



UNIVERSIDADE FEDERAL DO MARANHÃO

Bacharelado em Ciência da Computação

Francisco Victor Maciel Miranda Calvet

**Aplicação de Mineração de Dados a partir de  
ocorrências atendidas pelo Corpo de Bombeiros  
Militar do Maranhão**

São Luís

2019

Francisco Victor Maciel Miranda Calvet

# **Aplicação de Mineração de Dados a partir de ocorrências atendidas pelo Corpo de Bombeiros Militar do Maranhão**

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Ms. Francisco Glaubos Nunes Clímaco

São Luís

2019

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).  
Núcleo Integrado de Bibliotecas/UFMA

Calvet, Francisco Victor Maciel Miranda.

Aplicação de Mineração de Dados a partir de ocorrências atendidas pelo Corpo de Bombeiros Militar do Maranhão / Francisco Victor Maciel Miranda Calvet. - 2019.

49 f.

Orientador(a): Francisco Glaubos Nunes Clímaco.

Monografia (Graduação) - Curso de Ciência da Computação, Universidade Federal do Maranhão, 12/07/2019, 2019.

1. Mineração de Dados. 2. Orange. 3. Regras de Associação. 4. Weka. I. Clímaco, Francisco Glaubos Nunes. II. Título.

Francisco Victor Maciel Miranda Calvet

## **Aplicação de Mineração de Dados a partir de ocorrências atendidas pelo Corpo de Bombeiros Militar do Maranhão**

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

---

**Prof. Ms. Francisco Glaubos Nunes  
Clímaco**

Orientador

Universidade Federal do Maranhão

---

**Prof. Ms. Carlos Eduardo Portela  
Serra de Castro**

Examinador 1

Universidade Federal do Maranhão

---

**Prof.<sup>a</sup> Esp. Inez Cavalcanti Dantas**

Examinador 2

Universidade Federal do Maranhão

São Luís

2019

# Agradecimentos

Agradeço, primeiramente, ao Deus Todo-Poderoso pelo dom da vida. Sem Ele nada teria sido criado. Ele é o autor e consumidor da minha fé. As Suas misericórdias têm se renovado na minha vida a cada manhã.

Agradeço ao Senhor Jesus, que morreu na cruz em meu lugar e ao Espírito Santo, meu fiel amigo, Consolador, que tem guiado os meus passos e derramado da Sua sabedoria de forma extraordinária.

Agradeço à Santíssima Trindade por ter colocado em meu caminho pessoas tão especiais, que não mediram esforços em me ajudar durante a graduação e realização deste trabalho de conclusão de curso.

Agradeço também, especialmente, ao meu orientador, Prof. Ms. Francisco Glaubos Nunes Clímaco, que acreditou em mim e me deu todo suporte necessário, demonstrando grande paciência e sabedoria no norteamo de todas as etapas deste trabalho.

A todos os professores do curso de computação, pelos quais, desde o começo da minha graduação, sempre tive enorme respeito e admiração.

Aos meus pais, Francisco da Silva Calvet e Rosana Martins Miranda Calvet, pelo apoio e incentivo que me permitiram chegar até aqui, por serem incansáveis na dedicação em me guiar nos caminhos corretos.

À minha esposa, Raynara Araújo Calvet, que sempre se mostrou uma mulher sábia e sempre me apoiou, não medindo esforços para que eu chegasse até esta etapa de minha vida.

Ao meu precioso filho, Victor Daniel Araújo Calvet, de apenas 3 anos, que todas as noites dizia para eu estudar, se referindo a elaboração deste trabalho de conclusão de curso. Ele ficava brincando perto de mim enquanto eu redigia cada capítulo.

Aos meus amigos pelo apoio e paciência que tiveram todos os momentos. Especialmente, agradeço a Alex Eduardo de Aguiar Barbosa e a seu irmão Allan Fábio de Aguiar Barbosa, ambos graduados em Ciência da Computação pela UFMA. Eles foram meus maiores incentivadores para prestar vestibular para área de computação.

Não podia deixar de agradecer também aos meus colegas de serviço, Hilmar Nogueira, gestor de estatística e análise criminal da Secretaria de Segurança do Maranhão, e ao sargento do Corpo de Bombeiros Jailson Dias, que foram fundamentais na execução deste trabalho, fornecendo o banco de dados com as ocorrências do Corpo de Bombeiros Militar do Maranhão.

*"Tudo quanto fizerdes, fazei-o de todo o coração, como para o Senhor e não para homens."*

Colossenses 3: 23 ARA

# Resumo

O volume de dados armazenados em sistemas computacionais tem crescido nos mais diversos segmentos. Conseqüentemente, isso tem feito com que empresas e instituições desprendessem maiores esforços e investissem mais numa área da computação chamada Mineração de Dados. A Mineração de Dados tem como objetivo extrair conhecimentos de grandes bases de dados. Neste trabalho é utilizado uma técnica de Mineração de Dados conhecida como Regras de Associação, a fim de se obter conhecimento sobre as ocorrências atendidas pelo Corpo de Bombeiros Militar do Maranhão, entre os anos de 2013 e 2018. Para a extração desse conhecimento foram utilizados os softwares Weka e Orange, ambos de domínio livre.

**Palavras-chaves:** Mineração de Dados, Regras de Associação, Weka, Orange.

# Abstract

The volume of data stored in computer systems has grown in the most diverse segments. Consequently, this has led companies and institutions to make greater efforts and invest more in a field of computing called Data Mining. Data Mining aims to extract knowledge from large databases. This work uses a Data Mining technique known as Association Rules in order to obtain knowledge about the occurrences attended by the Military Fire Brigade of Maranhão, between the years of 2013 and 2018. For the extraction of this knowledge were used the softwares Weka and Orange, both of free domain.

**Keywords:** Data Mining, Association Rules, Weka, Orange.



# Lista de ilustrações

Figura 1 – Etapas do KDD. . . . .	16
Figura 2 – Algoritmo Apriori. . . . .	22
Figura 3 – Função Apriori-gen: passo de junção. . . . .	23
Figura 4 – Função Apriori-gen: passo de poda. . . . .	23
Figura 5 – Tela principal do Weka. . . . .	25
Figura 6 – Weka Explorer com destaque na região que contem as abas. . . . .	26
Figura 7 – Aba Associate no Weka. . . . .	26
Figura 8 – Exemplo de fluxo de trabalho no Orange. . . . .	27
Figura 9 – Base de dados disponibilizada no formato XLSX. . . . .	31
Figura 10 – Atributo BAIRRO com 854 valores distintos. . . . .	33
Figura 11 – Base de dados depois de adicionada a coluna BAIRROS_20_MAIS. . . . .	33
Figura 12 – Atributo TIPO com 141 valores distintos. . . . .	34
Figura 13 – Base de dados depois de adicionada a coluna TIPO_15_MAIS. . . . .	34
Figura 14 – Atributo SUBTIPO_500 com seus 19 valores distintos. . . . .	35
Figura 15 – Exemplo de registro onde BAIRRO está preenchido e AREA não. . . . .	35
Figura 16 – Mesmo bairro com diferentes grafias do nome. . . . .	35
Figura 17 – Os 10 bairros com maior número de ocorrências atendidas. . . . .	36
Figura 18 – Total de ocorrências atendidas por agencia/batalhão. . . . .	37
Figura 19 – Total de ocorrências atendidas por classe de ocorrência. . . . .	38
Figura 20 – Total de ocorrências atendidas por dia da semana. . . . .	38
Figura 21 – Ocorrências atendidas por turno. . . . .	39
Figura 22 – Ocorrências atendidas por classe de ocorrência e turno. . . . .	39
Figura 23 – Os 15 tipos de ocorrências mais frequentes. . . . .	40
Figura 24 – Os 10 tipos de incêndios mais frequentes. . . . .	40
Figura 25 – Incêndios florestais por mês. . . . .	41
Figura 26 – Parâmetros iniciais do <i>Apriori</i> . . . . .	41
Figura 27 – Regras geradas com 90% de confiança e suporte mínimo de 10%. . . . .	42
Figura 28 – Parte do experimento II, após remoção do atributo AGENCIA. . . . .	42
Figura 29 – Cinco melhores regras do experimento II. . . . .	43
Figura 30 – Regras geradas após o experimento III. . . . .	44
Figura 31 – Regras geradas após o experimento IV. . . . .	44

# Lista de tabelas

Tabela 1 – Registro de vendas de um supermercado. . . . .	20
Tabela 2 – Ocorrências por municípios. . . . .	37

# Lista de abreviaturas e siglas

1ª CIBM	1ª Companhia Independente Especializada de Bombeiros
1º BBBM	1º Batalhão de Bombeiros Militar
2º BBBM	2º Batalhão de Bombeiros Militar
BBA	Batalhão de Bombeiros Ambiental
BBEM	Batalhão de Bombeiros de Emergências Médicas
BBMAR	Batalhão de Bombeiros Marítimo
BBS	Batalhão de Busca e Salvamento
CBMMA	Corpo de Bombeiros Militar do Maranhão
CIOPS	Centro Integrado de Operações de Segurança do Estado
GPL	General Public License
KDD	Knowledge Discovery in Databases
NEAC	Núcleo de Estatística e Análise Criminal
SSP/MA	Secretaria de Segurança Pública do Maranhão

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
<b>1.1</b>	<b>Objetivos</b>	<b>13</b>
1.1.1	Objetivos Específicos	13
<b>1.2</b>	<b>Organização do Trabalho</b>	<b>13</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>15</b>
<b>2.1</b>	<b>Descoberta de Conhecimento em Base de Dados</b>	<b>15</b>
<b>2.2</b>	<b>Tarefas da Mineração de Dados</b>	<b>17</b>
2.2.1	Regras de Associação	19
2.2.2	O algoritmo <i>Apriori</i>	21
<b>3</b>	<b>FERRAMENTAS PARA MINERAÇÃO DE DADOS</b>	<b>24</b>
<b>3.1</b>	<b>Weka</b>	<b>24</b>
<b>3.2</b>	<b>Orange</b>	<b>26</b>
<b>4</b>	<b>METODOLOGIA E CONTEXTUALIZAÇÃO DO PROBLEMA</b>	<b>28</b>
<b>4.1</b>	<b>As ocorrências do Corpo de Bombeiros Militar do Maranhão</b>	<b>29</b>
<b>5</b>	<b>EXPERIMENTOS E ANÁLISE DOS RESULTADOS</b>	<b>31</b>
<b>5.1</b>	<b>Aplicação da Mineração de Dados no banco de ocorrências</b>	<b>31</b>
<b>5.2</b>	<b>Alguns dados estatísticos sobre as ocorrências do Corpo de Bombeiros Militar do Maranhão</b>	<b>36</b>
5.2.1	Os 10 bairros com maior número de ocorrências atendidas	36
5.2.2	Quantitativo de ocorrências atendidas na região metropolitana de São Luís	37
5.2.3	Comparativo de atendimentos por dia da semana e turno	38
5.2.4	Os 15 tipos de ocorrências mais frequentes	39
5.2.5	Os 10 tipos de incêndios mais frequentes	39
<b>5.3</b>	<b>Aplicação do algoritmo <i>Apriori</i></b>	<b>41</b>
<b>6</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS</b>	<b>46</b>
	<b>REFERÊNCIAS</b>	<b>48</b>

# 1 Introdução

Os recentes avanços tecnológicos têm permitido às empresas coletar e armazenar uma grande quantidade de dados referentes às suas atividades. Estes dados podem esconder conhecimentos valiosos e úteis para apoiar decisões estratégicas. Ao mesmo tempo, a utilização de computadores tem permitido a extração automática de conhecimento destes dados.

Estudos indicam que o volume de dados em uma organização dobra a cada cinco anos. A maioria das organizações não sofre de falta de dados, mas sim de uma abundância de dados redundantes e inconsistentes difíceis de usar para o suporte à decisão (SHEN; ZHANG; YANG, 2002). Neste contexto as empresas enfrentam o desafio de organizar os seus dados e extrair o conhecimento que se esconde nesta grande massa de dados.

De acordo com Silva, Peres e Boscarioli (2016, p. 10):

“‘Encontrar o que era desconhecido, o que estava escondido’ pode ser entendido como o ato de descobrir. Considerando que as bases de dados são geralmente volumosas e que o conhecimento pode estar implícito, faz-se necessário um trabalho de busca detalhado – o que, metaforicamente, diz-se ser um trabalho de ‘mineração’ – associado a um processo analítico, sistemático e, até onde possível, automatizado.”

É nesse cenário que, finalmente, a Mineração de Dados (do inglês, Data Mining) é definida. De forma simplificada, a mineração de dados provê um método automático para descobrir padrões em dados, sem a tendenciosidade e a limitação de uma análise baseada meramente na intuição humana (BRAGA, 2005).

Neste trabalho, foram utilizadas duas ferramentas de Mineração de Dados, *Weka*<sup>1</sup> e *Orange*<sup>2</sup>, em conjunto com a técnica de Regras de Associação, objetivando-se extrair conhecimento relevante de um banco de dados com ocorrências atendidas pelo Corpo de Bombeiros Militar do Maranhão, entre janeiro de 2013 e agosto de 2018, na região metropolitana de São Luís.

A aplicação da Mineração de dados tornará viável uma análise para se observar padrões e tendências existentes nos dados. Será possível saber, por exemplo, quais as ocorrências mais frequentes por bairros, quais as áreas com maior índice de determinadas ocorrências, quais os tipos de incêndios mais frequentes etc.

<sup>1</sup> Disponível para download em <<https://www.cs.waikato.ac.nz/ml/weka/downloading.html>>

<sup>2</sup> Disponível para download em <<https://orange.biolab.si/download/>>

## 1.1 Objetivos

O objetivo principal deste estudo é identificar, por meio da aplicação da tecnologia de mineração de dados, regras sobre as ocorrências atendidas pelo Corpo de Bombeiros Militar do Maranhão na região metropolitana de São Luís - MA, entre os anos de 2013 e 2018, a partir dos dados fornecidos pela Secretaria de Segurança Pública do Estado do Maranhão.

### 1.1.1 Objetivos Específicos

Para que seja possível atingir o objetivo principal descrito acima, é essencial que durante o processo, sejam alcançados os seguintes objetivos específicos:

- Aplicar técnicas de mineração de dados no banco relacionado de ocorrências atendidas pelo Corpo do Bombeiros Militar do Maranhão entre os anos de 2013 e 2018;
- Identificar regras de associação no banco relacionado de ocorrências atendidas pelo Corpo do Bombeiros Militar do Maranhão entre os anos de 2013 e 2018;
- Destacar as melhores regras associativas segundo a medida de interesse - Confidence (Confiança);
- Identificar, a partir dos dados analisados, bairros e áreas com maior densidade de ocorrências;
- Identificar os principais tipos de ocorrências atendidas.

## 1.2 Organização do Trabalho

Esta monografia está dividida em seis capítulos principais. Neste capítulo introdutório é apresentada uma explicação sobre o escopo deste trabalho, seguida dos objetivos, geral e específicos.

A partir do Capítulo 2 é apresentado o referencial teórico, explicando de forma detalhada o processo de Descoberta de Conhecimento em Base de Dados, também conhecida como KDD (termo proveniente do inglês, Knowledge Discovery in Databases) e como o mesmo é utilizado neste trabalho. Em seguida, são abordadas as tarefas da Mineração de Dados, de acordo com a definição de alguns autores. No mesmo capítulo, explica-se o que são as Regras de Associação e o funcionamento do algoritmo *Apriori*, método essencial para alcançar o objetivo geral anteriormente especificado.

O Capítulo 3 mostra o funcionamento das ferramentas aqui utilizadas, Weka e Orange. O Capítulo 4 apresenta a metodologia utilizada e contextualiza o problema,

explicando como e onde as ocorrências atendidas pelo Corpo de Bombeiros Militar do Maranhão são registradas.

