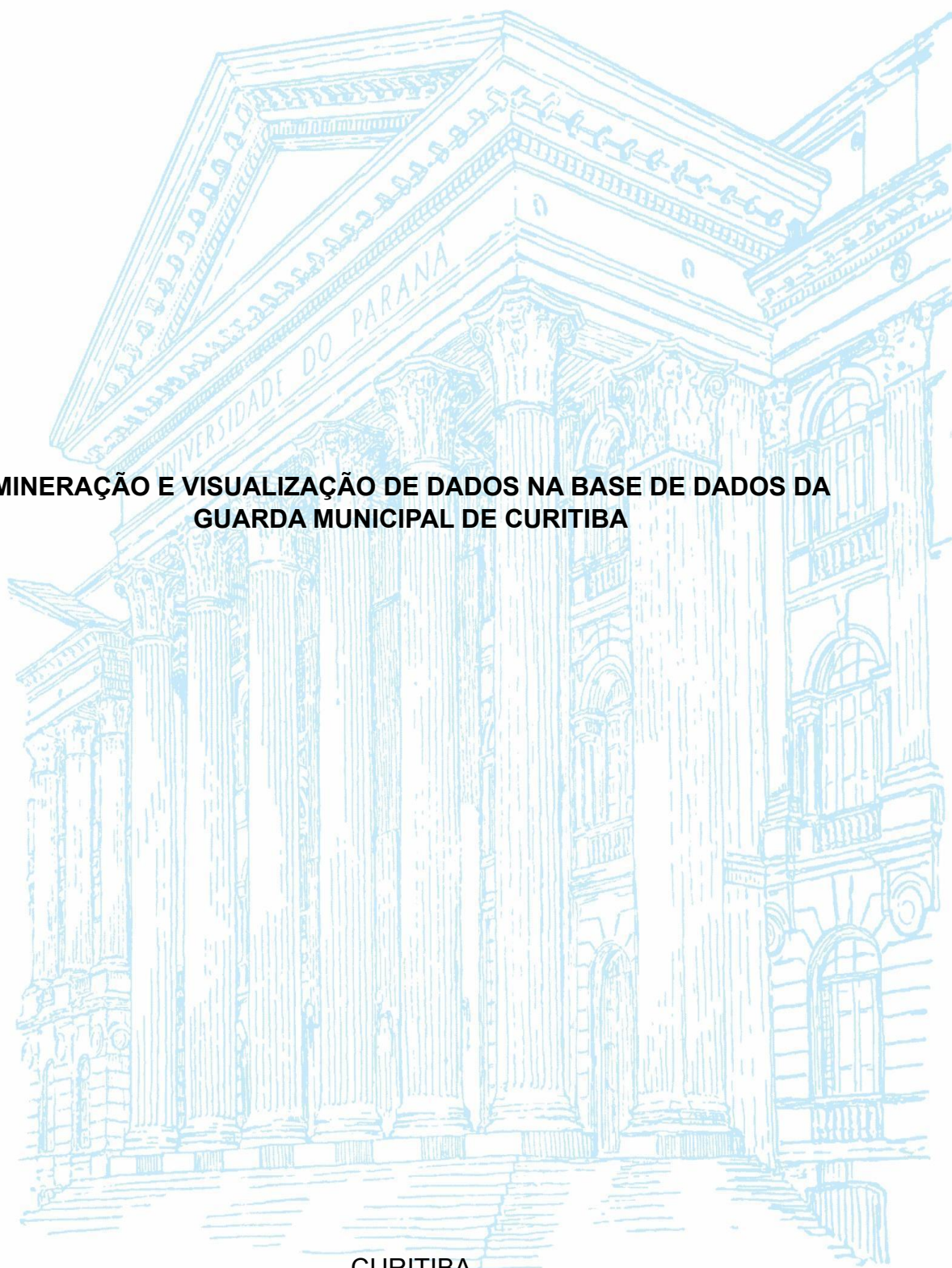


UNIVERSIDADE FEDERAL DO PARANÁ

ENZO ZIETEK SILOTO

**MINERAÇÃO E VISUALIZAÇÃO DE DADOS NA BASE DE DADOS DA
GUARDA MUNICIPAL DE CURITIBA**



CURITIBA

2022

ENZO ZIETEK SILOTO

**MINERAÇÃO E VISUALIZAÇÃO DE DADOS NA BASE DE DADOS DA
GUARDA MUNICIPAL DE CURITIBA**

Trabalho de Conclusão de Curso apresentado ao curso de Gestão da Informação, do Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Gestão da Informação.

Orientadora: Prof. Dr. Luciano Heitor Gallegos Marin

CURITIBA

2022

RESUMO

Nos tempos atuais onde a informação tomou um valor significativo a todos, o presente trabalho vem com a premissa de exemplificar uma análise para geração de informação e conhecimento fazendo o uso de técnicas de análise de dados na base de dados da SiGesGuarda responsável por contém os dados correspondentes às ocorrências atendidas pela Guarda Municipal de Curitiba, sendo realizado processos de predição por meio de mineração de dados através da ocorrência obtidas no período de 2009 até junho de 2022. Como norteador do processo foi utilizado a Knowledge Discovery in Databases (KDD) onde na sua na etapa de mineração de dados foram utilizados os algoritmos Naïve Bayes, Deep Learning e Decision Tree para a realização da análise da base de dados. Em relação aos resultados obtidos o modelo Naive Bayes obteve uma taxa de acerto de 36.2%, já em comparação o modelo Deep Learning obteve uma taxa de acerto de 36.9% e por fim o modelos Decision Tree apresentou uma taxa de acerto de 29.5%. Foi trazido como objetivo a identificação de padrões presentes na base de dados da Guarda Municipal de Curitiba, como forma de possibilitar a descoberta de informação e conhecimento, obtidas através das análise e desenvolvimento de estatística descritiva, já para a geração de conhecimento foi realizado através do processamento dos algoritmos e da manipulação e otimização dos resultados através dos resultados obtidos e simulador.

Palavras-chave: Análise de Dados. Guarda Municipal de Curitiba. Mineração de Dados. Naïve Bayes. *Deep Learning*. *Decision Tree*. *RapidMiner*.

ABSTRACT

In current times where the information has taken a significant value to everyone, this work comes with the premise of exemplifying an analysis for generating information and knowledge making use of data analysis techniques in the database of SiGesGuarda responsible for containing the data corresponding to the occurrences attended by the Municipal Guard of Curitiba, being performed prediction processes through data mining through the occurrence obtained in the period from 2009 to June 2022. The process was guided by the Knowledge Discovery in Databases (KDD), where in the data mining stage the Naïve Bayes, Deep Learning and Decision Tree algorithms were used to analyze the database. In relation to the results obtained, the Naive Bayes model obtained a hit rate of 36.2%, while in comparison the Deep Learning model obtained a hit rate of 36.9% and finally the Decision Tree model presented a hit rate of 29.5%. It was brought as an objective the identification of patterns present in the database of the Municipal Guard of Curitiba, as a way to enable the discovery of information and knowledge, obtained through the analysis and development of descriptive statistics, already for the generation of knowledge was performed through the processing of algorithms and the manipulation and optimization of the results through the results obtained and the simulator.

Keywords: Data Analysis. Municipal Guard of Curitiba. Data

Mineração de Dados. Naïve Bayes. Deep Learning. Decision Tree. RapidMiner.

LISTA DE FIGURAS

FIGURA 1 - MODELO PROCESSUAL DE ADMINISTRAÇÃO DA INFORMAÇÃO	11
FIGURA 2: ETAPAS QUE COMPÕEM O PROCESSO DO KDD	14
FIGURA 3 - DISTRIBUIÇÃO DE OCORRÊNCIAS POR ANO	31
FIGURA 4 - NÚMERO DE OCORRÊNCIAS POR BAIRRO	32
FIGURA 5 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM A NATUREZA	33
FIGURA 6 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM A SUBCATEGORIA	34
FIGURA 7 - NÚMERO DE OCORRÊNCIAS SEMANAIS	35
FIGURA 8 - NÚMERO DE OCORRÊNCIAS PERÍODO	36
FIGURA 9 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM O DIA	37
FIGURA 10 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM O MÊS	38
FIGURA 11 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM SUA ORIGEM	38
FIGURA 12 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM A REGIONAL	39
FIGURA 13 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM A SECRETARIA	40
FIGURA 14 - SELEÇÃO DOS ATRIBUTOS	41
FIGURA 15 - TIPOS DE MODELOS	42
FIGURA 16 - RESULTADOS OBTIDOS	45
FIGURA 17 - PESO ATRIBUTOS NAIVE BAYES	46
FIGURA 18 - PESO ATRIBUTOS DEEP LEARNING	47
FIGURA 19 - SIMULADOR VARIÁVEL DESCUMPRIMENTO LEI 15799/2021 COVID-19	48
FIGURA 20 - SIMULADOR VARIÁVEL ANIMAL	49
FIGURA 21 - SIMULADOR VARIÁVEL	49
FIGURA 22 - PESO ATRIBUTOS DECISION TREE	51

LISTA DE QUADROS

QUADRO 1 – DADOS, INFORMAÇÃO E CONHECIMENTO	12
QUADRO 2 - DICIONÁRIO DE DADOS DA SIGESGUARDA	17
QUADRO 3 - DESCRIÇÃO DOS ATRIBUTOS SELECIONADOS PARA INTEGRAR O PROCESSO DE MINERAÇÃO	19
QUADRO 4 - DESCRIÇÃO DOS ATRIBUTOS RETIRADOS DO PROCESSO DE MINERAÇÃO	20
QUADRO 5 - SIGLA DAS SECRETARIAS	22
QUADRO 6 - BAIRROS DE CADA REGIONAL DE CURITIBA	25
QUADRO 7 - INCOERÊNCIA REGIONAL BAIRRO NOVO	26
QUADRO 8 - CONVERSÃO DO ATRIBUTO OCORRENCIA_MES	28
QUADRO 9 - DISTRIBUIÇÃO DO ESPAÇO DE TEMPO EM PERÍODO	29
QUADRO 10 - RENOMEAÇÃO DE INSTÂNCIAS DO ATRIBUTO ORIGEM_CHAMADO_DESCRICAÇÃO	29
QUADRO 11 - CATEGORIZAÇÃO DOS ATRIBUTOS	30

SUMÁRIO

1 INTRODUÇÃO	5
1.1 PROBLEMATIZAÇÃO	6
1.2 OBJETIVO	7
1.2.1 Objetivos Geral	7
1.2.2 Objetivos Específico	7
1.3 JUSTIFICATIVA	7
1.4 LIMITAÇÕES DA PESQUISA	8
2 FUNDAMENTAÇÃO TEÓRICA	9
2.1 GUARDA MUNICIPAL DE CURITIBA	9
2.2 GESTÃO DA INFORMAÇÃO	10
2.3 DADOS LIVRES/LEI DE ACESSO À INFORMAÇÃO	12
2.4 KNOWLEDGE DISCOVERY IN DATABASE - KDD	13
3 PROCEDIMENTOS METODOLÓGICOS	15
3.1 CARACTERIZAÇÃO DA PESQUISA	15
3.2 AMBIENTE DE PESQUISA	15
3.3 FERRAMENTAS UTILIZADAS	16
3.4 DICIONÁRIO DE DADOS	16
4 PROCESSAMENTO E ANÁLISE DOS RESULTADOS	19
4.1 SELEÇÃO DOS DADOS	19
4.2 PRÉ-PROCESSAMENTO E LIMPEZA DOS DADOS	24
4.3 TRANSFORMAÇÃO DOS DADOS	27
4.4 ESTATÍSTICA DESCRITIVA	30
4.5 DATA MINING	40
4.6 INTERPRETAÇÃO E AVALIAÇÃO DOS RESULTADOS	42
4.6.1 Resultado do Modelo Naive Bayes	45
4.6.2 Resultado do Modelo Deep Learning	46
4.6.3 Resultado Do Modelo Decision Tree	50

5. CONSIDERAÇÕES FINAIS	52
5.1 ALCANCE DOS OBJETIVOS	53
5.2 TRABALHOS FUTUROS	53
REFERÊNCIAS	55
APÊNDICE	57

1 INTRODUÇÃO

O crescimento no volume de dados gerados pelas organizações e sua diversidade vem se transformando em um desafio na medida que cresce, necessitando assim um trabalho minucioso de gerenciamento, tratamento e análise de dados para conseguir extrair as informações consideradas valiosas para os processos estratégicos da organização sem quaisquer ruídos ou informações de baixa qualidade.

Esses grandes volumes de dados podem ser encontrados em base de dados privados ou em formato abertos, como é o caso da base da Guarda Municipal de Curitiba que decorrente da lei de acesso à informação disponibilizou o acesso livre a população dessa e de muitas outras bases governamentais.

Entretanto, grande parte da população não sabe que tem acesso a esses tipos de dados e/ou não sabe como fazer o tratamento e a extração da informação dessas bases, impossibilitando dessa forma a sua utilização por grande parte da população.

Como forma de dar continuidade ao trabalho anterior de SOUZA (2021) busca-se por meio da atualização da base de dados Guarda Municipal de Curitiba, utilizar de técnicas e ferramenta para o processo de mineração de dados, além de software de visualização de dados para facilitar o entendimento dos dados, tendo como principal objetivo prever tendências de ocorrências atendida pela Guarda Municipal de acordo com sua natureza da ocorrência atendida, podendo servir para a descoberta de informações e geração de conhecimentos úteis para o auxílio do processo de tomada de decisões estratégicas de organizações e do conhecimento público.

1.1 PROBLEMATIZAÇÃO

Com a modernização das tecnologias de comunicação e os novos registros de aplicações no uso dos dados gerados ao redor do mundo, além da diminuição do custo de armazenamento dos dados, atrelado ao interesse crescente das organizações em acompanhar as mudanças e as tendências do ambiente ao seu redor fez com que ocorresse uma escalabilidade na demanda por dados, juntamente com uma maior disponibilidade desses dados.

O processo de exploração e análise que através da busca de padrões, realizado através da associação dessas grandes quantidades de dados é denominado Mineração de Dados, embora seja uma forma mais sofisticada, complexa e difícil de analisar dados, traz benefícios para resultados obtidos por ela, além de proporcionar *insights* sobre o negócio que outras técnicas não são capazes de produzir.

No entanto a uma grande disponibilidade de bases públicas e privadas que têm seu acesso liberado para que a população possa utilizá-las e produzir novas informações e aplicações digitais para a sociedade, mas muitas vezes elas não acabam sendo utilizadas por desconhecimento da sua existência ou por falta de conhecimento das ferramentas e mecanismos para realizar a extração de informações e conhecimentos nelas contidas.

A Guarda Municipal de Curitiba recebe diariamente diversos tipos de ocorrências de todas as regiões de Curitiba gerando um grande volume de dados armazenados ao longo dos anos, esses dados podem vir a apresentar padrões e tendências de recorrência, que podem vir a serem transformados em informações úteis para a própria Guarda Municipal e para a sociedade. Considerando esse cenário, este trabalho busca identificar: **De que maneira as ocorrências atendidas pela Guarda Municipal de Curitiba podem ser previstas?**

1.2 OBJETIVO

Para responder à questão levantada na pesquisa foram definidos os objetivos a serem alcançados, sendo estes desmembrados em objetivo geral e específicos.

1.2.1 Objetivos Geral

O objetivo geral desta pesquisa é identificar padrões presentes na base de dados da Guarda municipal de Curitiba, possibilitando assim a descoberta de informação e conhecimento

1.2.2 Objetivos Específico

Os objetivos específicos deste trabalho são:

- Fazer a aplicação de métodos de mineração de dados para descoberta de conhecimento;
- Comparar e analisar os resultados alcançados por meio das análises estatísticas e mineração de dados;
- Definir as ferramentas que serão utilizadas para o processo de mineração e de visualização dos resultados;

1.3 JUSTIFICATIVA

Atualizar e dar continuidade ao trabalho de SOUZA (2021) trazendo uma perspectiva geral das ocorrências atendidas pela Guarda Municipal de Curitiba de acordo com o tipo de ocorrência e a região onde ela ocorreu, fazendo a utilização de ferramenta de mineração de dados diferente as usadas

anteriormente, e acrescentando a utilização de mapas geográficos interativos para o processo de visualização da informação.

Disponibilizar através desse trabalho as informações obtidas como forma de servir de ferramenta para a Guarda Municipal de Curitiba, Prefeitura de Curitiba e os demais interessados utilizarem a como uma forma de auxiliar o planejamento e a tomada de decisão futuras, melhorando e instigando estratégias e políticas de ações de forma mais assertiva.

Já no âmbito pessoal, a pesquisa se justifica por tratar-se de uma temática de interesse pessoal do autor, possibilitando assim o desenvolvimento de novos conhecimentos através da aplicação de ferramenta e métodos para o desenvolvimento do presente trabalho, além de visualizar a aplicação do conhecimento adquiridos ao longo do curso de Gestão da Informação, atrelados a área de interesse em ciência de dados.

1.4 LIMITAÇÕES DA PESQUISA

O trabalho tem como objetivo identificar padrões presentes na base de dados da Guarda Municipal de Curitiba, no período de janeiro de 2009 até junho de 2022, com o intuito de prevenir tendências de ocorrências de acordo com a localidade e a natureza da ocorrência.

