoritmo de aprendizagem, já que a partir de um mesmo radical, podemos formar várias classes de palavras. Contudo, o *stemming* pode diminuir a precisão das buscas na classificação de documentos. Por exemplo, as variações morfológicas aumentam a capacidade de discriminação entre documentos. Outro ponto que podemos considerar negativo dessa técnica é a ocorrência de *overstemming*, que ocorre quando parte do

radical é extraído junto com o afixo, e *understemming* que ocorre quando o afixo não é totalmente removido (JIVANI, 2011).

Para o exemplo acima, temos: 'perd', 'outr', 'jog', 'elimin', 'deix', 'pav'.

- ✓ Conversão dos Dados: O SVM é um modelo “*Bag of Words*” ou Bolsa de Palavras, é a representação de documentos de forma numérica mais utilizada para MT e Recuperação da Informação. Como o nome indica, neste modelo o documento é representado apenas pelo conjunto de palavras que o compõe, sem discriminação de ordem. Utilizamos esse modelo por resolver nosso problema nesse contexto e apresentar bons resultados. Ele é bastante útil para a montagem da matriz de termos e documentos. Esta matriz pode ser representada de formas mais robustas, como através de *WordEmbedding* (vetorização mais elaborada, que considera a ordem que os termos aparecem no texto, normalmente criada por modelos de redes neurais) ou da Ponderação TF-IDF, que será utilizada neste trabalho e brevemente apresentada nesta seção.

Em um documento de texto, nem todos os termos são igualmente relevantes para descrever o seu conteúdo. Palavras muito frequentes em geral são vagas, sendo um grande complicador para o processamento dos dados na hora de prever uma classificação por exemplo. Em contrapartida, palavras pouco utilizadas conseguem reduzir o conjunto de registros relacionados, facilitando uma possível classificação. Pensando na frequência com que cada palavra aparece em um documento de texto são definidos os pesos, quantificando a sua importância na descrição do documento. Consiste em uma representação de um documento através de um conjunto de suas palavras, desconsiderando a ordem e a gramática e considerando apenas a multiplicidade, utilizamos esse modelo por apresentar bons resultados para nosso contexto, e por resolver nosso problema. Por isso, esse modelo não aceita dados textuais, apenas dados numéricos (KO, 2012).

Foi necessário um passo a mais de pré-processamento, em que foi feita essa transformação das bases para dados numéricos.

Uma das técnicas utilizadas na mineração de texto é o processo de indexação de documentos, que consiste na organização de documentos por meio de índices. Isto possibilita maior eficiência no gerenciamento, principalmente quando é necessário manipular grandes massas de dados textuais. E uma das etapas da indexação é a atribuição de pesos aos termos de um documento a fim de quantificar sua frequência. Essa etapa é necessária porque o SVM necessita que os dados sejam gerados de forma independente e identicamente distribuída. De acordo com Gomes (2013), a modalidade de peso TF-IDF é uma fusão das modalidades TF e IDF. Na modalidade TF, são atribuídos pesos a cada termo de acordo com sua frequência.

Na Equação 2.1 mostramos o cálculo do TF.

Equação 2.1 – Cálculo do TF.

$$TF = \text{ocorrências do termo} / \text{total de termos} \quad \text{Equação 2.1}$$

Na modalidade IDF, é contabilizado o inverso da frequência dos termos. Diferente de outros pesos que são atribuídos em relação a um termo e documentos específicos, esse peso caracteriza apenas o termo, independente do documento.

É usado para medir se um termo é comum ou raro em uma coleção de documentos.

Aa Equação 2 mostra o cálculo do IDF. Equação 2 – Cálculo do IDF.

$$IDF = 1 + \log_e \frac{\text{número total de documentos}}{\text{quantidade de documentos nos quais o termo ocorre}}$$

Sendo a união dos 2 termos descrito acima, o TF-IDF é uma medida de peso global, pois o peso é ponderado considerando a frequência do termo tanto em relação ao documento quanto em relação à coleção completa de documentos.

A Equação 3 mostra o cálculo do TF-IDF.

Equação 3 – Cálculo do TF-IDF.

$$TF - IDF = TF * IDF$$

Equação 3 – Cálculo do TF-IDF.

O produto final dessa etapa é uma representação matricial da frequência de cada palavra dos registros.

### 5.3.3 MODELO

A terceira etapa é a criação do modelo ou modelagem, em que são aplicados os algoritmos. Esses algoritmos são aplicados na base já pré-processada, trata-se das bases totais processadas anteriormente, onde o atributo meta já é conhecido. O algoritmo recebe esse complemento da base e aprende com ela, extraindo as entidades, características e padrões, e associando-os às suas respectivas classes já existentes na verificação de algo novo ou diferente do que já encontramos (PERKINS, 2014). Para isso, são utilizadas algumas métricas de avaliação das classificações feitas pelo modelo e uma checagem pelos analistas do DD, responsáveis na aceitação ou não dessas novas palavras e entidades encontradas, que serão ou não utilizadas pelo modelo. Ao fim dessa etapa, o algoritmo apresenta uma listagem com as palavras e entidades encontradas pelo modelo, com o reconhecimento das novas palavras,

termos e entidades adquirido. Assim, o modelo está apto a identificar informações em textos de denúncias com essas palavras, termos e entidades reconhecidas em sua série histórica, podendo desta maneira ligar algo novo descoberto com toda a série histórica, proporcionando novas ligações de produção de conhecimento e assimilações aos textos das denúncias.

Este processo é cíclico e dinâmico a cada atualização da base. Denominamos assimilação na mineração de texto das denúncias ao arcabouço analítico. Tratando-se, portanto, da etapa de assimilação do conhecimento produzido a partir das denúncias e comparado com a série histórica na produção de novos conhecimentos antes não identificados.

### 5.3.4 MÉTRICAS DE AVALIAÇÃO

Nessa etapa escolhemos algumas métricas para a avaliação dos resultados encontrados para o trabalho. Foram escolhidas as seguintes métricas de avaliação: acurácia, precisão, *recall* e matriz de confusão, A matriz de confusão é uma estrutura que possibilita visualizar o desempenho de um algoritmo de classificação para todas as classes preditas.

Através da matriz da confusão, outras métricas podem ser obtidas, como a acurácia, a precisão, a sensibilidade (conhecida também como *recall*), métricas estas que foram utilizadas neste trabalho e que serão apresentadas a seguir:

A acurácia é o cálculo da proximidade entre o valor predito pelo algoritmo e o valor real do atributo. É a porcentagem de predições corretas. A fórmula da acurácia é mostrada na Equação 4.

Equação 4 – Cálculo da Acurácia

$$A = \frac{\text{número de predições corretas}}{\text{número total de amostras}}$$

A precisão calcula a proximidade entre o número de predições positivas corretas, classificadas pelo algoritmo, e o número total de predições positivas. Trata-se da porcentagem de predições positivas corretas. A fórmula da precisão é mostrada na Equação 5.

Equação 5 – Cálculo da Precisão

$$P = \frac{\text{número de predições positivas corretas}}{\text{número de predições positivas}}$$

O *recall* é o cálculo da proximidade entre o número de predições positivas corretas, classificadas pelo algoritmo, e o número total de amostras positivas. Ou seja, trata-se da porcentagem de predições positivas encontradas pelo classificador. A fórmula do *recall* é

mostrada na Equação 6.

Equação 6 – Cálculo do *Recall*

$$R = \frac{\text{número de predições positivas corretas}}{\text{número de amostras positivas}}$$

A matriz de classificação contabiliza para todos as categorias de classificação, a predição obtida pelo algoritmo e o valor real do atributo. A visualização desses dados ajuda a perceber quais categorias o algoritmo teve maior dificuldade de classificar, apontando em que local será necessário aplicar possíveis melhorias e ajustes a fim de melhorar o desempenho do modelo para a etapa de assimilação. A Tabela 2 abaixo, mostra uma matriz de confusão genérica de N classes.

**Tabela 2: Matriz de Classificação Genérica de N Classes**

		Classe Predita			
		Classe 1	Classe 2	Classe 3	Classe N
Classe Real	Classe 1	<Acertos C1>	Erros C2	Erros C3	Erros CN
	Classe 2	Erros C1	<Acertos C2>	Erros C3	Erros CN
	Classe 3	Erros C1	Erros C2	<Acertos C3>	Erros CN
	Classe N	Erros C1	Erros C2	Erros C3	<Acertos CN>

Fonte: Adaptada pelo autor

### 5.3.5 ANÁLISE DOS RESULTADOS

A quinta e última etapa do processo é a de análise das classificações feitas pelo algoritmo. Verifica-se as classificações estão, de fato, condizentes por meio da análise manual de uma pequena amostra dos registros classificados encontradas como resultados.

É importante ressaltar que a etapa de pré-processamento foi feita utilizando a biblioteca NLTK, uma biblioteca do *Python* para o processamento de linguagem natural (NLTK, 2017).

Em seguida ao pré-processamento, foi necessário um ajuste ao modelo. Nessa etapa utilizamos o algoritmo SVM de aprendizagem supervisionada. Esse algoritmo foi escolhido pela sua capacidade de lidar com bases desbalanceadas e registros com muitos atributos (CESAR, 2017), possuindo boas propriedades de generalização e respondendo bem para dados não linearmente separáveis.

#### **5.4 CONCLUSÕES DO CAPÍTULO**

Neste capítulo foram apresentados os módulos 3 e 4 do modelo executado na base de dados do DD. Demonstramos a aplicação da normalização dos dados, a remoção de *stopwords*, a aplicação da técnica de *stemmer*, o processo de *tokenização* e o processamento de textos. Apresentamos as etapas de coleta e pré-processamento, a modelagem, as métricas de avaliação e as análises.

## **6. MODELO GENERALIZADO E CONSIDERAÇÕES FINAIS**

Neste capítulo serão apresentados o modelo generalizado gerado a partir do estudo de caso e que pode ser aplicado em outras áreas e crimes, bem como os trabalhos futuros. As contribuições e as considerações finais.

## **6. MODELO GENERALIZADO:**

O serviço do Disque Denúncia é um exemplo bem-sucedido de como o uso inteligente da informação pode auxiliar o trabalho das autoridades policiais e áreas de inteligências para a redução dos índices de violência urbana, aplicando recursos tecnológicos e inteligência estratégica. O que propomos com este trabalho são técnicas de mineração de dados textuais sensível ao contexto, com textos de denúncias para auxiliar a extração de informações e produção de conhecimento a temáticas de crimes e delitos, podendo ser aplicado em diversos contextos criminais a partir do uso do modelo generalizado que apresentamos neste trabalho. Servindo para análise e visualização de dados e projetos de ciências de dados relacionados a crimes que vão desde a coleta, correção, processamento, armazenamento e recuperação das informações geradas pelo serviço de atendimento. Pode ser usado em todos os 187 temas ou as 19 classificações cadastradas de denúncias, concebendo-se em um verdadeiro potencial de uso em extração de informações relevantes a cada contexto, e permitindo um vasto campo para trabalhos futuros.

As tecnologias de *stemming* foram utilizadas para a preparação e análise dos dados textuais e de clusterização para extração de informações e produção de conhecimentos aplicados ao contexto.

Utilizamos ferramentas para classificar e relacionar um grande volume de dados textuais,. Dessa forma, tornou-se possível encontrar informações fornecidas nos relatos das denúncias de maneira confiáveis e que fizeram sentido quando empregadas nas análises relacionadas ao tema.

Com a utilização de técnicas de mineração de textos num modelo computacional proposto neste trabalho descrito a seguir, selecionamos e encontramos informações textuais relevantes as milícias a partir das denúncias podendo ser estendidas a novos contextos com a aplicação do modelo generalizado.

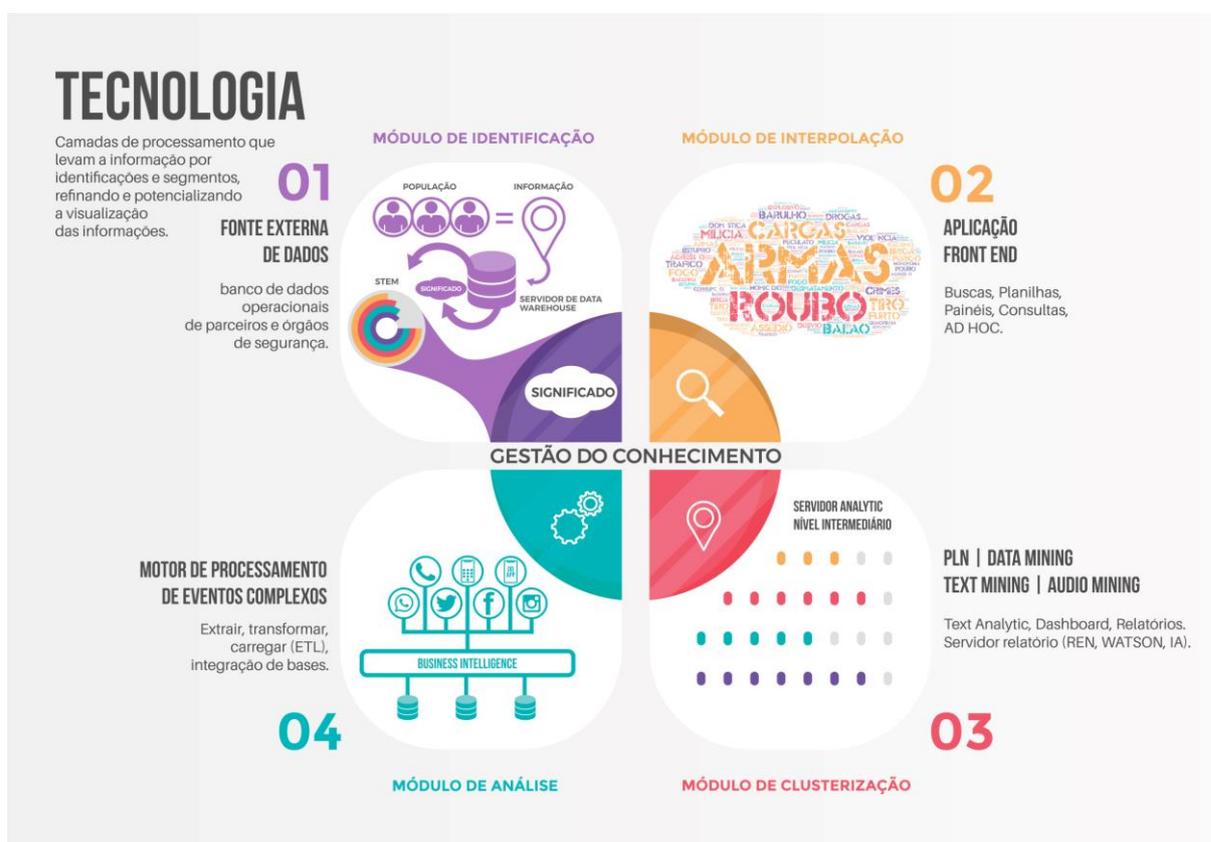
### **6.1 DESCRIÇÃO DO MODELO GENERALIZADO PROPOSTO**

O modelo generalizado proposto neste trabalho é composto de quatro módulos de funcionamento, descritos a seguir. Emprega a manifestação de entidades, palavras, expressões e semânticas utilizadas em relatos de denúncias na produção do conhecimento através do reconhecimento de entidades na mineração textual de denúncias, como forma de obtenção desses de dados e informações. O denominado *Arcabouço Analítico de Textos em Denúncias: Novas Tecnologias no Tratamento e Análises de Denúncias - Estudo de Caso Milícias* foi objeto

deste trabalho, mostrando a melhora da eficiência dos processos investigativos utilizando denúncias e demonstrando resultados promissores, permitindo desse modo a composição da construção do modelo generalizado, Figura 49, abaixo.

Estabelece com sua utilização a construção das relações das entidades, extração e descobertas de informações empregue na construção do conhecimento em diversos contextos, a análise e visualização dos dados para pesquisa, análises e estudos futuros, no decorrer do tempo, através das denúncias. Podendo utilizar esse *framework* em diferentes contextos, problemas e temáticas, de uma forma ampla e segura em trabalhos futuros.

**Figura 47: Modelo Generalizado Proposto**



Fonte: Figura criada pelo Autor

**1. Módulo de Identificação:** Através do texto fonte é produzida uma lista de palavras-chaves (*stems* chaves), aplica-se um algoritmo de *stemming* que transforma cada palavra no seu respectivo *stem* ou forma raiz. Algoritmos de *Stemming*, ou *stemmers* tem o objetivo de reduzir uma palavra ao seu *stem*, ou seja, à sua forma raiz. As palavras-chave de uma consulta ou de um documento são mais bem representadas por *stems* do que pelas palavras originais. Com isso

reduzimos o tamanho da ontologia, e com isso o número de termos distintos necessários para representar uma lista de documentos.