O Capítulo 5 apresenta os experimentos realizados, desde a seleção dos dados até das técnicas de mineração, bem como a análise dos resultados, mostrando as melhores regras geradas e relatórios estatísticos gerados através dos softwares Weka e Orange.

Por fim, o Capítulo 6 apresenta as considerações finais acerca do trabalho realizado, e sugestões para investigações futuras.

## 2 Referencial Teórico

A seguir são descritos os conceitos que fundamentam o trabalho. A Seção 2.1 explana sobre a Descoberta de Conhecimento em Base de Dados e como o mesmo é utilizado no trabalho. Em seguida, na Seção 2.2, são abordadas as tarefas da mineração de dados, de acordo com a definição de alguns autores. A Subseção 2.2.1 explica o que são as regras de associação, e por fim, a Subseção 2.2.2 detalha o funcionamento do algoritmo Apriori, base para alcançar o objetivo geral anteriormente especificado.

### 2.1 Descoberta de Conhecimento em Base de Dados

Com o rápido avanço da tecnologia, muitas são as informações registradas acerca dos mais variados domínios, e a quantidade de dados gerados tem crescido de forma exponencial. Um avião, por exemplo, durante um voo, envia centenas de dados para as torres de controle. Ao tirar a carteira de identidade, todos os dados pessoais do solicitante são registrados no banco de dados da Secretaria de Segurança do Estado. Ao entrar na emergência de um hospital, informações sobre o paciente são cadastradas. Independentemente da área, um grande volume de dados é gerado diariamente.

Uma grande quantidade de dados armazenados pode resultar numa sobrecarga de informação acompanhada de escassez de conhecimento, visto que, em geral, não existe a transformação destes dados em conhecimento útil para fins do processo de tomada de decisão (OLIVEIRA; FERREIRA, 2014).

E como resolver este problema? Como empresas e instituições podem fazer um melhor uso dessas informações brutas que fluem diariamente?

Abordar técnicas e ferramentas que buscam transformar os dados armazenados, sejam sobre uma partida de futebol, ocorrências criminais, hospitais, bancos, imobiliários, educação e outros, em conhecimento é o objetivo da área denominada Descoberta de Conhecimento em Base de Dados (Knowledge Discovery in Databases – KDD) (STEINER et al., 2004).

O KDD é um processo não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados. Ele envolve várias etapas distintas e iterativas, até que seu objetivo seja atingido (FAYYAD et al., 1996).

A Descoberta de Conhecimento em Base de Dados busca identificar padrões válidos, novos, facilmente compreensíveis pelos seres humanos, e que acrescentarão conhecimento útil para uma melhor tomada de decisão. É o processo de extrair conhecimento a partir de



um grande volume de informações através de técnicas, de forma automatizada e inteligente.

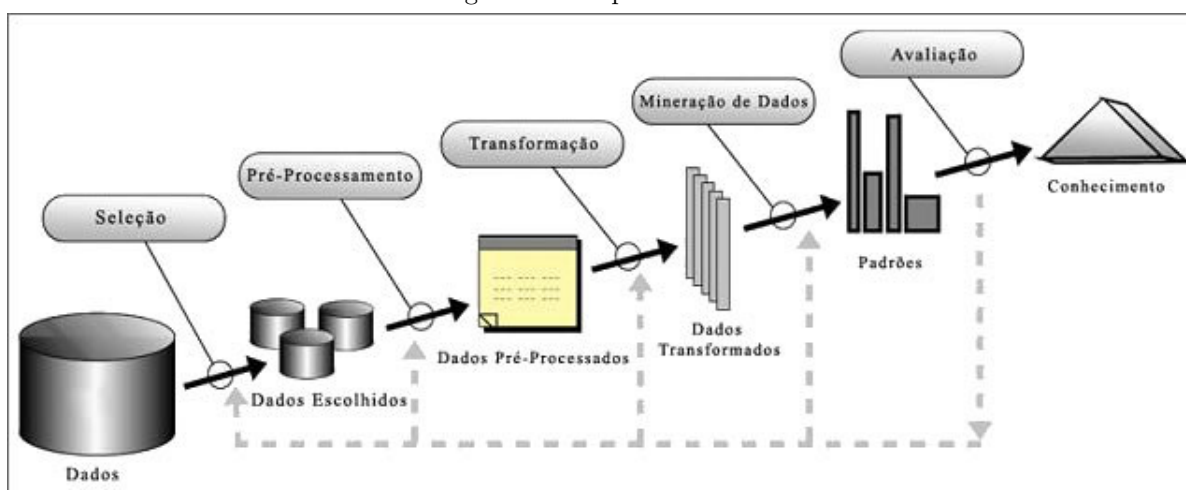
De acordo com Miller e Han (2009, p. 3):

“O KDD é mais conhecido através do termo mais popular ‘mineração de dados’. No entanto, a mineração de dados é apenas um componente (embora um componente central) do maior processo de KDD. Mineração de dados envolve a transformação de dados em informações ou fatos sobre o domínio descrito pelo banco de dados. O KDD é o processo de nível superior de obtenção de informações através da mineração de dados e da transformação dessas informações em conhecimento (ideias e crenças sobre o domínio) por meio da interpretação das informações e da integração com o conhecimento existente.”

Mineração de Dados e KDD são termos distintos entre si, mas intimamente ligados um ao outro. A Mineração de Dados é uma das etapas do processo da Descoberta de Conhecimento em Base de Dados, e consiste em analisar grandes volumes de dados sob diferentes perspectivas, a fim de descobrir informações úteis que normalmente não estão explícitas.

A Figura 1 ilustra, de forma esquematizada, todas as etapas do KDD, descritas a seguir:

Figura 1 – Etapas do KDD.



Fonte: Adaptado de (FAYYAD et al., 1996).

- Pré-processamento e limpeza dos dados: nesta fase são eliminados os dados redundantes e inconsistentes. Informações errôneas devem ser corrigidas a fim de não comprometer a qualidade dos modelos de conhecimento a serem extraídos ao final do KDD. Nesta fase também são utilizados métodos de redução ou transformação de atributos objetivando a diminuição dos atributos envolvidos no processo, o que melhora consideravelmente o desempenho do algoritmo de análise. Dados ausentes (missing values) e dados discrepantes, conhecidos como outliers (dados com valores extremos, atípicos ou com características bastante distintas dos demais registros), são tratados nesta fase;

- Transformação dos dados: nesta etapa os dados obtidos na fase de pré-processamento são analisados e reorganizados de forma específica para que possam ser interpretados na etapa seguinte, visto que os algoritmos de mineração possuem padrões que devem ser respeitados;
- Mineração de dados: aqui se escolhe o algoritmo de mineração de dados a ser utilizado de acordo com o objetivo do KDD. Nesta etapa são utilizados métodos para localizar padrões nos dados, além de definir quais modelos e parâmetros serão os mais adequados;
- Interpretação: Talvez a etapa mais complexa do processo de descoberta de conhecimento em base de dados. O objetivo desta fase é entender os resultados gerados após a mineração dos dados. Poderão surgir padrões, relacionamentos e descobertas de novos fatos, que podem ser utilizados em análises futuras, pesquisas, otimizações e outros.

Como visto acima, o KDD envolve etapas que geram conhecimentos a partir de dados relacionados, e sua principal característica é a extração de informações e conhecimentos implicitamente contidos em uma base de dados. Ele é iterativo e possui uma série de fases sequenciais, podendo haver retorno a etapas anteriores, ou seja, às descobertas realizadas (ou a falta delas).

## 2.2 Tarefas da Mineração de Dados

Mineração de Dados é um processo cujo objetivo é obter informações a partir de dados, em que são utilizadas diversas ferramentas e métodos. Como visto na seção anterior, ela é uma das etapas do processo de descoberta de conhecimento em base de dados.

Segundo [Fong \(1997\)](#), Mineração de Dados é um conjunto de técnicas utilizadas para descobrir padrões, grupos e modelos a partir de dados que são então renderizados em formas operacionais de fácil visualização e entendimento, além de automatizar a descoberta de conhecimento em base de dados.

A mineração de dados representa uma convergência de disciplinas. Como bem nos assegura [Kroenke \(2017\)](#), ela é a aplicação de técnicas oriundas de diversas áreas do conhecimento, tais como matemática, estatística e inteligência artificial, utilizadas para encontrar padrões entre dados, bem como para fins de classificação e previsão.

Conforme explicado acima, busca-se encontrar relações entre os dados, de tal modo que informações outrora escondidas, possam se tornar visíveis. Isto é possível através de técnicas e algoritmos, principalmente da inteligência artificial, que facilitam e auxiliam o

trabalho dos analistas de negócios das empresas e/ou organizações, tornando as mesmas mais competitivas.

Para [Gassenferth et al. \(2015\)](#), Mineração de Dados permite encontrar, automaticamente, padrões, anomalias e regras com o propósito de transformar dados, aparentemente ocultos, em informações úteis para a tomada de decisão e/ou avaliação de resultados.

O autor deixa claro, ainda, que a mineração de dados facilita o processo de extração de conhecimento a partir de um grande volume de registros, convencionais ou não, através de técnicas e algoritmos de diferentes áreas (estatística, banco de dados e inteligência artificial), que procuram relações de similaridade ou discordância entre dados.

Segundo [Braga \(2005, p. 12\)](#), “A mineração de dados compreende um conjunto de técnicas para descrição e predição a partir de grandes massas de dados”. Conforme mencionado pelo autor, as tarefas de extração de padrões podem ser agrupadas em atividades preditivas e descritivas.

Modelos descritivos estão relacionados a modelos de classificação não supervisionada, com base em observação e descoberta. No data mining não supervisionado, os analistas criam hipóteses após a análise para explicar os padrões encontrados. Já os modelos de classificação supervisionada são identificados como modelos de predição com argumento categórico. Um modelo é desenvolvido antes da análise e técnicas são aplicadas aos dados para se estimar os parâmetros do modelo.

São exemplos de atividades descritivas: clustering (utilizada para separar os registros de uma base de dados em subconjuntos), regras de associação (usada para detectar relações entre registros) e sumarização (procura identificar e indicar características comuns entre conjunto de dados).

Como atividades preditivas, temos: classificação (pretende descobrir uma função que seja capaz de mapear um conjunto de registros em um conjunto de variáveis anteriormente definidas) e a regressão (a estimativa é usada para definir um valor para alguma variável contínua desconhecida).

Conforme explicado acima, a aplicabilidade da mineração de dados é muito grande, podendo ser utilizada em vários contextos: saúde, educação, comércio, esporte, meteorologia, governo e outros.

No comércio, por exemplo, uma base de dados contendo os registros de vendas, submetida à mineração de dados, pode nos informar quais produtos são vendidos em conjunto e quais os mais frequentes.

Ainda para [Gassenferth et al. \(2015, p. 211\)](#):

“O Data Mining é formado por diversas técnicas e ferramentas que, através de algoritmos baseados em estatística e redes neurais, auxiliam na exploração de dados e na extração de evidências destes para emissão

de relatórios e decisões nos níveis superiores de uma organização. Nesse sentido, Mineração de Dados permite o desenvolvimento de novas estratégias para a organização através da informação escondida nas bases de dados da empresa.”

Evidentemente, a mineração de dados pode ser utilizada para contribuir de forma significativa no processo de descoberta de conhecimento.

O autor deixa claro que o Data Mining permite aos especialistas concentrarem esforços apenas nas partes mais relevantes dos dados. Uma vez esgotadas as alternativas iniciais de análise, como análises meramente visuais, com planilhas ou ainda com uso de estatística descritiva, usa-se a mineração de dados.

Logo, é importante compreender que na Mineração de Dados, o que gera valor de fato é o conjunto de ações que são realizadas a partir dos processamentos dos dados. Para isso, é preciso ter domínio do problema, saber onde aplicar as melhores técnicas e quais ferramentas de mineração são mais adequadas em cada situação.

### 2.2.1 Regras de Associação

Regra de associação, como atividade descritiva do Data Mining, é uma das mais diversas técnicas de mineração de dados para se extrair conhecimento útil.

Muitos artigos sobre algoritmos, medidas e problemas relacionados às regras de associação podem ser encontrados na literatura. Segundo Barro e Marin (2001), as regras de associação são um dos modelos mais bem estudados para aquisição de conhecimento na área de Data Mining.

E como funciona esta técnica? Qual a sua finalidade no processo de Mineração de Dados?

Bem, as regras de associação descrevem padrões de relacionamento entre itens de uma base de dados. São descritas em forma de regras do tipo [Se  $A$  então  $B$ ], ou [ $A \rightarrow B$ ], ( $A$  implica em  $B$ ), em que  $A$  e  $B$  são conjuntos de itens do banco de dados;  $A$  é o antecedente da regra e  $B$  é o conseqüente da regra, podendo existir qualquer número de itens em cada lado da regra.

Vamos entender este conceito com a ajuda de uma base de dados que contém os registros de vendas de um supermercado, conforme a Tabela 1.

A partir da Tabela 1, podemos tirar algumas conclusões simples: Quem compra leite quase sempre compra ovos; quem compra ovos e açúcar sempre compra leite; quem compra cerveja, sempre compra fraldas. Um exemplo de regra que poderia ser gerado através de um algoritmo de mineração seria: *ovos, açúcar  $\rightarrow$  leite. Se ovos e açúcar, então leite*. Interpretando essa regra, concluímos que quem compra ovos e açúcar, também compra leite.

Tabela 1 – Registro de vendas de um supermercado.

Transação	Itens
1	Leite, ovos, café, açúcar, fraldas, manteiga
2	Leite, café, farinha
3	Leite, ovos, açúcar
4	Café, açúcar
5	Fraldas
6	Manteiga, ovos, leite
7	Café, açúcar, leite, ovos
8	Farinha, manteiga, ovos
9	Manteiga, ovos, leite, café, açúcar
10	Fraldas, café, cerveja

Fonte: [Fagundes \(2009, p. 17\)](#).

Para [Ferrari e Silva \(2017\)](#), um bom algoritmo de mineração de regras de associação deve gerar relações que sejam estatisticamente relevantes para o universo representado pela base de dados. Há duas questões fundamentais na mineração de regras de associação: elaborar de forma eficiente as regras de associação e quantificar a significância de cada regra proposta. Conforme explicado acima, o produto da mineração de tais regras deve mostrar uma quantidade substancial de conhecimento extraído da base de dados, acompanhada de sua importância.

O significado de uma regra de associação está intimamente relacionado a dois parâmetros: suporte e confiança. O suporte de um conjunto de itens diz respeito a porcentagem de registros que contêm todos esses itens ([RAMAKRISHNAN; GEHRKE, 2008](#)). O autor deixa claro que o suporte da regra  $A \rightarrow B$  é o suporte do conjunto de itens  $A \cup B$ .

Já a confiança da regra  $A \rightarrow B$  é a porcentagem de tais transações que também contêm todos os itens em  $B$  ([RAMAKRISHNAN; GEHRKE, 2008](#)). De forma mais clara, seja  $sup(A)$  a porcentagem das transações que contêm  $A$  e seja  $sup(A \cup B)$  a porcentagem das transações que contêm  $A$  e  $B$ . Então, a confiança da regra  $A \rightarrow B$  é  $sup(A \cup B) / sup(A)$ .

Voltemos a Tabela 1. Considerando a regra *ovos, açúcar*  $\rightarrow$  *leite*, podemos constatar que os itens ovos, açúcar e leite foram comprados juntos em quatro das dez transações (transações 1, 3, 7 e 9). Neste caso, dizemos que a regra tem suporte de 40% (4/10).

Agora vamos analisar a confiança. Considerando a regra *ovos*  $\rightarrow$  *açúcar*, percebemos que a mesma possui confiança de 67% (4/6), visto que o conjunto *ovos, açúcar* aparece quatro vezes e o item ovos aparece seis vezes. Segundo [Quilici-Gonzalez e Zampiroli \(2014, p. 34\)](#), “Regras de associação são particularmente úteis para analisar o comportamento de clientes e propor vendas casadas”.