2 FUNDAMENTAÇÃO TEÓRICA

Essa seção irá trazer uma discussão acerca de conceitos relevantes à compreensão do presente trabalho de conclusão de curso. Constituem a revisão de literatura abordagens sobre a Guarda Municipal de Curitiba, Gestão da Informação, Dados livres/abertos, Processo de Descoberta de Conhecimento em Bases de Dados também conhecida como *Knowledge Discovery in Databases* (KDD) e Mineração de Dados.

2.1 GUARDA MUNICIPAL DE CURITIBA

No ano 1986, A cidade de Curitiba enfrentava um aumento nas depredações, despertando a necessidade da criação de um grupo distinto, tendo a proteção da população como prioridade. Com este intuito, em 17 de julho de 1986, o Prefeito Municipal Roberto Requião de Mello e Silva, sancionou após aprovação da Câmara Municipal de Curitiba, o Projeto de Lei n.º 56/84, implementando a Lei n.º 6867/1986, que criou o Serviço Municipal de Vigilância – VIGISERV. (A GUARDA MUNICIPAL, 2013).

Segundo Carvalho (2011, p. 18) Embora já existissem alguns decretos que tratavam o VIGISERV - Serviço Municipal de Vigilância - como Guarda Municipal, foi apenas em 05 de outubro de 1989, por meio da Lei n.º 7356/89, que a VIGISERV passou a denominar-se efetivamente Guarda Municipal de Curitiba.

A Guarda Municipal vem atuando em conjunto através de ações Integradas de Segurança Pública com as polícias federal, civil e militar e outras com diferentes órgãos públicos; ações de fiscalização urbana; de proteção ao transporte coletivo; de orientação e fiscalização do trânsito, de prevenção e combate a ocupações irregulares. Além de ser responsável por cuidar diretamente da segurança das escolas municipais, creches, postos de saúde, praças, bosques, parques, armazéns da família e dos demais equipamentos

públicos de Curitiba. Combatem pichação, atos de vandalismo contra o patrimônio público e outras situações de flagrante delito, se baseando no Decreto Municipal 1179/12 e na Lei Federal 13.022/14. (GUARDA MUNICIPAL-Central de Emergência: 153).

De acordo com Carvalho (2011, p. 251),

Cabe lembrar que a Guarda Municipal não está exclusivamente voltada para a segurança pública, conforme os moldes do Regime Militar, mas sim para atuação na área de defesa social que corresponde a uma parcela significativa da prestação de serviço à comunidade de maneira extensiva, o qual abrange segurança pública, defesa civil, entre outras ações do poder público.

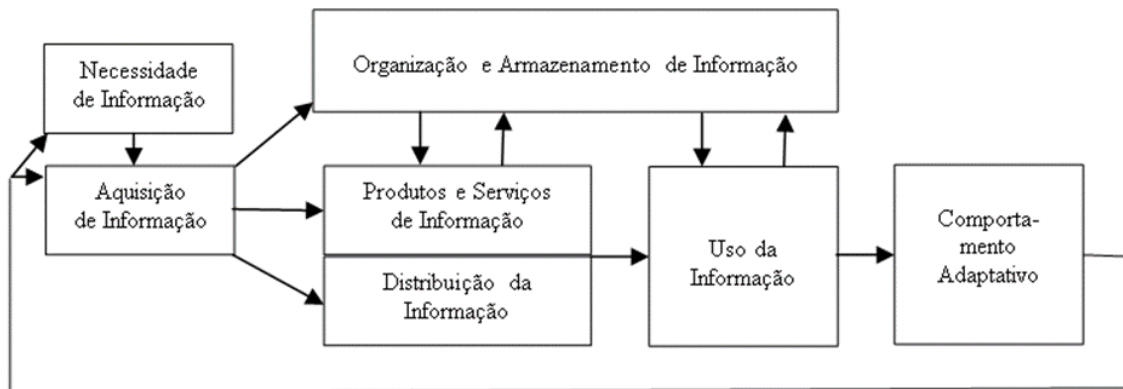
2.2 GESTÃO DA INFORMAÇÃO

A Gestão da Informação pode ser definida como o processo de integração dos diferentes tipos de informação, sendo executado através das atividades de coleta, armazenamento, gerenciamento e uso da informação.

Segundo Tarapanoff (2006, p. 22) “O principal objetivo da gestão da informação é identificar e potencializar recursos informacionais de uma organização ou empresa e sua capacidade de informação, ensinando-a a aprender e adaptar-se a mudanças ambientais”.

Choo (2003) Apresenta a visão da administração da informação através de uma rede de processos que adquirem, criam, organizam, distribuem e usam a informação. Define o conceito da organização do conhecimento como sendo capaz de organizar seus recursos e competências através da transformação da informação em compreensão e insight, sendo capaz de difundir esse conhecimento através de ações, que possibilitem o aprendizado e a adaptação da organização aos ambientes em constante mudança. Dessa forma Choo (2003) apresenta um modelo processual de administração da informação, que pode ser interpretado como um ciclo da gestão da informação, conforme apresentado na figura 1.

FIGURA 1 - MODELO PROCESSUAL DE ADMINISTRAÇÃO DA INFORMAÇÃO



FONTE:CHOO (2003, p. 404).

Para Davenport (1998) a gestão da Informação é compreendida como um conjunto estruturado de atividades que incluem o modo como as organizações obtêm, distribuem e utilizam a sua informação e conhecimento. Como forma de definir e distinguir o termo informação, dados e conhecimento é apresentado no Quadro 1 com as características de cada uma delas.

QUADRO 1 – DADOS, INFORMAÇÃO E CONHECIMENTO

Dados	Informação	Conhecimento
Simples observações sobre o estado do mundo	Dados dotados de relevância e propósito	Informação valiosa da mente humana - Inclui reflexão, síntese, contexto
<ul style="list-style-type: none"> • Facilmente estruturado • Facilmente obtido por máquinas • Frequentemente quantificado • Facilmente transferível 	<ul style="list-style-type: none"> • Requer unidade de análise • Exige consenso em relação ao Significado • Exige necessariamente a mediação humana 	<ul style="list-style-type: none"> • De difícil estruturação • De difícil captura em máquinas • Frequentemente tácito • De difícil transferência

FONTE: adaptado DAVENPORT (1998, p. 18).

Davenport (1998) ressalta que o termo informação em si envolve os três, além de servir como uma ponte entre os dados brutos e o conhecimento que pode vir a ser obtido.

2.3 DADOS LIVRES/LEI DE ACESSO À INFORMAÇÃO

Segundo Governo Digital (2019), Os Dados Abertos Governamentais é uma metodologia utilizada para a publicação de dados do governo em formatos reutilizáveis, visando o aumento da transparência e maior participação política por parte do cidadão, podendo gerar diversas aplicações desenvolvidas em colaboração com a sociedade. Esses dados abertos governamentais trazem três principais efeitos sobre as políticas pública, sendo eles a **Inclusão**, através do fornecimento de dados abertos, padronizados e acessíveis que permitem que os cidadãos utilizem e os adaptem para suas necessidades; a **Transparência**: disponibilizando dados e informações públicas de forma aberta

e acessível para serem utilizadas de maneira adequada aos seus propósitos; E a **Responsabilidade**: oferecendo diversos pontos de vista sobre o desempenho dos governo no cumprimento de suas metas em políticas públicas.

2.4 KNOWLEDGE DISCOVERY IN DATABASE - KDD

O Knowledge Discovery In Database (KDD) é formado através do processo de identificação de padrões pela análise de grandes conjuntos de dados, tendo como sua principal etapa o processo de mineração, na execução prática de análise e de algoritmos específicos que, sob limitações de eficiência computacionais toleráveis, são capazes de produzir uma relação particular de padrões a partir dos dados (FAYYAD; PIATETSKYSHAPIRO; SMYTH, 1996).

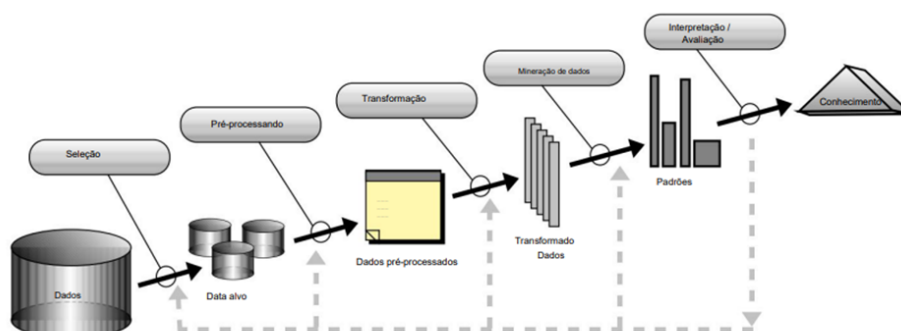
Para SILVA (2016.) o objetivo do KDD é encontrar padrões intrínsecos aos dados nela contidos, apresentando-os com o intuito de facilitar sua assimilação como conhecimento, sendo esse um processo analítico, sistemático e às vezes automatizado.

CASTRO e FERRARI (2016) Sintetiza o processo do KDD em quatro partes principais, sendo ela a **Base de dados**: coleção organizada dos dados contendo valores quantitativos ou qualitativos referentes a um conjunto de itens, que permite uma recuperação eficiente dos dados; **Pré-Processamento**: Visam preparar os dados para uma análise eficiente e eficaz. Essa etapa inclui a limpeza (remoção de ruídos e dados inconsistentes), a integração (combinação de dados obtidos a partir de múltiplas fontes), a seleção ou redução (escolha dos dados relevantes à análise) e a transformação (transformação ou consolidação dos dados em formatos apropriados para a mineração); **Mineração de dados**: Processo corresponde à aplicação de algoritmos capazes de extrair conhecimentos a partir dos dados pré-processados. Consiste em discutidas técnicas de análise descritiva (medidas de distribuição, tendência central e variância, e métodos de visualização), agrupamento (segmentação de bases de dados), predição

(classificação e estimação), associação (determinação de atributos que co-ocorrem) e detecção de anomalias; **Validação do conhecimento:** Avaliação dos resultados da mineração objetivando identificar conhecimentos verdadeiramente úteis e não triviais.

FAYYAD (1996), apresenta o KDD como sendo um conjunto composto de 5 (cinco) etapas, sendo representado na FIGURA 2, e descritas a seguir: **Seleção dos dados:** Etapa responsável por fazer a seleção do conjunto de dados ou focar em um subconjunto de variáveis ou amostras de dados, no qual será realizada a análise; **Pré-processamento e limpeza dos dados:** Etapa responsável por realizar a remoção de ruído se apropriado, coletando as informações necessárias para modelar ou contabilizar o ruído, decidir sobre estratégias para lidar com campos de dados ausentes, e contabilização de informações de sequência de tempo e mudanças conhecidas; **Transformação dos dados:** Etapa responsável por encontrar recursos úteis para representar os dados dependendo do objetivo da tarefa, utilizando de métodos de transformação que gerem dados relevantes a partir dos dados já existentes; **Data Mining:** Etapa responsável por buscar padrões de interesse em uma forma de representação particular ou um conjunto de tais representações, incluindo regras ou árvores de classificação, regressão e agrupamento; **Interpretação e Avaliação dos resultados:** Etapa responsável por avaliar o desempenho através da interpretação e demonstração dos padrões obtidos ao longo do process

FIGURA 2 - ETAPAS QUE COMPÕEM O PROCESSO DO KDD



FONTE: adaptado FAYYAD; PIATETSKYSHAPIRO; SMYTH, (1996, p. 41)

3 PROCEDIMENTOS METODOLÓGICOS

Esse capítulo tem o intuito de apresentar as ferramentas e método utilizado neste trabalho através das sessões: caracterização da pesquisa; ambiente de pesquisa; ferramentas utilizadas; dicionário de dados.

3.1 CARACTERIZAÇÃO DA PESQUISA

A presente pesquisa adota uma característica descritiva, buscando obter e comparar entre si as diferentes informações adquiridas ao longo do trabalho. Esse tipo de pesquisa tem como objetivo a descrição das características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis.

Em relação à finalidade da presente pesquisa é caracterizada por ser aplicada, através da descoberta de conhecimento obtida ao longo do desenvolvimento da pesquisa como forma de enriquecer o desenvolvimento do trabalho como um todo.

A abordagem da pesquisa será quantitativa, realizada através da análise obtida por meio de ferramentas estatísticas.

3.2 AMBIENTE DE PESQUISA

O trabalho utilizou a base de dados da Guarda Municipal de Curitiba contendo as ocorrências atendidas pela mesma.

A Guarda Municipal presta atendimento descentralizado aos 75 bairros de Curitiba dentro das 10 regionais que compõem Curitiba, através de patrulhamento preventivo e ostensivos dos prédios, logradouros e equipamentos públicos, por meio de efetivo fixo e através de rondas, além de atender diferentes tipos de ocorrências pela Central de Emergência da Guarda

Municipal (número 153) e pela Central de Atendimentos de Curitiba (número 156).

Por se tratar de um órgão público a base de dados está disponível no site da prefeitura de Curitiba na sessão de dados abertos licenciada sob uma Licença Creative Commons, disponível para os cidadãos conforme a lei de acesso à informação, facilitando o acesso e utilização dos dados para o desenvolvimento da presente pesquisa.

3.3 FERRAMENTAS UTILIZADAS

Para a realização do processo de mineração de dados foi escolhido o software RapidMiner, por ser uma ferramenta que oferece diversos recursos para realizar o pré-processamento e preparação dos dados para assim serem feitas as análises, além de possuir recursos para construir workflows de forma visual. O RapidMiner também oferece uma licença acadêmica de forma gratuita que disponibiliza linhas de dados ilimitadas, processadores lógicos ilimitados, suporte da comunidade e acesso a recursos automatizados.

3.4 DICIONÁRIO DE DADOS

A base de dados do Sistema de Gestão da Guarda Municipal de Curitiba (SiGesGuarda) contendo os dados das ocorrências atendidas pela Guarda Municipal de Curitiba de 2009 até o momento da sua extração no mês de junho (06) de 2022, contabilizou 403.325 registros para o desenvolvimento do presente trabalho, contendo 35 campos com seus respectivos atributos podendo ser observados no quadro 2. O acesso da base de dados e do dicionário de dados pode ser obtido em formato CSV através da seção de dados abertos do site¹ da Prefeitura Municipal de Curitiba.

¹ <https://www.curitiba.pr.gov.br/dadosabertos/busca/?grupo=12>

QUADRO 2 - DICIONÁRIO DE DADOS DA SIGESGUARDA

COLUNAS	FORMATO	CAMPO OBRIGATÓRIO?	DESCRIÇÃO DOS CAMPOS
ATENDIMENTO_ANO	int	SIM	Ano em que foi realizado ao atendimento
ATENDIMENTO_BAIRRO_NOME	varchar	SIM	Nome do bairro em que foi realizado o atendimento
EQUIPAMENTO_URBANO_NOME	varchar	NÃO	Nome do equipamento urbano
FLAG_EQUIPAMENTO_URBANO	char	SIM	Flag para identificar se é um equipamento urbano ou não
FLAG_FLAGRANTE	char	SIM	Flag para identificar se é um equipamento urbano ou não
LOGRADOURO_NOME	varchar	SIM	Nome do logradouro
NATUREZA_DEFESA_CIVIL	bit	NÃO	Flag para identificar se é uma natureza de defesa civil ou não (0 - Não, 1 - Sim)
NATUREZA_DESCRICAO	varchar	NÃO	Descrição da primeira natureza cadastrada na ocorrência
SUBCATEGORIA_DESCRICAO	varchar	NÃO	Descrição da sub-categoria da primeira natureza
OCORRENCIA_ANO	int	SIM	Ano de cadastro da ocorrência
OCORRENCIA_CODIGO	int	SIM	Código da ocorrência
OCORRENCIA_DATA	datetime	SIM	Data da ocorrência
OCORRENCIA_DIA_SEMANA	varchar	SIM	Dia da semana em que a ocorrência foi cadastrada
OCORRENCIA_HORA	varchar	SIM	Hora em que a ocorrência foi cadastrada
OCORRENCIA_MES	int	SIM	Mês em que a ocorrência foi cadastrada
OPERACAO_DESCRICAO	varchar	NÃO	Nome da operação que realizará o atendimento, caso haja

ORIGEM_CHAMADO_DESCRICA O	varchar	SIM	Local em que se originou a chamada
REGIONAL_FATO_NOME	varchar	SIM	Regional do fato (local em que ocorreu a ocorrência)
SECRETARIA_NOME	varchar	SIM	Nome da secretaria solicitante
SECRETARIA_SIGLA	varchar	SIM	Sigla da secretaria solicitante
SERVICO_NOME	varchar	SIM	Nome do serviço que realizará o atendimento
SITUACAO_EQUIPE_DESCRICA O	varchar	NÃO	Situação da equipe no momento atual
NUMERO_PROTOCOLO_156	int	NÃO	Número do protocolo quando a origem do chamado for 156

FONTE: adaptado SiGesGuarda (2022)

Os atributos NATUREZA_DESCRICA
O; NATUREZA_DEFESA_CIVIL e SUBCATEGORIA_DESCRICA
O, são apresentados originalmente em 5 estâncias cada um, sendo utilizados como campos complementares dos seus antecessores de acordo com o preenchimento de cada registro de ocorrência.