**2. Módulo de Interpolação:** A partir da lista de *stems* de cada denúncia, busca-se a lista de significados mapeados com suas respectivas estatísticas de frequência do banco.

**3. Clustering:** As denúncias são *clusterizadas* a partir dos significados extraídos de cada uma delas. Primeiro, agrupa-se as denúncias similares e adicionam-se informações estatísticas aos temas classificados. Cria-se um sumário executivo para posterior validação.

**4. Módulo Análise:** A partir dos tópicos classificados e da clusterização, utiliza-se a base de conhecimento em que está contida uma lista de medidas que serve para o direcionamento dos tópicos extraídos e para a realização de inferências do usuário.

## 6.2 CONSIDERAÇÕES FINAIS E CONTRIBUIÇÕES

A aplicação do arcabouço analítico de textos em denúncias voltada para o caso da milícia proporcionou a melhoria da qualidade dos serviços na central de atendimento, uma vez que se estende automaticamente sobre o processamento, pesquisa e análise, mostrando melhoria na classificação dos tipos de crimes, nas correções de dados e no aumento do reconhecimento de informações pertinentes a partir da coleta de dados oriundos das denúncias.

A seguir, lembramos as duas hipóteses que duas hipóteses conduziram este trabalho:

✓ Hipótese 1 – (H1): O uso de um ambiente analítico com novas tecnologias, aplicando mineração de dados textuais sensível ao contexto, em denúncias, configura-se como um instrumento de coleta e extração de informações eficientes na produção do conhecimento sobre milícias para a atividade de Inteligência.

✓ Hipótese 2 – (H2): A utilização de um arcabouço de mineração de dados textuais, sensível ao contexto em denúncias, e a cultura do valor da informação, orientada a dados para a visualização e extração de informações, que auxilie a tomada de decisão, agrega valor à produção do conhecimento a temática milícias.

Em relação a H1, podemos dizer que se destacam na mineração de dados a extração e associação das informações recebidas nas denúncias por temas, possibilitando formação de uma base de conhecimento e a criação de um dicionário específico e único de palavras, termos, expressões e semânticas utilizadas em relatos de crimes, através das denúncias para o tema

milícias. Isso proporciona oportunidades na extração de conhecimento e nas perspectivas úteis para a produção de conhecimento de milícias através de texto contidos em denúncias, aplicado na área de inteligência.

A transformação do processo de captura dessas denúncias, por meio da alteração do *script* de coleta utilizada na central de atendimento, proporcionou melhorias substanciais para a implementação do ambiente analítico denominado AATDM, o que viabiliza mecanismos aplicados na mineração de textos das denúncias, implementada por esse estudo. Favorece a melhoria na captura de informações, aumenta a qualidade dos dados capturados e cria melhores possibilidade de pesquisa e análises, ou seja, favorece o trabalho de analistas da área de inteligência. Os resultados encontrados mostraram-se bastante expressivos e eficientes até o momento, sendo utilizados pela área de inteligência e aplicados na busca e coleta de dados sobre milícias. Houve um aumento significativo do acesso, da busca e da coleta de dados por meio do ambiente analítico do arcabouço. Permitiu através de seu uso a melhoria em todas as etapas do processo visto neste estudo, e configurando-se desta maneira, como um instrumento eficiente na produção de conhecimento acerca do tema.

Podemos concluir que a metodologia e o modelo aplicados ao estudo de caso, mostraram-se relevantes e promissores aos objetivos estabelecidos a partir de sua aplicação no mundo real. Obviamente necessita de atualizações, acompanhamento e modificações contínuas para sua melhoria. O trabalho é constante, ininterrupto e precisa ser atualizado constantemente para sua melhor eficiência. Melhorias no código de georreferenciamento, dicionários e nuvens de palavras serão frequentes e sucessivas para obter sempre boas respostas.

A partir das análises no decorrer do tempo, foi possível examinar as principais modificações, seus contextos, seus temas de atuação e áreas geográficas apontadas, ocasionando a construção de conhecimento sobre as milícias. Até o momento, o ambiente analítico vem sendo utilizado em algumas operações.

As classificações e palavras utilizadas nas denúncias passaram por adequações linguísticas por apresentarem características específicas como diversidade cultural e social, interferências nas variantes linguísticas, regras comunicativas de um contexto específico, gírias, neologismo, estrangeirismo e abreviações. As palavras na denúncia representam características complexas de uma linguagem territorial habitualmente ligadas a indagações como nichos, gueto, grupos, facções, organização criminosa, milícias, crimes entre outros, que se relacionam de alguma maneira entre si ou em conjunto. E que foram assinalados e trabalhados com a metodologia proposta pelo modelo deste estudo.

As tecnologias de *stemming* para a preparação dos dados, clusterização para extração

e produção do conhecimento textual, a partir das denúncias, mostraram-se adequadas para a abordagem do problema apresentado, assim como todo o processo de limpeza e correções das palavras, através da mineração de dados em textos fez-se possível a produção de conhecimento relacionada a tema.

Em relação a H2 – que trata da utilização do modelo do arcabouço analítico de textos, baseado na mineração de dados textuais e na cultura do valor da informação orientada a dados para a tomada de decisão e agregar valor à produção do conhecimento a temática milícias, – podemos dizer que a cultura do valor da informação orientada a dados como suporte a tomada de decisão, denominada *data driven*, foi empregada com sucesso como forma de coleta de visualização de dados e informações neste estudo, e obteve resultados expressivos no seu propósito inicial que era extrair dados, informações e agregar valores ao processo de produzir conhecimentos, a partir da base de dados de denúncias do DD. Mostrou-se de uma enorme riqueza a possibilidade da mineração de textos unidos a visualização de dados, de forma a examinar o processo de denúncias para a temática milícia. A forma dinâmica de atualizações do arcabouço de texto permitiu o controle e o acompanhamento em tempo real, de fiscalizar informações importantes para a área de segurança e principalmente a verificação de ocorrências em áreas específicas, permitindo agir de forma rápida em relação as demandas necessárias.

Foi percebido o entusiasmo gerado a partir da visualização dos dados utilizados no arcabouço analítico de textos, na produção de conhecimento, na pesquisa e na produção de análise sobre as milícias pelas equipes que utilizaram o arcabouço. Constituiu-se um aumento significativo em todos os módulos e *dashboards* criados, acarretando a necessidade de construção de novas visualizações de dados a partir do uso com todos os ciclos de tratamentos históricos de denúncias ou com cortes específicos a cada demanda, concedendo novos filtros e possibilidades de aprimoramento de análises e melhoria de performance devido a sua utilização. Cada equipe é responsável por um tema ou carteira e possui experiência em analisar as informações de maneira muito própria ao tipo de trabalho solicitado em seu órgão ou agência e poder usar o arcabouço como uma ferramenta na concepção dessas informações tornou o processo enriquecedor para ambos os lados.

A mineração de dados textuais para a produção de conhecimento contidas nas denúncias necessita de características específicas e muitas vezes não percebidas nos relatos, como questões de adequações linguísticas, já mencionadas anteriormente e, por características próprias, denominadas como o idioma local, nativo, cheio de expressões e significados.

Desperta a necessidade do reconhecimento de uma linguagem não formal, representada por expressões e representações linguísticas, como gírias que necessitam ser compreendidas

nesse abismo social e racial em que se encontra o estado, onde as diferenças entre o asfalto e o morro tratando-se de segurança são vitais na compreensão do dia a dia, na complexidade dessas denúncias, no sentimento de apelo gerada por esse cidadão. A criação de um glossário de palavras que incluiu expressões, termos semânticos a cada tipo ou classe de denúncia no decorrer do tempo com foco em milícias, pressupôs um potencial estudo futuro sobre o tema, mesmo com todos os desafios do uso da língua portuguesa e das dificuldades oriundas das características específicas da própria linguagem utilizada que é encontrada nas denúncias.

Iniciando desde a coleta e perpassando a padronização e limpeza e correção de dados, extração de radicais temáticos – formados pelas principais palavras relacionadas às diferentes categorias de denúncia, transpassando pela construção dos dicionários especializados para apoio à extração de entidades, as pistas, características, localização e temas textuais que foram construídos a partir das denúncias até a construção de indicadores de performance, através da construção dos cenários que foram convertidos em *dashboards*, utilizando técnicas de visualização de dados *self-service*, permitindo que o analista possa pesquisar e visualizar toda a análise necessária a temática além da análise de vínculos, normalmente empregados aos estudos da área. As denúncias não são apenas utilizadas como fonte de dados, mas também como fonte de informações em tempo real na descoberta de novos conhecimentos ao tema milícias.

Como exemplo do que foi descoberto em nossas análises, a representação das palavras e sua abrangência linguística dentro do contexto milícias é de uma riqueza imensurável para a produção do conhecimento. O arcabouço analítico de texto usado como ferramenta a partir do relato da denúncia demonstra todo o clamor do cidadão e a triste realidade enfrentada. Características de envolvidos, territórios, movimentação, formas de franquia, áreas de domínio no decorrer do tempo entre outras podem ser analisadas e acompanhadas para aplicação de melhores investigações e monitoramento, garantindo dessa maneira uma das premissas iniciais deste trabalho, que era permitir acessos e construção de valores através dos dados para que pudessem ser utilizados pelos tomadores de decisão e possibilitassem a tomada de medidas cabíveis e mitigação de riscos, propondo soluções cabíveis a temática milícias.

A conversão dos elementos de tendência para a produção do conhecimento contida em seus princípios doutrinários foi utilizada e respeitada na implementação do arcabouço. Do mesmo modo, foi incorporada a técnica do ciclo de produção do conhecimento sugerida na DISPERJ para a atividades de inteligência, suportando assim toda a adequação necessária para sua utilização por todos os órgãos, agências e área de inteligência. A criação dos comparativos em forma de *dashboards* para acompanhar as variações, monitoramento, e as suas respectivas

nuvens de palavras, propiciando uma visão profunda da complexidade referente ao tema milícias, o que contribui para auxiliar o analista e demais profissionais, ampliarem sua capacidade de pesquisa, análise e produção do conhecimento, convertendo-se em novas oportunidades de enfrentamento as milícias.

### **6.3 CONTRIBUIÇÕES**

As contribuições desse trabalho partiram da classificação automática em denúncias realizada no trabalho de Rafael Adnet Pinho (PINHO *et al*, 2017) onde verificou-se que a quantidade de palavras que compõem uma denúncia caiu de 38 palavras em média em 2015, para cerca de 18 palavras em média em 2018, até a aplicação de técnicas de mineração de dados textuais e módulos de pré-processamento, correção, limpeza de dados e produção de conhecimento através de relatos de denúncias utilizadas no arcabouço analítico de texto para a temática Milícias, instrumento realizado neste trabalho. Isto demonstra que estudos em conjunto possibilitam ganho de conhecimento.

Propor a implementação das frações significativas neste trabalho tornou-se uma necessidade de pôr em prática, não apenas para resolver o problema identificado da diminuição de palavras nas denúncias, como similarmente um trabalho específico para capturar pistas relevantes a produção de conhecimento, e melhorar as informações capturadas nas denúncias, obtendo mais completas de informações deste modo. Diante da complexidade na captação e coleta de dados e com o propósito de aliar conceitos empregados da área de inteligência a esses dados, consideramos importante e necessária a implementação do emprego do conceito de frações significativas na captura das denúncias. Essa implementação aliada aos recursos tecnológicos executados pelo modelo mostrou-se bastante promissora na produção e ligação de conhecimentos escondidos nos dados, possibilitando a geração de informações em sua maioria relevantes ao tema. É necessário um acompanhamento constante devido à dinâmica e à velocidade dos conhecimentos produzidos na atualidade. É a ação constante de descoberta e retroalimentação de novos conceitos textuais e semânticos que está em constante produção de novas dinâmicas e interpretações.

**Tabela 3: Contribuições do Trabalho**

Proposta de Framework sensível ao contexto e seus Resultados	
Estudo de caso	Resultados
Implementação do sistema de BI de atendimento	Melhoria dos processos internos de atendimento e melhores resultados
Integração das Bases de Dados	Melhoria no processo de extração, cruzamento, e visualização de dados
Criação do ambiente seguro de coleta de Dados	Melhoria na coleta de dados a partir da criação do ambiente seguro
Criação, teste e adaptações na metodologia de <i>Data Cleaning</i>	Melhoria no processo de anomalias da base de dados
Atualizações da base de dados	Melhoria na limpeza e geração dos dados utilizados no arcabouço
Padronizações de Logradouros	Melhoria na identificação e localização dos logradouros
Geocodificação da base limpa de dados	Melhoria e preservação do anonimato das denúncias geocodificadas
Geração de Lat e Long das denúncias e entidades	Melhoria na geração das localidades envolvidas nas denúncias
Interpolação dos Dados	Melhoria no processo de clusterização dos dados para o arcabouço
Geração da base de conhecimento	Melhoria na captura das palavras e geração do dicionários de termos utilizados
Geração da base de Personas	Melhoria na identificação de suspeitos e envolvidos nas denúncias
Geração da base de rastro	Melhoria na construção da linha de tempo das denúncias
Clusterização e Agrupamento dos dados	Melhoria nos cruzamentos das informações e recursos computacionais
Implementação dos módulos do arcabouço analítico	Todos os Módulos apresentaram bons resultados ao experimento do estudo de caso
Implementação da mineração textual	Melhoria na captura, extração e produção de conhecimento ao tema
Implementação das métricas de avaliação	Melhoria dos resultados encontrados no estudo de caso
Generalização do Modelo a outros temas	Proposta de um Modelo Generalizado que pode ser aplicado a outros temas em estudos futuros

Fonte: elaborada pelo autor

Percebemos que a produção de conhecimento textual encontrado nas denúncias

Da mesma forma coletar palavras e texto direcionados a geração de informações que pudessem ser utilizadas ao estudo milícias, configura-se importante suporte na construção do conhecimento e que por suas características torna-se único na recepção dessas informações. O tratamento desses dados permitiu aplicar inteligências múltiplas proativas e especializadas para combater essa estrutura, e produzir conhecimento específico e refinado a partir de texto, gerando informações vitais para essas ações. Sem dúvida alguma, o emprego de tecnologias a essas necessidades é o fator decisivo para alcançar sucesso. Empregar tecnologia na atividade de inteligência deve ser compreendido como uma necessidade vital para diminuir conflitos e reduzir mortes, além dos outros indicadores relacionados à violência que são exibidos diariamente em nossos telejornais. Aliar a tecnologia no processo de construção, produção e gestão do conhecimento textual de pistas e informações em denúncias é importante para essas questões. Recursos tecnológicos que sejam capazes de otimizar, organizar e reduzir recursos humanos empregados em seu tratamento e produção podem ser utilizados dessa maneira a defesa do Estado, da sociedade e do cidadão.

Tratar as entidades como obtenção de pistas possibilitou descobertas importantes relacionadas ao tema milícias, e tornou-se uma necessidade imprescindível para a melhoria do serviço. A utilização desta técnica não foi para aumentar o número médio de palavras utilizadas nas denúncias apenas, mas para captar as melhores informações e pistas que podem compor aquela denúncia. Ou seja, a palavra, o texto importante aquela ação. Essas pistas e entidades, que representam pessoas, objetos, carros, facções, vulgos, classes, foragidos entre outras são todos os dados ou informações importantes que podem ser relacionados na composição dessas

informações sobre algo ou alguém.

Desta forma, vale lembrar que ações conjuntas de pesquisas ligadas ao mesmo tema e base de dados proporcionaram evoluções importantes a todos os trabalhos, compartilhando o aprendizado colaborativo do mesmo modo que os códigos gerados para a constante melhoria dos resultados vêm sendo construídos e estimulados atualmente junto a outras instituições, alunos e estudos.

O sentido da palavra, o termo, a frase, a expressão, a terminologia utilizados na denúncia precisam ser entendidas na coleta porque nos traz elementos simbólicos relevantes e necessários para compreender sobre a representação empregada ao território, o comportamento de violência de uma facção, seu comportamento de expansão ou domínio, a análise de sentimento (sentimento de apelo, medo, pavor, urgência, necessidade imediata de uma ação etc) que expressa aquela denúncia.