Assim, a informação de que clientes que compram o produto  $A$  geralmente compram

o produto  $B$  pode aumentar de modo considerável as vendas de um supermercado, já que toda vez que um cliente manifestar interesse em comprar o produto  $A$ , o supermercado pode também lhe oferecer o produto  $B$ .

Conforme explicado acima, as regras de associação podem, por exemplo, auxiliar na gestão de vendas de determinados produtos, fornecendo informações como quais produtos são mais vendidos, em que período do ano, qual o perfil dos clientes que mais compram um determinado produto etc.

De acordo com [Goldschmidt e Passos \(2005, p. 62\)](#):

“Um dos principais fatores de motivação para a tarefa de descoberta de regras associativas refere-se à possibilidade de incremento nas vendas de um determinado segmento comercial a partir de estratégias estabelecidas em função do conjunto de regras de associação extraído de grandes bases de dados históricas.”

O autor deixa claro na citação acima a importância da descoberta de regras associativas nos negócios, uma vez que podem ser extraídas informações sobre o comportamento de compras de cada cliente. Esses valiosos dados podem ser usados em diversas atividades, tais como: realização de promoções, reorganização dos produtos em gôndolas e prateleiras, análise de perdas, campanhas de marketing, e muitas outras ações.

Por fim, a mineração através de regras de associação é uma técnica usada na elaboração de relações sob a forma de regras entre itens de um conjunto de dados. Essas regras são de fácil entendimento para os seres humanos e podem ajudar a prever a presença, ou não, de outro conjunto de itens.

### 2.2.2 O algoritmo *Apriori*

O *Apriori* (Figura 2) é um dos algoritmos mais conhecidos na área de Mineração de Dados. Ele emprega busca em profundidade e gera conjuntos de itens candidatos (padrões) de  $k$  elementos a partir de conjuntos de itens de  $k - 1$  elementos. Os padrões não frequentes são eliminados. Toda a base de dados é rastreada e os conjuntos de itens frequentes são obtidos a partir dos conjuntos de itens candidatos.

1)  $F_k$  - conjunto de itens frequentes de tamanho  $k$  (conjunto com  $k$  elementos) que atende o suporte mínimo estabelecido. Cada membro deste conjunto tem dois campos. O primeiro é conjunto de itens e o segundo é um contador para o suporte.

2)  $C_k$  - Conjunto de itens candidatos de tamanho  $k$ . Cada membro deste conjunto tem dois campos. O primeiro é conjunto de itens e o segundo é um contador para o suporte.

O algoritmo principal (*Apriori*) faz uso de duas subrotinas: *apriori-gen*, para gerar o conjunto de itens candidatos (conjunto composto pelos valores correspondentes ao

Figura 2 – Algoritmo Apriori.

```

 $F_1 \leftarrow \{\text{Conjuntos de itens freqüentes de tamanho 1}\} \text{ /* Na}$ 
primeira passagem  $k = 1$  */
1 para  $k = 2$ ;  $F_{k-1} \neq \text{vazio}$ ;  $k++$  faça */
   /* Na segunda passagem  $k = 2$  */
2    $C_k \leftarrow \text{apriori-gen}(F_{k-1})$  /* Novos candidatos */
3   para todo transação  $t \in T$  faça
4      $C_t \leftarrow \text{subconjunto}(C_k, t)$  /* Candidatos contidos
em  $t$  */
5     para todo candidato  $c \in C_t$  faça
6        $c.\text{contagem}++$ 
7     fim
8      $F_k \leftarrow \{c \in C_k | c.\text{contagem} \geq \text{MinSup}\}$ 
9   fim
10 fim
11 Resposta  $F \leftarrow \text{Reunião de todos os } F_k$ 

```

Fonte: (AGRAWAL; IMIELINSKI; SWAMI, 1993).

suporte de cada item). Neste conjunto são considerados todos os itens, independente deles atenderem o suporte\_mínimo especificado e eliminar aqueles que não são freqüentes, e a subrotina subconjunto, utilizada para extrair as regras de associação. De forma geral, a sua meta é procurar por relações entre os dados enquanto eles são separados. Simultaneamente, o algoritmo calcula o valor correspondente à confiança e ao suporte.

O algoritmo trabalha sobre uma base de transações em busca de itens freqüentes, ou seja, aqueles que possuem suporte maior ou igual ao suporte mínimo. Desta forma, como entrada, é necessário fornecer um valor correspondente ao suporte\_mínimo e outro correspondente à confiança\_mínima, além de um arquivo de itens e transações.

O algoritmo é executado da seguinte forma:

- Na primeira passagem, o suporte para cada item individual (conjuntos-de-1-item) é contado e todos aqueles que satisfazem o suporte\_mínimo são selecionados, constituindo-se os conjuntos-de-1-item freqüentes ( $F_1$ ).
- Na segunda iteração, conjuntos-de-2-itens candidatos são gerados pela junção dos conjuntos-de-1-item (a junção é feita através da função *apriori-gen*) e seus suportes são determinados pela pesquisa no banco de dados, sendo, assim, encontrados os conjuntos-de-2-itens freqüentes.
- O algoritmo Apriori prossegue iterativamente, até que o conjunto-de-k-itens encontrado seja um conjunto vazio.

A função **apriori-gen** toma como argumento  $F_{(k-1)}$  (conjuntos-de-(k-1)-itens) e retorna o conjunto dos conjuntos de todos os conjuntos-de-k-itens. Primeiramente, no passo de junção, os elementos de  $F_{(k-1)}$  são combinados, de acordo com a Figura 3.

Figura 3 – Função Apriori-gen: passo de junção.

```

inserirEm  $C_k$ 
selecione  $p.item_1, p.item_2, \dots, p.item_{k-1}, q.item_{k-1}$ 
de  $F_{k-1}p, F_{k-1}q$ 
onde  $p.item_1 = q.item_1, \dots, p.item_{k-2} = q.item_{k-2}, p.item_{k-1} <$ 
 $q.item_{k-1}$ 

```

Fonte: (AGRAWAL; IMIELINSKI; SWAMI, 1993).

A seguir, ocorre o passo de poda, todos os conjuntos de itens  $c \in C_K$ , tal que algum conjunto-de-(k-1)-itens de  $c$  não está em  $F_{(k-1)}$ , são eliminados, de acordo com a Figura 4.

Figura 4 – Função Apriori-gen: passo de poda.

```

para todo conjuntos de itens  $c \in C_k$  faça
  | para todo subconjuntos-de-(k-1)-itens  $s$  de  $c$  faça
  | | se  $s \notin F_{k-1}$  então
  | | | remova  $c$  de  $C_k$ 
  | | fim
  | fim
fim

```

Fonte: (AGRAWAL; IMIELINSKI; SWAMI, 1993).

Após a determinação dos conjuntos de itens frequentes  $F$ , pode-se obter as regras de associação. Deve-se observar que cada item frequente é um conjunto de  $k$  itens. Para cada item frequente  $Y = I_1I_2\dots I_k$  de  $F$ , com  $k \geq 2$ , pode-se gerar todas as regras (no máximo  $k$  regras) que usam itens do conjunto  $I_1, I_2, \dots, I_k$ . O antecedente de cada uma das regras será o subconjunto  $X$  de  $Y$  tal que  $X$  tem  $k - 1$  itens, e o conseqüente será o item  $Y - X$ .



## 3 Ferramentas para Mineração de Dados

Com o crescimento da Mineração de Dados tanto no meio acadêmico quanto no meio corporativo, muitas ferramentas e bibliotecas voltadas para este processo foram desenvolvidas. Entre as suítes pagas, e de excelente qualidade, podemos citar: STATISTICA, SPSS, SAS, Microsoft Analysis Services, Clementine, entre outras.

No entanto, os softwares livres estão ganhando cada dia mais espaço devido seus poderosos recursos e a possibilidade total de customização em sua maioria. Entre os programas gratuitos de Mineração de Dados mais populares, Clesio (2012) cita:

- **R:** o Projeto R (ou somente “R”) é um software de estatística que possui variados pacotes, com funções matemáticas, estatísticas e econométricas. Mais precisamente, ele é um conjunto integrado de recursos de software para manipulação de dados, cálculo e exibição gráfica. Pode ser baixado em <<http://www.r-project.org>>.
- **RapidMiner:** voltada para equipes de análise, o RapidMiner une preparação dos dados, aprendizado de máquina e implantação de modelos preditivos. Tem como vantagem uma baixa curva de aprendizado, além de produzir gráficos de ótima qualidade. Maiores informações em <<https://rapidminer.com>>.
- **Weka:** desenvolvido pela Universidade de Waikato na Nova Zelândia, o Weka tem como principal característica ser uma ferramenta de fácil uso, o que pode até passar a impressão de que é trivial o processo de Mineração de Dados. O download do Weka está disponível em <<https://www.cs.waikato.ac.nz/ml/weka/downloading.html>>.
- **Orange:** outro software gratuito que tem crescido bastante. Através dessa ferramenta é possível realizar uma análise de dados simples com visualização de dados inteligente. Usando uma programação visual, é possível fazer uma exploração interativa de dados para análise qualitativa rápida com visualizações simples. O Orange está disponível através do link <<https://orange.biolab.si/>>.

### 3.1 Weka

O Weka é um dos softwares mais conhecidos atualmente para se trabalhar com Mineração de Dados. Ele foi desenvolvido em Java e é *open source*, ou seja, de código aberto, podendo ser utilizado gratuitamente sob a licença GPL (General Public License). Pode ser copiado, distribuído, mas não editado.

“Embora o acrônimo Weka signifique Waikato Environment for Knowledge Analysis, algo como Ambiente Waikato para Análise de Conhecimento, Weka é também o nome de um pássaro muito popular na Nova Zelândia, e por isso ele é também a mascote do software, tendo sua imagem na tela principal do sistema, além de um ícone que indica o status do processamento na parte inferior direita do utilitário Weka Explorer (AMARAL, 2016, p. 14).”

A Figura 5 apresenta a tela principal do Weka:

Figura 5 – Tela principal do Weka.



Fonte: Elaborada pelo autor.

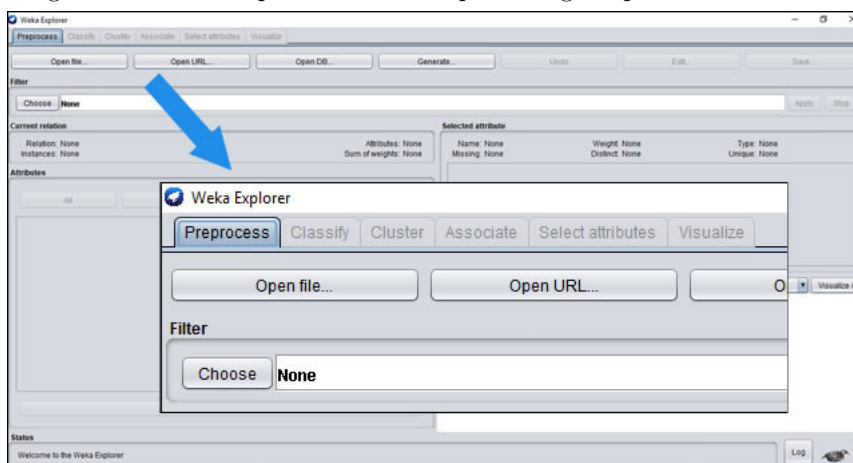
Mantido pela Universidade de Waikato, na Nova Zelândia, o software pode ser baixado através do link <<https://www.cs.waikato.ac.nz/~ml/weka/downloading.html>>.

Neste trabalho não será mostrado o processo de instalação desta ferramenta, visto que o *website* acima descrito apresenta o passo-a-passo para se executar tal atividade. Nesta seção focaremos em algumas funcionalidades do Weka, explanando o que esta poderosa ferramenta oferece para se extrair conhecimento de grandes volumes de dados.

Ao clicar no botão *Explorer*, ainda na Figura 5, uma nova janela se abrirá contendo algumas abas (Figura 6). Essas abas estão intimamente ligadas ao processo de KDD, permitindo com que o usuário possa fazer um pré-processamento dos dados, classificá-los, agrupá-los, trabalhar com regras de associação, regressão, além de poder visualizá-los.

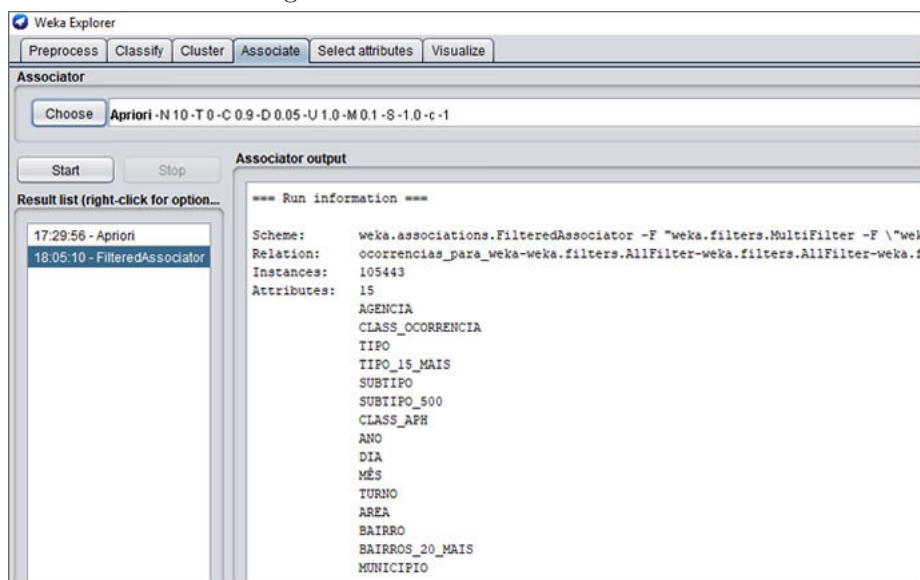
Utilizou-se apenas as abas *Preprocess* e *Associate*. Em *Preprocess* carregou-se a base de dados a ser estudada. Como explicado na Seção 2.1, é no pré-processamento que se faz a limpeza dos dados. Dados redundantes são eliminados, atributos não relevantes são removidos e atributos numéricos são transformados em discretos, por exemplo. Já a aba *Associate*, conforme Figura 7, permite a execução do algoritmo *Apriori*, a fim de encontrar as melhores regras de associação a partir dos dados previamente tratados.

Figura 6 – Weka Explorer com destaque na região que contém as abas.



Fonte: Elaborada pelo autor.

Figura 7 – Aba Associate no Weka.



Fonte: Elaborada pelo autor.

## 3.2 Orange

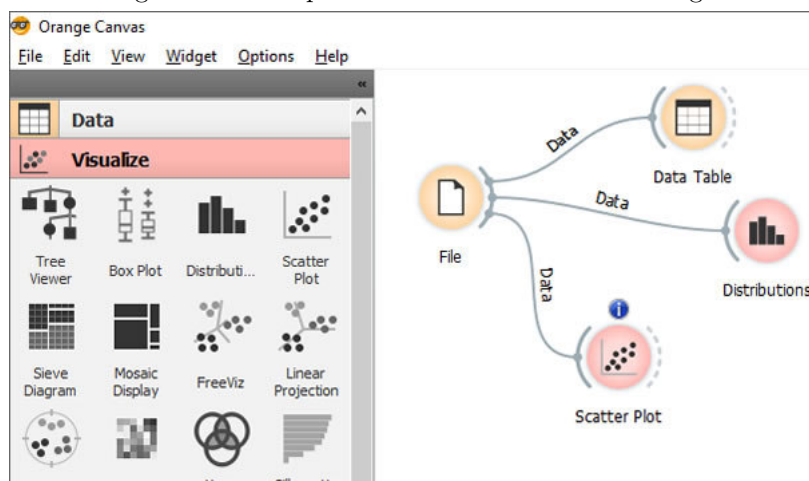
Orange é outro software usado em *Data Mining*, de código aberto e disponível gratuitamente através do link <https://orange.biolab.si/>. Ele fornece um vasto conjunto de ferramentas, permitindo com que atividades de pré-processamento, mineração e visualização de dados sejam realizadas facilmente.