4 PROCESSAMENTO E ANÁLISE DOS RESULTADOS

4.1 SELEÇÃO DOS DADOS

Para o presente trabalho foram selecionados 11 atributos considerados relevantes representativamente para integrar a pesquisa e o processo de mineração, podendo ser visualizados no Quadro 3 com suas respectivas descrições e exemplos.

QUADRO 3 - DESCRIÇÃO DOS ATRIBUTOS SELECIONADOS PARA INTEGRAR O PROCESSO DE MINERAÇÃO

ATRIBUTO	DESCRIÇÃO	EXEMPLO
ATENDIMENTO_ANO	Ano em que foi realizado ao atendimento	2018
ATENDIMENTO_BAIRRO_NOME	Nome do bairro em que foi realizado o atendimento	PORTÃO
NATUREZA1_DESCRICA O	Descrição da primeira natureza cadastrada na ocorrência	ALARMES
SUBCATEGORIA1_DESCRICA O	Descrição da sub-categoria da primeira natureza	DISPARO DE ALARME (VIOLAÇÃO)
OCORRENCIA_DATA	Data da ocorrência	2009-01-03 18:31:00.000
OCORRENCIA_DIA_SEMANA	Dia da semana em que a ocorrência foi cadastrada	QUINTA
OCORRENCIA_HORA	Hora em que a ocorrência foi cadastrada	17:45:00
OCORRENCIA_MES	Mês em que a ocorrência foi cadastrada	5
ORIGEM_CHAMADO_DESCRICA O	Local em que se originou a chamada	156
REGIONAL_FATO_NOME	Regional do fato (local em que ocorreu a ocorrência)	PINHEIRINHO
SECRETARIA_SIGLA	Sigla da secretaria solicitante	SMDT

FONTE: adaptado SiGesGuarda (2022)

Os demais 24 atributos não selecionados para compor a pesquisa são apresentados no quadro 4, juntamente com a descrição do motivo de sua remoção.

QUADRO 4 - DESCRIÇÃO DOS ATRIBUTOS RETIRADOS DO PROCESSO DE MINERAÇÃO

ATRIBUTO	DESCRIÇÃO DO MOTIVO DA EXCLUSÃO
EQUIPAMENTO_URBANO_NOME	Apresenta diversos campos vazios, além de se misturar com locais onde aconteceram a ocorrência e locais de encaminhamento de acordo com o tipo de ocorrência.
FLAG_EQUIPAMENTO_URBANO	Não corresponde ao foco da pesquisa
FLAG_FLAGRANTE	Não corresponde ao foco da pesquisa
LOGRADOURO_NOME	Apresenta diversas inconsistências em relação à acentuação, erros de escrita e erros/preenchimento de informação sem validade.
NATUREZA1_DEFESA_CIVIL	Impertinentes à pesquisa.
NATUREZA2_DEFESA_CIVIL	Campos vazios e impertinentes à pesquisa.
NATUREZA2_DESCRICAO	Campos de complementação, sendo em sua maioria vazios com descrições repetitivas e impertinentes à pesquisa.
NATUREZA3_DEFESA_CIVIL	Campos vazios e impertinentes à pesquisa.
NATUREZA3_DESCRICAO	Campos de complementação, sendo em sua maioria vazios com descrições repetitivas e impertinentes à pesquisa.
NATUREZA4_DEFESA_CIVIL	Campos vazios e impertinentes à pesquisa.
NATUREZA4_DESCRICAO	Campos de complementação, sendo em sua maioria vazios com descrições repetitivas e impertinentes à pesquisa.
NATUREZA5_DEFESA_CIVIL	Campos vazios e impertinentes à pesquisa.
NATUREZA5_DESCRICAO	Campos de complementação, sendo em sua maioria vazios com descrições repetitivas e impertinentes à pesquisa.
SUBCATEGORIA2_DESCRICAO	Campos vazios.
SUBCATEGORIA3_DESCRICAO	Campos vazios.

SUBCATEGORIA4_DESCRICAO	Campos vazios.
SUBCATEGORIA5_DESCRICAO	Campos vazios.
OCORRENCIA_ANO	Os valores se repetem ao atributo "ATENDIMENTO_ANO"
OCORRENCIA_CODIGO	Impertinentes à pesquisa.
OPERACAO_DESCRICAO	Impertinentes à pesquisa.
SECRETARIA_NOME	Impertinentes à pesquisa.
SERVICO_NOME	Impertinentes à pesquisa.
SITUACAO_EQUIPE_DESCRICAO	Impertinentes à pesquisa.
NUMERO_PROTOCOLO_156	Diversos campos vazios, impertinentes à pesquisa.

Fonte: O AUTOR (2022).

Os atributos (EQUIPAMENTO_URBANO_NOME e NUMERO_PROTOCOLO_156) foram os atributos que mais apresentaram campos vazios sendo removidos primordialmente por esse motivo.

Já os atributos (NATUREZA2_DEFESA_CIVIL; NATUREZA2_DESCRICAO; NATUREZA3_DEFESA_CIVIL; NATUREZA3_DESCRICAO; NATUREZA4_DEFESA_CIVIL; NATUREZA4_DESCRICAO; NATUREZA5_DEFESA_CIVIL; NATUREZA5_DESCRICAO; SUBCATEGORIA2_DESCRICAO; SUBCATEGORIA3_DESCRICAO; SUBCATEGORIA4_DESCRICAO; SUBCATEGORIA5_DESCRICAO) também apresentam diversos campos vazios, especialmente por se tratarem de atributos complementares de acordo com a estância numerica que as antecedem, e por não serem atributos de preenchimento obrigatório.

Para os atributos (FLAG_EQUIPAMENTO_URBANO; FLAG_FLAGRANTE; NATUREZA1_DEFESA_CIVIL; OCORRENCIA_CODIGO; OPERACAO_DESCRICAO; SERVICO_NOME; SITUACAO_EQUIPE_DESCRICAO) foram removidos por não corresponderem ao foco da pesquisa.

O atributo OCORRENCIA_ANO foi retirado por apresentar os mesmos dados do atributo ATENDIMENTO_ANO já escolhido anteriormente para compor a pesquisa. Para o atributo SECRETARIA_NOME foi optado por removê-lo da pesquisa, porém seus dados foram transcritos juntamente com suas abreviações (contida no atributo SECRETARIA_SIGLA) como forma de consulta apresentada no Quadro 5 .

QUADRO 5 - SIGLA DAS SECRETARIAS

SIGLA	NOME DA SECRETARIA
COHAB	COMPANHIA DE HABITAÇÃO DE CURITIBA
CSA	CURITIBA S/A
FCC	FUNDAÇÃO CULTURAL DE CURITIBA
FAS	FUNDAÇÃO DE AÇÃO SOCIAL
IMAP	INSTITUTO MUNICIPAL DE ADMINISTRAÇÃO PÚBLICA
SAM	SECRETARIA ANTIDROGAS MUNICIPAL
SGM	SECRETARIA DO GOVERNO MUNICIPAL
SEDPD	SECRETARIA ESPECIAL DOS DIREITOS DA PESSOA COM DEFICIENCIA
SMPC	SECRETARIA EXTRAORDINARIA MUNICIPAL DE PLANEJAMENTO E COORDENAÇÃO
SEPLAN	SECRETARIA EXTRAORDINARIA MUNICIPAL DE PLANEJAMENTO E COORDENAÇÃO
SICT	SECRETARIA MUN. DA IND. COM. E TURISMO
SMCS	SECRETARIA MUNICIPAL DA COMUNICAÇÃO SOCIAL
SMCR	SECRETARIA MUNICIPAL DA CRIANÇA
SME	SECRETARIA MUNICIPAL DA EDUCAÇÃO
SMS	SECRETARIA MUNICIPAL DA SAÚDE
SMAP	SECRETARIA MUNICIPAL DE ADMINISTRAÇÃO E DE GESTÃO DE PESSOAL
SMAM	SECRETARIA MUNICIPAL DE ASSUNTOS METROPOLITANOS
SMDT	SECRETARIA MUNICIPAL DE DEFESA SOCIAL E TRANSITO

SMOP	SECRETARIA MUNICIPAL DE OBRAS PÚBLICAS
SMF	SECRETARIA MUNICIPAL DE PLANEJAMENTO, FINANÇAS E ORÇAMENTO
SMRH	SECRETARIA MUNICIPAL DE RECURSOS HUMANOS
SMSA	SECRETARIA MUNICIPAL DE SANEAMENTO
SMSAN	SECRETARIA MUNICIPAL DE SEGURANÇA ALIMENTAR E NUTRICIONAL
SETRAN	SECRETARIA MUNICIPAL DE TRÂNSITO
SMTU	SECRETARIA MUNICIPAL DE TURISMO
SMELJ	SECRETARIA MUNICIPAL DO ESPORTE, LAZER E JUVENTUDE
SMTE	SECRETARIA MUNICIPAL DO TRABALHO E EMPREGO
SMU	SECRETARIA MUNICIPAL DO URBANISMO
SEPHA	SECRETARIA MUNICIPAL EXTRAORDINÁRIA DE POLÍTICA HABITACIONAL
SEA	SECRETARIA MUNICIPAL EXTRAORDINÁRIA ANTIDROGAS
SECOPA	SECRETARIA MUNICIPAL EXTRAORDINÁRIA DA COPA 2014
SMEM	SECRETARIA MUNICIPAL EXTRAORDINÁRIA DA MULHER
SEPE	SECRETARIA MUNICIPAL EXTRAORDINÁRIA DE PROJETOS ESPECIAIS
SERCOM	SECRETARIA MUNICIPAL EXTRAORDINÁRIA DE RELAÇÕES COM A COMUNIDADE
SERIN	SECRETARIA MUNICIPAL EXTRAORDINÁRIA DE RELAÇÕES INSTITUCIONAIS
SERIC	SECRETARIA MUNICIPAL EXTRAORDINÁRIA DE RELAÇÕES INTERNACIONAIS E CERIM
SMMA	SECRETARIA MUNICIPAL MEIO AMBIENTE
URBS	URBANIZAÇÃO DE CURITIBA

Fonte: O AUTOR (2022).

Por fim o atributo LOGRADOURO_NOME foi removido por apresentar diversas variações e inconsistências de escrita e acentuações, além de ser um

campo de preenchimento obrigatório tornando se inviável a sua correção e padronização devido ao grande volume de ocorrências despadronizadas.

4.2 PRÉ-PROCESSAMENTO E LIMPEZA DOS DADOS

A Partir da finalização da etapa de Seleção de Dados foi dado início ao processo de remoção e/ou adequação dos dados ausentes ou que apresentavam alguma inconsistência dentro dos atributos selecionados, em relação a seu preenchimento e ao foco da pesquisa, com o intuito de reduzir os ruídos encontrados dentro da base e eliminar dados considerados anormais e inconsistentes.

No atributo ATENDIMENTO_ANO, (39) ocorrências não apresentavam dados preenchidos, entretanto o campo OCORRENCIA_ANO continha os mesmos dados registrados, não sendo necessário realizar a exclusão das ocorrências faltantes, sendo complementadas os respectivos dados de acordo com o atributos alocados do campo OCORRENCIA_ANO, dessa forma as ocorrências permaneceram na análise.

No atributo REGIONAL_FATO_NOME, optou-se por retirar as ocorrências que foram registradas como “REGIÃO METROPOLITANA” por não corresponder a uma regional do município de Curitiba, resultando na exclusão de 418 registros como essa denominação, além de mais 11 ocorrências que não apresentavam dados no atributo REGIONAL_FATO_NOME.

Para o atributo ATENDIMENTO_BAIRRO_NOME foram removidos os de registro com denominados “INDICAÇÕES CANCELADA” que contava com 1 registro, o “BAIRRO FICTÍCIO” que apresentava 14 registros, o “BAIRRO NAO INFORMADO” com 9 registro e o BAIRRO NÃO LOCALIZAD com 2 registros. Também foram removidos mais 145 registros por não trazerem nenhuma informação consigo (campos vazios).

Os bairros da região metropolitana de Curitiba registrados incorretamente em relação a sua regional foram removidos no processo de

limpeza por não corresponder ao foco da pesquisa, sendo descritos no quadro 7 e nos apêndices A a I de acordo com a regional onde foram registrados.

Em relação a o processo de padronização e adequação dos dados apresentados o atributo ATENDIMENTO_BAIRRO_NOME contém o bairro “CIDADE INDUSTRIAL” e “CIC” constituindo o mesmo bairro, sendo necessário padronizá los como sendo apenas um, ficando assim apenas a nomenclatura de CIDADE INDUSTRIAL para ambas.

Após realizar buscas dentro de sites da Prefeitura Municipal de Curitiba e da Fundação de Ação Social (FAS) por informações referentes às regionais de Curitiba e seus respectivos bairros, foi notado uma discordância em relação a divisão dos bairros em alguns pontos de fronteira entre as regionais, assim como na e própria base de dados da SiGesGuarda onde alguns bairros faziam parte de diferentes regionais respectivamente.

Para resolver essas inconsistências encontradas o quadro 6 foi desenvolvido, baseado nas informações apresentadas no site² da Prefeitura Municipal de Curitiba que apresenta os novos administradores de cada regional com seus respectivos bairros que constituem cada regional.

QUADRO 6 - BAIRROS DE CADA REGIONAL DE CURITIBA

REGIONAIS	BAIRROS
BAIRRO NOVO	GANCHINHO; SÍTIO CERCADO; UMBARÁ
BOA VISTA	ABRANCHES; ATUBA; BACACHERI; BAIRRO ALTO; BARREIRINHA; BOA VISTA; CACHOEIRA; PILARZINHO; SANTA CÂNDIDA; SÃO LOURENÇO; TABOÃO; TARUMÃ; TINGUI
BOQUEIRÃO	ALTO BOQUEIRÃO; BOQUEIRÃO; HAUER; XAXIM
CAJURU	CAJURU; CAPÃO DA IMBUIA; GUABIROTUBA; JARDIM DAS AMÉRICAS; UBERABA

²<https://www.curitiba.pr.gov.br/noticias/veja-como-entrar-em-contato-com-as-regionais-e-qual-delas-cuida-do-seu-bairro/40912>

CIC	AUGUSTA; CIDADE INDUSTRIAL; RIVIERA; SÃO MIGUEL
MATRIZ	AHÚ; ALTO DA GLÓRIA; ALTO DA XV; BATEL; BIGORRILHO; BOM RETIRO; CABRAL; CENTRO; CENTRO CÍVICO; CRISTO REI; HUGO LANGE; JARDIM BOTÂNICO; JARDIM SOCIAL; JUVEVÊ; MERCÊS; PRADO VELHO; REBOUÇAS; SÃO FRANCISCO
PINHEIRINHO	CAPÃO RASO; FANNY; LINDÓIA; NOVO MUNDO; PINHEIRINHO;
PORTÃO	ÁGUA VERDE; FAZENDINHA; GUAÍRA; PAROLIN; PORTÃO; SANTA QUITÉRIA; SEMINÁRIO; VILA IZABEL
SANTA FELICIDADE	BUTIATUVINHA; CAMPINA DO SIQUEIRA; CASCATINHA; CAMPO COMPRIDO; LAMENHA PEQUENA; MOSSUNGUÊ/ ORLEANS; SANTA FELICIDADE; SANTO INÁCIO; SÃO BRAZ; SÃO JOÃO; VISTA ALEGRE
TATUQUARA	CAMPO DO SANTANA; CAXIMBA; TATUQUARA

Fonte: O AUTOR (2022).

Com base no Quadro 6, todas as ocorrências foram padronizadas em relação ao seu bairro e sua respectiva regional, corrigindo assim suas incoerências e removendo eventuais ruídos como apresentado do Quadro 7 e nos Apêndices A a I.

QUADRO 7 - INCOERÊNCIA REGIONAL BAIRRO NOVO

BAIRRO	TOTAL DE OCORRÊNCIAS	O'QUE FOI FEITO?
AHÚ	1	TRANSFERIDO PARA REGIONAL MATRIZ
ALTO BOQUEIRÃO	13	TRANSFERIDO PARA REGIONAL BOQUEIRÃO
AUGUSTA	3	TRANSFERIDO PARA REGIONAL CIC
BOM RETIRO	1	TRANSFERIDO PARA REGIONAL MATRIZ

BOQUEIRÃO	3	TRANSFERIDO PARA REGIONAL BOQUEIRÃO
BUTIATUVINHA	1	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
CAJURU	14	TRANSFERIDO PARA REGIONAL CAJURU
CAMPO DE SANTANA	3	TRANSFERIDO PARA REGIONAL TATUQUARA
CAXIMBA	1	TRANSFERIDO PARA REGIONAL TATUQUARA
CENTRO	1	TRANSFERIDO PARA REGIONAL MATRIZ
GANCHINHO	1808	PERMANECEU
PINHEIRINHO	22	TRANSFERIDO PARA REGIONAL PINHEIRINHO
PRADO VELHO	1	TRANSFERIDO PARA REGIONAL MATRIZ
SEMINÁRIO	1	TRANSFERIDO PARA REGIONAL PORTÃO
SÍTIO CERCADO	24869	PERMANECEU
TATUQUARA	1	TRANSFERIDO PARA REGIONAL TATUQUARA
UMBARÁ	2044	PERMANECEU
XAXIM	1	TRANSFERIDO PARA REGIONAL BOQUEIRÃO

Fonte: O AUTOR (2022).