O presente trabalho estabeleceu um referencial teórico representativo neste sentido, podendo ser útil para pesquisas futuras ligadas ao tema como a implementação de sistemas de informação integrados nas organizações de atividades de inteligência. Este arcabouço pode ser aplicado a trabalho futuros nos ramos de proteção como: as ameaças cibernéticas, *hackerativismo*, *big data* sobre terrorismo, espionagem digital, corrupção, lavagem de dinheiro e evasão de divisas, interferência externa, crime organizado, roubo e manipulação ilegal de dados sensíveis. É sobre todas essas questões que trata a utilização do arcabouço, ou seja, favorecer a realização de estudo analíticos em textos de denúncias que sirvam como fonte de coleta de dados para agentes e analistas da área de inteligência entenderem o fenômeno e dessa forma construir maneiras de repressão aos crimes.

Em relação a atuação de organizações criminosas, torna-se notório que o problema de milícias, crime organizado, tráfico de drogas e armas ligados a segurança da sociedade e do cidadão envolvem diversas variáveis que se influenciam mutuamente (pobreza, desemprego, crime, violência, homicídio, armas, tráfico de drogas, crescimento econômico entre outros). Mostrando que se trata de um problema complexo com características específicas e bastantes marcantes, com indicações, contextualizações e discursos ainda mais complexo para a produção do conhecimento, que envolvem questões de ordem públicas, políticas e diversas mazelas existentes em nossa sociedade, que perduram por longos tempos e que possuem interferências políticas marcantes no decorrer desses longos anos.

Para concluir, vale ressaltar que, não se trata de uma obra concluída e acabada, que responde a todos os anseios e premissas estabelecidas em relação as milícias. Continua como uma obra aberta e que necessita continuidade de estudos, pesquisas e colaboração tanto da

academia, quanto das forças de segurança, do cidadão se motivando a continuar denunciando e os resultados sendo mostrados na imprensa, uma retroalimentação ininterrupta a favor da paz.

Oferecer aparatos tecnológicos e acolher as necessidades reais enfrentadas por essas áreas foi um fator de motivação contínuo em todo o trabalho. Ou modificamos o modo de atuação diante de tamanha complexidade gerada pela violência no Rio de Janeiro, auxiliando de alguma maneira essas instituições a conseguirem exercer seus papéis de uma forma digna e melhor, ou estaremos refém das organizações criminosas. É importante que a sociedade civil auxilie as forças de segurança na garantia da paz e da preservação das vidas, possibilitando a proteção de seus cidadãos, ou não teremos saída para essas questões.

Demonstramos com este trabalho, um repertório completo de possibilidades de pesquisas dentro de uma linha ampla e abrangente a diferentes temas e oportunidades acadêmicas, não apenas como esta que demonstra a contribuição da computação no processo, mas também em áreas onde a realidade dessas questões é importante, como os aspectos econômicos ligados a milícias, o comportamento de indivíduos diante dessa realidade, a evasão de dinheiro gerada por esses grupos entre tantos outros aspectos. Abre-se uma oportunidade a partir desse trabalho, em conjunto com a academia através de acordos de cooperação como por exemplo, com Programa de Pós-graduação em Economia da UFRJ, que desenvolve estudo sobre aspectos econômicos, violência doméstica e violência contra a mulher, buscando outras visões sobre o tema. É preciso favorecer pesquisas, buscando modificar e agir em prol das temáticas voltadas para a segurança. A aproximação da academia aos problemas da sociedade e a busca de melhores soluções e enfrentamento dessas realidades é benéfica para ambos.

Com esse trabalho conseguimos atingir resultados positivos em algumas das questões levantadas na pesquisa e alcançamos contribuições como:

- ✓ o aumento da qualidade na captação de informações;
- ✓ o aumento do número de denúncias relevantes por atendentes;
- ✓ o aumento da capacidade de armazenamento das informações;
- ✓ o aumento da capacidade de tratamento das informações, produção de conhecimento em análises customizáveis ao tema milícias;
- ✓ O aumento da capacidade de otimização e envio de informações aos órgãos competentes;
- ✓ O aumento da integração das bases e automatizar os processos de carga, limpeza e troca de informações (ETL);
- ✓ O aumento de difusão de denúncias e informações para órgãos de inteligência;
- ✓ O maior monitoramento das informações difundidas;

- ✓ O aumento de respostas aos denunciantes no acompanhamento das denúncias;
- ✓ O aumento de cooperação entre os órgãos de inteligências e polícias.

Questões importantes foram abordadas diante deste trabalho tais como:

A questão da violência em áreas específicas dominadas pelas milícias e o aumento da criminalidade nessas áreas e seus entornos e como é esta realidade diária onde uma grande parte da população é violentada, sendo noticiado diariamente nos telejornais locais. Torna-se indispensável e inadiável o desenvolvimento de aparatos científicos e tecnológicos, que utilizem tecnologias no tratamento de informações e ferramentas que possibilitem pesquisas e informações em denúncias que permitam ações efetivas de produção de conhecimentos necessárias para tomadores de decisão planejem ações específicas no combate à violência e ao crime organizado.

Assessorar a tomada de decisões com base na análise de informações geradas na produção do conhecimento através das denúncias foi um dos princípios norteadores deste trabalho, bem como uma de suas contribuições. Ao propormos o uso de ferramentas de análise e visualização de dados vinculados a uma abordagem que aprimorasse o processo de criação de valor, fornecesse uma estrutura de avaliação, monitoramento e alerta de impacto resultante dessa ferramenta, baseado na visualização e análise dessas informações, podemos dizer que o objetivo foi alcançado no trabalho e agora o seu próprio uso gerar novas oportunidades, despertar novas necessidades e novos olhares.

Todo o armazenamento e diagnóstico diferenciado dessas denúncias, assim como a integração com outras bases de dados, na produção do conhecimento ao estudo de caso milícia foi vital para o resultado encontrado. Sem essa possibilidade não teríamos alcançados resultados positivos seja pelo dado ou pelas necessidades que foram sendo geradas a cada implementação. Isso permite que a atividade de Inteligência sirva ao propósito de apoiar o processo decisório, a partir da produção de conhecimentos, e uma atuação colaborativa para a gestão de políticas públicas e a preservação do Estado Democrático de Direito na aplicação de ações de segurança para curto, médio e longo prazos que possam ser acompanhados no decorrer do tempo com o intuito de preservação da ordem pública.

As diferentes formas de entrada e canais utilizados na captura dessas denúncias e o aumento dos canais digitais incorporados ao atendimento mostraram-se crescentes e promissores, tanto em relação ao aumento da quantidade, originadas nestas novas formas de comunicação, quanto na qualidade e precisão dessas informações. Por isso, foi necessário a adaptação e o desenvolvimento de um arcabouço analítico de textos combinando todas as

tecnologias que envolviam todos os canais existentes para o processo de coleta e visualização de dados, técnicas de processamento de linguagem natural, *data mining* e mineração de textos em denúncias para seu tratamento, integração e produção do conhecimento que acompanhem o estudo e análise de dessas informações

É por meio da transformação do modo de como esses dados são visualizados e trabalhados que é possível a produção do conhecimento, empregado na área de Inteligência. Medir os resultados produzidos, planejar o monitoramento de dados, informações e suspeitos, bem como a incorporação de novas bases ao estudo, permitindo identificar e produzir novas informações, desenvolvendo novas visões e, principalmente, entender seus significados. Gerar conhecimento a partir dessas descobertas e usá-las no arcabouço foram as ações desenvolvidas por este trabalho, relacionadas ao tema milícia. O caminho foi longo e a aprendizagem contínua na compreensão e entendimento doutrinários necessários, bem como na adaptação da construção do arcabouço a regras rígidas e necessárias para o seu uso. O acesso às informações provenientes ao tema, a geração de conhecimentos territoriais, todo o processo de limpeza e padronização necessária, a integração dos dados com outras fontes, a visualização desses dados em seus *dashboards*, a construção da metodologia de análise textual empregada para a produção do conhecimento com base nas denúncias exigiu dedicação. Por isso, é uma felicidade poder executar alguns dos ensinamentos apreendidos, muitos dos quais de natureza sensíveis e sigilosos e perceber que foi possível alcançar esse resultado.

Desta forma, o estudo auxiliou na conversão de dados brutos de denúncias em *insight* promissores, auxiliando a área de inteligência com informações necessárias aos seus estudos e análises. Através desse estudo é possível entender o que representa o fenômeno milícia; quais as demandas reprimidas que são supridas por essa organização criminosa; quais as oportunidades de lucros que ela propõe-se a empreender através de suas franquias; onde e de qual forma estão localizados os conflitos que ela gera; a forma de aumento de expansão e monopólio e, principalmente de que maneira isso afeta a comunidade que se expressa através das denúncias; identificar as características e o que representa o retorno da normalidade após os conflitos nesses territórios e; quais os fatores que afetam seu faturamento e formas de negócios entre outros.

Analizamos a aplicação direta de técnicas de linguagem natural nos textos de denúncias para a produção de conhecimento. A partir dessas descobertas foi possível fazer pesquisas e encontrar indícios ligados ao tema das milícias. Vale lembrar que, a continuidade de aplicação dessas técnicas torna-se importante a estudos futuros. Os temas de estudos referente a violência pelo mundo preocupam-se em entender os objetivos claros de maneira mais rápida das

representações envolvidas a essas violências e ameaças, de maneira constante nas sociedades modernas por representar ameaças a soberania e como forma de antecipação a esses fatos. Para isso, necessitamos obter dados e informações para: analisar, interpretar e disseminar conhecimentos para ações conjuntas entre os órgãos responsáveis pela segurança pública e seus locais. O radicalismo ideológico, o racismo e diversos conflitos vêm crescendo em toda parte do mundo e hoje é um tema de preocupação mundial.

Lamentavelmente, no Brasil, a realidade enfrentada por órgãos de segurança e áreas de inteligência, vem encarando a algum tempo a falta de recursos empregadas no setor, e tornou em diversos momentos esse estudo bastante complicado na sua realização e implementação.

Diversas autorizações e termos de cooperações burocráticos entre órgãos precisaram ser revistos e assinados diante da necessidade de desenvolvimento a temática milícias e dos protocolos de segurança dos dados produzidos a partir dessas informações. A burocracia entre os órgãos e áreas precisam ser revista para a melhoria e compartilhamento de iniciativas como essa. Após a extinção da secretaria de segurança todos os órgãos tornaram-se isolados e reconfiguraram suas equipes e necessidades, dificultando em muito a troca de informações entre as áreas. Para a continuidade de trabalhos futuros, esses acordos e termos precisam ser revistos, não podemos permitir que movimentos políticos ou burocráticos enfraqueçam o desenvolvimento de estudos futuros apenas pela descontinuidade de ações. Isto ocorreu por diversas razões durante a realização deste trabalho, incluindo problemas ou modificações políticas, troca de comandos enfrentados diversas vezes, durante todo o trabalho, uma Intervenção Federal Militar em 2018, ocasionando recomeço a cada troca. Isso dificultou o crescimento das ações implementadas, bem como a modificação tanto de etapas do projeto, como das equipes envolvidas que normalmente também são trocadas em descontinuidade do trabalho, o que causa perda de recursos, tempo, conhecimento e desenvolvimento. É preciso existir mecanismos de planejamento de mudanças para que projetos e recursos não sejam prejudicados e descontinuados sucessivamente. Estes problemas foram enfrentados em diferentes fases deste trabalho o que o tornou no mínimo emocionante, mas sobretudo enriquecedor em permitir lidar com uma realidade mutável a cada momento.

Outro exemplo de dificuldade enfrentada que pode ser citada foi a busca de parceiros e representantes de softwares já consolidados e disponíveis no mercado, e tentávamos modular e desenvolver em conjunto uma solução baseada no modelo proposto por este estudo. As instituições mostravam-se interessadas na iniciativa e chegaram a contribuir inicialmente. No entanto, em um curto prazo de desenvolvimento e teste, essas parcerias não deram certo, seja por não ser permitido ceder a licença de uso por um tempo determinado, seja por tentarem se

apropriar dos códigos fontes desenvolvidos nas soluções empregadas ou propriedade intelectual de desenvolvida por este trabalho, procurando utilizar-se de seu saber prático, isto é, as habilidades adquiridas pelas experiências desse projeto e seduzindo-se a modificá-las para aplicar de maneira comercial, objetivando venda de seus produtos. Isso é algo bastante desgastante e gerou a necessidade de aparatos legais e debate sobre quebra de contrato, ocasionando a suspensão de parceria por objetivos distintos ao acordado inicialmente. Com consequência, perdeu-se a licença de uso dos aplicativos. Foram momentos estressantes, mas proporcionaram grande experiência, pois nos forçaram a tomar a decisão de não ficarmos presos a uma única tecnologia ou único fornecedor. Cabe ressaltar que, todas essas dificuldades na aplicação dessa pesquisa ocorrem por tratar-se de um cenário real em que toda essa complexidade faz parte do contexto e da realidade da área. São dados e informações da vida real e são dados sensíveis e que podem trazer uma certa notoriedade – que no meu ponto de vista, não podem ser comerciais. Infelizmente ou felizmente, nem sempre os trabalhos acadêmicos são pensados a partir de possibilidades e realidades como essas.

Estes são aspectos aparentemente sem controle ou de urgências, por exemplo, estamos habilitados e treinados a enfrentar questões técnicas quanto a escolha deste ou daquele algoritmo, qual o melhor e mais rápido resultado obtido, o porquê da escolha disso ou daquilo. Foi muito importante para mim vivenciar essa realidade, com essa possibilidade de aplicação da minha pesquisa, utilizando o arcabouço para apoiar o levantamento e coleta de dados reais, entre os órgãos e áreas envolvidas; poder testar na prática o que deveria e poderia ser mantido ou modificado para tentar obter melhores respostas em um ambiente real, vivido intensamente por essas pessoas na busca de melhores informações, e não apenas em um ambiente controlado e fictício para coleta dessas informações. A possibilidade de teste, com certeza, foi um diferencial significativo na realização e resultados deste trabalho, e poder empregar o *framework* no estudo de caso do arcabouço a essa realidade pulsante de informação tornou-se para mim a maior contribuição deste trabalho. Possibilitando dessa forma, a criação da proposta do modelo generalizado sensível a contexto, que poderá ser empregado em trabalhos futuros.

As necessidades de adaptações e o enfrentamento dos problemas a partir das possibilidades diante da realidade encontrada nessas instituições precisam ser trabalhadas como são, a vida como ela é, com os órgãos, agências e pessoal que temos, com as polícias que temos e com a criminalidade que precisa ser combatida e que bate em nossas portas.

Os *scripts* e os códigos produzidos necessitam ser revisitados constantemente, atualizados, melhorados e compreendido a cada novo período de tempo, pois, como vimos, o processo de construção de conhecimento e entrada de dados é muito complexo e dinâmico

tornando-se uma obra viva ao tema advindo através de informações das denúncias, algo rico, significativo e não encontrado em nenhuma outra base de dados a não ser na base do DD. Trata-se de um fenômeno mutável no decorrer do tempo, necessitando desta forma de constante atualização em sua compreensão e ações. Destaca-se que é uma iniciativa inovadora do emprego da tecnologia ao tema milícias.

#### **6.4 POSSIBILIDADES DE TRABALHOS FUTUROS:**

Indica-se a necessidade de trabalhos futuros referentes a continuidade do trabalho realizado. Apontaremos algumas sugestões a seguir que julgamos importantes para os desenvolvimentos de trabalhos relacionados ao tema.

Talvez a principal delas seja a utilização do modelo criado para detecção de denúncias em outros textos, como por exemplo o emprego do modelo para classificação de denúncias em textos de redes sociais, como já foi comprovado um aumento significativo de denúncias através desses dispositivos em redes sociais pelo DD, um estudo aplicando o modelo a esse canal seria bastante significativo como pesquisa, bem como a utilização de *chatbot* e o uso de *WhatsApp*, pois tratam-se de um público mais jovem que opta por esses mecanismos e canais de denúncia. A utilização de um modelo SVM aplicado aos canais de redes sociais na detecção de possíveis relatos de denúncias de crimes como forma de comentários por exemplo, monitorando alguns temas específicos seria relevante como trabalho futuro, bem como a utilização do *Word2vec* na aplicação deste contexto.

Outro ponto que poderia ser melhorado especificamente em um novo estudo seria na reconstrução e atualização do dicionário utilizado, uma vez que, além de mais palavras poderem ser adicionadas aos tipos de classificação de crimes, utilizando por exemplo, uma base coloquial mais abrangente. Isso ocorre porque não encontramos nenhuma base com todas as características necessárias aos relatos de denúncias ou crimes disponíveis, e tivemos que buscar nos históricos das denúncias as palavras para compor o dicionário utilizado neste estudo e como a classificação por tipo e classes de crimes para compor o dicionário utilizado, o que se configura uma das principais contribuições desse trabalho. No entanto, já encontra-se disponível uma nova versão do dicionário base disponível, chamada *Portuguese (new sp)*, que talvez possa trazer maior relevância para a etapa de correção ortográfica e limpeza dos dados.