O grande diferencial desta ferramenta, segundo Coleman e Tsongalis (2017), em relação ao demais softwares de mineração, é a sua interface de programação visual intuitiva. A interface gráfica do usuário permite que, tanto iniciantes quanto especialistas em dados, se concentrem na análise exploratória de dados em vez de trabalhar em codificação trabalhosa.

No Orange, a análise de dados é feita empilhando componentes em fluxos de

trabalho. Cada componente, chamado de *widget*, incorpora algumas tarefas de recuperação de dados, pré-processamento, visualização, modelagem ou avaliação. A combinação de diferentes *widgets* em um fluxo de trabalho permite a criação de esquemas abrangentes de análise de dados, conforme a Figura 8.

Figura 8 – Exemplo de fluxo de trabalho no Orange.



Fonte: Elaborada pelo autor.

Como descreve [Khosrow-Pour \(2018\)](#), o Orange oferece diversas funcionalidades divididas em nove categorias: operações com dados, visualização, classificação, regressão, avaliação, aprendizagem de máquina não supervisionada, associação, visualização usando QT<sup>1</sup> e implementações de protótipos. Ademais, outras funcionalidades podem ser facilmente adicionadas com a instalação de novos *widgets* a partir da própria ferramenta.

<sup>1</sup> O QT é um tipo de framework voltado para o desenvolvimento de interfaces gráficas criado pela empresa Trolltech.

## 4 Metodologia e contextualização do problema

Segundo Rampazzo (2002), a pesquisa é de fundamental importância diante de um problema, independente da área de conhecimento, visto que ela permite encontrar novos fatos, dados e soluções por meio dos processos do método científico.

A análise dos dados nos experimentos relacionados neste trabalho é essencialmente quantitativa. De acordo com Souza, Santos e Dias (2003, p. 72), “Na pesquisa quantitativa, a fim de comprovar as hipóteses, os recursos de estatística nos dirão se os resultados obtidos são significativos ou mero fruto do acaso”.

Conforme citado acima, a pesquisa quantitativa é baseada em critérios estatísticos. Nesse sentido, devido a pesquisa ter como base um banco de dados contendo registros de ocorrências do CBMMA, será utilizada uma abordagem quantitativa e como modo de análise, o raciocínio hipotético-dedutivo, através de testes e experimentações.

Segundo Alegria et al. (2012, p. 110), “... a pesquisa aplicada tem propósito prático e específico para gerar conhecimento ou avanço do conhecimento e tem uma esfera de atuação mais abrangente, seja científica, industrial, econômica ou acadêmica”. Devido aos fins práticos deste trabalho, utilizaremos a pesquisa aplicada.

De acordo com Ciribelli (2003, p. 54), “A Pesquisa Exploratória é o primeiro passo de qualquer Trabalho Científico. É também denominada Pesquisa Bibliográfica. Proporciona maiores informações sobre o tema que o pesquisador pretende abordar”.

A fim de atingir o objetivo da pesquisa, foram utilizadas diversas fontes bibliográficas. Assim, a pesquisa realizada neste trabalho se classifica como pesquisa exploratória, o que permitiu uma melhor delimitação do tema proposto, ajudou na definição dos objetivos, bem como proporcionou uma forma original para desenvolver o assunto.

Como instrumento para coleta de dados utilizou-se registros institucionais, mais especificamente, um banco de dados contendo informações sobre as ocorrências atendidas pelo Corpo de Bombeiros Militar do Maranhão, entre janeiro de 2013 e agosto de 2018. Ademais, foram utilizados também livros de maior relevância sobre o assunto, periódicos, artigos e textos em *websites*, a fim de se obter uma melhor apreciação do conteúdo apresentado no trabalho.

De início, foi solicitado ao NEAC - Núcleo de Estatística e Análise Criminal, setor subordinado à Secretaria de Segurança Pública do Estado do Maranhão, acesso aos registros de ocorrências atendidas pelo CBMMA, sendo disponibilizado tal banco de

dados, em formato *XLSX*<sup>1</sup>. De posse dessas informações, e dada a possibilidade de se extrair conhecimento dessa base de dados, optou-se por trabalhar com Mineração de Dados utilizando a técnica de Regras de Associação.

Sendo assim, os dados foram tratados, pré-processados, para então aplicar os algoritmos de Mineração de Dados, a fim de alcançar os objetivos deste trabalho. O software Weka foi utilizado para se obter relevantes regras de associação, enquanto o Orange foi usado para gerar gráficos e relatórios de fácil entendimento e visualização.

## 4.1 As ocorrências do Corpo de Bombeiros Militar do Maranhão

De acordo com [Cbmma \(2018\)](#), o Corpo de Bombeiros Militar do Maranhão é uma instituição centenária que tem estado presente nas principais cidades do Estado, tais como São Luís (Capital do Estado), Imperatriz, Bacabal, Açailândia, Balsas, Timon, São José de Ribamar e outras.

Hoje o CBMMA conta com 16 batalhões distribuídos na capital e no interior do Estado, e um total de mil e quinhentos militares. Homens e mulheres altamente capacitados que um dia juraram pela própria vida defender a sociedade. Só na cidade de São Luís existem dois batalhões específicos para incêndios urbanos (1º BBM - 1º Batalhão de Bombeiros Militar e o 2º BBM - 2º Batalhão de Bombeiros Militar), um especializado em busca e salvamento (BBS - Batalhão de Busca e Salvamento), um batalhão especializado em salvamento aquático (BBMAR - Batalhão de Bombeiros Marítimo), outro que trabalha com captura de animais e corte de árvores (BBA - Batalhão de Bombeiros Ambiental), um especializado em emergências médicas (BBEM - Batalhão de Bombeiros de Emergências Médicas), além de uma companhia independente especializada no combate a incêndios em aviões, localizada no Aeroporto Internacional de São Luís - Marechal Cunha Machado (1ª CIBM - 1ª Companhia Independente Especializada de Bombeiros).

Compete ao Corpo de Bombeiros Militar do Maranhão prestar socorro nos casos de inundações, alagamentos, desabamentos e/ou catástrofes, sempre que houver vítimas ou pessoas em iminente perigo de vida. Compete, ainda, realizar serviços de busca e salvamento de pessoas, animais, bens e haveres; realizar prevenção no meio aquático e serviço de guarda-vidas; realizar serviços de atendimento e transporte pré-hospitalar em vias e logradouros públicos, entre outros. ([BRASIL, 2015](#))

As ocorrências atendidas na grande ilha de São Luís (formada pela capital e pelos municípios de São José de Ribamar, Raposa, Paço do Lumiar, Alcântara, Bacabeira, Rosário, Santa Rita, Icatu, Morros, Presidente Juscelino, Axixá, e Cachoeira Grande) são registradas em um sistema no Centro Integrado de Operações de Segurança do Estado

<sup>1</sup> O formato de arquivo *XLSX* é implementado por versões mais recentes do Microsoft Excel para os seus documentos de saída de planilha.

(CIOPS), órgão integrante da Secretaria de Segurança Pública do Maranhão (SSP/MA).

“O Centro Integrado de Operações de Segurança é a reunião das instituições envolvidas no atendimento de emergências na área de segurança pública e defesa da cidadania em um mesmo local, através do telefone 190, compartimentando os mais modernos recursos tecnológicos. (SSPMA, 2018)”

Quando um cidadão precisa, por exemplo, do serviço do Corpo de Bombeiros, ele liga 190, ou 193, e essa ligação cai justamente no CIOPS, onde primeiramente, um agente de *call center* recebe as primeiras informações do solicitante, verifica a natureza da ocorrência para saber se é de natureza policial ou não, e então direciona a mesma para o setor responsável (polícia ou bombeiro). Em seguida, um oficial verifica a disponibilidade das viaturas, bem como a prioridade de cada chamada e aciona a unidade relacionada com aquele tipo de ocorrência.

Com relação às ocorrências geradas a partir de cidades fora da região metropolitana de São Luís, atualmente, estas são registradas tanto em livros físicos chamados Livros de Ocorrências, quanto cadastradas em outros sistemas disponíveis localmente nas unidades do interior. Isso se deve ao fato de ainda haver problemas com relação ao acesso à Internet em boa parte dos municípios no interior do Estado. Mensalmente, um relatório contendo todas as ocorrências daquela unidade é enviado ao Núcleo de Estatística e Análise Criminal da SSP/MA, que mescla as informações com os dados fornecidos pelo CIOPS e produz relatórios estatísticos voltados para o Alto Comando do CBMMA, bem como para o Secretário de Segurança do Estado.

## 5 Experimentos e análise dos resultados

### 5.1 Aplicação da Mineração de Dados no banco de ocorrências

A base de dados disponibilizada para este trabalho continha somente ocorrências atendidas na região metropolitana de São Luís, mais especificamente, dados da capital, São José de Ribamar, Paço do Lumiar e Raposa, totalizando 108.073 ocorrências, referentes ao período de 01/01/2013 a 31/08/2018, conforme a Figura 9.

Figura 9 – Base de dados disponibilizada no formato XLSX.

ID	AGENCIA	UNIDADE	CD	DESCRICAO	STATUS	LOCAL
108055	GEM	GD1	E257006	A24 AVERIGUACAO	ATEND. PRE HOSPITALAR	A24A DE PESSOA
108056	GEM	GD1	E257007	TR2 TRANSPORTE PRE-HOSPITALAR	ATEND. PRE HOSPITALAR	A04 EMERGENCIA
108057	GEM	GD1	E257008	TR2 TRANSPORTE PRE-HOSPITALAR	ATEND. PRE HOSPITALAR	A04 EMERGENCIA
108058	GEM	GD1	E257009	CP121-1 TENTATIVA DE HOMICIDIO	ATEND. PRE HOSPITALAR	default
108059	GEM	GD1	E257010	TR2 TRANSPORTE PRE-HOSPITALAR	ATEND. PRE HOSPITALAR	A04 EMERGENCIA
108060	GEM	GD1	E257011	TR2 TRANSPORTE PRE-HOSPITALAR	ATEND. PRE HOSPITALAR	A04 EMERGENCIA
108061	GEM	GD1	E257012	TR2 TRANSPORTE PRE-HOSPITALAR	ATEND. PRE HOSPITALAR	A02 CRIANCA
108062	GEM	GD1	E257013	TR2 TRANSPORTE PRE-HOSPITALAR	ATEND. PRE HOSPITALAR	A04 EMERGENCIA
108063	GEM	GD1	E257042	TR2 TRANSPORTE PRE-HOSPITALAR	ATEND. PRE HOSPITALAR	A20 AVC/ACIDENTE VASCULAR CER
108064	GEM	GD1	E257047	CTB9 COLISAO	ATEND. PRE HOSPITALAR	CTB9D VITIMA NAO FATAL
108065	GEM	GD1	E257048	CTB9 COLISAO	ATEND. PRE HOSPITALAR	CTB9D VITIMA NAO FATAL
108066	GEM	GD1	E257049	TR2 TRANSPORTE PRE-HOSPITALAR	ATEND. PRE HOSPITALAR	A04 EMERGENCIA
108067	GEM	GD1	E257050	A1 ACHADO DE CADAVER	ATEND. PRE HOSPITALAR	A1A EM LOCAL ACESSIVEL
108068	GEM	GD1	E257051	TR2 TRANSPORTE PRE-HOSPITALAR	ATEND. PRE HOSPITALAR	A04 EMERGENCIA
108069	GEM	GD1	E257052	CP129 LESAO CORPORAL	ATEND. PRE HOSPITALAR	CP129E VITIMA DE FAB (FERIMENTO P
108070	GEM	GD1	E257054	CP250-14 INCENDIO EM VEGETACAO/FLORESTAL	ATEND. PRE HOSPITALAR	CP250A SOMENTE DANOS MATERIAIS
108071	GEM	GD1	E257055	CP250-14 INCENDIO EM VEGETACAO/FLORESTAL	ATEND. PRE HOSPITALAR	CP250A SOMENTE DANOS MATERIAIS
108072	GEM	GD1	E257063	CP122 SUICIDIO	ATEND. PRE HOSPITALAR	CP122C TENTADO
108073	GEM	GD1	E257067	TR1 TRANSPORTE PSQUIATRICO	ATEND. PRE HOSPITALAR	A02 NAO-AGRESSIVO

Fonte: Elaborada pelo autor.

Contudo, muitos registros estavam incompletos, com diversas células não preenchidas e muitos dados inconsistentes. A partir de então, iniciou-se a fase de seleção, seguida do pré-processamento e limpeza dos dados, conforme o processo de descoberta de conhecimento estudado na Seção 2.1.

A planilha possuía 31 campos (atributos), sendo que somente 12 foram aproveitados, visto que os demais não teriam notória significância nos experimentos a serem realizados neste trabalho. Os campos utilizados para a pesquisa foram:

- **AGENCIA:** informa se aquela ocorrência foi atendida pelo BBEM (Batalhão de Bombeiros de Emergências Médicas) ou outra unidade, que nesta base é representada pelo valor CBMMA;
- **CLASS\_OCORRENCIA:** registra a classe da ocorrência (atendimento pré-hospitalar, auxílio a comunidade, busca/salvamento, incêndio, prevenção, produtos perigosos);



- TIPO: descreve o tipo da ocorrência (incêndio em veículo, choque, vazamento de gás, transporte psiquiátrico, capotamento, incêndio em vegetação/florestal, lavagem de pista em caso de risco, colisão, incêndio em estabelecimento comercial, captura/resgate de animais, e outros);
- SUBTIPO: contêm o subtipo da ocorrência (agressivo, vítima não fatal presa em ferragens, somente danos materiais, sem fogo, parturiente, desmaio etc.);
- CLASS\_APH: classe de atendimento pré-hospitalar (clínico, obstétrico, prevenção, psiquiátrico, trauma);
- ANO: ano da ocorrência registrada (2013, 2014, 2015, 2016, 2017 ou 2018);
- DIA: dia da semana em que ocorreu o fato (segunda-feira, terça-feira, quarta-feira, quinta-feira, sexta-feira, sábado ou domingo);
- MÊS: mês em que se registrou a ocorrência (jan, fev, mar, abr, mai, jun, jul, ago, set, out, nov, dez);
- TURNO: período do dia cadastrado (manhã, tarde, noite ou madrugada);
- ÁREA: região/zona da cidade onde aconteceu o fato (norte, sul, leste, oeste);
- BAIRRO: nome do bairro de uma determinada ocorrência, associado com sua zona (Maiobão-ZL, Anil-ZO, Vinhais-ZN, São Cristovão-ZS etc.);
- MUNICÍPIO: cidade onde se deu a ocorrência (São Luís, São José de Ribamar, Paço do Lumiar ou Raposa).

Visto que o atributo BAIRRO possuía um elevado número de valores distintos (854), sendo inviável a visualização gráfica destes valores no Weka (Figura 10), com a ajuda do Orange se verificou quais os 20 bairros mais frequentes e se criou mais uma coluna (BAIRROS\_20\_MAIS) na base de dados.

As células nesta nova coluna foram preenchidas de acordo com o valor existente no campo BAIRRO de cada registro. As células que ficaram vazias identificam a linha cujo valor no atributo BAIRRO não se encontra entre os 20 bairros mais frequentes (Figura 11).

O mesmo procedimento foi realizado com o atributo TIPO, que possuía 141 valores distintos (Figura 12), criando-se uma nova coluna, chamada TIPO\_15\_MAIS, preenchida com os 15 tipos de ocorrências mais frequentes (Figura 13).