4.3 TRANSFORMAÇÃO DOS DADOS

O atributo OCORRENCIA_DATA continha os dados referentes ao ano, mês, dia e horário das ocorrências, porém o atributo ATENDIMENTO_ANO já trazia os dados referentes ao ano da ocorrência, e o atributo OCORRENCIA_HORA já trazia as informações referente ao horário da ocorrência e o atributo OCORRENCIA_MES já trazia dados referentes ao mês da ocorrência, a partir daí foi necessário realizar a filtragem dos dados referentes ao dia das ocorrências. Por meio do excel foi utilizado o recurso de

dividir o texto em diferentes colunas (texto para colunas), dando origem a mais 1 novo atributo sendo ele denominado OCORRENCIA_DIA que contém os dados referentes ao dia da ocorrência.

Para o atributo OCORRENCIA_MES foi feita uma conversão dos dados numéricos para transformá-los em dados textuais como mostrado no quadro 8.

QUADRO 8 - CONVERSÃO DO ATRIBUTO OCORRENCIA_MES

MÊS EM NÚMEROS	MÊS EM TEXTUAL
1	JANEIRO
2	FEVEREIRO
3	MARÇO
4	ABRIL
5	MAIO
6	JUNHO
7	JULHO
8	AGOSTO
9	SETEMBRO
10	OUTUBRO
11	NOVEMBRO
12	DEZEMBRO

Fonte: O AUTOR (2022).

Com a criação dos atributos OCORRENCIA_DIA e como os demais atributos (OCORRENCIA_MES; OCORRENCIA_HORA; ATENDIMENTO_ANO) já conterem os mesmo dados possibilitou a exclusão do atributo OCORRENCIA_DATA da composição de atributos selecionados para compor o trabalho.

Em relação ao atributo OCORRENCIA_HORA optouse por transformar seus dados numéricos em quatro períodos do dia sendo eles apresentados conforme o Quadro 9.

QUADRO 9 - DISTRIBUIÇÃO DO ESPAÇO DE TEMPO EM PERÍODOS

PERÍODO	ESPAÇO DE TEMPO
MANHÃ	06:00 ÀS 11:59
TARDE	12:00 ÀS 17:59
NOITE	18:00 AS 23:59
MADRUGADA	00:00 ÀS 05:59

Fonte: O AUTOR (2022).

Com a nova transformação realizada nos espaços de tempo em períodos pré definidos no atributo OCORRENCIA_HORA optouse por realizar a renomeação do mesmo para OCORRENCIA_PERIODO.

Para o tratamento do atributo ORIGEM_CHAMADO_DESCRICAO foi realizado uma renomeação das instâncias que são apresentadas com o número de origem do chamado, sendo apresentadas as novas instâncias no Quadro 10.

QUADRO 10 - RENOMEAÇÃO DE INSTÂNCIAS DO ATRIBUTO ORIGEM_CHAMADO_DESCRICAO

INSTÂNCIA	NOVA INSTÂNCIA
153	GUARDA MUNICIPAL
156	CENTRAL DE ATENDIMENTO DA PREFEITURA DE CURITIBA
199	DEFESA CIVIL
CIOSP (190)	CIOSP
VAZIO "NÃO CADASTRAR" ANTIGO AOS GMS	ANTIGO AOS GMS
NÃO CADASTRAR "ANTIGO SIGA"	SIGA

Fonte: O AUTOR (2022).

4.4 ESTATÍSTICA DESCRITIVA

Após realizar a escolha, agrupamento, padronização e correção dos dados pertencentes a cada atributos, podemos analisar por meio do Quadro 11 os tipos de variáveis e suas variações de acordo com cada atributo.

QUADRO 11 - CATEGORIZAÇÃO DOS ATRIBUTOS

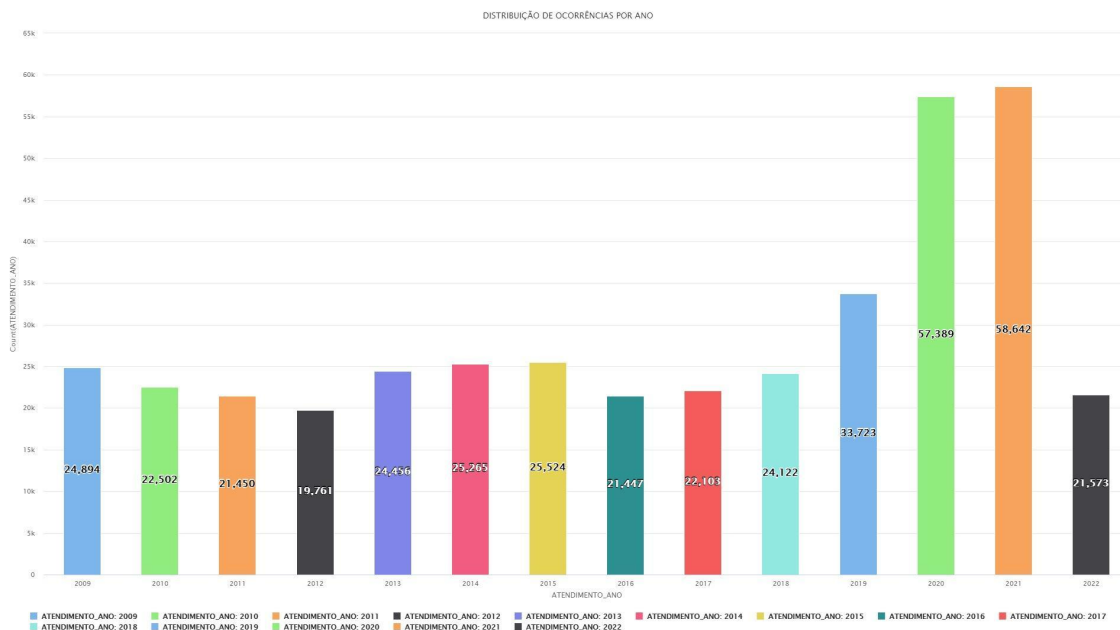
ATRIBUTO	VARIÁVEL	VARIAÇÃO
ATENDIMENTO_ANO	QUANTITATIVA CONTÍNUAS	DE 2009 ATÉ JUNHO DE 2022
ATENDIMENTO_BAIRRO_NOME	QUALITATIVA NOMINAIS	75 BAIRROS DE CURITIBA
NATUREZA1_DESCRICAO	QUALITATIVA NOMINAIS	186 DESCRIÇÕES
SUBCATEGORIA1_DESCRICAO	QUALITATIVA NOMINAIS	210 SUBCATEGORIAS
OCORRENCIA_DIA_SEMANA	QUALITATIVA ORDINAIS	TODOS OS DIAS DA SEMANA
OCORRENCIA_PERIODO	QUALITATIVA ORDINAIS	PERÍODOS (MANHÃ; TARDE; NOITE; MADRUGADA)
OCORRENCIA_DIA	QUANTITATIVA CONTÍNUAS	DE 1 ATÉ 31
OCORRENCIA_MES	QUALITATIVA ORDINAIS	12 MESES
ORIGEM_CHAMADO_DESCRICAO	QUALITATIVA NOMINAL	31 ORIGENS
REGIONAL_FATO_NOME	QUALITATIVA NOMINAL	10 REGIONAIS
SECRETARIA_SIGLA	QUALITATIVA NOMINAL	38 SECRETARIAS

Fonte: O AUTOR (2022).

Tomando como apoio o Quadro 11 nota-se que 9 atributos são categorizados como sendo variáveis qualitativos sendo destes 6 nominais e 3 ordinais. Já os outros 2 atributos são categorizados como sendo variáveis quantitativas contínuas.

O primeiro atributo a ser analisado foi o ATENDIMENTO_ANO, onde é possível examinar os anos e seus respectivos números de ocorrências registradas, podendo ser visualizado no Figura 3 .

FIGURA 3 - DISTRIBUIÇÃO DE OCORRÊNCIAS POR ANO

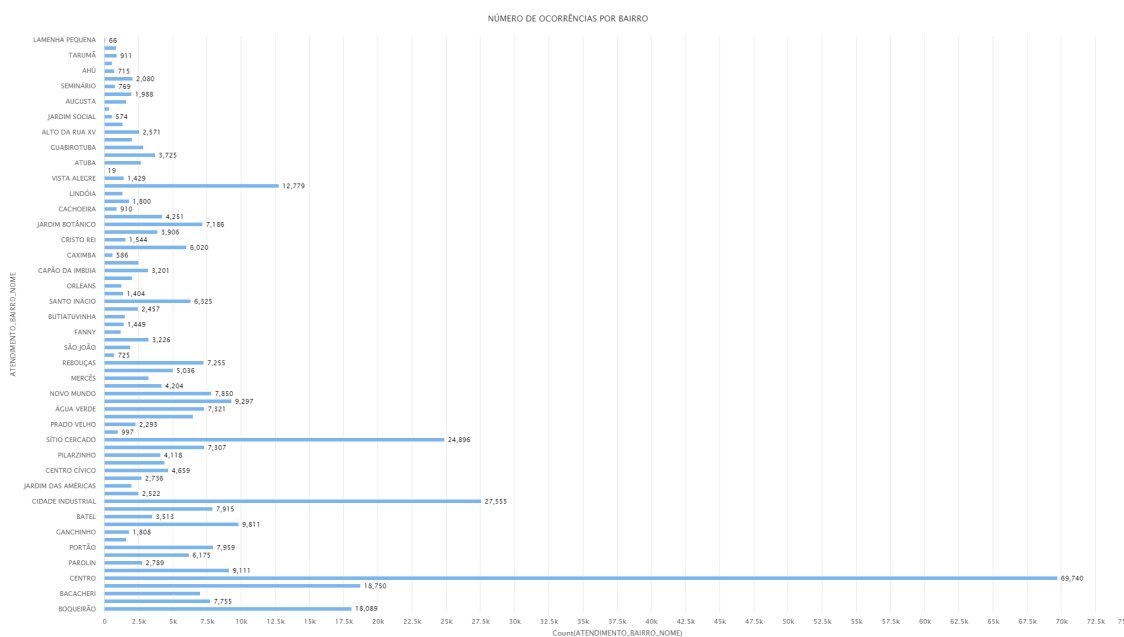


Fonte: O AUTOR (2022).

Podemos observar através da figura 3 que os anos que apresentaram um maior número de registros de ocorrências foram 2020 com (57.389) e 2021 com (58.642) registros, já o ano que apresentou o menor número de ocorrências registradas foi 2012 com (19.761) registros. Apresentando uma variação de 29 mil positivos e 9 mil negativos em relação a média de 29,329 mil registros de 2019 a 2021 (o ano de 2022 não entrou na média por não estar finalizado até o presente momento).

O segundo atributo analisado foi o ATENDIMENTO_BAIRRO_NOME, que apresenta os 75 bairros que compõem o Município de Curitiba, podendo ser visualizado na figura 4.

FIGURA 4 - NÚMERO DE OCORRÊNCIAS POR BAIRRO

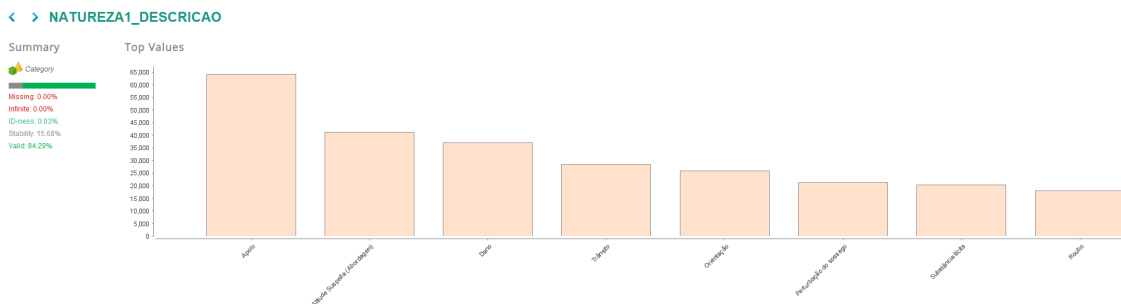


Fonte: O AUTOR (2022).

Observando a figura 4 os 3 bairros que apresentaram o maior número de registros estão o Centro com 69.740 registros, seguido por Cidade Industrial com 27.555 registros e Sítio Cercado com 24.896 registros, já o bairro que apresentou o menor número de registros foi o Riviera com 19 registros.

A terceira análise realizada foi em relação ao atributo NATUREZA1_DESCRICA0, que apresenta 186 variações em relação ao tipo de natureza relatada na ocorrência, podendo ser visualizado na figura 5 as que obtiveram o maior número de registros.

FIGURA 5 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM A NATUREZA



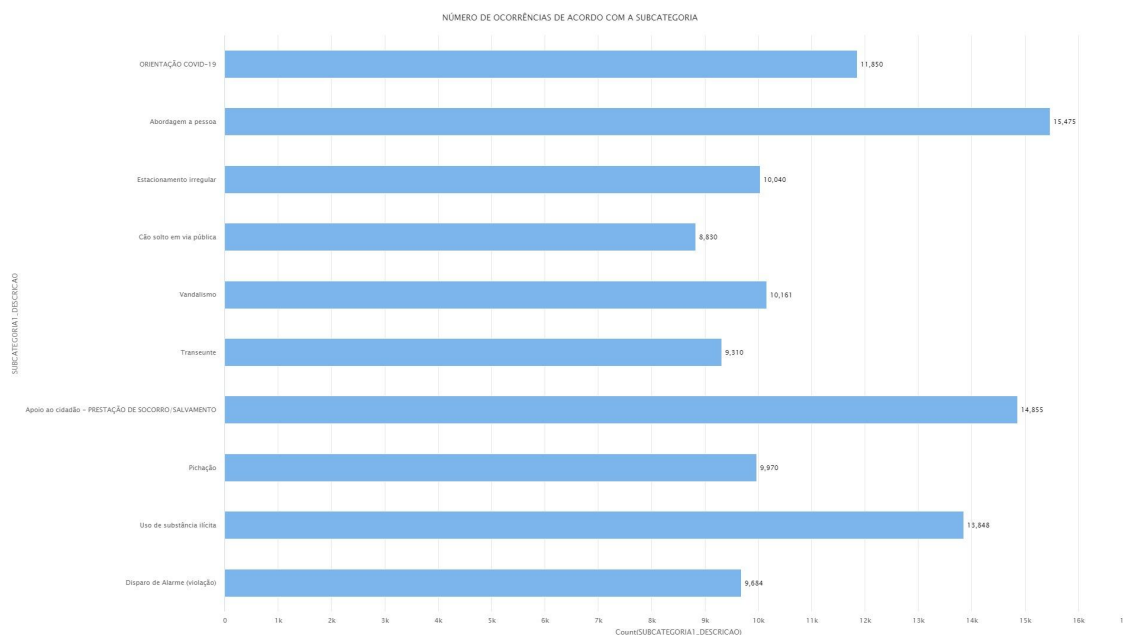
Fonte: O AUTOR (2022).

Com base na figura 5 podemos verificar que os 8 tipos de natureza de ocorrência que mais tiveram números de registros foram Apoio com (64.124) representando 15.92%, Atitude Suspeita (Abordagem) com (41.267) representando 10.24%, Dano com (37.093) representando 9.21%, Trânsito com (28.369) representando 7.04%, Orientação com (25.966) representando 6.45%, Perturbação do sossego com (21.214) representando 5.27%, Substância ilícita com (20.339) representando 5.05% e roubo com (18.099) representando 4.49%. As demais variáveis apresentavam registro inferiores a (15.000) individualmente.

A quarta análise realizada foi referente ao atributo SUBCATEGORIA1_DESCRICAO, que dentre os 402.851 registros selecionados para compor a pesquisa e o único atributo que apresenta dados vazios, devido principalmente ao fato de ser um atributo onde seu campo de preenchimento não é obrigatório, contabilizado 126.539 registros sem dados.