A construção do dicionário ainda demanda termos mais coloquiais e mais distantes da norma culta. Isso poderia render um trabalho de pesquisa nesse sentido. A dimensão do problema enfrentado e a dinâmica de palavras, termos e expressões que surgem a cada tipo de

ocorrência denunciada necessita ser melhor estudada. Em especial, destaca-se a necessidade de criação de um dicionário característico a cada tipo de canal, pois as formas de comunicação em cada canal e suas particularidades e traços de escrita são diferentes tanto no uso de aplicativos como em canais de internet. As pessoas costumam reproduzir marcas de regionalismos, além de vícios de linguagem comuns na comunicação coloquial, o que está longe da norma culta padrão da língua. Uma sugestão nessa vertente poderia ser um estudo de adequação de linguagem em dispositivos de denúncias que pudessem levar em considerações algumas questões como diversidade cultural e social que interferem nas variantes linguísticas ao realizar uma denúncia, levantando as regras comunicativas de um contexto específico ligado a um tipo de crime, por exemplo, violência a mulher. É preciso considerar se o emprego de gírias, neologismo, estrangeirismo e abreviações representam alguma nova forma de comunicação entre ou intra organizações criminosas, se o tipo de abordagem empregado pelo atendente está adequando ao contexto do vocabulário do denunciante, portanto, se é um fator determinante de eficácia, e quais seriam as adaptações linguísticas necessárias entre outros. Deste modo, podemos observar um vasto campo de pesquisa e trabalhos futuros relevantes a continuidade da temática enfrentada, bem como um rico material de possibilidades em trabalhos significativos que possam ser desenvolvidos empregando tecnologia para questões ainda sem resposta nesse trabalho.

Outra vertente não menos importante a ser explorada trata do aprimoramento detecção, limpeza e georreferenciamento dos locais das denúncias. Este trabalho é constante e já vem sendo utilizado por outros alunos no melhoramento do código gerado. Agir continuamente no aprimoramento dessas localidades é extremamente necessário para a melhoria dos resultados, e também no monitoramento dessas áreas. Cabe lembrar que as características de localização da maioria dessas áreas são extremamente específicas porque muitas vezes são locais constituídos por becos, vielas, pontos de referências e nomes ou apelidos de áreas dentro de uma comunidade, gerando a necessidade de uma atualização constante na identificação e limpeza dos dados para a sua real geocodificação. Agir e cuidar sempre da identificação precisa desses locais é sempre um ponto a ser melhorado. A base do DD já possui estas informações, mas além do endereço formal existem em vários relatos um vasto conjunto de referências a locais pouco conhecidos, como regiões dentro de comunidades. A identificação precisa dos locais e seu mapeamento poderia auxiliar as forças policiais na investigação de determinada ocorrência de crimes e de alguma forma melhorar a realidade dessas áreas, mantendo o cuidado de garantia do anonimato do denunciante. As questões sobre georreferenciamento e melhoria do código de limpeza identicamente são temas sugeridos como possíveis trabalhos futuros.

Um ponto bastante importante de estudos futuros trata de pesquisas e análises de dados oriundos de redes sociais. Os relatos nas redes muitas vezes chegam como apelo, carregados de desespero, angústia, inconformidades, indignação e clamor. São campos que podem ser usados em estudos como: análise de sentimento dos relatos das denúncias por tipo de crime, a possibilidade de criação de modelo preditivo de ocorrências criminais advindas pelas redes sociais, o emprego do reconhecimento facial das imagens que são enviadas para identificação de foragidos da justiça e supostos criminosos entre outros.

As imagens, vídeo, e áudios enviadas pelos aplicativos e canais de redes sociais não foram utilizadas nesse estudo. Um trabalho futuro que pudesse utilizar de todo esse material seria de grande valor para o serviço. Seria interessante desenvolver novos estudos para a classificação e a criação de glossário de tema e sentidos de expressões empregadas a cada tipo de classificação de denúncias em outros formatos, não apenas os de textos, mas também a todas as formas de fotos, vídeo e áudio recebidas com características, dados, e informações, que não foi utilizada neste estudo informações relevantes, bem como os vínculos existentes nas centenas ou milhares de informações digitais relacionando as análises textuais poderiam ser buscadas, ocasionando a produção de pesquisas e análises que contemplem o uso de multiformatos nessa produção de conhecimento, que permanece com sugestão de trabalho futuro.

Outro assunto que vem sendo abordado na literatura e que tem total aderência como trabalho futuro são os chamados relatórios de texto em linha. Trata-se de levantamento na linha do tempo e que necessita de acompanhamento constante dos dados de forma transversal ao tempo. Por exemplo, a sequência de denúncias sobre um alvo específico no tempo, suas características e informações levantadas a partir das denúncias, possui como característica o tempo e o crescimento explosivo da coleta de suas informações que são, na sua maioria, de natureza não-estruturada. Muitas técnicas de última geração envolvendo abordagens supervisionadas, não supervisionadas ou semi-supervisionadas foram desenvolvidas nos últimos anos para que a agregação automática destes relatórios pudessem ser realizadas e alguns trabalhos, poderíamos utilizar esse tipo de ferramenta para a geração dos rastros desse alvo e identificar toda a sua movimentação ao longo do tempo de uma maneira mais estruturada. A anotação de relatórios de crimes em linha é uma tarefa desafiante, uma vez que vários tipos de relatórios de crimes são frequentemente gerados ao longo do tempo. Identifica-se inicialmente as entidades nomeadas e seleciona apenas as palavras de contexto dentro de um par de entidades como uma frase. Assim, cada relatório é descrito por uma coleção de frases. As frases são vetorizadas e um algoritmo de agrupamento baseado em gráficos é aplicado ao agrupamento de todas as frases recolhidas. As frases dentro de um agrupamento são consideradas como o tipo

semelhante de frases, chamadas paráfrases e cada relatório é representado por um vector binário de dimensão igual ao número de agrupamentos obtidos. Já foi publicado por Priyanka Das (2019) uma primeira tentativa de agrupamento incremental de grupos de relatórios de crime integrando a rede neural e a teoria de conjuntos aproximados e seria bastante interessante a realização de um trabalho contemplando essa técnica em denúncias.

## REFERÊNCIAS

- ADOMAVICIUS, G.; TUZHILIN, A., Toward the next generation of recommender systems: a survey of the state-of-the-art e possible extensions. **IEEE Transactions on Knowledge e Data Engineering**, Piscataway, v.17 n.6, Jun. 2005, p.734-749. Disponível em: <https://ieeexplore.ieee.org/document/1423975>. Acesso em: 24 jun. 2020.
- ALMEIDA, Cleibson Aparecido de et al. Melhoria na qualidade de dados com a aplicação de "data cleaning" na base de dados de acidentes aeronáuticos da aviação civil brasileira. **Atoz: novas práticas em informação e conhecimento**, [S.l.], v. 5, n. 2, p. 72-79, dec. 2016. Disponível em: <https://revistas.ufpr.br/atoz/article/view/47303/29516>. Acesso em: 14 maio 2020.
- ALVARI H.; SARKAR, S.; SHAKARIAN, P. Detection of Violent Extremists in Social Media. In: INTERNATIONAL CONFERENCE ON DATA INTELLIGENCE AND SECURITY, 2., 2019. **Proccedings [...]**. [S. l.: s. n.], 2019.
- BARHAMGI, M., MASMOUDI, A., LARA\_CABRERA, R., Camacho, D. Social Networking data analysys with semantics: application to the radicalization problem. *Journal of Ambient Intelligence and Humanized Computing* (2018).
- ANISTIA INTERNACIONAL. **Você matou meu filho!**: Homicídios cometidos pela Polícia Militar na cidade do Rio de Janeiro. Rio de Janeiro: Anistia Internacional, 2015.
- ALERJ. 14 de novembro de 2008. RELATÓRIO FINAL DA COMISSÃO PARLAMENTAR DE INQUÉRITO DESTINADA A INVESTIGAR A AÇÃO DE MILÍCIAS NO ÂMBITO DO ESTADO DO RIO DE JANEIRO, Rio de Janeiro, 2008. Disponível em: [http://www.nepp-dh.ufrj.br/relatorio\\_milicia.pdf](http://www.nepp-dh.ufrj.br/relatorio_milicia.pdf). Acesso em: 8 jul. 2020.
- BARHAMGI, M.; MASMOUDI, A.; CABRERA, L. R.; CAMACHO, D. Social Networking data analysys with semantics: application to the radicalization problem. **Journal of Ambient Intelligence and Humanized Computing** (2018).
- BASTOS, P.M. **A Gestão do conhecimento no ambiente organizacional**: estudo da ferramenta idseg-dígitro, UFSC 2009.
- BELLINGER, G. **Knowledge Management**. Consortium benchmarking study. Final report. American Productivity & Quality Center, 1996.
- BETTENCOURT, L. M. A. The origin of scaling in cities. **Science**, v. 340,n. 6139, p. 1438-1441, 2013a.
- BETTENCOURT, Luís. The Origins of Scaling in Cities *Science* 21 Jun 2013b: V. 340, Issue 6139, pp. 1438-1441. Disponível em: <https://science.sciencemag.org/content/340/6139/1438/tab-pdf> Acesso em 29 mai. 2020.
- BOCCARA, N. **Modeling complex systems**. Germany: Springer, 2004.
- BOWEN, J.E. An expert System for Police Investigators of Economic crimes. **Expert**

**Systems with Applications**. Vol. 7 N. 2., 1994, p. 225-248.

BRASIL. **Decreto nº 4376, de 13 de setembro de 2002**. Dispõe sobre a organização e o funcionamento do Sistema Brasileiro de Inteligência, instituído pela Lei no 9.883, de 7 de dezembro de 1999, e dá outras providências. [S. l.], 2002. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/decreto/2002/D4376compilado.htm](http://www.planalto.gov.br/ccivil_03/decreto/2002/D4376compilado.htm). Acesso em: 13 jul. 2020.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 18 dez. 2011c. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/112527.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm). Acesso em 29 mai 2020.

BRITO, THIAGO DA SILVA; OLIVEIRA, RAFAEL DA SILVA. **Solução de Business Intelligence utilizando a plataforma Microsoft na área da Segurança Pública**. Orientador: ASTÉRIO KIYOSHI TANAKA. 2017. 71 p. Trabalho de Conclusão de Curso (Graduação) Sistemas de Informação - Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2017. Disponível em: <https://bsi.uniriotec.br/tcc/textos/201711ThiagoBritoRafaelOliveira.pdf>. Acesso em: 13 jul. 2020.

BRITO, Walkir Alexandre Toscano de; MOTTA, Claudia L. R.; BORGES, Zeca. Análise semântica a partir de Sistemas Complexos na base de dados do Disque Denúncia. **Cadernos de Segurança pública**, Rio de Janeiro, ano 10, ed. 10, p. 1-16, 2018. Disponível em: <http://www.isprevista.rj.gov.br/download/Rev20181006.pdf>. Acesso em: 5 jul. 2020.

BURGOS, M. B. **A utopia da comunidade**: Rio das Pedras, uma favela carioca. Edições Loyola, 2002

CANO, Ignacio; DUARTE, Thais. **“No sapatinho”**: a evolução das milícias no Rio de Janeiro (2008-2011). Rio de Janeiro: Fundação Heinrich Böll, 2012.

CESAR, M. V. G. **Classificação de falhas de equipamentos de unidade de intervenção em construção de poços marítimos por meio de mineração textual**. 2017. Dissertação (Mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2017.

CHAUDHURI et al. An Overview of Business Intelligence Technology. **ACM**, N. 8., Vol. 54, 2011, 88-98.

CHEN, H. CHUNG, W. XU, J.J. QIN, GWY, CHAU, M. Crime data Mining: A general framework and some examples. **IEEE Computer Society**, 2004, p. 50-56.

CLAUDINO, Nilton. MINHA DOR NÃO SAI NO JORNAL. **Revista Piauí**, Rio de Janeiro, v. 59, n. 00, p. 00-00, ago. 2011. Disponível em: <https://piaui.folha.uol.com.br/materia/minha-dor-nao-sai-no-jornal/>. Acesso em: 30 jun. 2020.

CRAWFORD, R. **Na era do capital humano**: o talento, a inteligência e o conhecimento como forças econômicas - seu impacto nas empresas e nas decisões de investimento. São

Paulo: Atlas, 1994.

DAMASCENO et Al. SimCleaner - Sistema de Padronização de Bases de Dados utilizando Funções de Similaridade. Laboratório de Planejamento de Redes de Alto Desempenho. Conference: XIV Semana de Informática (SEMINF) e Escola Regional de Informática Norte (ERIN) At: Belém – PA. Disponível em:  
[https://www.researchgate.net/profile/Elton\\_Moutinho/publication/228778258\\_SimCleaner-Sistema\\_de\\_Padronizacao\\_de\\_Bases\\_de\\_Dados\\_utilizando\\_Funcoes\\_de\\_Similaridade/links/00b7d52f4a474b21ce000000/SimCleaner-Sistema-de-Padronizacao-de-Bases-de-Dados-utilizando-Funcoes-de-Similaridade.pdf](https://www.researchgate.net/profile/Elton_Moutinho/publication/228778258_SimCleaner-Sistema_de_Padronizacao_de_Bases_de_Dados_utilizando_Funcoes_de_Similaridade/links/00b7d52f4a474b21ce000000/SimCleaner-Sistema-de-Padronizacao-de-Bases-de-Dados-utilizando-Funcoes-de-Similaridade.pdf). Acesso em: 30 jun. 2020.

DAS, Priyanka *et al.* Group incremental adaptive clustering based on neural network and rough set theory for crime report categorization. **Neurocomputing**, [s. l.], 2019. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0925231219315723?via%3Dihub>. Acesso em: 13 jul. 2020.

DAVENPORT, T. H. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, 1998.

DILEK, Selma; ÇAKIR, Hüseyin; AYDIN; Mustafa. APPLICATIONS OF ARTIFICIAL INTELLIGENCE TECHNIQUES TO COMBATING CYBER CRIMES: A REVIEW. **International Journal of Artificial Intelligence & Applications**, [s. l.], ano 2015, v. 6, ed. 1, p. 21-39, 21 abr. 2018. Disponível em:  
<https://arxiv.org/ftp/arxiv/papers/1502/1502.03552.pdf>. Acesso em: 7 jul. 2020.

DRUCKER, P. F. **Administrando para o futuro**: os anos 90 e a virada do século. São Paulo: Enio Matheus Guazzelli & Cia Ltda., 1992.

DRUCKER, P. **Sociedade pós-capitalista**. São Paulo: Pioneira, 1993.

EDUCAÇÃO EM ALVO. **O efeito da violência armada na sala de aula**, FGV DAPP/Fogo cruzado. Relatório 2017. Disponível em:  
<http://fogocruzado.org.br/educac%CC%A7a%CC%83o-em-alvo-os-efeitos-da-viole%CC%82ncia-armada-nas-salas-de-aula-do-rio-de-janeiro/>. Acesso em: 29 nov. 2018.

ERDI, P. **Complexity explained**. Germany: Springer, 2008.

ÉRNICA, M., BATISTA, A.A.G. A Escola, a metrópole e a vizinhança vulnerável. **Cad. Pesquisa**, [S. l.], v. 42, n. 146, p. 640-666, 2012. Disponível em:  
[http://www.scielo.br/scielo.php?pid=S0100-15742012000200016&script=sci\\_abstract&tlng=pt](http://www.scielo.br/scielo.php?pid=S0100-15742012000200016&script=sci_abstract&tlng=pt). Acesso em: 29 nov. 2018.

FIALHO, F. A. P. **Ciências da cognição**. Florianópolis: Insular, 2001.

FONSECA, Pedro. O que é milícia? 2019. Disponível em:  
<https://www.bol.uol.com.br/listas/o-que-e-milicia.htm>. Acesso em: 8 jul. 2020.

FUENTES, D.; BARDELI, R.; J.A. ORTEGA. L. GONZALEZ-ABRIL. A similarity measure between videos using alignment graphical; and speech feat. *Expert System with Applications*, 38, 2012.

FURTADO, B. A.; SAKOWSKI, P.A.M.; TÓVOLLI, M. H. **Modelagem de Sistemas Complexos para políticas Públicas**. 2015. Disponível em: [http://www.ipea.gov.br/portal/index.php?option=com\\_content&id=25860&Itemid=383](http://www.ipea.gov.br/portal/index.php?option=com_content&id=25860&Itemid=383). Acesso em: 29 nov. 2018.