De forma análoga, a coluna SUBTIPO\_500 foi adicionada, porém, com base nos subtipos de ocorrências com frequência acima de 500 registros. Sendo assim, enquanto o

Figura 10 – Atributo BAIRRO com 854 valores distintos.

No.	Label	Count	Weight
1	ANIL-ZO	1683	1683.0
2	MAIOBAO-ZL	2687	2687.0
3	COHAMA-ZN	1445	1445.0
4	CALHAU-ZN	1997	1997.0
5	VINHAI-ZN	846	846.0
6	V. ESPERANCA-ZS	434	434.0
7	OUTEIRO DA CRUZ-ZO	388	388.0
8	MONTE CASTELO-ZO	1084	1084.0
9	CIDADE OPERARIA-ZL	2414	2414.0

Fonte: Elaborada pelo autor.

Figura 11 – Base de dados depois de adicionada a coluna BAIRROS\_20\_MAIS.

ANO	DIA	MÊS	TURNO	AREA	BAIRRO	BAIRROS_20_MAIS	MUNICIPIO
ANO 2014	Segunda-Feira	Nov	Manha	OESTE	ANIL-ZO	ANIL-ZO	SAO LUIS
ANO 2015	Terça-Feira	Jan	Manha	LESTE	MAIOBAO-ZL	MAIOBAO-ZL	PACO DO LUMIAR
ANO 2017	Sábado	Abr	Tarde	NORTE	COHAMA-ZN	COHAMA-ZN	SAO LUIS
ANO 2013	Sexta-Feira	Mar	Manha	OESTE	ANIL-ZO	ANIL-ZO	SAO LUIS
ANO 2013	Sábado	Jun	Noite	NORTE	CALHAU-ZN	CALHAU-ZN	SAO LUIS
ANO 2013	Segunda-Feira	Jun	Tarde	NORTE	VINHAI-ZN		SAO LUIS
ANO 2013	Domingo	Out	Noite	SUL	V. ESPERANCA-ZS		SAO LUIS
ANO 2014	Terça-Feira	Out	Manha	OESTE	ANIL-ZO	ANIL-ZO	SAO LUIS
ANO 2015	Sexta-Feira	Fev	Manha	NORTE	CALHAU-ZN	CALHAU-ZN	SAO LUIS
ANO 2015	Quarta-Feira	Jul	Manha	OESTE	OUTEIRO DA CRUZ-ZO		SAO LUIS
ANO 2015	Quinta-Feira	Nov	Manha	OESTE	MONTE CASTELO-ZO		SAO LUIS
ANO 2016	Quarta-Feira	Mar	Manha	LESTE	CIDADE OPERARIA-ZL	CIDADE OPERARIA-ZL	SAO LUIS
ANO 2016	Sábado	Jul	Noite	LESTE	JARDIM AMERICA-ZL		SAO LUIS
ANO 2017	Sábado	Jan	Tarde	LESTE	MAIOBAO-ZL	MAIOBAO-ZL	PACO DO LUMIAR
ANO 2017	Sábado	Abr	Manha	NORTE	RENASCENCA I-ZN		SAO LUIS
ANO 2017	Quarta-Feira	Mai	Tarde	NORTE	COHAB ANIL II-ZN		SAO LUIS
ANO 2018	Domingo	Jan	Madrugada	LESTE	MAIOBAO-ZL	MAIOBAO-ZL	PACO DO LUMIAR
ANO 2018	Sábado	Fev	Manha	NORTE	COMATRAC III-ZN		SAO LUIS

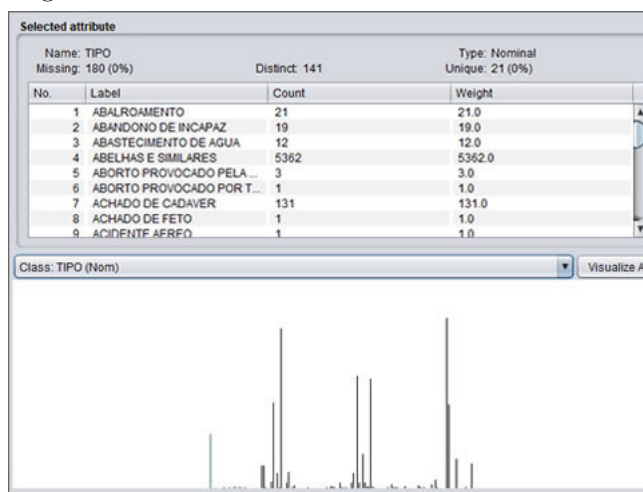
Fonte: Elaborada pelo autor.

atributo SUBTIPO possuía 83 valores distintos, o atributo SUBTIPO\_500 ficou somente com 19. A Figura 14 apresenta os 19 valores distintos para o atributo SUBTIPO\_500.

A base possuía alguns registros com informação no campo BAIRRO, mas que não continha nenhum dado no campo AREA. Nestes casos, as células vazias foram preenchidas com base no nome dos bairros, uma vez que cada bairro (na grande maioria) possuía em sua terminação a sigla da zona correspondente (Figura 15).

Outro problema que foi resolvido ainda nesta etapa, foi o fato de existirem ocorrências que foram geradas na mesma localidade, e vários nomes referentes ao mesmo bairro foram registrados na base. Por exemplo, a Figura 16 apresenta quatro grafias para o

Figura 12 – Atributo TIPO com 141 valores distintos.



Fonte: Elaborada pelo autor.

Figura 13 – Base de dados depois de adicionada a coluna TIPO\_15\_MAIIS.

AGENCIA	CLASS_OCORRENCIA	TIPO	TIPO_15_MAIIS	SUBTIPO
5314	GEM	ATEND. PRE HOSPITALAR	INCENDIO EM VEGETACAO/FLORESTAL	SOMENTE DANOS MATERIAIS
5315	GEM	ATEND. PRE HOSPITALAR	INCENDIO EM VEGETACAO/FLORESTAL	SOMENTE DANOS MATERIAIS
5316	GEM	ATEND. PRE HOSPITALAR	INCENDIO EM VEGETACAO/FLORESTAL	SOMENTE DANOS MATERIAIS
5317	GEM	ATEND. PRE HOSPITALAR	INCENDIO EM VEGETACAO/FLORESTAL	SOMENTE DANOS MATERIAIS
5318	GEM	ATEND. PRE HOSPITALAR	INCENDIO EM VEGETACAO/FLORESTAL	SOMENTE DANOS MATERIAIS
5319	GEM	ATEND. PRE HOSPITALAR	INCENDIO EM VEGETACAO/FLORESTAL	SOMENTE DANOS MATERIAIS
5320	GEM	ATEND. PRE HOSPITALAR	INCENDIO EM VEGETACAO/FLORESTAL	SOMENTE DANOS MATERIAIS
5321	GEM	ATEND. PRE HOSPITALAR	INCENDIO EM VEGETACAO/FLORESTAL	SOMENTE DANOS MATERIAIS
5322	GEM	ATEND. PRE HOSPITALAR	INCENDIO EM VEGETACAO/FLORESTAL	SOMENTE DANOS MATERIAIS
5323	CBMMA	INCENDIO	INCENDIO EM VEICULO	SOMENTE DANOS MATERIAIS
5324	CBMMA	INCENDIO	INCENDIO EM VEICULO	SOMENTE DANOS MATERIAIS
5325	CBMMA	INCENDIO	INCENDIO EM VEICULO	SOMENTE DANOS MATERIAIS
5326	CBMMA	INCENDIO	INCENDIO EM VEICULO	SOMENTE DANOS MATERIAIS
5327	CBMMA	INCENDIO	INCENDIO EM VEICULO	SOMENTE DANOS MATERIAIS
5328	CBMMA	INCENDIO	INCENDIO EM VEICULO	SOMENTE DANOS MATERIAIS
5329	CBMMA	INCENDIO	INCENDIO EM VEICULO	SOMENTE DANOS MATERIAIS
5330	CBMMA	INCENDIO	INCENDIO EM VEICULO	SOMENTE DANOS MATERIAIS
5331	CBMMA	INCENDIO	INCENDIO EM VEICULO	SOMENTE DANOS MATERIAIS

Fonte: Elaborada pelo autor.

bairro Vila Embratel (EMBRATEL-ZS, V. EMBRATEL, V. EMBRATEL-ZS e VILA EMBRATEL). Ou seja, as ocorrências aconteceram no bairro Vila Embratel, e foram registradas em quatro lugares diferentes.

Além dos problemas acima descritos, foram encontrados também na base uma grande quantidade de registros repetidos, e muitos com apenas três ou quatro campos preenchidos. Todas essas linhas repetidas ou com um número ínfimo de atributos preenchidos foram excluídas. Foram removidas também os registros cujo campo MUNICIPIO estava com valor totalmente diverso dos valores normais para aquele campo (relativo a nome de cidade). Por exemplo, haviam registros com número de telefone, ou nome de bairro, no campo MUNICIPIO.

Após as devidas correções e exclusões das linhas que apresentavam algum tipo de problema, a base de dados ficou com 105.271 registros e um novo arquivo foi gerado com

Figura 14 – Atributo SUBTIPO\_500 com seus 19 valores distintos.

Selected attribute			
Name: SUBTIPO_500		Type: Nominal	
Missing: 24830 (24%)		Distinct: 19	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	VÍTIMA NAO FATAL	19786	19786.0
2	SOMENTE DANOS MATERIAIS	18704	18704.0
3	VITIMA NAO FATAL PRESA EM FERRAGENS	547	547.0
4	DE PESSOA	2150	2150.0
5	EMERGENCIA	10505	10505.0
6	VITIMA DE PAF (PROJETIL DE ARMA DE FOGO)	2545	2545.0
7	A MULHER	794	794.0
8	VÍTIMA DE FAB (FERIMENTO P/ARMA BRANCA)	3099	3099.0
9	POR OUTROS MEIOS	4321	4321.0
10	CRIANCA	858	858.0
11	VÍTIMA DE QUEDA DA PROPRIA ALTURA	1112	1112.0
12	VÍTIMA DE QUEDA DE ALTURA ELEVADA	2208	2208.0
13	CONVULSAO	667	667.0
14	DESMAIO	862	862.0
15	PARTURIENTE	1394	1394.0
16	AGRESSIVO	6109	6109.0
17	NAO-AGRESSIVO	1173	1173.0
18	ARMADO	1026	1026.0
19	SEM FOGO	2753	2753.0

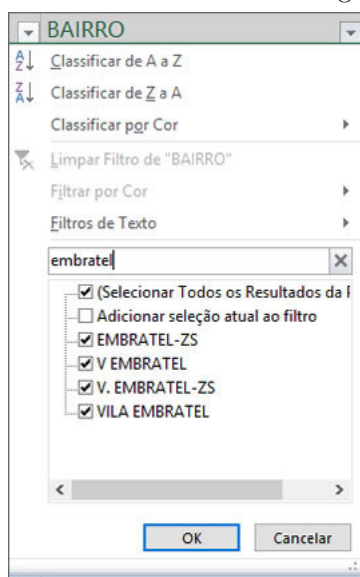
Fonte: Elaborada pelo autor.

Figura 15 – Exemplo de registro onde BAIRRO está preenchido e AREA não.

AREA	BAIRRO
LESTE	CIDADE OPERARIA-ZL
NORTE	LITORANEA-ZN
LESTE	SAO_BRAZ_DOS_MACACOS-ZL
NORTE	COHAFUMA-ZN
NORTE	COHATRAC-ZN
	MARACANA-ZL

Fonte: Elaborada pelo autor.

Figura 16 – Mesmo bairro com diferentes grafias do nome.



Fonte: Elaborada pelo autor.

a extensão *CSV*<sup>1</sup> (ocorrencias\_cbmma\_2013\_a\_2018.csv), formato de arquivo este, que

<sup>1</sup> Os arquivos Comma-separated values, também conhecido como CSV, são arquivos de texto que faz uma ordenação de bytes ou um formato de terminador de linha, separando valores com vírgulas.

é lido tanto no Weka, quanto no Orange.

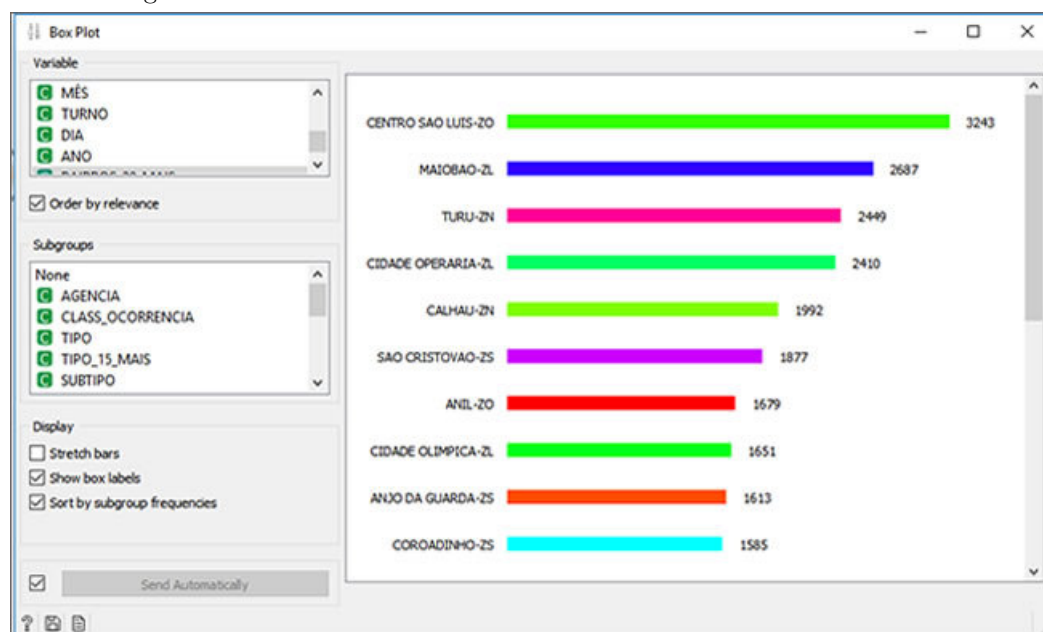
## 5.2 Alguns dados estatísticos sobre as ocorrências do Corpo de Bombeiros Militar do Maranhão

De posse do arquivo no formato *CSV*, e após carregá-lo nas ferramentas de mineração de dados aqui utilizadas, já foi possível a extração de conhecimentos relevantes, mesmo sem a aplicação de algoritmos específicos, como o *Apriori*.

As subseções a seguir apresentam alguns gráficos e resultados obtidos apenas com o carregamento da base de dados no Weka, e com a ajuda de alguns *widgets* no Orange.

### 5.2.1 Os 10 bairros com maior número de ocorrências atendidas

Figura 17 – Os 10 bairros com maior número de ocorrências atendidas.



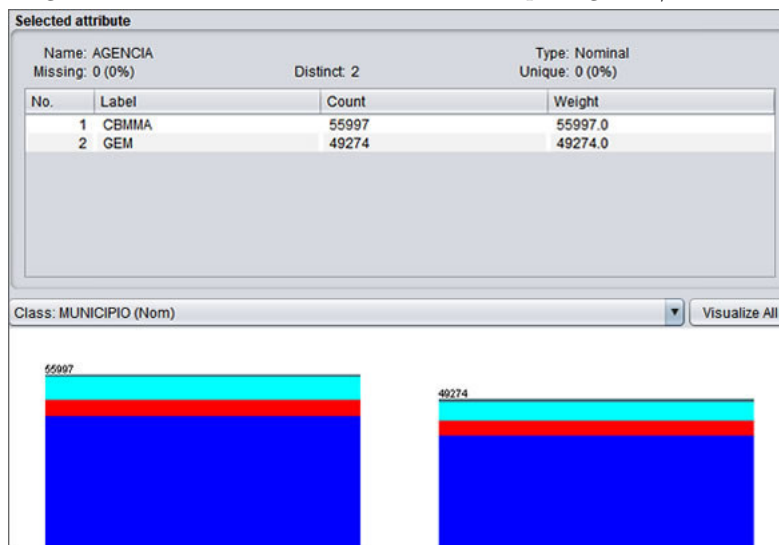
Fonte: Elaborada pelo autor.