Em relação aos seus dados registrados o atributo SUBCATEGORIA1_DESCRICAO apresenta 210 variações em relação às descrições de subcategorias registradas,

FIGURA 6 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM A SUBCATEGORIA

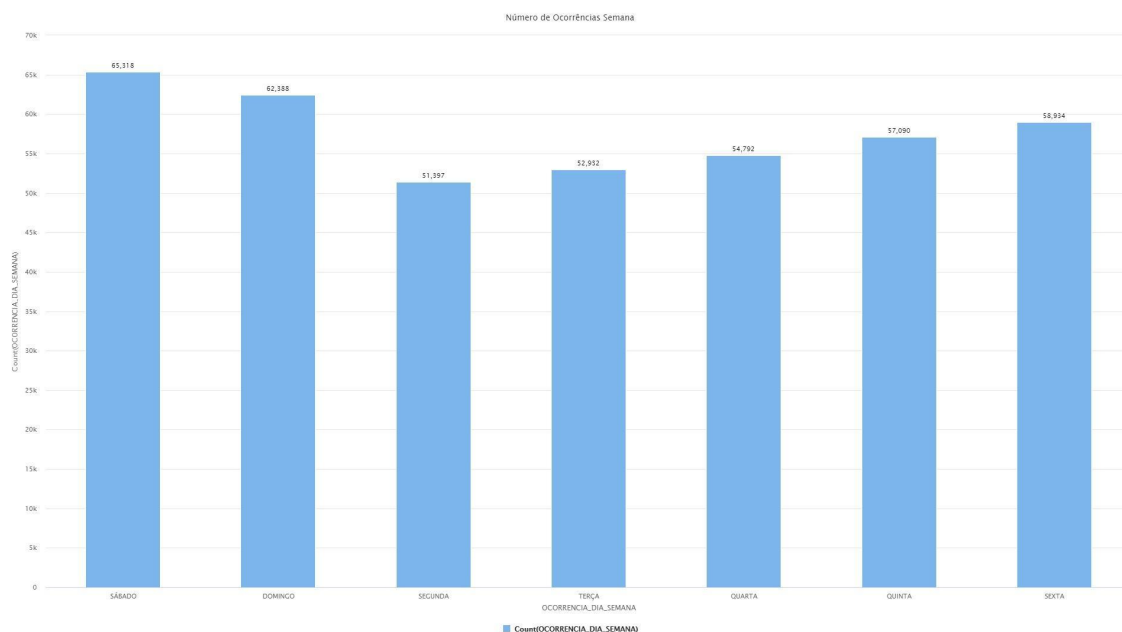


Fonte: O AUTOR (2022).

Através da figura 6 é possível ver as 10 subcategorias que mais apresentaram registros, sendo as 5 que mais tiveram ocorrências registradas a Abordagem a pessoa com (15.475), Apoio ao cidadão com (14.855), Uso de Substância Ilícita com (13.848), Orientação Covid-19 com (11.850), e Vandalismo com (10.161).

A quinta análise realizada apresenta a distribuição do número de ocorrências durante os dias da semana do atributo OCORRENCIA_DIA_SEMANA.

FIGURA 7 - NÚMERO DE OCORRÊNCIAS SEMANAIS

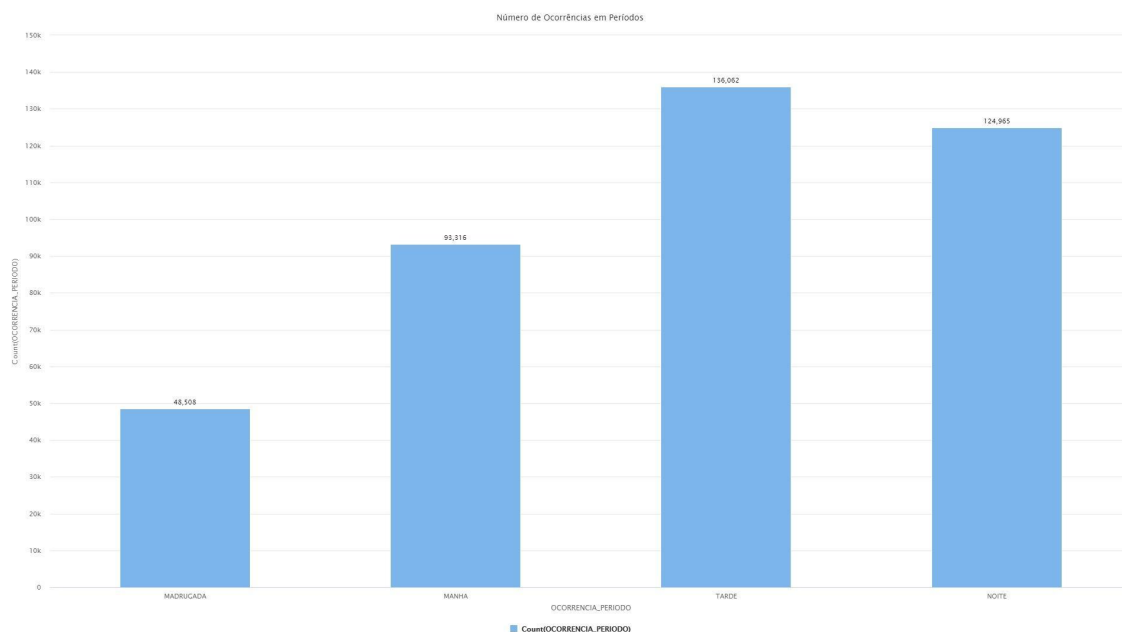


Fonte: O AUTOR (2022).

A partir da figura 7 é notável que o número de ocorrências registradas apresenta sua menor concentração de registros na SEGUNDA com (51.397), e vai subindo ao longo da semana, sendo contabilizados na TERÇA (52.932), na QUARTA (54.792), na QUINTA (57.090) e na SEXTA (58.934), sendo o final de semana a maior concentração de registros sendo o SÁBADO com (65.318) e o DOMINGO com (62.388) registros.

A sexta análise realizada mostra a distribuição dos ocorrências através dos períodos temporais definidos anteriormente para o atributo OCORRENCIA_PERIODO.

FIGURA 8 - NÚMERO DE OCORRÊNCIAS PERÍODO

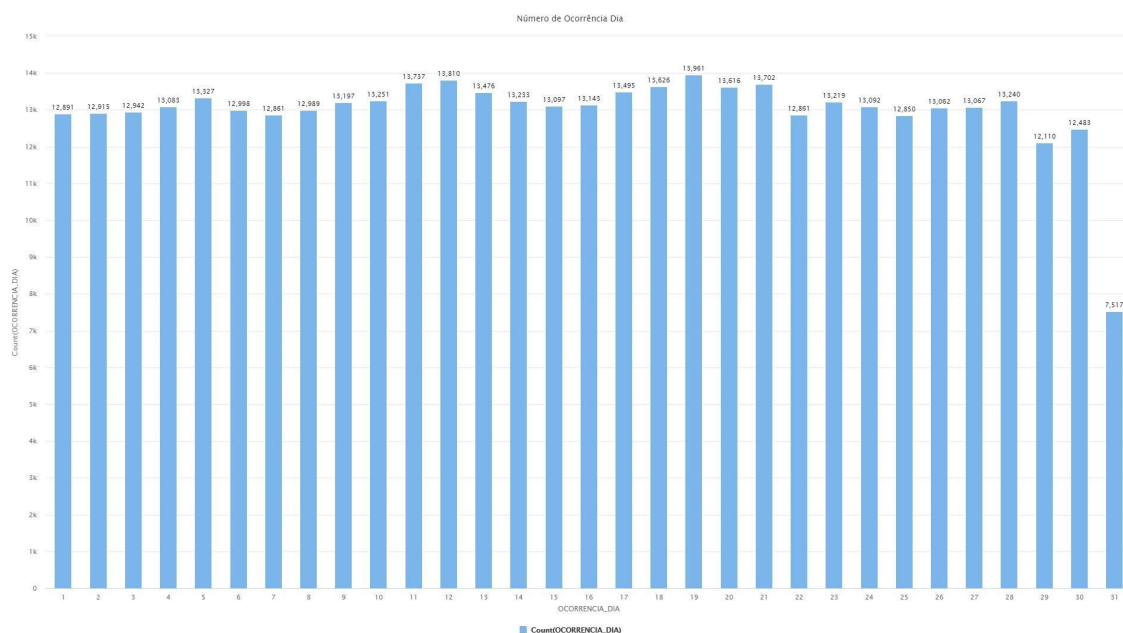


Fonte: O AUTOR (2022).

Observando a figura 8 vemos que a maior concentração de ocorrências está entre os períodos da TARDE com (136.062) representando 33.77% e da NOITE com (124.965) representando 31.02%, já os períodos que apresentam uma menor concentração está a MANHÃ COM (93.316) representando 23.16% e a MADRUGADA com (48.508) representando 12.04%.

A sétima análise realizada mostra a distribuição dos ocorrências durante os dias do mês para o atributo OCCORRENCIA_DIA.

FIGURA 9 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM O DIA

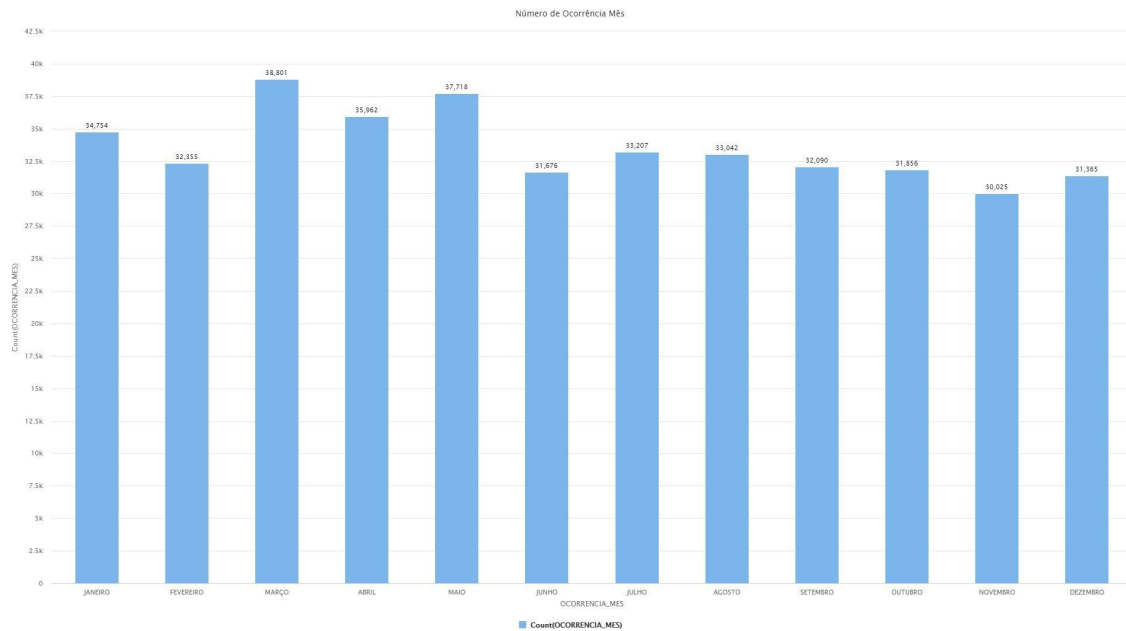


Fonte: O AUTOR (2022).

Por meio da figura 9 podemos perceber que não há variações expressivas em relação a distribuição de ocorrência de acordo com o dia, a não ser pelo dia 31, devido a não existência do mesmo em alguns meses do ano. Porém sua utilização no processo de análise pode vir a ser útil quando comparado a outros atributos, onde podem ter alguma relação de interesse ou foco específico para a realização de análises.

A oitava análise realizada mostra a distribuição das ocorrências durante o mês para o atributo OCORRENCIA_MES.

FIGURA 10 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM O MÊS

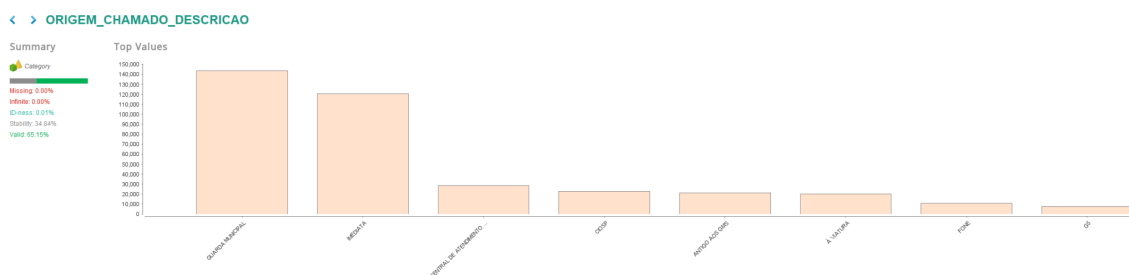


Fonte: O AUTOR (2022).

Através da Figura 10 podemos observar um aumento no número de registros nos primeiros 6 meses, o que é explicado devido a utilização das ocorrências do ano de 2022 até o sexto mês antes da sua extração. Entretanto, os meses não apresentam nenhum resultado discrepante entre si.

A nona análise realizada apresenta a distribuição das ocorrências entre os diferentes tipos de origem o atributo ORIGEM_CHAMADO_DESCRICAÇÃO.

FIGURA 11 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM SUA ORIGEM

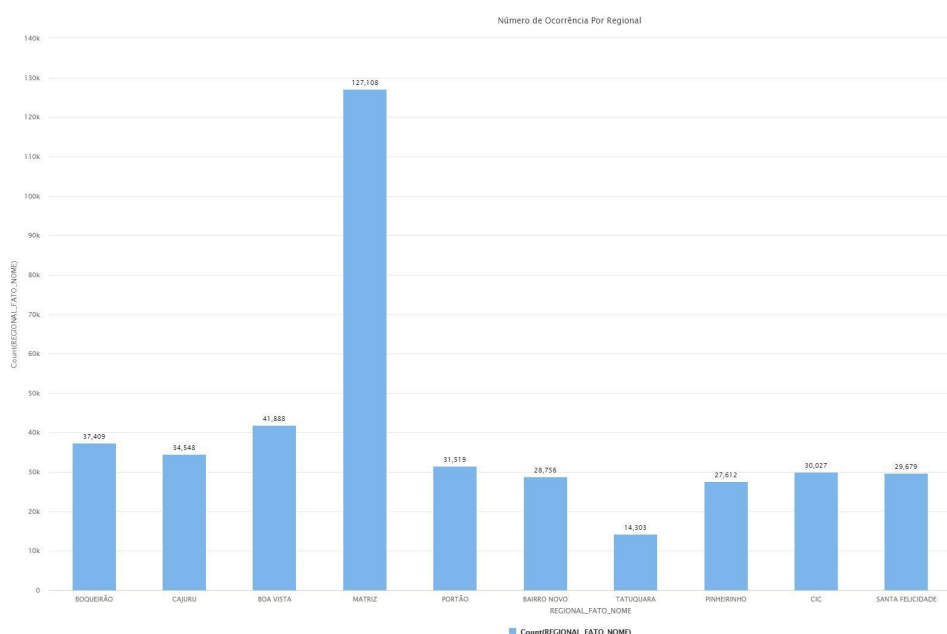


Fonte: O AUTOR (2022).

Podemos notar na figura 11, uma grande diferença entre a distribuição da origem do chamado das ocorrências, tendo o variável GUARDA MUNICIPAL com (143.425) ocorrências correspondendo sozinho a 35.60% do número de ocorrências, e a variável IMEDIATA com (120,783) ocorrências correspondente a 29.98%. Isso se deve pelo motivo dos dois principais meios de origem do chamado serem diretamente ligados a Central da Guarda Municipal e o Policial que está na rua e é solicitado para o cumprimento da sua função.

A décima análise realizada apresenta a distribuição das ocorrências entre as regionais de Curitiba através do atributo REGIONAL_FATO_NOME.

FIGURA 12 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM A REGIONAL



Fonte: O AUTOR (2022).

Observando a figura 12 notamos que em comparação às demais regionais a regional MATRIZ se destaca entre as demais apresentando (127.108) registros, que correspondem a 31.55%. Já a regional TATUQUARA apresenta o menor número de registro com (14.303) correspondente a 3.55%.

A décima primeira e última análise realizada apresenta a distribuição das ocorrências entre as secretarias do atributo SECRETARIA_SIGLA.

FIGURA 13 - NÚMERO DE OCORRÊNCIAS DE ACORDO COM A SECRETARIA



Fonte: O AUTOR (2022).

Através da figura 13 observamos uma grande disparidade na distribuição, a secretaria SMDT com (259.997) ocorrências correspondendo sozinha a 64.54% dos registros. Em seguida aparece a secretaria SMMA com (56.866) ocorrências, correspondendo a 14.12% dos registros.

4.5 DATA MINING


O atual capítulo de Data Mining tem o objetivo de demonstrá-las as etapas de elaboração do processo de mineração da base de dados da Guarda Municipal de Curitiba através da aplicação do algoritmo Naïve Bayes, Deep Learning e Decision Tree.

Como já apresentado no capítulo de metodologia, foi utilizado para a realização do presente trabalho e aplicação dos algoritmos a ferramenta RapidMiner através de sua versão estudantil.

Depois de realizar a instalação e autenticação da sua licença se iniciaram os trabalhos dentro da ferramenta, onde optou-se por utilizar uma extensão do RapidMiner Studio denominada “Auto Model” que acelera o processo de construção e validação de modelos, criando modelos que podem ser verificados ou modificados caso desejado.

Após realizar a importação e carregamento da base de dados já tratada para dentro da extensão Auto Model foi selecionado o atributo chave NATUREZA1_DESCRICAO para ser realizada a sua predição.