GIL, Antônio Carlos. Como classificar as pesquisas. S/D. Disponível em: <http://www.madani.adv.br/aula/Frederico/GIL.pdf>. Acesso em: 8 jul. 2020.

GILL, L.; STATISTICS, G. B. O. for N. Book. Methods for automatic record matching and linkage and their use in national statistics. London : National Statistics, 2001.

GOMES, N. O. **Categorização de textos - estudo de caso**: documentos de pedidos de patente no idioma português. 2013. Tese (Doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, 2013.

GRANDIN, Felipe *et al.* Franquia do crime: 2 milhões de pessoas no rj estão em áreas sob influência de milícias. **G1**. Rio de Janeiro. 14 mar. 2018. Disponível em: <https://g1.globo.com/rj/rio-de-janeiro/noticia/franquia-do-crime-2-milhoes-de-pessoas-no-rj-estao-em-areas-sob-influencia-de-milicias.ghtml>. Acesso em: 30 jun. 2020.

HEWITT, A.N. BEAUREGARD, E., ANDERSEN, M.A. BRANTINGHAM, P.L. Identifying the nature of risky places for sexual crimes: the applicability of crime pattern and social disorganization theories in Canadian context. **Journal of Justice** 57 (2018) 35-46. Disponível em [https://books.google.com.br/books?hl=en&lr=&id=KGIfiiP1i4C&oi=fnd&pg=PR5&dq=Natural+language+processing+with+Python:analyzing+text+with+the+natural+language+toolkit&ots=Y3Alv8MGH3&sig=qVmMsCwsIso9UO5XM22uttVYZ\\_s#v=onepage&q=Natural%20language%20processing%20with%20Python%3A%20analyzing%20text%20with%20the%20natural%20language%20toolkit&f=false](https://books.google.com.br/books?hl=en&lr=&id=KGIfiiP1i4C&oi=fnd&pg=PR5&dq=Natural+language+processing+with+Python:analyzing+text+with+the+natural+language+toolkit&ots=Y3Alv8MGH3&sig=qVmMsCwsIso9UO5XM22uttVYZ_s#v=onepage&q=Natural%20language%20processing%20with%20Python%3A%20analyzing%20text%20with%20the%20natural%20language%20toolkit&f=false). Acesso em: 29 nov. 2018.

HSM MANAGEMENT. Muito além de um modismo. Ano 4, n.22, set./out. 2000. [https://books.google.com.br/books?hl=en&lr=&id=KGIfiiP1i4C&oi=fnd&pg=PR5&dq=Natural+language+processing+with+Python:+analyzing+text+with+the+natural+language+toolkit&ots=Y3Alv8MGH3&sig=qVmMsCwsIso9UO5XM22uttVYZ\\_s#v=onepage&q=Natural%20language%20processing%20with%20Python%3A%20analyzing%20text%20with%20the%20natural%20language%20toolkit&f=false](https://books.google.com.br/books?hl=en&lr=&id=KGIfiiP1i4C&oi=fnd&pg=PR5&dq=Natural+language+processing+with+Python:+analyzing+text+with+the+natural+language+toolkit&ots=Y3Alv8MGH3&sig=qVmMsCwsIso9UO5XM22uttVYZ_s#v=onepage&q=Natural%20language%20processing%20with%20Python%3A%20analyzing%20text%20with%20the%20natural%20language%20toolkit&f=false). Acesso em: 29 nov. 2020.

HUSSAIN, Amir. Cognitive computation: a case study in cognitive control of autonomous systems and some future directions. In: Neural Networks (IJCNN), The 2013 International Joint Conference on. IEEE, 2013. p. 1-6.

JAKOB, Alberto Augusto Eichman; YOUNG, Andrea Ferraz. O uso de métodos de interpolação espacial de dados nas análises sociodemográficas. Anais do XV Encontro Nacional de Estudos Populacionais, Caxambu, p. 1-22, 2006. XV Encontro Nacional de Estudos Populacionais, ABEP, realizado em Caxambu – MG – Brasil, de 18 a 22 de setembro de 2006. Disponível em: <http://www.abep.org.br/publicacoes/index.php/anais/article/view/1530/1494>. Acesso em 08 jul. 2020.

JIVANI, A. G. A comparative study of stemming algorithms. **Int. J. Comp. Tech. Appl.**, 2(6), 2011, p. 930-1938

JURAFSKY, D.; MARTIN, J. H. **Speech and language processing**: An introduction to natural language processing, computational linguistics, and speech recognition, Prentice Hall, Pearson Education International, 2009.

LAURIANO, Carolina. **Juíza assassinada sofreu emboscada e levou 21 tiros, diz delegado no Rio**. G1, Rio de Janeiro, p. 0-00, 12 ago. 2011. Disponível em: <http://g1.globo.com/rio-de-janeiro/noticia/2011/08/juiza-assassinada-sofreu-emboscada-e-levou-21-tiros-diz-delegado-no-rio.html>. Acesso em: 13 jul. 2020.

KANT DE LIMA, Roberto; EILBAUM, Lucía; MEDEIROS, Flavia (orgs). **Casos de repercussão**: perspectivas antropológicas sobre rotinas burocráticas e moralidades. Rio de Janeiro: Consequência, 2017.

KO, Y. A study of term weighting schemes using class information for text classification. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, 2012 p. 1029-1030.

KU, C. H. & LEROY, G. A. Decision support system: automated crime report analysis and classification for e-government. Elsevier, 2014. Disponível em: <https://pdfs.semanticscholar.org/c186/7cace312aa54c5b089570289d4edc4c2cbd2.pdf>. Acesso em 29 nov. 2018.

KU, C.H., LEROY, G.A. Crime Reports Analysis System to Identify Related Crimes. **Journal of the American Society for Information Science and Technology** 62 (8):1533-1547, 2011.

LASLEY, J.R. When Crime Reporting goes High-Tech: An experimental test of computerized citizen respons to crime. **Journal of Criminal Justice** Vol. 23 N: 6, 1995, p. 519-529.

LEMGRUBER, Julita; MUSUMECI, Leonarda; CANO, Ignácio. **Quem vigia os vigias?** Rio de Janeiro: Record, 2003.

LÉVY, P. **As tecnologias da inteligência**: o futuro do pensamento na era da informática. Rio de Janeiro: Editora 34, 1994.

LOECK, J.F., MAGNUS, T.M, MONTENEGRO, D., MATTOS, E.V., Gestão da Informação e governos locais: experiências do observatório de Segurança Pública de canoas (RS) e novas possibilidades. **Revista Brasileira de Segurança Pública**, São Paulo, v. 10. n.2, 88-103 Ago/Set 2016. Disponível em: <http://www.forumseguranca.org.br/publicacoes/gestao-da-informacao-e-governos-locais-experiencias-do-observatorio-de-seguranca-publica-de-canoas-rs-e-novas-possibilidades/>. Acesso em: 29 nov. 2018.

KAUARK, Fabiana da Silva; MANHÃES, Fernanda Castro; MEDEIROS, Carlos Henrique. **METODOLOGIA DA PESQUISA**: Um guia prático. Itabuna: VIA LITTERARUM EDITORA, 2010. Disponível em: [http://www.pgcl.uenf.br/arquivos/livrode Metodologia da pesquisa 2010\\_011120181549.pdf](http://www.pgcl.uenf.br/arquivos/livrode Metodologia da pesquisa 2010_011120181549.pdf).

Acesso em: 13 jul. 2020.

MANSOUR, Samah. Social Media Analysis of User's Responses to Terrorism Using Sentiment Analysis and Text Mining. **Procedia Computer Science**, [s.l.], v. 140, p. 95-103, 2018.

MARTINS, Isnard Thomas. **DESCOBERTA DE CONHECIMENTO EM HISTÓRICOS CRIMINAIS**: algoritmos e sistemas. 2009. 201 f. Tese (Doutorado) - Curso de Engenharia de Produção, Puc Rio, Rio de Janeiro, 2009. Disponível em: <https://www.maxwell.vrac.puc-rio.br/colecao.php?strSecao=resultado&nrSeq=14011@1>. Acesso em: 29 mai. 2020.

MENDONÇA, Alba Valéria. Solução de crimes depende de ajuda da população. **G1**. Rio de Janeiro, p. 00-00. 10 mar. 2007. Disponível em: <http://g1.globo.com/Noticias/Rio/0,,MUL9408-5606,00-SOLUCAO+DE+CRIMES+DEPENDENTE+AJUDA+DA+POPULACAO.html>. Acesso em: 30 jun. 2020.

MILLER, J. H.; PAGE, S. E. Complex adaptive systems: an introduction to computational models of social life. New Jersey: Princeton Studies in Complexity, 2007.

MITCHELL, M. Complexity a guided tour. Oxford: Oxford University Press, 2009.

MONTEIRO, J., ROCHA, R., Tráfico de drogas e desempenho escolar no Rio de Janeiro, FGV/IBRE novembro 2013. Disponível em: <http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/11716/Trafico%20de%20drogas%20e%20desempenho%20escolar.pdf?sequence=1>. Acesso em 29 nov. 2018.

MORAES, Luciane. **"DISQUE-DENÚNCIA: A ARMA DO CIDADÃO"** UM ESTUDO SOBRE OS PROCESSOS DE CONSTRUÇÃO DA VERDADE A PARTIR DAS EXPERIÊNCIAS DA CENTRAL DISQUE-DENÚNCIA DO RIO DE JANEIRO – 2006. 155 f.: 30 cm. Dissertação (mestrado) – Universidade Federal Fluminense, Niterói, 2006. Disponível em <http://ppgantropologia.sites.uff.br/wp-content/uploads/sites/16/2016/07/LUCIANE-PATR%C3%8DCIO-BRAGA-DE-MORAES-1.pdf>. Acesso em: 30 jun. 2020.

MORAES, Luciane. **“FALAR, OUVIR E ESCUTAR”**: ETNOGRAFIA DOS PROCESSOS DE PRODUÇÃO DE DISCURSOS E DE CIRCULAÇÃO DA PALAVRA NOS RITUAIS DE PARTICIPAÇÃO DOS CONSELHOS COMUNITÁRIOS DE SEGURANÇA. - 2011. 311f.: il. 30 cm. Tese (doutorado) - Universidade Federal Fluminense, Niterói, 2011. Disponível em: <http://ppgantropologia.sites.uff.br/wp-content/uploads/sites/16/2016/07/LUCIANE-PATR%C3%8DCIO-BRAGA-DE-MORAES.pdf>. Acesso em: 30 jun. 2020.

NETO, F.P.F., CASTRO, A.F. Business Intelligence applied in decision support for public safety. *Sistemas y Tecnologías de Informacion*. 9 Conferencia Ibérica de Sistemas y Tecnologías de informacion. Barcelona, CISTI, 2014, p. 198-203.

NLTK 3.2.5 Documentation. Disponível em: <http://www.nltk.org/>. Acesso em: 09 dez. 2017

NONAKA, I.; KONNO. The concept of “Ba”. *California Management Review*, Berkeley,

v.40, n.3, p.40-54, spring, 1998.

NONAKA, I.; TAKEUCHI, H. Criação do conhecimento na empresa: como as empresas japonesas geram a dinâmica da inovação. Rio de Janeiro: Campus, 1997.

PÁSSARO, S., KLEIN, E. & LOPER, E. Natural Language Processing with Python: analyzing text with the natural language toolkit, 2009.

PAULA, G. de, DANDOLINI, G. A. & SOUZA, J. A. Tecnologia da Informação e Comunicação e as atividades de Inteligência. Revista Ordem Pública, Vol. 5, N.1, Semestre I, 2012. Disponível em: <https://rop.emnuvens.com.br/rop/article/view/48>. Acessado em 29 nov. 2018.

PEDREGOSA, F. Et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, 2011, p. 2825–2830. Disponível em: <http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>. Acesso em 24 jun. 2020.

PERKINS, J.. Python 3 Text Processing with NLTK 3 Cookbook. Packt Publishing Ltd, 2014.

PETRELLI, Marcelo Bastos. **A Gestão do conhecimento no ambiente organizacional:** estudo da ferramenta idseg-dígitro. 2009. 51 f. TCC (Graduação) - Curso de Administração, Universidade Federal de Santa Catarina, Florianópolis, 2009. Disponível em: <https://repositorio.ufsc.br/handle/123456789/120335>. Acesso em: 24 jun. 2020.

PINHO et al. Automatic Crime Report Classification through a weightless Neural Networks. ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium). Disponível em: <https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2017-42.pdf>. Acesso em: 30 jun. 2020.

PINHO, R. A., BRITO, W. A. T., MOTTA, C. L. R. & LIMA, P. V. Automatic Crime Report Classification through a weightless Neural Networks. ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017. Disponível em: <https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2017-42.pdf>. Acessado em 29/11/2018.

PRIYANKA DAS, ASIT KUMAR DAS, JANMENJOY NAYAK, DANILO PELUSI, WEIPING DING. Group Incremental Adaptive Clustering based on Neural Network and Rough Set Theory for Crime Report Categorization, Neurocomputing, 2019.

PONCHIROLLI, O . A importância do capital humano na gestão estratégica do conhecimento sob a perspectiva da teoria do agir comunicativo. 2003. Tese (Doutorado em Engenharia de Produção) - Universidade Federal de Santa Catarina, Florianópolis, 2003. HSM MANAGEMENT. Muito além de um modismo. Ano 4, n.22, set./out. 2000.

QUICK, M. Crime-general and crime-specific spatial patterns: A multivariate spatial analysis of four crime types at the small-area scale. *Journal of Criminal Justice*, 58, 2018, p. 22-32.

RAHM, E.; DO, H. H.. Data cleaning: Problems and current approaches. *IEEE Data Engineering Bulletin*, 23(4), 2000, p. 3 – 13. Disponível [https://www.betterevaluation.org/sites/default/files/data\\_cleaning.pdf](https://www.betterevaluation.org/sites/default/files/data_cleaning.pdf). Acesso em 24 jun. 2020.

RAJESWARI, C., BASU, D., MAURYA, N. Comparative Study of Big data Analytics Tools: R and Tableau. 2017. Disponível em: [https://www.researchgate.net/publication/321483078\\_Comparative\\_Study\\_of\\_Big\\_data\\_Analytics\\_Tools\\_R\\_and\\_Tableau](https://www.researchgate.net/publication/321483078_Comparative_Study_of_Big_data_Analytics_Tools_R_and_Tableau) . Acesso em: 24 jun. 2020.

RANDALL, S. M.; FERRANTE, A. M.; SEMMENS, J. The effect of data cleaning on record linkage quality. *BMC Medical Informatics and Decision Making*, v. 13, 06, 2013.  
SCARTON, C. E. *WordNet.Br*. 2010. Disponível em: <http://www.nilc.icmc.usp.br/wordnetbr/index.html>>. Acesso em: 30 jun. 2020.

RIO DE JANEIRO (Estado). Doutrina abril 2005. *DOCTRINA DE INTELIGÊNCIA DE SEGURANÇA PÚBLICA DO ESTADO DO RIO DE JANEIRO*. [S. l.], 2005.

SANTOS, Maria Isabel Araújo Silva dos. **A SEGURANÇA DO SEGREDO: PROPOSTA DE FRAMEWORK DE APLICAÇÃO DOS INSTRUMENTOS DE PROTEÇÃO DO SEGREDO NO AMBIENTE DE INOVAÇÃO DA BASE INDUSTRIAL DE DEFESA**. Orientador: Prof. Dr. Luiz Otávio Pimentel. 2016. 307 f. Tese (Doutorado) - PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E GESTÃO DO CONHECIMENTO, Florianópolis, 2016. Disponível em: <https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/178302/347163.pdf?sequence=1&isAllowed=y>. Acesso em: 13 jul. 2020.

SANTOS, Isabella; BARBOSA, Cátia. CIÊNCIA DA INFORMAÇÃO E INTELIGÊNCIA DE ESTADO: RELAÇÕES ENTRE A GESTÃO DA INFORMAÇÃO E DO CONHECIMENTO E A PROTEÇÃO DO CONHECIMENTO SENSÍVEL. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, XVII., 2016, Salvador, Bahia. DESCOBRIMENTOS DA CIÊNCIA DA INFORMAÇÃO: DESAFIOS DA MULTI, INTER E TRANSDISCIPLINARIDADE (MIT)*. Salvador, Bahia: 2016. Disponível em: [http://200.20.0.78/repositorios/bitstream/handle/123456789/3359/2016\\_GT4-CO\\_28.pdf?sequence=1](http://200.20.0.78/repositorios/bitstream/handle/123456789/3359/2016_GT4-CO_28.pdf?sequence=1). Acesso em: 13 jul. 2020.