Com o uso do *widget* Box Plot no Orange, foi possível, facilmente, identificar os 10 bairros com mais ocorrências atendidas pelo CBMMA durante todo o período analisado. Como pode ser visto na Figura 17, o Centro de São Luís é o bairro com maior número de ocorrências, totalizando 3.243, seguido do bairro Maiobão, localizado no município Paço do Lumiar, com 2.687 ocorrências.

## 5.2.2 Quantitativo de ocorrências atendidas na região metropolitana de São Luís

A Figura 18 apresenta as 105.271 ocorrências divididas em dois grupos (CBMMA e GEM), sendo que 49.274 ocorrências (46,8%) foram atendidas pelo GEM (atualmente chamado de BBEM, Batalhão de Bombeiros de Emergências Médicas) e as demais, 55.997 ocorrências (53,2%), por um batalhão especializado diverso do BBEM.

Figura 18 – Total de ocorrências atendidas por agencia/batalhão.



Fonte: Elaborada pelo autor.

Todas as ocorrências da base de dados utilizada estão distribuídas em quatro municípios, de acordo com a Tabela 2.

Tabela 2 – Ocorrências por municípios.

Município	Quantidade de ocorrências
SÃO LUÍS	84.274
SÃO JOSÉ DE RIBAMAR	11.146
PAÇO DO LUMIAR	8.310
RAPOSA	1.541

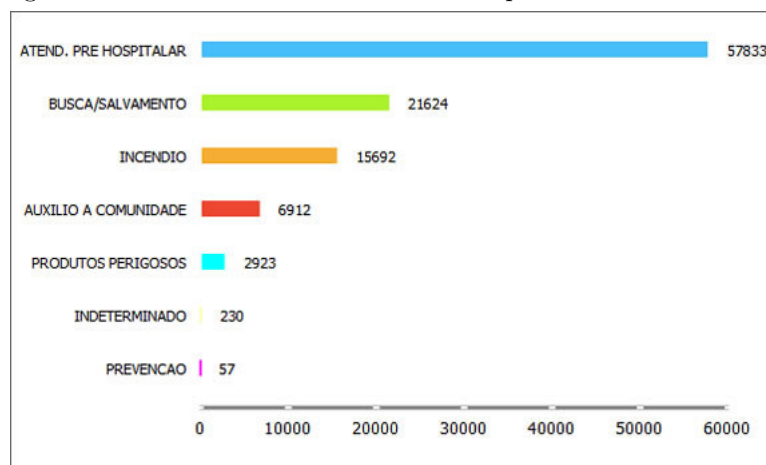
Fonte: Elaborada pelo autor.

No tocante às classes de ocorrências, a Figura 19 apresenta todos os registros da base de dados separados por classe de ocorrência (*Atendimento Pré-Hospitalar*, *Busca/Salvamento*, *Incêndio*, *Auxílio a Comunidade*, *Produtos Perigosos*, *Indeterminado* e *Prevenção*).

Visualizando a Figura 19, observa-se, claramente, a predominância de ocorrências relacionadas à classe *Atendimento Pré-Hospitalar*, totalizando 57.833 registros, o equivalente a 55% de todas as ocorrências encontradas na base de dados.

É importante ressaltar que na classe *Indeterminado* se encontram todas as ocorrências não enquadradas nas demais classes, por exemplo: acidentes envolvendo viaturas

Figura 19 – Total de ocorrências atendidas por classe de ocorrência.

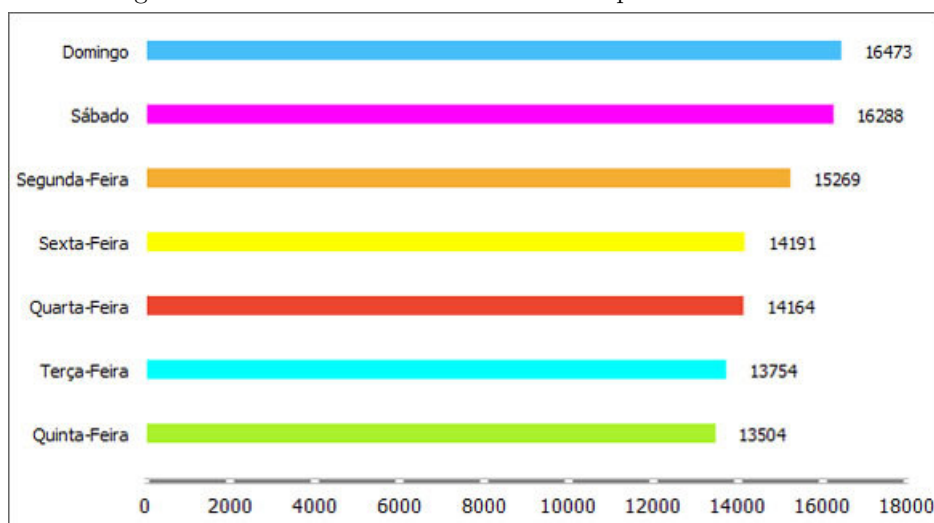


Fonte: Elaborada pelo autor.

militares, acidentes de trabalho, desacato, desobediência, porte ilegal de arma de fogo, roubo etc., desde que atendidas pelo Corpo de Bombeiros.

### 5.2.3 Comparativo de atendimentos por dia da semana e turno

Figura 20 – Total de ocorrências atendidas por dia da semana.

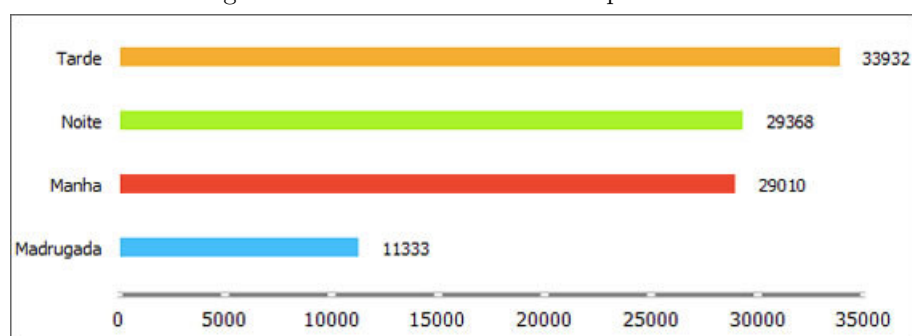


Fonte: Elaborada pelo autor.

Analisando a Figura 20, observa-se que nos dias *Sábado* e *Domingo* há um número maior de ocorrências registradas. Ainda utilizando o *widget* Box Plot (Figura 21), rapidamente se verifica qual o período do dia mais frequente na base de dados: *Tarde*, com 33.932 registros.

Agora correlacionando os atributos *CLASS\_OCORRENCIA* e *TURNO*, e fazendo uso do *widget* Distributions, observou-se que houve um maior registro de ocorrências da classe *Atendimento Pré-hospitalar* nos turnos vespertino (*Tarde* - das 12h às 18h) e

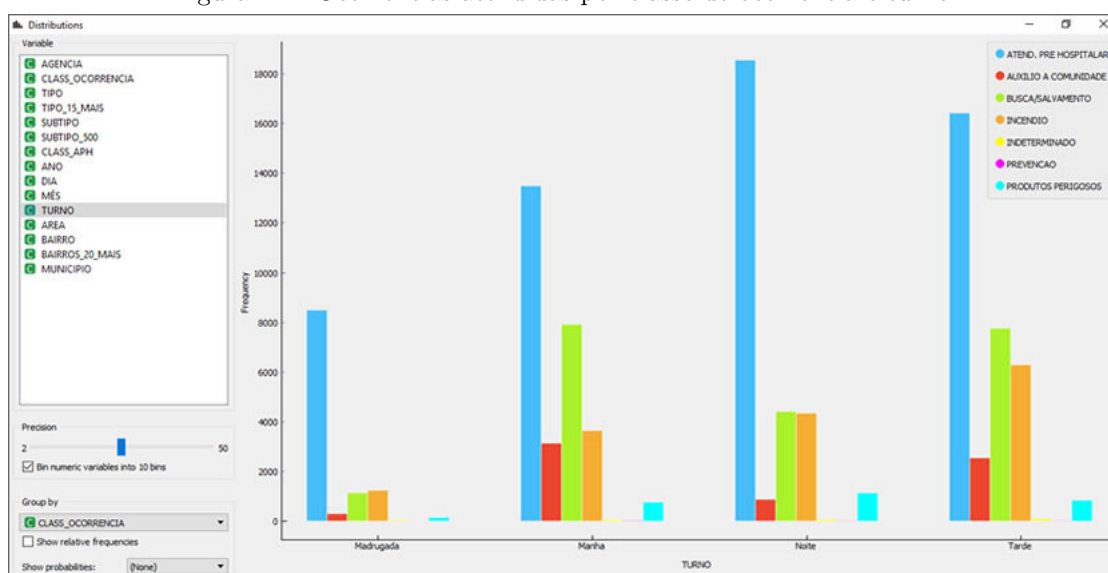
Figura 21 – Ocorrências atendidas por turno.



Fonte: Elaborada pelo autor.

noturno (*Noite* - das 18h às 00h), enquanto a classe *Incêndio* esteve mais presente no período vespertino (*TURNO: Tarde*), conforme Figura 22.

Figura 22 – Ocorrências atendidas por classe de ocorrência e turno.



Fonte: Elaborada pelo autor.

## 5.2.4 Os 15 tipos de ocorrências mais frequentes

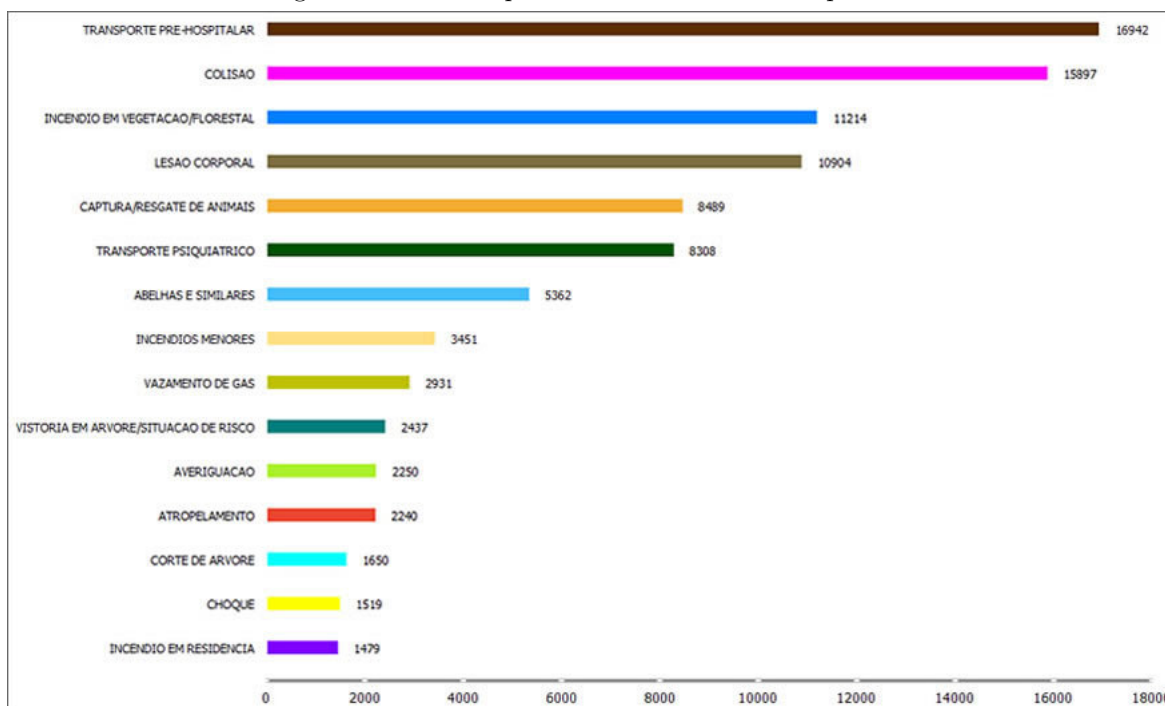
A Figura 23, também gerada através do *widget* Distributions, no Orange, apresenta os 15 tipos de ocorrências mais frequentes na base de dados. Analisando a mesma, percebe-se que o tipo de ocorrência *Transporte Pré-Hospitalar* é o mais frequente, com 16.942 registros, seguido do tipo *Colisão*, com 15.897 registros e do tipo *Incêndio em Vegetação (Incêndio Florestal)*, com 11.214 registros.

## 5.2.5 Os 10 tipos de incêndios mais frequentes

A Figura 24 apresenta os 10 tipos de incêndios mais frequentes entre janeiro de 2013 e agosto de 2018, ou seja, em toda a base de dados. Das 11.214 ocorrências de *Incêndio*



Figura 23 – Os 15 tipos de ocorrências mais frequentes.



Fonte: Elaborada pelo autor.

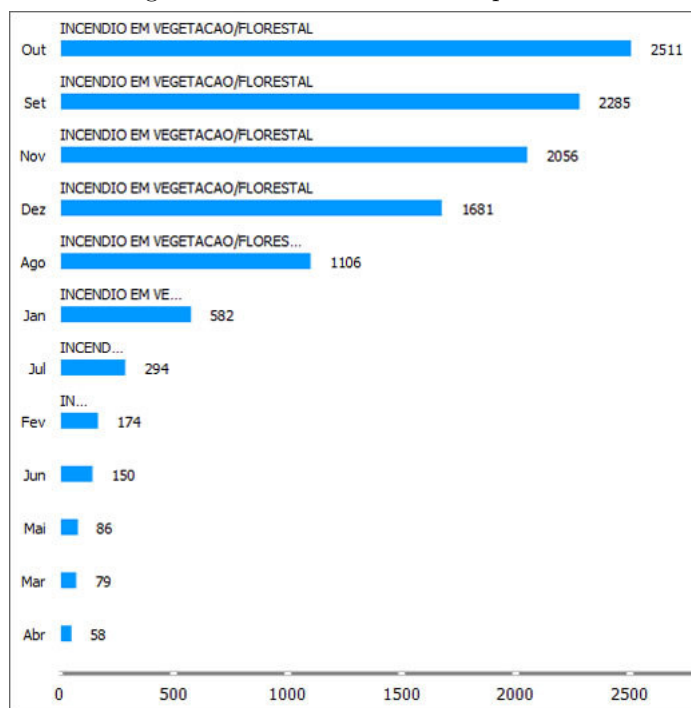
em *Vegetação*, 8.533 (equivalente a 76% do total) foram registradas nos meses de setembro, outubro, novembro e dezembro (Figura 25), meses estes nos quais se verifica uma maior incidência dos raios solares no Estado do Maranhão (CRESESB, 2019).

Figura 24 – Os 10 tipos de incêndios mais frequentes.



Fonte: Elaborada pelo autor.

Figura 25 – Incêndios florestais por mês.



Fonte: Elaborada pelo autor.

### 5.3 Aplicação do algoritmo *Apriori*

Após a extração de conhecimento efetuada com a ajuda dos *widgets* no Orange, foram realizados alguns experimentos na ferramenta Weka a fim de descobrir as melhores regras de associação aplicando-se o algoritmo *Apriori*. A análise de todas as regras, assim como a identificação das melhores, foi realizada com a ajuda de um profissional da área, um oficial do Corpo de Bombeiros Militar do Maranhão.