FIGURA 14 - SELEÇÃO DOS ATRIBUTOS



Selected: 10 / Total: 10

Selectid	Status	Quality	Name	Correlation	D-ness	Stability	Missing	Text-ness
<input checked="" type="checkbox"/>	●		ATENDIMENTO_ABO	?	0.00%	11.54%	0.00%	0.00%
<input checked="" type="checkbox"/>	●		ATENDIMENTO_SABIDO_NOME	?	0.02%	17.27%	0.00%	19.01%
<input checked="" type="checkbox"/>	●		SUBCATEGORIA_DESCRICAO	?	0.01%	1.00%	31.41%	36.70%
<input checked="" type="checkbox"/>	●		OCCORRENCIA_DIA_SEMANA	?	0.00%	16.27%	0.00%	2.57%
<input checked="" type="checkbox"/>	●		OCCORRENCIA_PERIODO	?	0.00%	23.89%	0.00%	2.44%
<input checked="" type="checkbox"/>	●		OCCORRENCIA_DIA	?	0.11%	3.84%	0.00%	0.20%
<input checked="" type="checkbox"/>	●		OCCORRENCIA_MES	?	0.00%	8.51%	0.00%	2.82%
<input checked="" type="checkbox"/>	●		ORIGEM_CHAMADO_DESCRICAO	?	0.11%	15.52%	0.00%	24.48%
<input checked="" type="checkbox"/>	●		REGIONAL_FATO_NOME	?	0.00%	31.64%	0.00%	11.80%
<input checked="" type="checkbox"/>	●		SECRETARIA_SIGLA	?	0.11%	64.48%	0.00%	1.74%

Fonte: O AUTOR (2022).

Depois de selecionado o atributo chave da predição o Auto Model apresenta uma resumo das características dos atributos, e aponta quais deles apresentam estar de acordo para o processamento, como a base de dados já passou anteriormente pelas etapas anteriores de seleção, limpeza e transformação seus atributos aparecem em verde, sinalizando uma situação positiva em relação aos dados dos atributos.

Após selecionados os atributos que irão compor a predição e selecionado onde será realizado o processamento dos algoritmos, para esse trabalho foi realizado o processamento através de um computador pessoal com capacidade de processamento de 6 cores e 6 Threads e com 16 GBytes de memória RAM.

FIGURA 15 - TIPOS DE MODELOS

The image shows a software interface with three main sections: Models, Data Preparation, and Column Analysis.

Models:

- Naive Bayes
- Generalized Linear Model
- Use Regularization Calculate p-Values
- Logistic Regression
- Fast Large Margin
- Automatically Optimize
- Deep Learning - *Warning: long computation time on this data!*
- Decision Tree - *Warning: long computation time on this data!*
- Automatically Optimize Maximal Depth:
- Random Forest
- Automatically Optimize Number of Trees: Maximal Depth:
- Gradient Boosted Trees
- Automatically Optimize Number of Trees: Maximal Depth: Learning Rate:
- Support Vector Machine
- Automatically Optimize

Data Preparation:

- Remove Columns with Too Many Values
- Maximum Number of Values:
- Extract Date Information
- Extract Text Information
- Select Text Columns (0)
- Number of Extracted Features:
- Automatic Feature Selection
- Additional Minutes (Maximum):
- Final Feature Set should be:
- Automatic Feature Generation
- Function Complexity can be:

Column Analysis:

- Correlations between Columns
- Importance of Columns
- Explain Predictions - *Warning: long computation time on this data!*

Fonte: O AUTOR (2022).

Em seguida são selecionados os modelos que serão executados para a realização da predição, além de opções de preparação de dados e análise de colunas assim como mostrado na Figura 15. Foram selecionados os algoritmos Naïve Bayes, Deep Learning e Decision Tree, além da opção de Explain Predictions onde o Auto Model calcula as previsões para um conjunto de testes e todas as linhas com um valor em falta para a coluna de destino, se este parâmetro for verificado, o Auto Modelo irá calcular adicionalmente quais os valores de coluna que tiveram maior impacto nas previsões de cada modelo específico em cada linha. Em seguida é dado início ao processamentos dos algoritmos um por vez, assim que é finalizado o primeiro algoritmo os seus resultados são apresentados para consulta, enquanto o próximo algoritmo é executado.

4.6 INTERPRETAÇÃO E AVALIAÇÃO DOS RESULTADOS

Ao realizar a extração da base de dados Guarda Municipal de Curitiba através do Portal de Dados Abertos de Curitiba foi obtido os registros de ocorrências do período de 2009 até o mês de junho (06) de 2022, contabilizando (403.325) registros de ocorrência com (35) atributos ao todo. Porém através dos processos de Seleção, Pré-processamento/Limpeza dos

dados e Transformação dos dados fizeram com que o número de atributos reduzisse para (11) e o número de registros para (402.851).

No capítulo 4.4 é possível identificar alguns padrões dentro da base de dados. Como observado que a maior concentração de ocorrências está entre os anos de 2020 com (57.389) e 2021 com (58.642) registros. A principal localidade da ocorrência se originou no bairro Centro com (69.740) registros, seguido por Cidade Industrial com (27.555) registros e Sítio Cercado com (24.896) registros, porém em comparação o bairro Riviera apresentou apenas (19) registros. Em relação a Natureza da ocorrência a variável Apoio com (64.124), Dano com (37.093), Trânsito com (28.369) e Orientação com (25.966) correspondendo juntas a praticamente metade de todas as ocorrências. Já para a Subcategoria dos registros foi decidido manter os (126.539) registros que se encontravam vazios para preservar os dados dos demais atributos, já em relação aos suas variáveis que foram mantidas se destacam a Abordagem a pessoa com (15.475), Apoio ao cidadão com (14.855), Uso de Substância Ilícita com (13.848), Orientação Covid-19 com (11.850) e Vandalismo com (10.161). As ocorrências ao longo da semana apresentaram um padrão crescente ao longo da semana, se iniciando na segunda com (51.397) registros e se concentrando aos finais de semana no sábado com (65.318) e no domingo com (62.388) registros. Para os registros definidos através dos 4 períodos de tempo observamos que a tarde com (136.062) representando 33.77% e a noite com (124.965) representando 31.02% se destacam, o período da manhã com (93.316) representando 23.16% e a madrugada com (48.508) representando 12.04%, porém como percebido é mostrado na exemplificação da figura 19, 20 e 21 determinados tipos de ocorrência tem uma tendência de ocorrerem no períodos específicos. Em relação a variação da distribuição dos dias ao longo do mês (do dia 1 ao dia 31) observamos que não há uma discrepância a não ser o dia 31 que é explicada por não ocorrer em determinados meses do ano, porém quando relacionados com os demais atributos da pesquisa podemos analisar datas específicas e obter informações úteis. Para a distribuição ao longo dos meses podemos observar que os 6 primeiros meses do ano tiveram uma leve aumento, devido a utilização dos

mesmos seis meses do ano de 2022, onde se deu a extração da base. A origem do chamado se destacam as variáveis guarda municipal com (143.425) e imediata com (120,783) por serem duas variáveis ligadas diretamente a guarda municipal de Curitiba, e não origem secundárias de chamados repassada para a mesma fazer a averiguação. Em relação às regionais que compõem a cidade de Curitiba, se destacam a matriz por corresponder sozinha a (127.108) registros, que correspondem a 31.55% e a regional Tatuquara que apresenta o menor número de registro com (14.303) correspondente a 3.55%. Por fim as secretarias que se destacaram estão a SMDT com (259.997) ocorrências correspondendo sozinha a 64.54% dos registros, seguida pela secretaria SMMA com (56.866) ocorrências, correspondendo a 14.12% dos registros totais.

Ao final do processo de mineração foi obtido os resultados entre os modelos como apresentado na figura 16, onde em comparação entre eles se destacaram o modelo Naive Bayes pelo seu maior ganho de informação e o modelo Deep Learning por ter sido o que teve a maior taxa de acerto e o melhor desempenho em relação ao tempo de processamento. O modelo Decision Tree embora seja de fácil interpretação a sua visualização dos seus resultados através do desenvolvimento de seu tronco até sua folha acabou se tornando inviável, devido ao seu enorme tamanho, não sendo possível apresentar seus resultados por meio da visualização da árvore gerada.

FIGURA 16 - RESULTADOS OBTIDOS



Fonte: O AUTOR (2022).

A Seguir será apresentado os resultados obtidos para cada um dos três algoritmos utilizados no processo de mineração.

4.6.1 Resultado do Modelo Naive Bayes

Esse operador gera um modelo de classificação Naive Bayes através do RapidMiner Studio Core. O Naive Bayes é um classificador de alta polaridade e baixa variação, baseado no teorema de Bayes utilizando de densidades de probabilidade gaussianas para modelar os dados do Atributo. Em relação aos casos típicos de utilização estão a categorização de texto, incluindo detecção de spam, análise de sentimentos, e sistemas de recomendação. O pressuposto fundamental de Naive Bayes é que, dado o valor da etiqueta (a classe), o valor de qualquer atributo é independente do valor de qualquer outro atributo. A rigor, esta suposição raramente é verdadeira (é "ingênua!"), a sua utilização mostra que o classificador Naive Bayes funciona frequentemente bem.

Em relação a performance e resultados o modelo Naive Bayes obteve uma Accuracy de 36.2%, ou seja o modelo obteve uma taxa de acerto de 36.2% das previsões a partir dos resultados obtidos através da matriz de confusão. Além disso, o modelo obteve o maior Gains (46,414) entre os modelos comparados, que corresponde ao ganho de informação obtida e

melhorada através dos nós gerados. Em relação ao tempo total de processamento, o modelo levou ao todo 60 horas, 6 minutos e 18 segundos para chegar ao seu resultado final.

FIGURA 17 - PESO ATRIBUTOS NAIVE BAYES

Naive Bayes - Weights

Attribute	Weight
SECRETARIA_SIGLA	0.137
ORIGEM_CHAMADO_DESCRICAO	0.124
REGIONAL_FATO_NOME	0.062
OCORRENCIA_PERIODO	0.046
OCORRENCIA_MES	0.043
ATENDIMENTO_ANO	0.038
OCORRENCIA_DIA_SEMANA	0.038
OCORRENCIA_DIA	0.027

Fonte: O AUTOR (2022).

Através da Figura 17 os 3 atributo que apresentarem maior peso são respectivamente a SECRETARIA_SIGLA com (0.137); ORIGEM_CHAMADO_DESCRICHÃO com (0.124) e REGIONAL_FATOR_NOME com (0.062).

4.6.2 Resultado do Modelo Deep Learning

O Deep Learning através do cluster H2O 3.30.0.1. multinível para aprender relacionamentos não lineares baseia-se numa rede neural artificial multinível para aprender relacionamentos não lineares por meio da alimentação de várias camadas que são treinadas com descida de gradiente estocástico usando retro propagação. A rede pode conter um grande número de camadas ocultas que consistem em neurônios com funções tanh, rectificador e maxout activação. Apresenta características avançadas tais como taxa de aprendizagem adaptativa, recozimento de taxa, treino de impulso, desistência e regularização L1 ou L2 permitem uma elevada precisão preditiva. Cada nó de computação treina uma cópia dos parâmetros do modelo global nos seus dados locais com multithreading (assíncrono), e contribui periodicamente para

o modelo global através da média do modelo em toda a rede. O operador inicia um cluster H2O local de 1 nó e executa o algoritmo sobre o mesmo. Embora utilize um nó, a execução é paralela. É possível definir o nível de paralelismo alterando as definições/Preferências/Gerais/Número de fios. Por defeito, utiliza o número recomendado de threads para o sistema.

Em relação a performance do modelo Deep Learning obteve em comparação aos demais modelos testados a maior accuracy de 36.9%, ou seja o modelo acertou 36.9% das previsões por meio dos resultados gerados pela matriz de confusão. O modelo obteve um Gains de (29,994) que corresponde ao ganho de informação obtido através dos nós gerados. Em relação ao tempo de processamento, o modelo Deep Learning foi o mais rápido entre os modelos testados, durando ao todo 33 horas, 8 minutos e 46 segundos até a obtenção do seu resultado final.

FIGURA 18 - PESO ATRIBUTOS DEEP LEARNING

Deep Learning - Weights

Attribute	Weight ↓
SECRETARIA_SIGLA	0.161
ATENDIMENTO_ANO	0.130
ORIGEM_CHAMADO_DESCRICAÇÃO	0.111
REGIONAL_FATO_NOME	0.071
OCORRENCIA_DIA	0.059
OCORRENCIA_DIA_SEMANA	0.050
OCORRENCIA_PERIODO	0.048
OCORRENCIA_MES	0.044

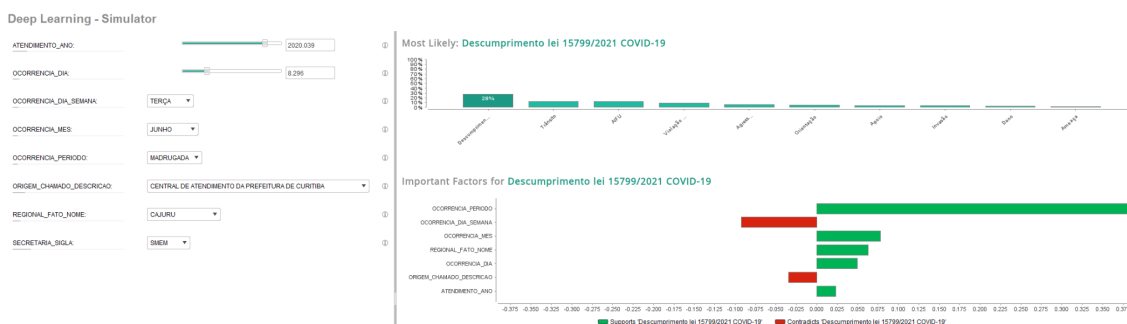
Fonte: O AUTOR (2022).

Como mostrado na Figura 18 os 3 atributo que apresentaram maior peso de acordo com o modelo são SECRETARIA_SIGLA com (0.161); ATENDIMENTO_ANO com (0.130) e ORIGEM_CHAMADO_DESCRICAÇÃO com (0.111).

Por ter apresentado uma maior *accuracy* e um menor tempo de processamento entre os outros modelos testados obtuse por fazer 3 exemplos através de um simulador disponível que utiliza os resultados obtidos através do modelo selecionado. Por meio de uma opção de otimização disponível dentro

do simulador, e após selecionar qual classe deseja ser otimizada dentro do atributo NATUREZA1_DESCRICAO. Todos testes realizados dentro do simulador foram definidos para maximizar a confiança do atributo selecionado, deixando o desvio padrão dentro de 2 em torno da média, e sem utilizar limitador de tempo para execução até o encontro das entradas ideais.

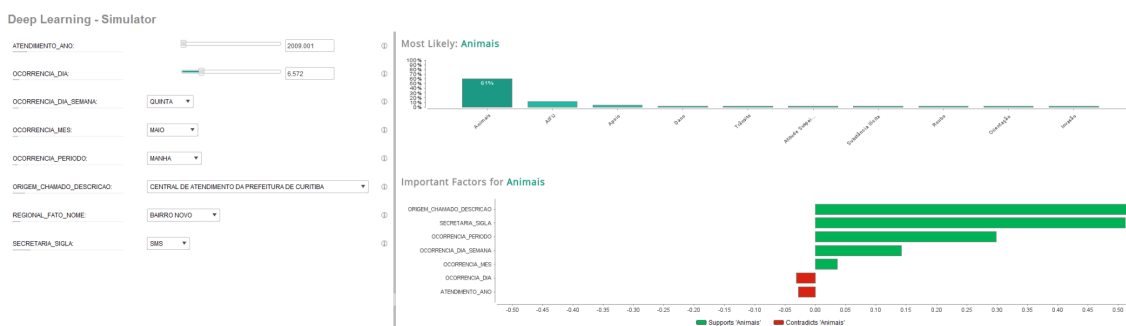
FIGURA 19 - SIMULADOR VARIÁVEL DESCUMPRIMENTO LEI 15799/2021 COVID-19



Fonte: O AUTOR (2022).

Através da análise da figura 18 podemos observar que o simulador apresentou como sendo o maior apoio para a variável (Descumprimento da lei 15799/2021 COVID-19) o atributo OCORRENCIA_PERIODO através da variável (MADRUGADA), o atributo OCORRENCIA_MES com a variável (JUNHO) e REGIONAL FATOR NOME com a variável (CAJURU) também apresentam um apoio embora mais sucinto, já o atributo OCORRENCIA_DIA_SEMANA com a variável (TERÇA) e o atributo ORIGEM_CHAMADO_DESCRICAO com o atributo (CENTRAL DE ATENDIMENTO DA PREFEITURA DE CURITIBA) contradizem a sua ocorrência.