SÊMOLA, Marcos. *Gestão da Segurança da Informação: uma visão executiva* – Ed. Campus, 2014.

SENGE, P. M. *A quinta disciplina: arte e prática da organização da aprendizagem*. São Paulo: Best Seller, 1990.

SHEPTYCKI, J. The Police intelligence Division-of-Labour. *Policing and Society*, 2017 Vol. 27 N. 6, 620-635.

SHÖN, D. *Theory in practice*. San Francisco: Jossey-Bass, 1983.

SILVA, Alexandra Valeria Vicente da. *A Polícia Militar do Estado do Rio de Janeiro e seus*

Mitos: repetindo, rompendo e recomeçando. *Cadernos de Segurança pública*, Rio de Janeiro, n. 04, 2013. Disponível em: <http://www.isprevista.rj.gov.br/download/Rev20130404.pdf>. Acesso em: 29 maio 2020.

SILVA, Francielle Amaral da; VIEIRA JÚNIOR, Nilson Carlos. **Manual para elaboração das ilustrações: tabelas, gráficos e figuras**. Foz do Iguaçu: Unila, 2014. Disponível em: <https://www.unila.edu.br/sites/default/files/files/Manual%20para%20Elaboracao%20das%20Ilustracoes%20Final.pdf>. Acesso em: 13 jul. 2020.

SOUZA, Anderson José de. Comparação entre abordagens de Drift Detection baseados em conjuntos 2013 de classificadores: um estudo de caso para previsão de crimes – 2013. 81 f. : il. ; 30 cm Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2013.

STEWART, T. A. *Capital intelectual: a nova vantagem competitiva das empresas*. Rio de Janeiro: Campus, 1998.

SVEIBY, K. E. *A nova riqueza das organizações*. Rio de Janeiro: Campus, 1998.

TESENG, Y.H. HO, Z.P., YANG, K.S. CHEN, C.C. Mining term Networks from Text Collections for Crime Investigation. *Expert Systems with Applications* 39, 2012, p. 10082-10090.

THIOLLENT, M. *Metodologia da pesquisa-ação*. 17 ed. São Paulo: Cortez. 2009.

TOGNOLLI, Claudio Julio *et al.* Saiba tudo sobre o guardião o software da PF que grampeou todos os corruptos: análise do cenário internacional e fatores determinantes. Claudio Tognolli: *Jornalismo e Informação*, São Paulo, 21 abr. 2018. Disponível em: <http://claudiotognolli.com.br/saiba-tudo-sobre-o-guardiao-o-software-da-pf-que-grampeou-todos-os-corruptos/>. Acesso em: 7 jul. 2020.

TOURE, I.; GANGOPADHYAY, A.. Real time big data analytics for predicting terrorist incidents. *2016 IEEE Symposium on Technologies for Homeland Security (HST)*, Waltham, MA, 2016, pp. 1-6, doi: 10.1109/THS.2016.7568906. Disponível em <https://ieeexplore.ieee.org/document/7568906>. Acesso em 02 jul. 2020.

VENTURA, Zuenir. *Cidade Partida*. 1. ed. [S. l.]: Companhia das letras, 1994.

VILAÇA, Márcio Luiz Corrêa. Pesquisa e ensino: considerações e reflexões. *Revista do Curso de Letras da Uniabeu, Nilópolis*, v. 1, n. 2, p.59-74, maio-ago 2010. Disponível em: . Acesso em: 13 jul. 2020.

VOMFELL, L. HÄRDLE, W.K., LESSMANN, S. Improving Crimes Count Forecasts using Twitter and Taxi data. *Decision Support System*. 113, 2018, p. 73-85.

WANG et al. Cognitive informatics and Computational Intelligence: From Information Revolution to Intelligence Revolution. In: INFORMATION RESOURCES MANAGEMENT ASSOCIATION. *Intelligent Systems: Concepts, Methodologies, Tools, and Applications*. Editora: Engineering Science Reference. Edição: 4 Volume Set, 2018, p. 278-296.

WANG, G.A.; CHEN, H., XU, J.J.; ATABAKHSH, H. Automatically Detecting Criminal Identify Deception: An Adaptive Detection Algorithm, IEEE Vol. 36 N.5 IEEE Transactions on System Man and Cybernetics, 2006.

WARDEN, J. Bi Intelligence for the Business of Crime Fighting, IEEE International Conference on Intelligence and Security Informatics, 2010 p. 159-161.

WU, H. C., LUK, R. W. P., WONG, K. F. & KWOK, K. L. Interpreting tf-idf term weights as making relevance decisions. ACM Transactions on Information Systems (TOIS), 26(3), 13, 2008. Disponível em: <<https://dl.acm.org/citation.cfm?id=1361686>>. Acesso em: 24 jun. 2018.

XU, J.J., QIN, G.W.Y., CHAU, M. Crime Data Mining: A general Framework and Some Examples, 2004.

XU, J.J., CHEN, H. Fighting organized crimes: using shortest path algorithms to identify associations in criminal networks, 2003. Decision Support Systems 38, 2004, p. 473-487. Yin, Robert. Case Study Research: Design and Methods. Thousand Oaks, London, New Delhi: Sage, 1994.

ZALUAR, A., LEAL, M.C. Violência extra e intramuros. Revista Brasileira Ciências Sociais Vol. 16 número 45 fevereiro/ /de 2001. Disponível em: <[http://www.scielo.br/scielo.php?pid=S0102-69092001000100008&script=sci\\_abstract&tlng=pt](http://www.scielo.br/scielo.php?pid=S0102-69092001000100008&script=sci_abstract&tlng=pt)>. Acesso em: 29 nov. 2018.

ZALUAR, Alba; CONCEIÇÃO, Isabel Siqueira. Favelas sob o controle das Milícias no Rio de Janeiro que paz? São Paulo em Perspectiva, v. 21, n. 2, p. 89-101, jul./dez. 2007 . Disponível em [http://produtos.seade.gov.br/produtos/spp/v21n02/v21n02\\_08.pdf](http://produtos.seade.gov.br/produtos/spp/v21n02/v21n02_08.pdf) Acesso em 24 jun 2020.

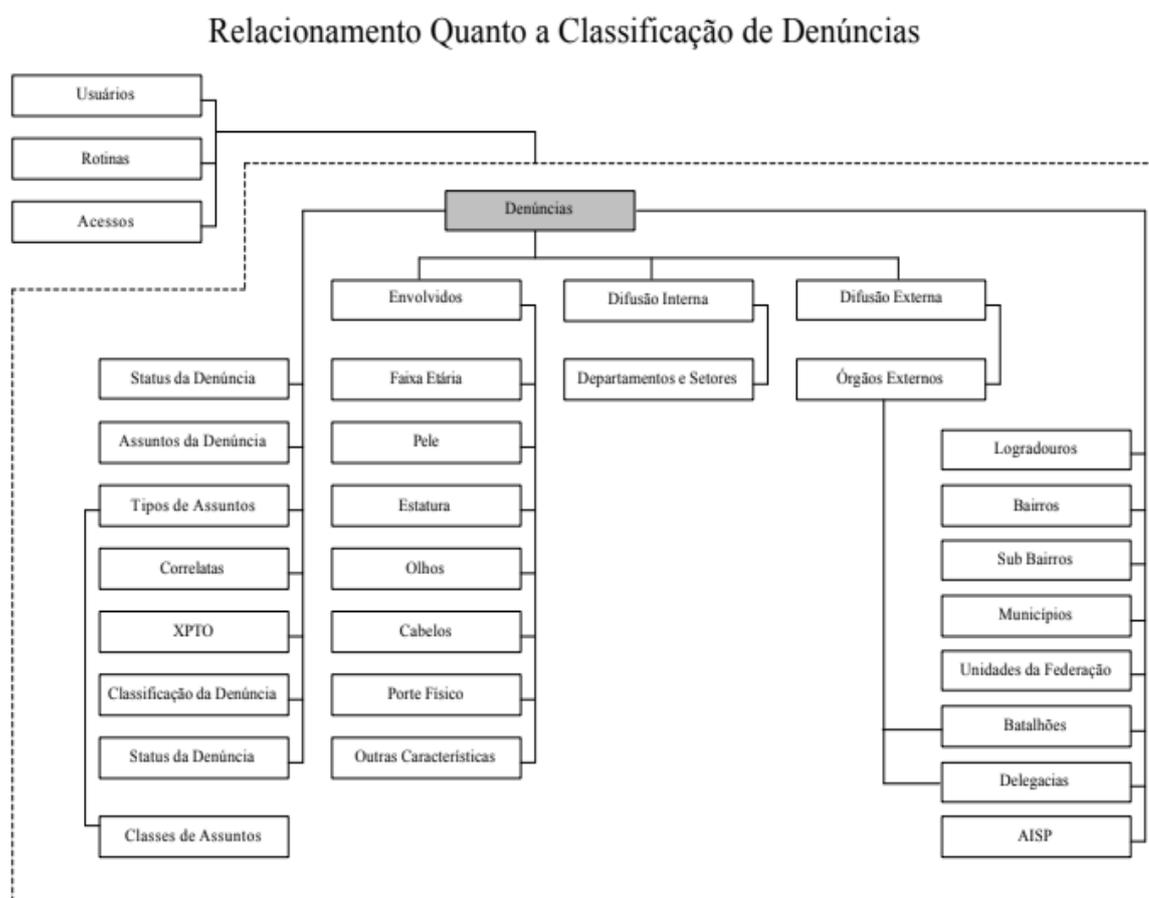
ZANASI, A. Discovering Data Mining. Prentice Hall, 1997.

## APÊNDICE A - PILARES DO SERVIÇO DO DISQUE DENÚNCIA



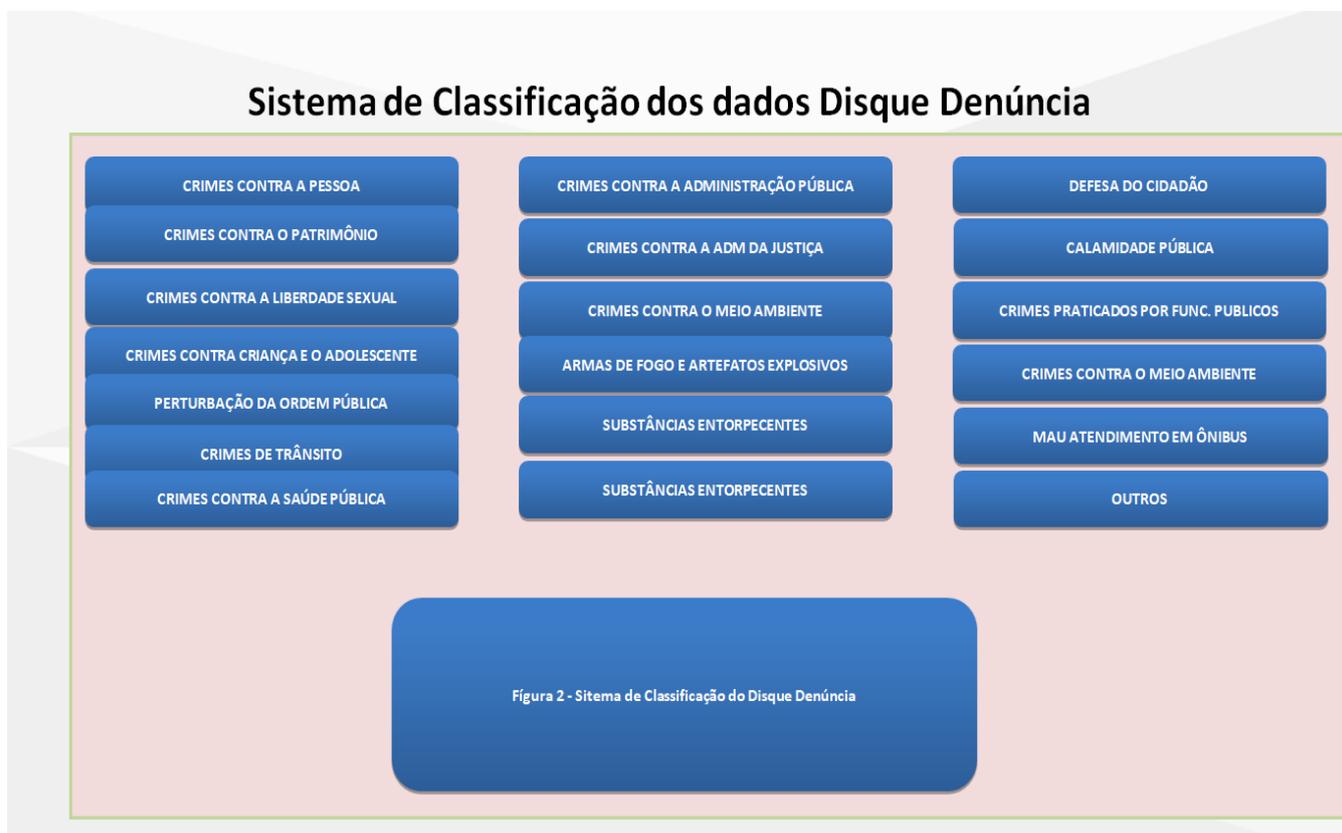
Fonte: elaboração do autor.

## APÊNDICE B - RELAÇÃO QUANTO A CLASSIFICAÇÃO DE DENÚNCIAS



Fonte: elaboração própria.

## APÊNDICE C – SISTEMA DE CLASSIFICAÇÃO DO DISQUE DENÚNCIA



Fonte: elaboração do autor

**APÊNDICE D - QUADRO DE DENÚNCIAS CADASTRADAS, QUE CONTÉM O XPTO “MILÍCIA” (OU A ESCRITA “MILÍCIA”) NO RELATO, DE ACORDO COM OS DEZ TIPOS DE ASSUNTO MAIS DENUNCIADOS, NOS ANOS DE 2003 A 2008**

<u>2003 – Denúncias por tipo de assunto (qtd de denúncias)</u>	<u>2004 – Denúncias por tipo de assunto (qtd de denúncias)</u>	<u>2005 – Denúncias por tipo de assunto (qtd de denúncias)</u>
<u>Concussão (15)</u>	<u>Concussão (5)</u>	<u>Maus tratos a menores (2)</u>
<u>Desvio de conduta (8)</u>	<u>Desvio de conduta (4)</u>	<u>Veículos abandonados (2)</u>
<u>Tráfico de Drogas (6)</u>	<u>Tráfico de Drogas (2)</u>	<u>Tráfico de drogas (2)</u>
<u>Homicídio consumado (4)</u>	<u>Roubo/furto a transeuntes (2)</u>	<u>Receptação/comercialização de produtos roubados (1)</u>
<u>Corrupção (3)</u>	<u>Violência à mulher (1)</u>	<u>Grupos de extermínio (1)</u>
<u>Ameaça (2)</u>	<u>Estelionato (1)</u>	<u>Desvio de conduta (1)</u>
<u>Abuso de autoridade (2)</u>	<u>Abuso de autoridade (1)</u>	<u>Furto de veículos (1)</u>
<u>Tentativa de homicídio (2)</u>	<u>Extorsão simples (1)</u>	
<u>Barulho (1)</u>	<u>Tráfico de Drogas (1)</u>	
<u>Criança e adolescente infrator (1)<sup>18</sup></u>	<u>Obras audiovisuais/literárias/artísticas falsificadas (1)</u>	

Fonte Elaboração própria com base nos dados do Disque Denúncia / RJ

**Quadro De Denúncias Cadastradas, Que Contém O Xpto “Milícia” (Ou A Escrita “Milícia”) No Relato, De Acordo Com Os Dez Tipos De Assunto Mais Denunciados, Nos Anos De 2006 A 2008**

<u>2006 – denúncias por tipo de assunto (qtd de denúncias)</u>	<u>2007 – denúncias por tipo de assunto (qtd de denúncias)</u>	<u>2008 – denúncias por tipo de assunto (qtd de denúncias)</u>
<u>Desvio de conduta (28)</u>	<u>Extorsão simples (661)</u>	<u>Extorsão simples (1920)</u>
<u>Extorsão simples (18)</u>	<u>Desvio de conduta (404)</u>	<u>Desvio de conduta (397)</u>
<u>Tráfico de drogas (17)</u>	<u>Tráfico de drogas (232)</u>	<u>Homicídio consumado (304)</u>
<u>Grupos de extermínio (12)</u>	<u>Homicídio consumado (201)</u>	<u>Posse ilícita de armas de fogo (253)</u>

<sup>18</sup> além desses 10 assuntos, no ano de 2003 somam-se também Roubo de veículos; Posse ilícita de armas de fogo; Pessoas desaparecidas; Maus tratos a menores; Mau atendimento em órgãos públicos e privados, com 01 denúncia cada.