Inicialmente, foi aplicado o *Apriori* com todos os atributos presentes na base e assumindo os valores já estabelecidos por padrão (Figura 26) no algoritmo, ou seja, confiança de 90%, suporte mínimo de 10%, e número de regras igual a 10. O resultado da execução do algoritmo pode ser visualizado na Figura 27.

Figura 26 – Parâmetros iniciais do *Apriori*.

lowerBoundMinSupport	0.1
metricType	Confidence
minMetric	0.9
numRules	10

Fonte: Elaborada pelo autor.

Conforme a Figura 27, foram geradas 10 regras com 100% de confiança cada. Ou seja, em cada regra, o número de registros contendo o termo no antecedente era igual ao número de registros contendo o termo no consequente da regra.

Figura 27 – Regras geradas com 90% de confiança e suporte mínimo de 10%.

```

Best rules found:
1. AGENCIA=GEM 49274 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 49274 <conf:(1)> lift:(1.82) lev:(0.21) [22204] conv:(22204.22)
2. AGENCIA=GEM MUNICIPIO=SAO LUIS 39545 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 39545 <conf:(1)> lift:(1.82) lev:(0.17) [17820] conv:(17820.06)
3. AGENCIA=GEM CLASS_APH=CLINICO 24955 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 24955 <conf:(1)> lift:(1.82) lev:(0.11) [11245] conv:(11245.41)
4. CLASS_OCORRENCIA=BUSCA/SALVAMENTO 21624 ==> AGENCIA=CBMMA 21624 <conf:(1)> lift:(1.88) lev:(0.1) [10121] conv:(10121.51)
5. AGENCIA=GEM CLASS_APH=CLINICO MUNICIPIO=SAO LUIS 20124 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 20124 <conf:(1)> lift:(1.82) lev:(0.09) [9068] conv:(9068.43)
6. AGENCIA=GEM CLASS_APH=TRAUMA 19501 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 19501 <conf:(1)> lift:(1.82) lev:(0.08) [8787] conv:(8787.69)
7. AGENCIA=GEM AREA=LESTE 18889 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 18889 <conf:(1)> lift:(1.82) lev:(0.08) [8511] conv:(8511.9)
8. CLASS_OCORRENCIA=BUSCA/SALVAMENTO MUNICIPIO=SAO LUIS 17044 ==> AGENCIA=CBMMA 17044 <conf:(1)> lift:(1.88) lev:(0.08) [7977] conv:(7977.75)
9. TIPO_15_MAIS=TRANSPORTE FRE-HOSPITALAR 16942 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 16942 <conf:(1)> lift:(1.82) lev:(0.07) [7634] conv:(7634.53)
10. AGENCIA=GEM TIPO_15_MAIS=TRANSPORTE FRE-HOSPITALAR 16483 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 16483 <conf:(1)> lift:(1.82) lev:(0.07) [7427] conv:(7427.69)

```

Fonte: Elaborada pelo autor.

Este primeiro experimento não trouxe nenhuma informação interessante. Após análise por parte do especialista, verificou-se que a maioria das regras eram óbvias. Em sete regras, o antecedente continha  $AGENCIA = GEM$  e o consequente  $CLASS\_OCORRENCIA = ATEND. PRE HOSPITALAR$ , o que significa que o Batalhão de Bombeiros de Emergências Médicas, aqui representado por  $AGENCIA = GEM$ , atendeu ocorrências da classe *Atendimento Pré-Hospitalar*, sendo que este batalhão só atende essa classe de ocorrência.

A regra n° 4 ( $CLASS\_OCORRENCIA = BUSCA/SALVAMENTO 21624 ==> AGENCIA = CBMMA 21624$ ) também é irrelevante, visto que o Batalhão de Bombeiros de Emergências Médicas não atende ocorrências da classe *Busca/Salvamento*, restando apenas a opção  $AGENCIA = CBMMA$  para as ocorrências relativas a esta classe. O mesmo entendimento se extrai a partir da regra n° 8 ( $CLASS\_OCORRENCIA = BUSCA/SALVAMENTO MUNICIPIO = SAO LUIS 17044 ==> AGENCIA = CBMMA 17044$ ).

Tendo em vista a ausência de conhecimento útil no primeiro experimento, removeu-se o atributo  $AGENCIA$  na aba *Preprocess* e se ajustou o parâmetro *numRules* no algoritmo para 500, mantendo os mesmos valores para suporte e confiança. Parte do resultado pode ser visualizado na Figura 28.

Figura 28 – Parte do experimento II, após remoção do atributo  $AGENCIA$ .

```

Best rules found:
1. TIPO_15_MAIS=TRANSPORTE FRE-HOSPITALAR 16942 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 16942 <conf:(1)> lift:(1.82) lev:(0.07) [7634] conv:(7634.53)
2. CLASS_OCORRENCIA=INCENDIO SUBTIPO_500=SOMENTE DANOS MATERIAIS 19465 ==> CLASS_APH=CLINICO 19465 <conf:(1)> lift:(2.11) lev:(0.08) [8151] conv:(8151.11)
3. TIPO_15_MAIS=COLISAO SUBTIPO_500=VITIMA NAO FATAL 15156 ==> CLASS_APH=TRAUMA 15156 <conf:(1)> lift:(2.84) lev:(0.09) [9828] conv:(9828.49)
4. TIPO_15_MAIS=TRANSPORTE FRE-HOSPITALAR MUNICIPIO=SAO LUIS 13803 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 13803 <conf:(1)> lift:(1.82) lev:(0.06) [6220]
5. TIPO_15_MAIS=TRANSPORTE FRE-HOSPITALAR CLASS_APH=CLINICO 13716 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 13716 <conf:(1)> lift:(1.82) lev:(0.06) [6180]
6. TIPO_15_MAIS=COLISAO SUBTIPO_500=VITIMA NAO FATAL MUNICIPIO=SAO LUIS 12183 ==> CLASS_APH=TRAUMA 12183 <conf:(1)> lift:(2.84) lev:(0.08) [7900] conv:(7900)
7. CLASS_OCORRENCIA=INCENDIO SUBTIPO_500=SOMENTE DANOS MATERIAIS MUNICIPIO=SAO LUIS 12054 ==> CLASS_APH=CLINICO 12054 <conf:(1)> lift:(2.11) lev:(0.06) [635]
8. TIPO_15_MAIS=TRANSPORTE FRE-HOSPITALAR CLASS_APH=CLINICO MUNICIPIO=SAO LUIS 11201 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 11201 <conf:(1)> lift:(1.82)
9. TIPO_15_MAIS=INCENDIO EM VEGETACAO/FLORESTAL CLASS_APH=CLINICO 11163 ==> SUBTIPO_500=SOMENTE DANOS MATERIAIS 11163 <conf:(1)> lift:(5.63) lev:(0.09) [917]
10. TIPO_15_MAIS=INCENDIO EM VEGETACAO/FLORESTAL SUBTIPO_500=SOMENTE DANOS MATERIAIS 11163 ==> CLASS_APH=CLINICO 11163 <conf:(1)> lift:(2.11) lev:(0.06) [588]
11. TIPO_15_MAIS=LESAO CORPORAL 10904 ==> CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR 10904 <conf:(1)> lift:(1.82) lev:(0.05) [4913] conv:(4913.64)
12. AREA=SUL 15709 ==> MUNICIPIO=SAO LUIS 15705 <conf:(1)> lift:(1.25) lev:(0.03) [3129] conv:(626.65)
13. AREA=OESTE 19280 ==> MUNICIPIO=SAO LUIS 19273 <conf:(1)> lift:(1.25) lev:(0.04) [3838] conv:(480.69)
14. CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR AREA=OESTE 10850 ==> MUNICIPIO=SAO LUIS 10846 <conf:(1)> lift:(1.25) lev:(0.02) [2160] conv:(432.82)
15. CLASS_OCORRENCIA=INCENDIO CLASS_APH=CLINICO 15481 ==> SUBTIPO_500=SOMENTE DANOS MATERIAIS 15465 <conf:(1)> lift:(5.62) lev:(0.12) [12714] conv:(748.85)
16. CLASS_OCORRENCIA=INCENDIO CLASS_APH=CLINICO MUNICIPIO=SAO LUIS 12069 ==> SUBTIPO_500=SOMENTE DANOS MATERIAIS 12054 <conf:(1)> lift:(5.62) lev:(0.09) [590]
17. TIPO_15_MAIS=INCENDIO EM VEGETACAO/FLORESTAL 11214 ==> SUBTIPO_500=SOMENTE DANOS MATERIAIS 11163 <conf:(1)> lift:(5.6) lev:(0.09) [9170] conv:(177.34)
18. TIPO_15_MAIS=INCENDIO EM VEGETACAO/FLORESTAL 11214 ==> CLASS_APH=CLINICO 11163 <conf:(1)> lift:(2.1) lev:(0.06) [5859] conv:(113.66)
19. TIPO_15_MAIS=INCENDIO EM VEGETACAO/FLORESTAL 11214 ==> SUBTIPO_500=SOMENTE DANOS MATERIAIS CLASS_APH=CLINICO 11163 <conf:(1)> lift:(5.68) lev:(0.09) [915]
20. CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR SUBTIPO_500=VITIMA NAO FATAL MUNICIPIO=SAO LUIS 10852 ==> CLASS_APH=TRAUMA 10732 <conf:(0.99)> lift:(2.81) lev:(0.0)
21. SUBTIPO_500=VITIMA NAO FATAL MUNICIPIO=SAO LUIS 15953 ==> CLASS_APH=TRAUMA 15973 <conf:(0.99)> lift:(2.81) lev:(0.1) [10165] conv:(57.16)
22. CLASS_OCORRENCIA=ATEND. FRE HOSPITALAR SUBTIPO_500=VITIMA NAO FATAL 13480 ==> CLASS_APH=TRAUMA 13325 <conf:(0.99)> lift:(2.81) lev:(0.08) [8586] conv:(56)
23. SUBTIPO_500=VITIMA NAO FATAL 19786 ==> CLASS_APH=TRAUMA 19545 <conf:(0.99)> lift:(2.81) lev:(0.12) [12589] conv:(53.02)
24. SUBTIPO_500=SOMENTE DANOS MATERIAIS 18704 ==> CLASS_APH=CLINICO 18455 <conf:(0.99)> lift:(2.09) lev:(0.09) [9609] conv:(39.43)

```

Fonte: Elaborada pelo autor.

Com estas mudanças, o algoritmo conseguiu gerar 43 regras. Destas, cinco já

apresentaram algum conhecimento relevante (Figura 29).

Figura 29 – Cinco melhores regras do experimento II.

<p>17. TIPO_15_MAIS=INCENDIO EM VEGETACAO/FLORESTAL 11214 ==&gt; SUBTIPO_500=SOMENTE DANOS MATERIAIS 11163 &lt;conf:(1)&gt; lift:(5.6) lev:(0.09) [9170] conv:(177.34)</p> <p>27. CLASS_OCORRENCIA=INCENDIO 15692 ==&gt; SUBTIPO_500=SOMENTE DANOS MATERIAIS 15465 &lt;conf:(0.99)&gt; lift:(5.55) lev:(0.12) [12676] conv:(56.6)</p> <p>32. CLASS_OCORRENCIA=INCENDIO MUNICIPIO=SAO LUIS 12251 ==&gt; SUBTIPO_500=SOMENTE DANOS MATERIAIS 12054 &lt;conf:(0.98)&gt; lift:(5.54) lev:(0.09) [9877] conv:(50.88)</p> <p>36. TIPO_15_MAIS=COLISAO MUNICIPIO=SAO LUIS 12777 ==&gt; SUBTIPO_500=VÍTIMA NAO FATAL 12183 &lt;conf:(0.95)&gt; lift:(5.07) lev:(0.09) [9781] conv:(17.44)</p> <p>38. TIPO_15_MAIS=COLISAO 15897 ==&gt; SUBTIPO_500=VÍTIMA NAO FATAL 15156 &lt;conf:(0.95)&gt; lift:(5.07) lev:(0.12) [12168] conv:(17.4)</p>
--

Fonte: Elaborada pelo autor.

A regra nº 17 (TIPO\_15\_MAIS = INCENDIO EM VEGETACAO/FLORESTAL 11214 ==> SUBTIPO\_500 = SOMENTE DANOS MATERIAIS 11163) significa que, dos 11.214 registros de ocorrências de incêndio em vegetação, 11.163 apresentaram somente danos materiais (11.163/11.214 = 0.995 [99,54%], o algoritmo arredondou para 100% de confiança). Os 51 registros restantes (11.214 - 11.163) são relativos a incêndios em vegetação com SUBTIPO\_500 = VÍTIMA NÃO FATAL (45 registros) e 6 registros onde este campo não foi preenchido. Esta última informação foi possível obter verificando diretamente a planilha com a base de dados.

A regra nº 27 (CLASS\_OCORRENCIA = INCENDIO 15692 ==> SUBTIPO\_500 = SOMENTE DANOS MATERIAIS 15465) diz que, com 99% de confiança, quando houver uma ocorrência de incêndio, haverá somente danos materiais.

A regra nº 32 (CLASS\_OCORRENCIA = INCENDIO MUNICIPIO=SAO LUIS 12251 ==> SUBTIPO\_500 = SOMENTE DANOS MATERIAIS 12054) é muito semelhante a anterior. A diferença é que ela apresenta informações sobre os incêndios ocorridos apenas ao município de São Luís. Com 98% de confiança, toda vez que houver um incêndio em São Luís, tal incêndio terá como consequência apenas danos materiais.

A regra nº 36 TIPO\_15\_MAIS=COLISAO MUNICIPIO = SAO LUIS 12777 ==> SUBTIPO\_500 = VÍTIMA NAO FATAL 12183 apresenta informações acerca das ocorrências do tipo colisão no município de São Luís. Com 95% de confiança, sempre que houver uma colisão em São Luís, não haverá vítima fatal.

Por fim, a regra nº 38 (TIPO\_15\_MAIS = COLISAO 15897 ==> SUBTIPO\_500 = VÍTIMA NAO FATAL 15156) informa que, com relação a todos os municípios existentes na base de dados, sempre que houver uma colisão, não haverá vítima fatal, com 95% de confiança.

No terceiro experimento, foram removidos os atributos CLASS\_APH e MUNICIPIO, na aba *Preprocess*, além de alterados os parâmetros *minMetric* para 0.8 e *numRules* para 1000, ou seja, confiança de 80% e geração de 1000 regras. O algoritmo retornou apenas seis regras conforme a Figura 30. Das seis regras geradas, as cinco primeiras já haviam sido apresentadas no experimento anterior e a sexta não trouxe nenhum conhecimento relevante.

Figura 30 – Regras geradas após o experimento III.

```

Minimum support: 0.1 (10527 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 29

Size of set of large itemsets L(2): 17

Best rules found:

1. TIPO_15_MAIS=TRANSPORTE PRE-HOSPITALAR 16942 ==> CLASS_OCORRENCIA=ATEND. PRE HOSPITALAR 16942 <conf:(1)>
2. TIPO_15_MAIS=LESAO CORPORAL 10904 ==> CLASS_OCORRENCIA=ATEND. PRE HOSPITALAR 10904 <conf:(1)>
3. TIPO_15_MAIS=INCENDIO EM VEGETACAO/FLORESTAL 11214 ==> SUBTIPO_500=SOMENTE DANOS MATERIAIS 11163 <conf:(1)>
4. CLASS_OCORRENCIA=INCENDIO 15692 ==> SUBTIPO_500=SOMENTE DANOS MATERIAIS 15465 <conf:(0.99)>
5. TIPO_15_MAIS=COLISAO 15897 ==> SUBTIPO_500=VITIMA NAO FATAL 15156 <conf:(0.95)>
6. SUBTIPO_500=SOMENTE DANOS MATERIAIS 18704 ==> CLASS_OCORRENCIA=INCENDIO 15465 <conf:(0.83)>

```

Fonte: Elaborada pelo autor.