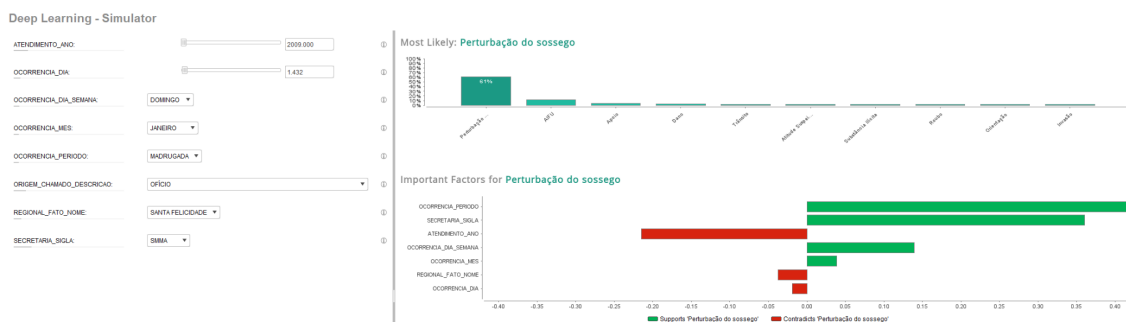
FIGURA 20 - SIMULADOR VARIÁVEL ANIMAL



Fonte: O AUTOR (2022).

Para a análise da figura 20 podemos observar que o simulador apresentou como sendo o maior apoio para a variável (Animal) o atributo ORIGEM_CHAMADO_DESCRICAO através da variável (CENTRAL DE ATENDIMENTO DA PREFEITURA DE CURITIBA), o atributo SECRETARIA_SIGLA com o variável (SMS), o atributo OCORRENCIA_PERIODO com a variável (MANHA), e o atributo OCORRENCIA_DIA_SEMANA com a variável (QUINTA).

FIGURA 21 - SIMULADOR VARIÁVEL



Fonte: O AUTOR (2022).

Com a análise da figura 21 podemos observar que o simulador apresentou como sendo o maior apoio para a variável (Perturbação do sossego) o atributo OCORRENCIA_PERIODO através da variável (MADRUGADA), o atributo SECRETARIA_SIGLA através da variável (SMMS), o atributo OCORRENCIA_DIA_SEMANA através da variável (DOMINGO). Já o

atributo ATENDIMENTO_ANO através da variável (2009) contradiz a sua ocorrência.

4.6.3 Resultado Do Modelo Decision Tree

Este Operador gera um modelo de Decision Tree, que pode ser utilizado para classificação e regressão, e construída uma árvore como uma colecção de nós destinada a criar uma decisão sobre a afiliação de valores a uma classe ou uma estimativa de um valor alvo numérico. Cada nó representa uma regra de divisão para um atributo específico. Para classificação esta regra separa valores pertencentes a diferentes classes, para regressão separa-os a fim de reduzir o erro de uma forma óptima para o critério do parâmetro selecionado. A construção de novos nós é repetida até que o critério de paragem seja cumprido. Uma previsão para a etiqueta de classe atributo é determinada dependendo da maioria dos exemplos que atingiram esta folha durante a geração, enquanto uma estimativa para um valor numérico é obtida através da média dos valores de uma folha. Este Operador pode processar conjuntos de exemplos contendo atributos tanto nominais como numéricos. Após geração, o modelo de árvore de decisão pode ser aplicado a novos exemplos utilizando o operador de aplicação do modelo. Cada exemplo segue os ramos da árvore de acordo com a regra de divisão até que uma folha seja alcançada.

Em relação a performance do modelo Decision Tree em comparação com os outros modelos obteve a menor *accuracy* de 29.5%, obtida através dos resultados gerados pela matriz de confusão. Em relação ao ganho de informação seu Gains obteve (19,438) através dos seus nós gerados. O tempo de processamento total do seu processamento levou ao todo 34 horas, 15 minutos e 55 segundos até a obtenção dos seus resultados finais.

FIGURA 22 - PESO ATRIBUTOS DECISION TREE

Decision Tree - Weights

Attribute	Weight
ORIGEM_CHAMADO_DESCRICAÇÃO	0,102
OCORRENCIA_PERIODO	0,065
REGIONAL_FATO_NOME	0,061
OCORRENCIA_DIA	0,058
ATENDIMENTO_ANO	0,049
SECRETARIA_SIGLA	0,043
OCORRENCIA_MES	0,035
OCORRENCIA_DIA_SEMANA	0,035

Fonte: O AUTOR (2022).

Como mostrado na Figura 22 os 3 atributo que apresentaram maior peso são ORIGEM_CHAMADO_DESCRICAÇÃO com (0.102); OCORRENCIA_PERIODO com (0.065) e REGIONAL_FATO_NOME com (0.061).

5. CONSIDERAÇÕES FINAIS

O crescimento do número de dados vem aumentando constantemente a todo momento através de diversas formas de tecnologia, muitas vezes esses dados são gerados de forma automática através dos sistemas que os concebem, ou ainda de forma híbrida combinando dados gerados a partir de informações obtidas através do sistema e dos usuários do mesmo, através de processo que pode variar de acordo para que a função que o sistema foi criado. Esses dados observados separadamente não apresentam nenhum significado nem valor, não tendo sentido algum, porém quando organizados e ordenados com o intuito de transmitir algum significado dentro de um determinado contexto se tornam informações. Muitas bases de dados acabam sendo utilizadas apenas como conjunto de registros gerados por meio de uma atividade corriqueira, podendo ser utilizados a partir de um objetivo para obter informações e até mesmo conhecimentos úteis para seus usuários e/ou público em geral. A lei de dados abertos possibilitou uma maior transparência e acesso a dados gerados pelas organizações públicas do país, proporcionando o desenvolvimento de pesquisas a partir desses dados, como o caso do presente trabalho, que busca responder através da elaboração do mesmo “De que maneira as ocorrências atendidas pela Guarda Municipal de Curitiba podem ser previstas?”. A desenvolvimento da pesquisa proporcionou a utilização dos conceitos e passo do KDD através da sua utilização prática no trabalho, por meio da Seleção dos dados e atributos que comporão o trabalho, na Limpeza dos dados realizada através das correções, padronizações e adequação dos dados além da remoção de ruídos encontrados, na Transformação dos dados para a realização de uma melhor análise e na criação de novos campos a partir do aproveitamento de dados pertencentes de outros atributos, no processo de Mineração de Dado através da utilização da plataforma RapidMiner que proveu o meio de processamento dos algoritmo junto com a visualização dos mesmos para a realização da predição e Interpretação dos Resultados alcançados, respondendo assim o problema levantado, através de técnicas de mineração de dados aplicadas a base de dados da SiGesGuarde e demonstradas ao longo do capítulo 4.6.

5.1 ALCANCE DOS OBJETIVOS

O objetivo geral do trabalho focou-se na identificação de padrões presentes na base de dados da Guarda Municipal de Curitiba, a fim de possibilitar a descoberta de informação e conhecimento na mesma. Esse objetivo foi alcançado como mostrado no capítulo 4.4 através das informações obtidas nos gráficos gerados, e no capítulo 4.6 através das informações apresentadas e das exemplificações demonstradas através dos resultados gerados. Em relação aos objetivos específicos também foram atendidos.

O primeiro objetivo refere a utilização da aplicação de métodos de mineração de dados para descoberta de conhecimento, que foi utilizada e demonstrada do trabalho, através a utilização do KDD para o desenvolvimento do trabalho e da utilização dos algoritmos Naïve Bayes, Deep Learning e Decision Tree.

O segundo objetivo se referia a comparação e analisar os resultados alcançados por meio das análises estatísticas e mineração de dados, demonstrados através do capítulo 4.6 por meio da comparação dos resultados obtidos entre os três algoritmos testados.

O terceiro objetivo referia-se a definir a ferramenta responsável por realizar o processo de mineração e de visualização dos resultados, que foi a utilização da ferramenta RapidMiner.

5.2 TRABALHOS FUTUROS

Embora tenha sido feita uma análise prévia dos atributos e variáveis que melhor explicassem cada ocorrência, foi notado que dentro do atributo como por exemplo NATUREZA_DESCRIBÇÃO algumas variáveis poderiam se juntar ou serem incorporadas uma nas outras por se aproximarem do tipo de ocorrência, como a variável Animais e a variável Ataque de cão feroz.

Outra sugestão refere se a inserção de dados referentes a latitude e longitude dos bairros de Curitiba, como uma forma de implementar a utilização de mapas interativos para agregar uma maior valor a apresentação dos resultados, facilitando a interpretação do mesmo, podendo ser utilizados através de como *Power BI*, *Tableau* e até mesmo o próprio *Rapidminer*.

Também é recomendado a utilizar outros algoritmos e ferramentas de ciência de dados como meio de comparação entre os resultados obtidos podendo gerar análises diversificadas.

REFERÊNCIAS

A GUARDA MUNICIPAL. **Secretaria Municipal da Defesa Social**, 2013, Disponível em: <http://www.defesasocial.curitiba.pr.gov.br/guarda-municipal.html>. Acesso em: 10 de abr. de 2022.

CARVALHO, C. F. **Guarda municipal: O que você precisa saber sobre guarda municipal e nunca teve a quem perguntar**. 3. ed. Curitiba: Edição do autor, 2011.

CASTRO, Leandro Nunes; FERRARI, Daniel Gomes. **Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações**. São Paulo: Editora Saraiva, 2016.

CHOO, Chun Wei. **A organização do conhecimento: Como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões**. Tradução Eliana Rocha. - São Paulo: Editora Senac São Paulo, 2003.

DAVENPORT, Thomas H. **Ecologia da Informação: porque só a tecnologia não basta para o sucesso na era da informação**. 2. ed. São Paulo: Futura, 1998.

SILVA, Leandro Augusto; PERES, Sarajane M.; BOSCARIOLI, Clodis. **Introdução à Mineração de Dados - Com Aplicações em R**. [Digite o Local da Editora]: Grupo GEN, 2016. 9788595155473. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788595155473/>. Acesso em: 24 abr. 2022.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, [S. l.], v. 17, n. 3, p. 37, 1996. DOI: 10.1609/aimag.v17i3.1230. Disponível em: <https://ojs.aaai.org/index.php/aimagazine/article/view/1230>. Acesso em: 24 apr. 2022.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas S.a, 2008. 220 p.

GOVERNO DIGITAL. **Dados Abertos Governamentais**, 2019. Disponível em: <https://www.gov.br/governodigital/pt-br/dados-abertos/dados-abertos-governamentais>>. Acesso em: 13 mar. 22.

GUARDA MUNICIPAL - Central de Emergência: 153. **Prefeitura Municipal De Curitiba**, Disponível em: <https://www.curitiba.pr.gov.br/servicos/guarda-municipal-central-de-emergencia-153/2>. Acesso em: 10 de abr. de 2022.

SOUZA, Angela Ferreira de. **Análise de dados em base da Guarda Municipal de Curitiba**. Orientadora: Prof.^a Dr.^a Denise Fukumi Tsunoda. 2021. 160 f. TCC (Graduação) – Curso de Gestão da Informação, Departamento de Gestão da Informação, Universidade Federal do Paraná, Curitiba, 2021.

Disponível em: <https://acervodigital.ufpr.br/handle/1884/70255>. Acessado em: 03 abr. 2022.

TARAPANOFF, Kira (Org.). **Inteligência, informação e conhecimento**. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia: Ict; UNESCO, 2006. 453 p.

PORTAS ABERTAS - Veja como entrar em contato com as Regionais e qual delas cuida do seu bairro. **Prefeitura Municipal de Curitiba**, Curitiba. Disponível em: <https://www.curitiba.pr.gov.br/noticias/veja-como-entrar-em-contato-com-as-regionais-e-qual-delas-cuida-do-seu-bairro/40912>>. Acesso em: 19 de jul. de 2022.

APÊNDICE

APÊNDICE A - INCOERÊNCIA REGIONAL BOA VISTA

BAIRRO	TOTAL DE OCORRÊNCIAS	O QUE FOI FEITO?
ABRANCHES	2016	PERMANECEU
AHÚ	8	TRANSFERIDO PARA REGIONAL MATRIZ
ALTO DA RUA XV	1	TRANSFERIDO PARA REGIONAL MATRIZ
ATUBA	2666	PERMANECEU
BACACHERI	7010	PERMANECEU
BAIRRO ALTO	4409	PERMANECEU
BARREIRINHA	3224	PERMANECEU
BELAS AGUAS	1	REMOVIDO
BOA VISTA	5023	PERMANECEU
BOM RETIRO	1	TRANSFERIDO PARA REGIONAL MATRIZ
CABRAL	12	TRANSFERIDO PARA REGIONAL MATRIZ
CACHOEIRA	908	PERMANECEU
CAPÃO DA IMBUIA	1	TRANSFERIDO PARA REGIONAL CAJURU
CENTRO	1	TRANSFERIDO PARA REGIONAL MATRIZ
GUABIROTUBA	1	TRANSFERIDO PARA REGIONAL CAJURU
HUGO LANGE	1	TRANSFERIDO PARA REGIONAL MATRIZ
JARDIM BELA VISTA	1	REMOVIDO
JARDIM BOTÂNICO	1	TRANSFERIDO PARA REGIONAL MATRIZ

JARDIM SOCIAL	3	TRANSFERIDO PARA REGIONAL MATRIZ
JR TAISA	1	REMOVIDO
JUVEVÊ	3	TRANSFERIDO PARA REGIONAL MATRIZ
LOTEAMENTO SÃO GERÔN	1	REMOVIDO
MARACANÃ	1	REMOVIDO
MAUA	1	REMOVIDO
PILARZINHO	4086	PERMANECEU
SANTA CÂNDIDA	6166	PERMANECEU
SANTA FELICIDADE	3	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
SÃO FRANCISCO	4	TRANSFERIDO PARA REGIONAL MATRIZ
SÃO JOÃO	3	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
SÃO LOURENÇO	1798	PERMANECEU
TABOÃO	1986	PERMANECEU
TARUMÃ	491	PERMANECEU
TINGUI	1592	PERMANECEU
VISTA ALEGRE	6	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE

Fonte: O AUTOR (2022).

APÊNDICE B - INCOERÊNCIA REGIONAL BOQUEIRÃO

BAIRRO	TOTAL DE OCORRÊNCIAS	O'QUE FOI FEITO?
AHÚ	1	TRANSFERIDO PARA

		REGIONAL MATRIZ
ALTO BOQUEIRÃO	9110	PERMANECEU
ALTO DA GLÓRIA	1	TRANSFERIDO PARA REGIONAL MATRIZ
BARRO PRETO	2	REMOVIDO
BOQUEIRÃO	18079	PERMANECEU
CAJURU	2	TRANSFERIDO PARA REGIONAL CAJURU
CENTRO	4	TRANSFERIDO PARA REGIONAL MATRIZ
CIDADE INDUSTRIAL	1	TRANSFERIDO PARA REGIONAL CIC
FANNY	3	TRANSFERIDO PARA REGIONAL PINHEIRINHO
GUABIROTUBA	2	TRANSFERIDO PARA REGIONAL CAJURU
HAUER	3718	PERMANECEU
JARDIM BOTÂNICO	1	TRANSFERIDO PARA REGIONAL MATRIZ
JARDIM DAS AMÉRICAS	1	TRANSFERIDO PARA REGIONAL CAJURU
LINDÓIA	1	TRANSFERIDO PARA REGIONAL PINHEIRINHO
PAROLIN	9	TRANSFERIDO PARA REGIONAL PORTÃO
PINHEIRINHO	3	TRANSFERIDO PARA REGIONAL PINHEIRINHO
PRADO VELHO	4	TRANSFERIDO PARA REGIONAL MATRIZ
REBOUÇAS	2	TRANSFERIDO PARA REGIONAL MATRIZ
SÃO MIGUEL	1	TRANSFERIDO PARA REGIONAL CIC

SÍTIO CERCADO	9	TRANSFERIDO PARA REGIONAL BAIRRO NOVO
TATUQUARA	1	TRANSFERIDO PARA REGIONAL TATUQUARA
UBERABA	3	TRANSFERIDO PARA REGIONAL CAJURU
UMBARÁ	1	TRANSFERIDO PARA REGIONAL BAIRRO NOVO
XAXIM	6437	PERMANECEU

Fonte: O AUTOR (2022).

APÊNDICE C - INCOERÊNCIA REGIONAL CAJURU

BAIRRO	TOTAL DE OCORRÊNCIAS	O'QUE FOI FEITO?
BACACHERI	1	TRANSFERIDO PARA REGIONAL BOA VISTA
BAIRRO ALTO	3	TRANSFERIDO PARA REGIONAL BOA VISTA
BOQUEIRÃO	5	TRANSFERIDO PARA REGIONAL BOQUEIRÃO
CACHOEIRA	1	TRANSFERIDO PARA REGIONAL BOA VISTA
CAJURU	18698	PERMANECEU
CAMPO COMPRIDO	1	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
CAPÃO DA IMBUIA	3198	PERMANECEU
CIDADE INDUSTRIAL	1	TRANSFERIDO PARA REGIONAL CIC
CRISTO REI	6	TRANSFERIDO PARA REGIONAL MATRIZ
FANNY	1	TRANSFERIDO PARA REGIONAL PINHEIRINHO

GUABIROTUBA	2861	PERMANECEU
HAUER	2	TRANSFERIDO PARA REGIONAL BOQUEIRÃO
JARDIM BOTÂNICO	3	TRANSFERIDO PARA REGIONAL MATRIZ
JARDIM DAS AMÉRICAS	1976	PERMANECEU
JARDIM SANTA MÔNICA	1	REMOVIDO
JARDIM SOCIAL	8	TRANSFERIDO PARA REGIONAL MATRIZ
PORTÃO	1	TRANSFERIDO PARA REGIONAL PORTÃO
PRADO VELHO	3	TRANSFERIDO PARA REGIONAL MATRIZ
SANTA FELICIDADE	1	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
SÃO FRANCISCO	2	TRANSFERIDO PARA REGIONAL MATRIZ
SEMINÁRIO	1	TRANSFERIDO PARA REGIONAL PORTÃO
SÍTIO CERCADO	2	TRANSFERIDO PARA REGIONAL BAIRRO NOVO
TABOÃO	1	TRANSFERIDO PARA REGIONAL BOA VISTA
TARUMÃ	417	TRANSFERIDO PARA REGIONAL BOA VISTA
UBERABA	7753	PERMANECEU
UMBARÁ	1	TRANSFERIDO PARA REGIONAL BAIRRO NOVO

Fonte: O AUTOR (2022).