<u>Concussão (5)</u>	<u>Concussão (118)</u>	<u>Tráfico de drogas (251)</u>
<u>Jogos de azar (2)</u>	<u>Posse ilícita de armas de fogo (92)</u>	<u>Ameaça (169)</u>
<u>Sequestro e cárcere privado (2)</u>	<u>Ameaça (89)</u>	<u>Rádio/tv/telefonía clandestina (168)</u>
<u>Corrupção (2)</u>	<u>Grupos de extermínio (79)</u>	<u>Concussão (155)</u>
<u>Posse ilícita de armas de fogo (2)</u>	<u>Rádio/tv/telefonía clandestina (55)</u>	<u>Localização de foragidos da justiça (93)</u>
<u>Guarda/comércio de animais silvestres (1)</u> <sup>19</sup>	<u>Obstrução de vias públicas (41)</u> <sup>6</sup>	<u>Grupos de extermínio (77)</u> <sup>7</sup>

Fonte Elaboração própria com base nos dados do Disque Denúncia / RJ

<sup>19</sup> No ano de 2006 foram cadastradas 95 denúncias com os parâmetros analisados. Além dessas denúncias apontadas na tabela, existem outras 06 cadastradas, dispersas em 06 assuntos diferentes, que receberam 01 denúncia cada.<sup>6</sup> No ano de 2007 foram cadastradas 2373 denúncias com os parâmetros analisados. Os quantitativos demonstrados acima representam apenas um recorte acerca dos 10 assuntos mais denunciados nesse período com base nos parâmetros filtrados.

<sup>7</sup>No ano de 2008 foram cadastradas 4663 denúncias com os parâmetros analisados. Os quantitativos demonstrados acima representam apenas um recorte acerca dos 10 assuntos mais denunciados nesse período com base nos parâmetros filtrados.

## APÊNDICE E - TIPO DE ERRO PARA OS OUTROS MUNICÍPIOS: ARMAÇÃO DOS BÚZIOS E CABO FRIO

### ARMAÇÃO DOS BUZIOS

Bairro Corrigido	Den Logr Bairro
ALTOS DE BUZIOS	ALTOS DE BUZIOS
BOSQUE	BOSQUE
BUZIOS	BUZIOS
	Búzios
CAMPO DE POUSO	CAMPO DE POUSO
CAPAO	CAPAO
CEM BRACAS	C BRACAS
	CEM BRACA
	CEM BRACAS
	CEM BRAÇAS
	CEM BRACOS
	CEM BRAÇOS
	CEM PRAÇAS
	SEM BRAÇA
	SEM BRACAS
CENTRO	CENTRO
GERIBA	Geribá
	GERIBÁ
	GERIBA
	JERIBA
JOSE GONCALVES	JOSE GONÇALVES
	JOSE GONCALVES
MANGUINHOS	MANGUINHO
	MANGUINHOS
	MANGUNHOS
RASA	P RASA
	PR RASA
	Rasa
	RASA
VILA LUIZA	VILA LUIZA
VILLAGE DE BUZIOS	VILAGE DE BUZIOS
	VILAGE DE BÚZIOS
	VILLAGE DE BUZIOS

### CABO FRIO

Bairro Corrigido	Den Logr Bairro
ALGODOAL	ALGODOAL
	ALGODUAL
BRAGA	BRAGA
CAJUEIRO	Cajueiro
	CAJUEIRO
CAMINHO DE BUZIOS	CAMINHO DE BÚZIOS
	CAMINHO DE BUZIOS
	CAMINHO E BUZIOS
CAMINHO VERDE	CAM VERDE
	CAMINHO VERDE
CENTRO	CENTRINHO
	CENTRO
DUNAS DO PERO	DUNAS DE PERO
	DUNAS DO PERO
FOGUETE	FOGUETE
	PR FOGUETE
GAMBOA	GAMBOA
GUARANI	GUARANI
ITAJURU	ITAJURU
JACARE	JACARE
	JACARÉ
JARDIM CAICARA	Jardim Caiçara
	JARDIM CAICARA
	JARDIM CAIÇARA
JARDIM ESPERANCA	JARDIM ESPERANACA
	JARDIM ESPERANÇA
	JARDIM ESPERANCA
	KARDIM ESPERANÇA
JARDIM EXCELCIOR	JARDIM EXCELCIOR
	JARDIM EXCELSIOR
JARDIM NAUTILUS	JARDIM NAUTILUS
	JARDIM NÁUTILUS
	JARDIM NAUTILUS

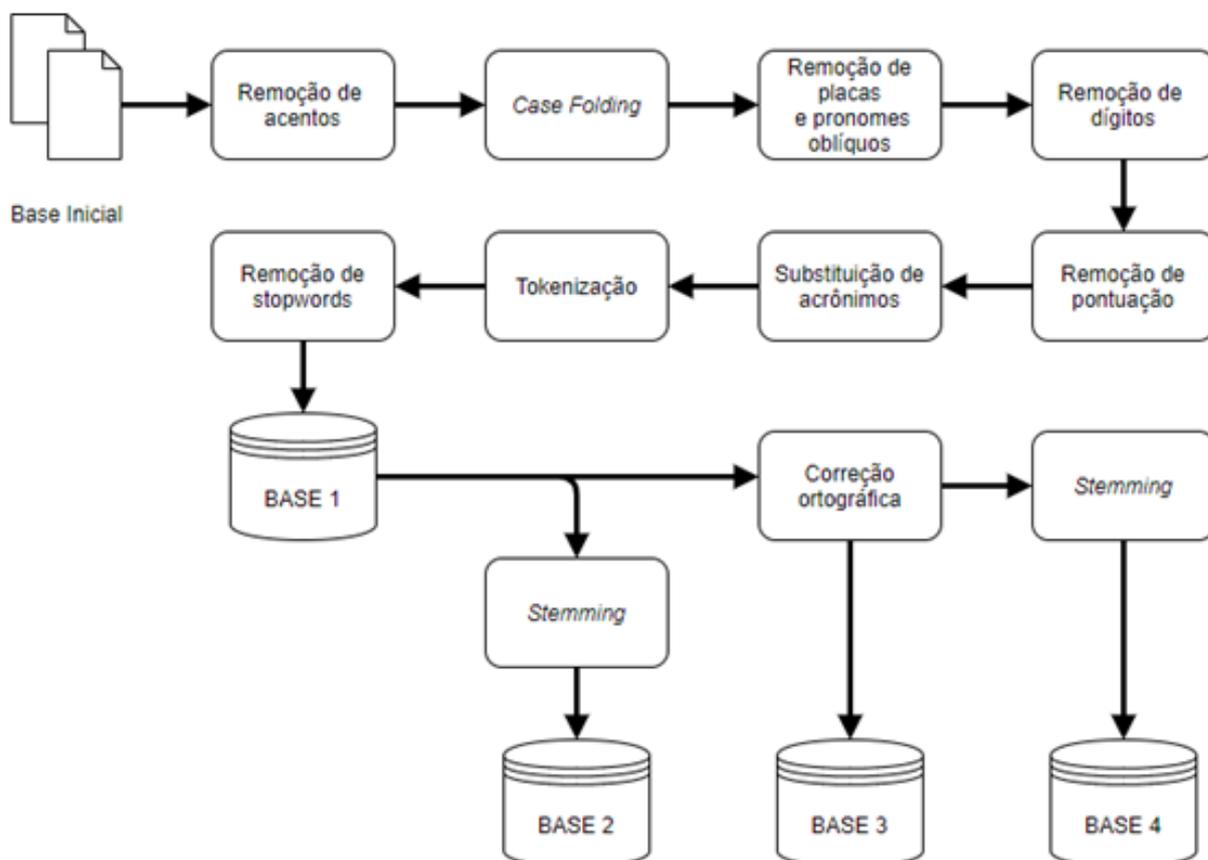
Fonte: figura adaptada pelo autor.

### APÊNDICE F - TIPO DE ERRO PARA O BAIRRO VILA KENNEDY

Den Logr Bairro	Den Dt Rec											Total ..
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	
VILA KENNEDY		2	2	20	16	5	64	122	264	210	22	727
VL KENNEDY	10	58	56	21	31	10	15	4	1	3	2	211
V KENNEDY	5	9	18	16	25	10						83
VL KENEDY		11	18	10			5			1		45
VILA KENEDY					3			5	5		4	17
V KENEDY		10	5									15
V KENNEY			11									11
V KENEDDY		1		2	4		2					9
VILA KENEDDY		2			4					3		9
V KENEDI			4			1						5
VILA KENNEDY								2	2			4
VL KENNIDY			4									4
VILA KENNEY					3							3
VL KENNEDY			2		1							3
VL KENEDI		2	1									3
VL KENNEYD						3						3
V KENNEDI			2									2
VILA PRESIDENTE KENNED..		2										2
VILLA KENNEDY									2			2
VL KENEDY			2									2
VL KENEDDI			1		1							2
VL KENEDDY			2									2
VL KENNEY				2								2
VL KENNED			2									2
VL KENNEY			2									2
VILA KENNEDDY											1	1
VILA KENNEDUY								1				1
VL KENENY				1								1
VL KENNEDI			1									1
VL KENNYDI			1									1
Total geral	15	97	134	72	88	29	86	134	274	217	29	1.175

Fonte:figura adaptada pelo autor.

**APÊNDICE G - ESTRUTURA E ETAPAS DE *DATA CLEANING* PARA PRÉ-PROCESSAMENTO E GERAÇÃO DE BASES UTILIZADAS NO MODELO**



Fonte: Figura adaptada pelo autor.

## APÊNDICE H - *DATA CLEANING*: IDENTIFICAÇÃO E CORREÇÕES

den_logr_uf	den_logr_mun	den_logr_tp	den_logr_ds	den_logr_num	den_logr_bairro	den_texto	em	
OGAS	RJ	JAPERI	R	GENEZIO PEREIRA VILELA		Não Informado	INFORMA QUE NA RUA MENCIONADA, EM FRENTE AO ...	De
OGAS	RJ	RIO DE JANEIRO	R	IMBE		COELHO NETO	INFORMA QUE NA ESQUINA DA RUA CITADA, COM A R...	
OGAS	RJ	RIO DE JANEIRO	R	PEDRO LABATUT	66	HONORIO GURGEL	INFORMA QUE NO ENDEREÇO MENCIONADO, PRÓXIM...	JAT
OGAS	RJ	RIO DE JANEIRO	R	PEDRO LABATUT	66	HONORIO GURGEL	INFORMA QUE NO ENDEREÇO MENCIONADO, PRÓXIM...	JAT
OGAS	RJ	RIO DE JANEIRO	R	PEDRO LABATUT	66	HONORIO GURGEL	INFORMA QUE NO ENDEREÇO MENCIONADO, PRÓXIM...	NIL
OGAS	RJ	RIO DE JANEIRO	R	PEDRO LABATUT	66	HONORIO GURGEL	INFORMA QUE NO ENDEREÇO MENCIONADO, PRÓXIM...	NIL
OGAS	RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	CA
OGAS	RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	CA
OGAS	RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	FRJ
OGAS	RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	FRJ
OGAS	RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	JO5
OGAS	RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	JO5
OGAS	RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	MÁ
OGAS	RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	MÁ
OGAS	RJ	RIO DE JANEIRO	R	SERVIDAO		RECREIO DOS BANDEIRANTES	INFORMA QUE NA RUA MENCIONADA, DENTRO DA FA...	De
OGAS	RJ	RIO DE JANEIRO	R	MOREIRA PINTO	144	SANTO CRISTO	INFORMA QUE NO ENDEREÇO CITADO, FUNDOS DE U...	REI
OGAS	RJ	NITEROI	TV	SEIS	21	IA CONCEICAO	INFORMA QUE NO ENDEREÇO CITADO, NO MORRO D...	ALU
OGAS	RJ	NITEROI	TV	SEIS	21	IA CONCEICAO	INFORMA QUE NO ENDEREÇO CITADO, NO MORRO D...	CR
OGAS	RJ	RIO DE JANEIRO	N.I.	37ª DELEGACIA POLICIAL		ILHA DO GOVERNADOR	INFORMA QUE NA DELEGACIA CITADA, ENCONTRA-SE...	CA

Fonte: Figura adaptada pelo autor.

## APÊNDICE I - TRANSFORMAÇÃO DOS CARACTERES EM MAIÚSCULOS

den_logr_uf	den_logr_mun	den_logr_tp	den_logr_ds	den_logr_num	den_logr_bairro	den_texto	em
OGAS RJ	JAPERI	R	GENEZIO PEREIRA VILELA		Não Informado	INFORMA QUE NA RUA MENCIONADA, EM FRENTE AO ...	De
OGAS RJ	RIO DE JANEIRO	R	IMBE		COELHO NETO	INFORMA QUE NA ESQUINA DA RUA CITADA, COM A R...	
OGAS RJ	RIO DE JANEIRO	R	PEDRO LABATUT	66	HONORIO GURGEL	INFORMA QUE NO ENDEREÇO MENCIONADO, PRÓXIM...	JAT
OGAS RJ	RIO DE JANEIRO	R	PEDRO LABATUT	66	HONORIO GURGEL	INFORMA QUE NO ENDEREÇO MENCIONADO, PRÓXIM...	JAT
OGAS RJ	RIO DE JANEIRO	R	PEDRO LABATUT	66	HONORIO GURGEL	INFORMA QUE NO ENDEREÇO MENCIONADO, PRÓXIM...	NIL
OGAS RJ	RIO DE JANEIRO	R	PEDRO LABATUT	66	HONORIO GURGEL	INFORMA QUE NO ENDEREÇO MENCIONADO, PRÓXIM...	NIL
OGAS RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	CA
OGAS RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	CA
OGAS RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	FR
OGAS RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	FR
OGAS RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	JO
OGAS RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	JO
OGAS RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	MA
OGAS RJ	RIO DE JANEIRO	FV	TIMBAU		RAMOS	INFORMA QUE O TRAFICANTE "CARLOS ANDRÉ", VUL...	MA
OGAS RJ	RIO DE JANEIRO	R	SERVIDAO		RECREIO DOS BANDEIRANTES	INFORMA QUE NA RUA MENCIONADA, DENTRO DA FA...	De
OGAS RJ	RIO DE JANEIRO	R	MOREIRA PINTO	144	SANTO CRISTO	INFORMA QUE NO ENDEREÇO CITADO, FUNDOS DE U...	REI
OGAS RJ	NITEROI	TV	SEIS	21	IA CONCEICAO	INFORMA QUE NO ENDEREÇO CITADO, NO MORRO D...	ALI
OGAS RJ	NITEROI	TV	SEIS	21	IA CONCEICAO	INFORMA QUE NO ENDEREÇO CITADO, NO MORRO D...	CR
OGAS RJ	RIO DE JANEIRO	N.I.	37ª DELEGACIA POLICIAL		ILHA DO GOVERNADOR	INFORMA QUE NA DELEGACIA CITADA, ENCONTRA-SE...	CA

Fonte: Figura adaptada pelo autor.

## APÊNDICE J - RETIRANDO ACENTUAÇÃO

```
# limpando string
drug_traf$den_logr_bairro = toupper(drug_traf$den_logr_bairro)
drug_traf$den_logr_bairro = gsub("Ê", "E", drug_traf$den_logr_bairro, fixed = T)
drug_traf$den_logr_bairro = gsub("Í", "I", drug_traf$den_logr_bairro, fixed = T)
drug_traf$den_logr_bairro = gsub("Ç", "C", drug_traf$den_logr_bairro, fixed = T)
drug_traf$den_logr_bairro = gsub("Ó", "O", drug_traf$den_logr_bairro, fixed = T)
drug_traf$den_logr_bairro = gsub("Õ", "O", drug_traf$den_logr_bairro, fixed = T)
drug_traf$den_logr_bairro = gsub("Â", "A", drug_traf$den_logr_bairro, fixed = T)
drug_traf$den_logr_bairro = gsub("Á", "A", drug_traf$den_logr_bairro, fixed = T)
drug_traf$den_logr_bairro = gsub("Ú", "U", drug_traf$den_logr_bairro, fixed = T)
drug_traf$den_logr_bairro = gsub("É", "E", drug_traf$den_logr_bairro, fixed = T)
drug_traf$den_logr_bairro = gsub("Ã", "A", drug_traf$den_logr_bairro, fixed = T)
```

Fonte: Figura adaptada pelo autor.