Devido a dificuldade na geração de regras com os parâmetros anteriores, decidiu-se fazer um quarto experimento alterando a métrica confiança para 0.5 (*minMetric* = 0.5). O resultado pode ser visualizado na Figura 31.

Figura 31 – Regras geradas após o experimento IV.

```

1. TIPO_15_MAIS=TRANSPORTE PRE-HOSPITALAR 16942 ==> CLASS_OCORRENCIA=ATEND. PRE HOSPITALAR 16942 <co
2. TIPO_15_MAIS=LESAO CORPORAL 10904 ==> CLASS_OCORRENCIA=ATEND. PRE HOSPITALAR 10904 <conf:(1)> lif
3. TIPO_15_MAIS=INCENDIO EM VEGETACAO/FLORESTAL 11214 ==> SUBTIPO_500=SOMENTE DANOS MATERIAIS 11163
4. CLASS_OCORRENCIA=INCENDIO 15692 ==> SUBTIPO_500=SOMENTE DANOS MATERIAIS 15465 <conf:(0.99)> lift:
5. TIPO_15_MAIS=COLISAO 15897 ==> SUBTIPO_500=VITIMA NAO FATAL 15156 <conf:(0.95)> lift:(5.07) lev:(
6. SUBTIPO_500=SOMENTE DANOS MATERIAIS 18704 ==> CLASS_OCORRENCIA=INCENDIO 15465 <conf:(0.83)> lift:
7. SUBTIPO_500=VITIMA NAO FATAL 19786 ==> TIPO_15_MAIS=COLISAO 15156 <conf:(0.77)> lift:(5.07) lev:(
8. SUBTIPO_500=VITIMA NAO FATAL 19786 ==> CLASS_OCORRENCIA=ATEND. PRE HOSPITALAR 13480 <conf:(0.68)>
9. TURNO=Noite 29368 ==> CLASS_OCORRENCIA=ATEND. PRE HOSPITALAR 18555 <conf:(0.63)> lift:(1.15) lev:
10. SUBTIPO_500=SOMENTE DANOS MATERIAIS 18704 ==> TIPO_15_MAIS=INCENDIO EM VEGETACAO/FLORESTAL 11163
11. AREA=LESTE 37905 ==> CLASS_OCORRENCIA=ATEND. PRE HOSPITALAR 22238 <conf:(0.59)> lift:(1.07) lev:(
12. AREA=OESTE 19280 ==> CLASS_OCORRENCIA=ATEND. PRE HOSPITALAR 10850 <conf:(0.56)> lift:(1.02) lev:(
13. ANO=ANO 2015 22552 ==> CLASS_OCORRENCIA=ATEND. PRE HOSPITALAR 12446 <conf:(0.55)> lift:(1) lev:(0
14. ANO=ANO 2016 24583 ==> CLASS_OCORRENCIA=ATEND. PRE HOSPITALAR 13504 <conf:(0.55)> lift:(1) lev:(-

```

Fonte: Elaborada pelo autor.

Como pode ser verificado na Figura 31, foram geradas 14 regras. A regra nº 9 (TURNO = Noite 29368 ==> CLASS\_OCORRENCIA = ATEND. PRE HOSPITALAR 18555) apresenta uma importante informação: se a ocorrência é atendida no período noturno (TURNO = Noite), então a classe da ocorrência será *Atendimento Pré-Hospitalar*, com 63% de confiança.

Informação semelhante a esta foi obtida, graficamente, através do Orange na Figura 22 (Subseção 5.2.3), no qual foi verificado que a maioria das ocorrências atendidas a noite foram da classe *Atendimento Pré-Hospitalar*. Observando a base de dados,

no tocante aos registros com  $TURNO = Noite$ , verificou-se que 108 ocorrências com  $CLASS\_OCORRENCIA = ATEND. PRE HOSPITALAR$ , foram atendidas por um batalhão diverso do Batalhão de Bombeiros de Emergências Médicas, enquanto 13.372 ocorrências foram atendidas pelo BBEM ( $AGENCIA = GEM$ ). Isso é um número elevado de ocorrências atendidas por um único batalhão, que dispõe somente de duas ambulâncias para cobrir toda a região metropolitana de São Luís, sendo que uma delas se encontra em um posto avançado na cidade de Bacabeira, fato este verificado *in loco*.

A regra n° 11 ( $AREA = LESTE 37905 \implies CLASS\_OCORRENCIA = ATEND. PRE HOSPITALAR 22238$ ) pode ser interpretada como: se  $AREA = LESTE$ , então  $CLASS\_OCORRENCIA = ATEND. PRE HOSPITALAR$ , com 59% de confiança. Isto significa que, se um chamado de ocorrência for gerado para os bairros pertencentes a zona leste ( $AREA = LESTE$ ) da região metropolitana de São Luís, provavelmente será uma ocorrência da classe *Atendimento Pré-Hospitalar*, com 59% de confiança. Os principais bairros que estão localizados na zona leste da região metropolitana de São Luís são: Alto Paranã, Andiroba, Bob Kenedy, Campo de Perizes, Centro de São José de Ribamar, Cidade Olímpica, Cidade Operária, São Raimundo, Estrada da Mata, Jardim Tropical, Maiobão, Centro de Panaquatira, Centro de Paço do Lumiar e Paranã.

A regra n° 12 ( $AREA = OESTE 19280 \implies CLASS\_OCORRENCIA = ATEND. PRE HOSPITALAR 10850$ ), semelhantemente a regra anterior, pode ser interpretada como: se  $AREA = OESTE$ , então  $CLASS\_OCORRENCIA = ATEND. PRE HOSPITALAR$ , com 56% de confiança. Isto significa que, se um chamado de ocorrência for gerado para os bairros pertencentes a zona oeste ( $AREA = OESTE$ ) da região metropolitana de São Luís, provavelmente será uma ocorrência da classe *Atendimento Pré-Hospitalar*, com 56% de confiança. Os principais bairros da zona oeste em questão são: Alemanha, Anil, Areinha, Bairro de Fátima, Camboa, Centro de São Luís, Caratatiua, Filipinho, João Paulo, Vila Palmeira, Apeadouro, Coroadó, Cruzeiro do Anil e Monte Castelo.

Vale destacar que o Batalhão de Bombeiros de Emergências Médicas se localiza no bairro Cohab Anil IV, zona norte da região metropolitana de São Luís, que abrange os bairros Cohab, Cohatrac, Angelim, Bequimão, Barramar, Calhau, Araçagi, Ponta D'Areia, Renascença, Cohama, Cohajap, Vinhais, Cohafuma, e os maiores índices de ocorrências de *Atendimento Pré-Hospitalar*, como comprovado anteriormente, concentram-se nos bairros pertencentes às zonas leste e oeste.

## 6 Conclusões e trabalhos futuros

A Mineração de Dados foi fundamental para a descoberta de informações relevantes na base de dados de ocorrências atendidas pelo Corpo de Bombeiros Militar do Maranhão, visto que, além de ter sido um trabalho inédito usando dados reais da corporação, foi possível extrair conhecimentos úteis, que poderão ser usados para melhorar o serviço público prestado ao cidadão pelo Estado.

A ferramenta Orange se mostrou bastante eficiente, permitindo, de forma prática, a visualização gráfica de informações outrora ocultas e de difícil entendimento existentes na base de dados.

Por meio do Orange, foi possível verificar: os 10 bairros com maior número de ocorrências atendidas; o percentual de ocorrências atendidas pelo Batalhão de Bombeiros de Emergências Médicas; a quantidade de ocorrências por município; um comparativo de atendimentos por dia da semana, sendo *Sábado* e *Domingo* os dias com maior registro de ocorrências; o período do dia com maior registro de ocorrências (*Tarde*); as classes de ocorrências mais frequentes; quais classes de ocorrências foram observadas por turno, sendo *Atendimento Pré-Hospitalar* a classe com maior registro no período noturno (TURNO = Noite); os 15 tipos de ocorrências mais frequentes; e, os 10 tipos de incêndios mais observados.

Com relação à ferramenta Weka, foram realizados quatro experimentos, alterando os parâmetros do algoritmo *Apriori*, e se observou uma dificuldade no tocante à geração das regras de associação. Isso se deu devido aos diversos problemas apresentados na base de dados: células vazias, células preenchidas com valores fora do padrão, dados repetidos, dados inconsistentes etc. Porém, as relações geradas pelo software e definidas como “melhores regras”, apresentaram conhecimentos relevantes para o Corpo de Bombeiros Militar do Estado do Maranhão.

De posse do conhecimento obtido neste trabalho, é possível sugerir, à Secretaria de Segurança Pública do Estado do Maranhão, a implantação de novos batalhões especializados em atendimento pré-hospitalar nas áreas com maior índice de ocorrências da classe *Atendimento Pré-Hospitalar*, aumentando o número de ambulâncias e, conseqüentemente, melhorando o tempo de resposta de cada ocorrência, fator decisivo em prol da vida.

Enfim, os objetivos desta pesquisa foram alcançados e propõe-se como trabalho futuro, devido principalmente ao baixo número de regras de associação geradas, a implantação de um novo sistema para registro das ocorrências, a fim de se obter maiores detalhes em cada ocorrência, tais como tempo de resposta, quantidade de vítimas na ocorrência, mensuração de danos, fatores que deram origem à ocorrência, emprego de

técnicas adequadas no atendimento, identificação de qual batalhão atendeu determinada ocorrência etc. Posteriormente, almeja-se a aplicação da Mineração de Dados e a extração de conhecimentos na nova base de dados.



# Referências

- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. *Mining Association Rules between Sets of Items in Large Databases*. New York: ACM Press, 1993. Citado 2 vezes nas páginas 22 e 23.
- ALEGRIA, R. et al. *Teoria e Prática da Pesquisa Aplicada*. Rio de Janeiro: Elsevier, 2012. 504 p. Citado na página 28.
- AMARAL, F. *Aprenda Mineração de Dados: Teoria e prática*. Rio de Janeiro: Alta Books Editora, 2016. 220 p. Citado na página 25.
- BARRO, S.; MARIN, R. *Fuzzy Logic in Medicine*. New York: Springer Science Business Media, 2001. v. 83 (Studies in Fuzziness and Soft Computing). 310 p. Citado na página 19.
- BRAGA, L. P. V. *Introdução à Mineração de Dados*. 2<sup>a</sup>. ed. Rio de Janeiro: E-papers, 2005. 212 p. Citado 2 vezes nas páginas 12 e 18.
- BRASIL. Lei nº 10.230, de 23 de abril de 2015. dispõe sobre a organização básica do corpo de bombeiros militar do maranhão e dá outras providências. *Diário Oficial [do] Estado do Maranhão*, São Luís, MA, 24 abr. 2015. Disponível em: <<http://www.defesacivil.ma.gov.br/files/2015/04/Lei-de-Organiza%C3%A7%C3%A3o-B%C3%A1sica-do-CBMMA.pdf>>. Acesso em: 17 dez. 2018. Citado na página 29.
- CBMMA. *Nossa História - Corpo de Bombeiros*. 2018. Disponível em: <<https://cbm.ssp.ma.gov.br/index.php/cbmma/institucional/nossa-historia/>>. Acesso em: 17 dez. 2018. Citado na página 29.
- CIRIBELLI, M. C. *Como elaborar uma dissertação de mestrado através da pesquisa científica*. Rio de Janeiro: 7Letras, 2003. 222 p. Citado na página 28.
- CLESIO, F. *Mineração de Dados com Software Livre*. 2012. Disponível em: <<https://mineracaodedados.wordpress.com/tag/statistica/>>. Acesso em: 18 dez. 2018. Citado na página 24.
- COLEMAN, W. B.; TSONGALIS, G. J. *Molecular Pathology: The Molecular Basis of Human Disease*. 2. ed. London: Academic Press, 2017. 802 p. Citado na página 26.
- CRESESB. *Base de dados de irradiação solar incidente*. 2019. Disponível em: <<http://www.cresesb.cepel.br/>>. Acesso em: 25 mai. 2019. Citado na página 40.
- FAGUNDES, L. G. *Mineração de Dados em arquivos de logs de um sistema de monitoração de Contact Centers*. São José dos Campos, 2009. 125 p. Citado na página 20.
- FAYYAD, U. M. et al. *Advances in Knowledge Discovery and Data Mining*. 1. ed. Califórnia: Menlo Park, 1996. 611 p. Citado 2 vezes nas páginas 15 e 16.
- FERRARI, D. G.; SILVA, L. N. d. C. *Introdução a mineração de dados*. São Paulo: Saraiva, 2017. 406 p. Citado na página 20.

- FONG, J. *Data Mining, Data Warehousing and Client-Server Databases: Proceedings of the 8th International Hong Kong Computer Society Database Workshop*. 1. ed. EUA: Springer, 1997. 332 p. Citado na página 17.
- GASSENFERTH, W. et al. *Gestão de negócios e sustentabilidade*. 1. ed. Rio de Janeiro: Brasport, 2015. 368 p. Citado na página 18.
- GOLDSCHMIDT, R.; PASSOS, E. *Data mining: um guia Prático*. Rio de Janeiro: Elsevier, 2005. 262 p. Citado na página 21.
- KHOSROW-POUR, M. *Advanced Methodologies and Technologies in Network Architecture, Mobile Computing, and Data Analytics*. Hershey PA: IGI Global, 2018. 1857 p. Citado na página 27.
- KROENKE, D. M. *Sistemas de informação gerenciais*. 1. ed. São Paulo: Saraiva, 2017. 583 p. Citado na página 17.
- MILLER, H. J.; HAN, J. *Geographic Data Mining and Knowledge Discovery*. Second edition. London: CRC Press, 2009. 486 p. Citado na página 16.
- OLIVEIRA, R. C.; FERREIRA, J. S. *Investigação operacional em ação: casos de aplicação*. Coimbra: Coimbra, Imprensa da Universidade de, 2014. 774 p. Citado na página 15.
- QUILICI-GONZALEZ, J. A.; ZAMPIROLI, F. d. A. *Sistemas inteligentes e mineração de dados*. Santo André: Triunfal Grafica e Editora, 2014. 148 p. Citado na página 20.
- RAMAKRISHNAN, R.; GEHRKE, J. *Sistemas de gerenciamento de banco de dados*. 3. ed. Porto Alegre: AMGH, 2008. 905 p. Citado na página 20.
- RAMPAZZO, L. *Metodologia científica*. São Paulo: Edicoes Loyola, 2002. 141 p. Citado na página 28.
- SHEN, Y.; ZHANG, Z.; YANG, Q. Objective-oriented utility-based association mining. IEEE International Conference on Data Mining, Maebashi, Japan, 2002. Acesso em: 15 mai. 2019. Citado na página 12.
- SILVA, L. A. d.; PERES, S. M.; BOSCARIOLI, C. *Introdução à Mineração de Dados: Com Aplicações em R*. Rio de Janeiro: Elsevier Brasil, 2016. 296 p. Citado na página 12.
- SOUZA, G. S. d.; SANTOS, A. R. d.; DIAS, V. B. *Metodologia da pesquisa científica: a construção do conhecimento e do pensamento científico no processo de aprendizagem*. Porto Alegre: Animal, 2003. 164 p. Citado na página 28.
- SSPMA. *CIOPS - SSP/MA - Secretaria de Segurança Pública do Maranhão*. 2018. Disponível em: <<https://www.ssp.ma.gov.br/ciops/>>. Acesso em: 17 dez. 2018. Citado na página 30.
- STEINER, M. T. A. et al. Data mining como suporte à tomada de decisões - uma aplicação no diagnóstico médico. XXXVI SBPO — O Impacto da Pesquisa Operacional nas Novas Tendências Multidisciplinares, São João Del Rei, MG, 2004. Disponível em: <<http://http://www.din.uem.br/sbpo/sbpo2004/pdf/arq0009.pdf>>. Acesso em: 23 jun. 2019. Citado na página 15.