APÊNDICE D - INCOERÊNCIA REGIONAL CIC

BAIRRO	TOTAL DE	O'QUE FOI FEITO?
--------	----------	------------------

	OCORRÊNCIAS	
ÁGUA VERDE	3	TRANSFERIDO PARA REGIONAL PORTÃO
AUGUSTA	1600	PERMANECEU
BACACHERI	1	TRANSFERIDO PARA REGIONAL BOA VISTA
BARIGUI	1	REMOVIDO
BATEL	2	TRANSFERIDO PARA REGIONAL MATRIZ
BIGORRILHO	1	TRANSFERIDO PARA REGIONAL MATRIZ
BOA VISTA	1	TRANSFERIDO PARA REGIONAL BOA VISTA
BOM RETIRO	2	TRANSFERIDO PARA REGIONAL MATRIZ
CABRAL	2	TRANSFERIDO PARA REGIONAL MATRIZ
CAJURU	2	TRANSFERIDO PARA REGIONAL CAJURU
CAMPINA DO SIQUEIRA	8	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
CAMPO COMPRIDO	196	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
CAMPO DE SANTANA	1	TRANSFERIDO PARA REGIONAL TATUQUARA
CAPÃO RASO	3	TRANSFERIDO PARA REGIONAL PINHEIRINHO
CENTRO	4	TRANSFERIDO PARA REGIONAL MATRIZ
CENTRO CÍVICO	10	TRANSFERIDO PARA REGIONAL MATRIZ
CIDADE INDUSTRIAL	27477	PERMANECEU

FANNY	2	TRANSFERIDO PARA REGIONAL PINHEIRINHO
FAZENDINHA	23	TRANSFERIDO PARA REGIONAL PORTÃO
GUABIROTUBA	1	TRANSFERIDO PARA REGIONAL CAJURU
GUAÍRA	1	TRANSFERIDO PARA REGIONAL PORTÃO
HAUER	2	TRANSFERIDO PARA REGIONAL BOQUEIRÃO
JARDIM BOTÂNICO	1	TRANSFERIDO PARA REGIONAL MATRIZ
MERCÊS	3	TRANSFERIDO PARA REGIONAL MATRIZ
NOVO MUNDO	11	TRANSFERIDO PARA REGIONAL PINHEIRINHO
ORLEANS	4	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
PINHAIS	1	REMOVIDO
PINHEIRINHO	3	TRANSFERIDO PARA REGIONAL PINHEIRINHO
PORTÃO	1	TRANSFERIDO PARA REGIONAL PORTÃO
REBOUÇAS	1	TRANSFERIDO PARA REGIONAL MATRIZ
RIVIERA	19	PERMANECEU
SANTA FELICIDADE	4	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
SANTA QUITÉRIA	1	TRANSFERIDO PARA REGIONAL PORTÃO
SÃO MIGUEL	852	PERMANECEU
SÍTIO CERCADO	2	TRANSFERIDO PARA

		REGIONAL BAIRRO NOVO
TATUQUARA	4	TRANSFERIDO PARA REGIONAL TATUQUARA
UMBARÁ	1	TRANSFERIDO PARA REGIONAL BAIRRO NOVO
XAXIM	1	TRANSFERIDO PARA REGIONAL BOQUEIRÃO

Fonte: O AUTOR (2022).

APÊNDICE E - INCOERÊNCIA REGIONAL MATRIZ

BAIRRO	TOTAL DE OCORRÊNCIAS	O'QUE FOI FEITO?
ÁGUA VERDE	55	TRANSFERIDO PARA REGIONAL PORTÃO
AHÚ	715	PERMANECEU
ALTO BOQUEIRÃO	1	TRANSFERIDO PARA REGIONAL BOQUEIRÃO
ALTO DA GLÓRIA	2080	PERMANECEU
ALTO DA RUA XV	2570	PERMANECEU
ATUBA	1	TRANSFERIDO PARA REGIONAL BOA VISTA
BACACHERI	3	TRANSFERIDO PARA REGIONAL BOA VISTA
BAIRRO ALTO	1	TRANSFERIDO PARA REGIONAL BOA VISTA
BARREIRINHA	1	TRANSFERIDO PARA REGIONAL BOA VISTA
BATEL	3496	PERMANECEU
BIGORRILHO	4088	PERMANECEU
BOM RETIRO	723	PERMANECEU
BOQUEIRÃO	3	TRANSFERIDO PARA REGIONAL BOQUEIRÃO

CABRAL	2457	PERMANECEU
CACHOEIRA	1	TRANSFERIDO PARA REGIONAL BOA VISTA
CAJURU	48	TRANSFERIDO PARA REGIONAL CAJURU
CAMPINA DO SIQUEIRA	4	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
CAPÃO DA IMBUIA	2	TRANSFERIDO PARA REGIONAL CAJURU
CASCATINHA	3	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
CAXIMBA	1	TRANSFERIDO PARA REGIONAL TATUQUARA
CENTRO	69732	PERMANECEU
CENTRO CÍVICO	4659	PERMANECEU
CIDADE INDUSTRIAL	5	TRANSFERIDO PARA REGIONAL CIC
COLÔNIA FARIA	1	REMOVIDO
CRISTO REI	1545	PERMANECEU
GANCHINHO	1	TRANSFERIDO PARA REGIONAL BAIRRO NOVO
GUABIROTUBA	1	TRANSFERIDO PARA REGIONAL CAJURU
HAUER	1	TRANSFERIDO PARA REGIONAL BOQUEIRÃO
HUGO LANGE	556	PERMANECEU
JARDIM BOTÂNICO	7186	PERMANECEU
JARDIM DAS AMÉRICAS	1	TRANSFERIDO PARA REGIONAL CAJURU
JARDIM SANTA MÔNICA	1	REMOVIDO

JARDIM SOCIAL	574	PERMANECEU
JUVEVÊ	997	PERMANECEU
LINDÓIA	1	TRANSFERIDO PARA REGIONAL PINHEIRINHO
MERCÊS	3222	PERMANECEU
NOVO MUNDO	3	TRANSFERIDO PARA REGIONAL PINHEIRINHO
PAROLIN	20	TRANSFERIDO PARA REGIONAL PORTÃO
PINHEIRINHO	2	TRANSFERIDO PARA REGIONAL PINHEIRINHO
PORTÃO	1	TRANSFERIDO PARA REGIONAL PORTÃO
PRADO VELHO	2289	PERMANECEU
REBOUÇAS	7215	PERMANECEU
SANTA CÂNDIDA	4	TRANSFERIDO PARA REGIONAL BOA VISTA
SANTA FELICIDADE	2	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
SANTA MÔNICA	1	REMOVIDO
SANTA QUITÉRIA	1	TRANSFERIDO PARA REGIONAL PORTÃO
SANTO INÁCIO	3	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
SÃO FRANCISCO	12777	PERMANECEU
SÃO LOURENÇO	2	TRANSFERIDO PARA REGIONAL BOA VISTA
SEMINÁRIO	3	TRANSFERIDO PARA REGIONAL PORTÃO
TARUMÃ	1	TRANSFERIDO PARA REGIONAL BOA VISTA

TINGUI	2	TRANSFERIDO PARA REGIONAL BOA VISTA
UBERABA	1	TRANSFERIDO PARA REGIONAL CAJURU
VILA IZABEL	2	TRANSFERIDO PARA REGIONAL PORTÃO
VISTA ALEGRE	4	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
XAXIM	2	TRANSFERIDO PARA REGIONAL BOQUEIRÃO

Fonte: O AUTOR (2022).

APÊNDICE F - INCOERÊNCIA REGIONAL PINHEIRINHO

BAIRROS	TOTAL DE OCORRÊNCIAS	O'QUE FOI FEITO?
ABRANCHES	1	TRANSFERIDO PARA REGIONAL BOA VISTA
ÁGUA VERDE	3	TRANSFERIDO PARA REGIONAL PORTÃO
BIGORRILHO	1	TRANSFERIDO PARA REGIONAL MATRIZ
BUTIATUVINHA	1	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
CAJURU	2	TRANSFERIDO PARA REGIONAL CAJURU
CAMPO DE SANTANA	1281	TRANSFERIDO PARA REGIONAL TATUQUARA
CAPÃO DA IMBUIA	1	TRANSFERIDO PARA REGIONAL CAJURU
CAPÃO RASO	9289	PERMANECEU
CAXIMBA	213	TRANSFERIDO PARA REGIONAL TATUQUARA

CENTRO	2	TRANSFERIDO PARA REGIONAL MATRIZ
CIDADE INDUSTRIAL	31	TRANSFERIDO PARA REGIONAL CIC
FANNY	670	PERMANECEU
FAZENDINHA	1	TRANSFERIDO PARA REGIONAL PORTÃO
GUAÍRA	4	TRANSFERIDO PARA REGIONAL PORTÃO
HAUER	2	TRANSFERIDO PARA REGIONAL BOQUEIRÃO
LINDÓIA	705	PERMANECEU
NOVO MUNDO	4417	PERMANECEU
PAROLIN	9	TRANSFERIDO PARA REGIONAL PORTÃO
PINHEIRINHO	7912	PERMANECEU
PORTÃO	15	TRANSFERIDO PARA REGIONAL PORTÃO
PRADO VELHO	3	TRANSFERIDO PARA REGIONAL MATRIZ
SANTO INÁCIO	1	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
SÍTIO CERCADO	14	TRANSFERIDO PARA REGIONAL BAIRRO NOVO
TARUMÃ	1	TRANSFERIDO PARA REGIONAL BOA VISTA
TATUQUARA	3682	TRANSFERIDO PARA REGIONAL TATUQUARA
UMBARÁ	4	TRANSFERIDO PARA REGIONAL BAIRRO NOVO
XAXIM	43	TRANSFERIDO PARA REGIONAL BOQUEIRÃO

Fonte: O AUTOR (2022).

APÊNDICE G - INCOERÊNCIA REGIONAL PORTÃO

BAIRROS	TOTAL DE OCORRÊNCIAS	O'QUE FOI FEITO?
ÁGUA VERDE	7321	PERMANECEU
ALTO DA RUA XV	1	TRANSFERIDO PARA REGIONAL MATRIZ
BACACHERI	1	TRANSFERIDO PARA REGIONAL BOA VISTA
BATEL	16	TRANSFERIDO PARA REGIONAL MATRIZ
BOA VISTA	12	TRANSFERIDO PARA REGIONAL BOA VISTA
CAMPINA DO SIQUEIRA	11	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
CAMPO COMPRIDO	1002	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
CAPÃO RASO	6	TRANSFERIDO PARA REGIONAL PINHEIRINHO
CENTRO	5	TRANSFERIDO PARA REGIONAL MATRIZ
CIDADE INDUSTRIAL	8	TRANSFERIDO PARA REGIONAL CIC
FANNY	544	TRANSFERIDO PARA REGIONAL PINHEIRINHO
FAZENDINHA	7306	PERMANECEU
GUABIROTUBA	3	TRANSFERIDO PARA REGIONAL CAJURU
GUAÍRA	2522	PERMANECEU
JARDIM BOTÂNICO	1	TRANSFERIDO PARA REGIONAL MATRIZ

LINDÓIA	631	TRANSFERIDO PARA REGIONAL PINHEIRINHO
MAUÁ	1	REMOVIDO
MERCÊS	2	TRANSFERIDO PARA REGIONAL MATRIZ
NOVO MUNDO	3433	TRANSFERIDO PARA REGIONAL PINHEIRINHO
PAROLIN	2790	PERMANECEU
PINHEIRINHO	3	TRANSFERIDO PARA REGIONAL PINHEIRINHO
PORTÃO	7958	PERMANECEU
PRADO VELHO	1	TRANSFERIDO PARA REGIONAL MATRIZ
REBOUÇAS	40	TRANSFERIDO PARA REGIONAL MATRIZ
SANTA CÂNDIDA	3	TRANSFERIDO PARA REGIONAL BOA VISTA
SANTA QUITÉRIA	1448	PERMANECEU
SÃO MIGUEL	1	TRANSFERIDO PARA REGIONAL CIC
SEMINÁRIO	457	PERMANECEU
SÍTIO CERCADO	1	TRANSFERIDO PARA REGIONAL BAIRRO NOVO
TINGUI	11	TRANSFERIDO PARA REGIONAL BOA VISTA
VILA IZABEL	1404	PERMANECEU
XAXIM	1	TRANSFERIDO PARA REGIONAL BOQUEIRÃO

Fonte: O AUTOR (2022).

APÊNDICE H - INCOERÊNCIA REGIONAL SANTA FELICIDADE

BAIRROS	TOTAL DE OCORRÊNCIAS	O'QUE FOI FEITO?
ABRANCHES	1	TRANSFERIDO PARA REGIONAL BOA VISTA
BARREIRINHA	1	TRANSFERIDO PARA REGIONAL BOA VISTA
BATEL	1	TRANSFERIDO PARA REGIONAL MATRIZ
BIGORRILHO	115	TRANSFERIDO PARA REGIONAL MATRIZ
BOM RETIRO	2	TRANSFERIDO PARA REGIONAL MATRIZ
BOQUEIRÃO	1	TRANSFERIDO PARA REGIONAL BOQUEIRÃO
BUTIATUVINHA	1515	PERMANECEU
CAMPINA DO SIQUEIRA	2736	PERMANECEU
CAMPO COMPRIDO	4248	PERMANECEU
CAMPO LARGO	1	REMOVIDO
CAPÃO RASO	1	TRANSFERIDO PARA REGIONAL PINHEIRINHO
CASCATINHA	340	PERMANECEU
CENTRO	1	TRANSFERIDO PARA REGIONAL MATRIZ
CIDADE INDUSTRIAL	26	TRANSFERIDO PARA REGIONAL CIC
COLONIA SAO VENANCIO	1	REMOVIDO
LAMENHA PEQUENA	67	PERMANECEU
MERCÊS	36	TRANSFERIDO PARA REGIONAL MATRIZ
MOSSUNGUÊ	1357	PERMANECEU
NOVO MUNDO	1	TRANSFERIDO PARA

		REGIONAL PINHEIRINHO
ORLEANS	1266	PERMANECEU
PILARZINHO	32	TRANSFERIDO PARA REGIONAL BOA VISTA
SANTA CÂNDIDA	2	TRANSFERIDO PARA REGIONAL BOA VISTA
SANTA FELICIDADE	6020	PERMANECEU
SANTA QUITÉRIA	1	TRANSFERIDO PARA REGIONAL PORTÃO
SANTO INÁCIO	6325	PERMANECEU
SÃO BRAZ	2490	PERMANECEU
SÃO FRANCISCO	2	TRANSFERIDO PARA REGIONAL MATRIZ
SÃO GERONIMO	1	REMOVIDO
SÃO JOÃO	1885	PERMANECEU
SEMINÁRIO	313	TRANSFERIDO PARA REGIONAL PORTÃO
TABOÃO	1	TRANSFERIDO PARA REGIONAL BOA VISTA
TINGUI	4	TRANSFERIDO PARA REGIONAL BOA VISTA
UBERABA	1	TRANSFERIDO PARA REGIONAL CAJURU
VILA GRAZIELA	1	REMOVIDO
VISTA ALEGRE	1429	PERMANECEU

Fonte: O AUTOR (2022).

APÊNDICE I - INCOERÊNCIA REGIONAL TATUQUARA

BAIRROS	TOTAL DE OCORRÊNCIAS	O'QUE FOI FEITO?
BACACHERI	1	TRANSFERIDO PARA

		REGIONAL BOA VISTA
BOQUEIRÃO	1	TRANSFERIDO PARA REGIONAL BOQUEIRÃO
CAMPO COMPRIDO	3	TRANSFERIDO PARA REGIONAL SANTA FELICIDADE
CAMPO DE SANTANA	3906	PERMANECEU
CAPÃO RASO	1	TRANSFERIDO PARA REGIONAL PINHEIRINHO
CAXIMBA	586	PERMANECEU
CIDADE INDUSTRIAL	8	TRANSFERIDO PARA REGIONAL CIC
FAZENDINHA	1	TRANSFERIDO PARA REGIONAL PORTÃO
PORTÃO	1	TRANSFERIDO PARA REGIONAL PORTÃO
TARUMÃ	1	TRANSFERIDO PARA REGIONAL BOA VISTA
TATUQUARA	9811	PERMANECEU
UMBARÁ	2	TRANSFERIDO PARA REGIONAL BAIRRO NOVO