## APÊNDICE K - CORRIGINDO ERROS IDENTIFICADOS COM AS MAIORES FREQÜÊNCIAS NA BASE

```

94 drug_traf$den_logr_bairro = gsub("BRAZ", "BRAS", drug_traf$den_logr_bairro, fixed = T)
95 drug_traf$den_logr_bairro = gsub("JD", "JARDIM", drug_traf$den_logr_bairro, fixed = T)
96 drug_traf$den_logr_bairro = gsub("PRQ", "PARQUE", drug_traf$den_logr_bairro, fixed = T)
97 drug_traf$den_logr_bairro = gsub("VL", "VILA", drug_traf$den_logr_bairro, fixed = T)
98 drug_traf$den_logr_bairro = gsub("3", "III", drug_traf$den_logr_bairro, fixed = T)
99 drug_traf$den_logr_bairro = gsub("2", "II", drug_traf$den_logr_bairro, fixed = T)
100 drug_traf$den_logr_bairro = gsub("BOAVISTA", "BOA VISTA", drug_traf$den_logr_bairro, fixed = T)
101 drug_traf$den_logr_bairro = gsub("BRAS DE P INA", "BRAS DE PINA", drug_traf$den_logr_bairro, fixed = T)
102 drug_traf$den_logr_bairro = gsub("CABOCU", "CABUCU", drug_traf$den_logr_bairro, fixed = T)
103 drug_traf$den_logr_bairro = gsub("CMPINHO", "CAMPINHO", drug_traf$den_logr_bairro, fixed = T)
104 drug_traf$den_logr_bairro = gsub("CORIA", "COREIA", drug_traf$den_logr_bairro, fixed = T)
105 drug_traf$den_logr_bairro = gsub("CUMPRIDO", "COMPRIDO", drug_traf$den_logr_bairro, fixed = T)
106 drug_traf$den_logr_bairro = gsub("EST ", "ESTRADA ", drug_traf$den_logr_bairro, fixed = T)
107 drug_traf$den_logr_bairro = gsub("ESTR ", "ESTRADA ", drug_traf$den_logr_bairro, fixed = T)
108 drug_traf$den_logr_bairro = gsub("GAO PARA", "GRAO PARA", drug_traf$den_logr_bairro, fixed = T)
109 drug_traf$den_logr_bairro = gsub("GRANDE CABUCU", "CABUCU", drug_traf$den_logr_bairro, fixed = T)
110 drug_traf$den_logr_bairro = gsub("H GUEGEL", "HONORIO GURGEL", drug_traf$den_logr_bairro, fixed = T)

```

Fonte: Figura adaptada pelo autor.

## APÊNDICE L - IDENTIFICAÇÃO REALIZANDO CORREÇÃO DOS BAIRROS

	V1
1	Não Informa
2	COELHO NETO
3	HONORIO GURGEL
4	RAMOS
5	RECREIO DOS BANDEIRANTES
6	SANTO CRISTO
7	IA CONCEICAO
8	ILHA DO GOVERNADOR
9	REALENGO
10	GUADALUPE
11	LAGOINHA
12	RIO COMPRIDO
13	VL SAPE
14	VL BRASILIA
15	VL MARTINS
16	FREGUESIA
17	BONSUCESSO
18	COPACABANA
19	CHS ARCAMPO
20	S AI FIXO

Showing 1 to 21 of 8,799 entries, 1 total columns

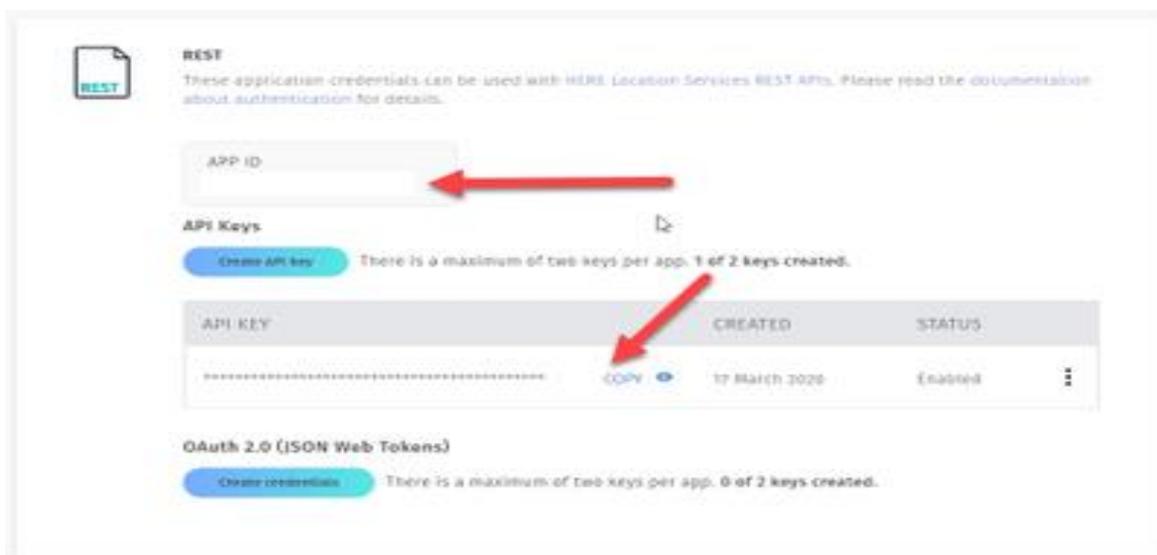
Fonte:figura adaptada pelo autor.

### APÊNDICE M - BASE NOVA ATUALIZADA E CORRIGIDA

den_logr_mun	den_logr_tp	den_logr_ds	den_logr_num	den_logr_bairro
NITEROI	TV	SEIS	21	ILHA DA CONCEICAO
RIO DE JANEIRO	N.I.	37ª DELEGACIA POLICIAL		ILHA DO GOVERNADOR
RIO DE JANEIRO	R	OLAVO SOUSA AGUIAR		REALENGO
RIO DE JANEIRO	R	MAIAMA		GUADALUPE
NOVA IGUACU	R	O		LAGOINHA
RIO DE JANEIRO	R	COSTA FERRAZ		RIO COMPRIDO
RIO DE JANEIRO	EST	ENGENHO NOVO, DO		REALENGO
DUQUE DE CAXIAS	R	AFONSO PENA		VILA SAPE
VOLTA REDONDA	TV	DEOLINDO MIGUEL	896	VILA BRASILIA
RIO DE JANEIRO	R	MARQUES DE MURITIBA		ILHA DO GOVERNADOR
QUEIMADOS	R	ALVARO MARTINS		VILA MARTINS
RIO DE JANEIRO	EST	TRES RIOS, DOS	480	FREGUESIA (JACAREPAGUA)
RIO DE JANEIRO	R	IVANILDO ALVES		BONSUCESSO
DUQUE DE CAXIAS	R	LEME		COPACABANA
DUQUE DE CAXIAS	AV	LARANJAL, DO		CHACARAS ARCAMPO
MAGE	R	DUAS BARRAS		SANTO ALEIXO
RIO DE JANEIRO	R	GUAXINDIBA	165	COELHO NETO
RIO DE JANEIRO	R	ALZIRA ROSA DE JESUS		PACIENCIA
RIO DE JANEIRO	R	GALDINO PIMENTEL	50	MEIER

Fonte: Figura adaptada pelo autor.

## APÊNDICE N - *HERE* CONFIGURAÇÃO ID DO APLICATIVO E A API KEY



**REST**  
These application credentials can be used with HERE Location Services REST APIs. Please read the documentation about authentication for details.

APP ID

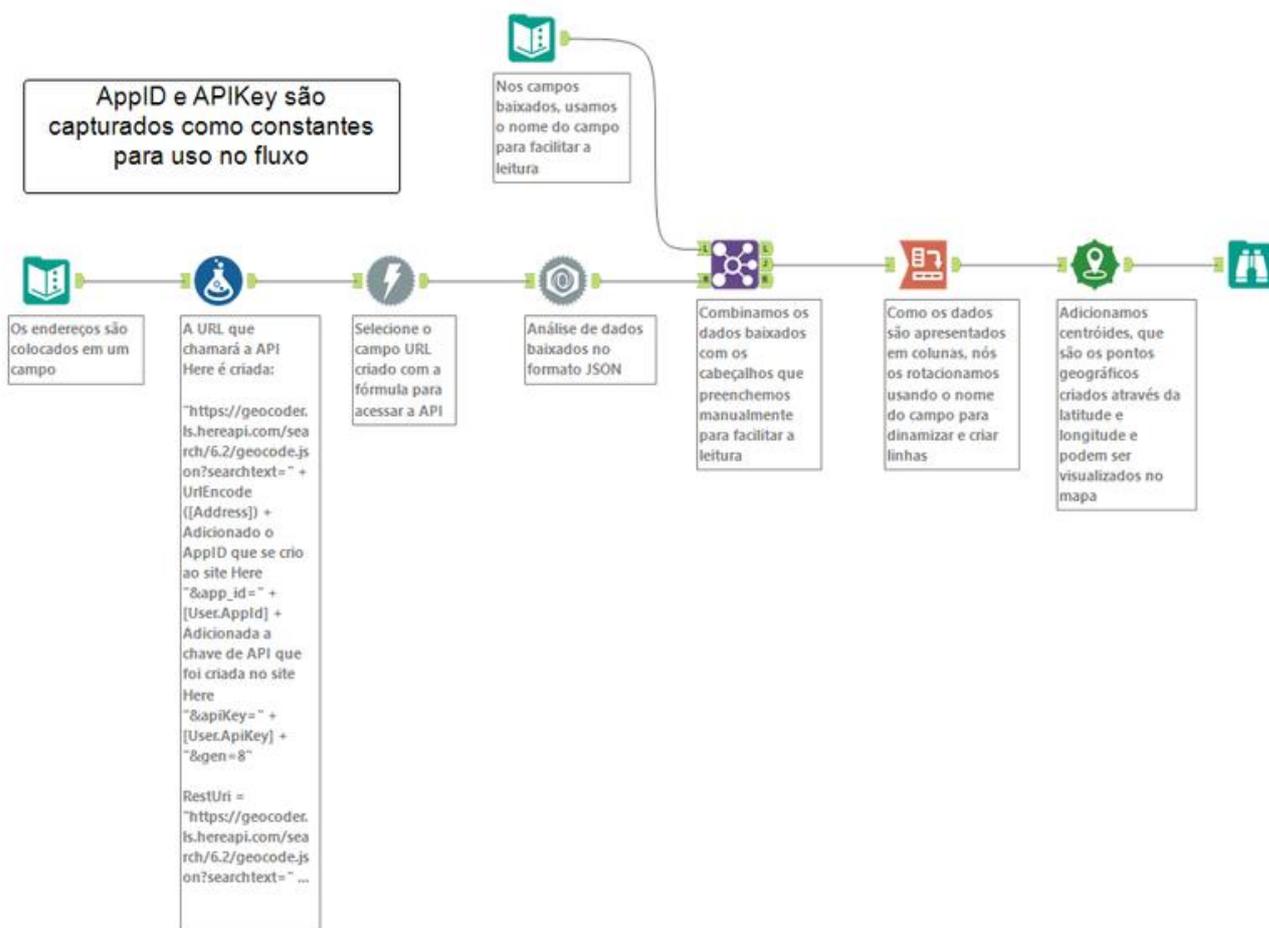
**API Keys**  
Create API key There is a maximum of two keys per app. 1 of 2 keys created.

API KEY	CREATED	STATUS
.....	17 March 2020	Enabled

**OAuth 2.0 (JSON Web Tokens)**  
Create credentials There is a maximum of two keys per app. 0 of 2 keys created.

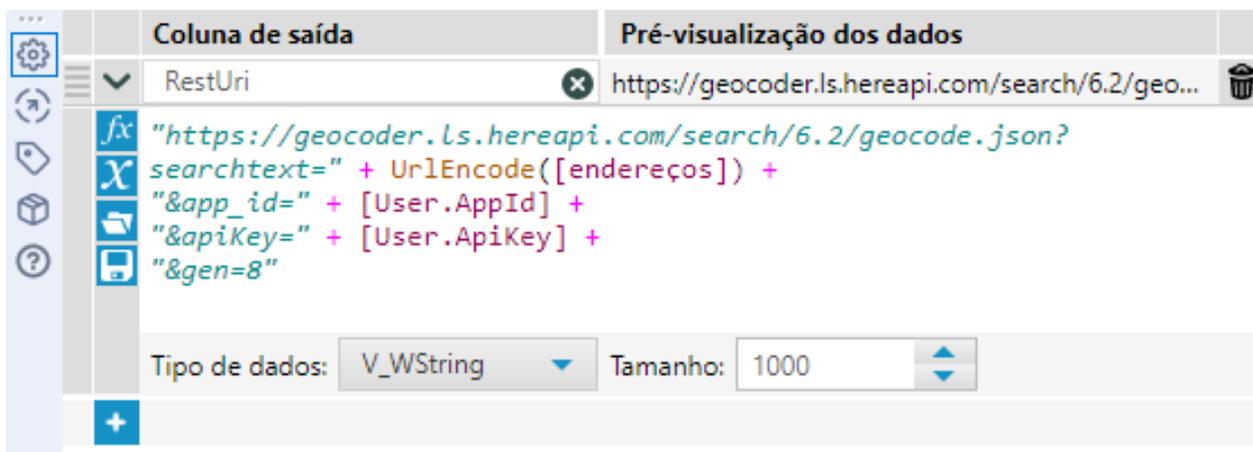
Fonte: Figura adaptada pelo autor.

## APÊNDICE O - FLUXO CRIADO E UTILIZADO NO PROCESSO DO ALTERYX



Fonte: Figura adaptada pelo autor.

## APÊNDICE P - PADRONIZAÇÃO DA SAÍDA E VISUALIZAÇÃO DOS DADOS FINAIS



Fonte: Figura adaptada pelo autor.

## APÊNDICE Q - FLUXO DE TRABALHO CONFIGURAÇÃO DA APLICAÇÃO APP ID E O APIKEY

Tela Fluxo de trabalho Tempo de execução Eventos Metainformações

Caminho: C:\Alteryx\Samples\Spatial\Here Geocoding\HERE Geocoding v1 PT.yxmd

Tipo

Fluxo de trabalho padrão

Aplicativo analítico

Macro: Macro padrão

Constantes:

	Tipo	Nome	Valor	#
▶	Engine	TempFilePath	C:\ProgramData\Alteryx\Engine	<input type="checkbox"/>
	Engine	Version	2020.1.1.23790	<input type="checkbox"/>
	Engine	WorkflowDirectory	C:\Alteryx\Samples\Spatial\Here Geocoding	<input type="checkbox"/>
	Engine	WorkflowFileName	HERE Geocoding v1 PT.yxmd	<input type="checkbox"/>
	Engine	GuiInteraction	True	<input type="checkbox"/>
	User	AppId	AppID	<input type="checkbox"/>
	User	ApiKey	API Key	<input type="checkbox"/>

Fonte: Figura adaptada pelo autor.

## APÊNDICE R - API GOOGLE UTILIZADA NO PROCESSO

```

        $prepAddr = str_replace(' ', '+', $address);

        $geocode=@file_get_contents
('https://maps.google.com/maps/api/geocode/json?
key=AIzaSyB7VHPwcbE3iYG8LbKIZ_Md7lv3NSUwWSI&address=.'.$prepAddr.'&sensor=false');

        $output= json_decode($geocode);

        try{
            @$latitude = $output->results[0]->geometry->location->lat;
            @$longitude = $output->results[0]->geometry->location-
>lng;

        }catch(Exception $e){
            echo 'Exceção capturada: ', $e->getMessage(), "\n";
            continue;
        }

        $sqlUpdateLatLong = "update denuncia set den_gps_lat='".
$latitude."', den_gps_long='". $longitude.'" where den_cd = ".$fetchGetDenuncias["den_cd"];
        $resUpdateLatLong = mssql_query($sqlUpdateLatLong);

```

Fonte: Figura adaptada pelo autor.

## APÊNDICE S - JSON COM LATITUDE E LONGITUDE

```
[{"den_cd":2084909,"latitude":null,"longitude":null}, {"den_cd":2084908,"latitude":null,"longitude":null}, {"den_cd":2084907}, {"den_cd":2084903,"latitude":null,"longitude":null}, {"den_cd":2084902,"latitude":null,"longitude":null}, {"den_cd":2084901,'
```

Fonte: figura adaptada pelo autor.